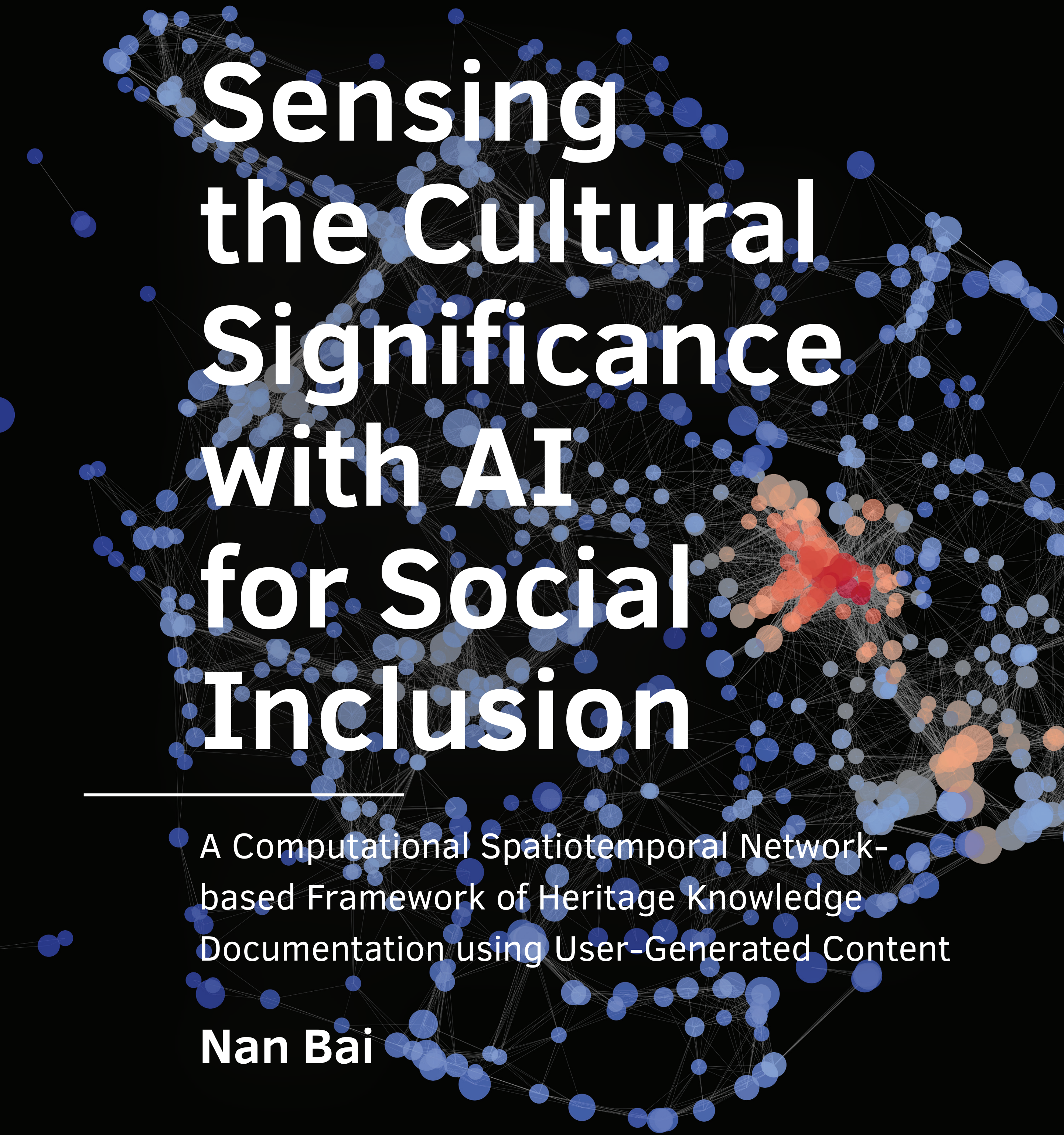


Sensing the Cultural Significance with AI for Social Inclusion

A Computational Spatiotemporal Network-based Framework of Heritage Knowledge Documentation using User-Generated Content

Nan Bai

Sensing the Cultural Significance with AI for Social Inclusion

A Computational Spatiotemporal Network-based Framework of Heritage Knowledge Documentation using User-Generated Content

Nan Bai

A+BE | Architecture and the Built Environment | TU Delft BK

23#17

Design | Sirene Ontwerpers, Véro Crickx

ISBN 978-94-6366-749-4
ISSN 2212-3202

This dissertation is open access at https://doi.org/10.7480/abe.2023.17

Sensing the Cultural Significance with AI for Social Inclusion

A Computational Spatiotemporal Network-based Framework of Heritage Knowledge Documentation using User-Generated Content

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Thursday 05, October 2023 at 12:30 o'clock by

Nan BAI
Master of Architecture,
Tsinghua University, China
Master of Science in Architektur,
Die Technische Universität Berlin, Germany
born in Gansu, China

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. A. R. Pereira Roders	Delft University of Technology, promotor
Dr. P. Nourian	University of Twente, copromotor

Independent members:

Prof. dr.ing. S. Nijhuis	Delft University of Technology
Prof. dr. L. Zhang	Tsinghua University, China
Prof. dr. R. da Silva Torres	Wageningen University and Research
Prof. arch. M. Turner	Bezalel Academy of Arts, Israel
Prof. dr. L.C.M. Itard	Delft University of Technology, reserve member

Other members:

Prof. dr. T. Cheng	University College London, United Kingdom

The PhD research is within the framework of the Heriland-Consortium. HERILAND is funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813883.

To my caring family, lifelong mentors, and beloved friends.

Nan Bai

Acknowledgement

When I was still a very little child, a close relative from my family, uncle LU, always called me “Dr. BAI”, partly, if not mainly, for fun, despite I knew nothing about what that meant. He kept calling me that until I started attending middle school. This friendly joke seemed to become a charm and this charm seems to get working.

Before applying for the PhD position at TU Delft, I tasted very divergent things throughout my journey of bachelor’s and master’s study, including but not limited to architectural [re]design, rural vitalization, psychology, behavioural science, big data analysis, social network analysis, architecture and art history, historic architecture study and heritage conservation (or “Historische Bauforschung und Denkmalpflege” in German for a better context). All these experience together has shaped me as I currently am in terms of expertise and interests. Because of the complexity in my experience and interests, I felt extremely lucky to have found and got selected by this current PhD position about “inclusive heritage management processes” at TU Delft, offering me a unique opportunity to make use of everything that I was already good at and that I would definitely love to learn. I cannot wait to express my greatest gratitude to my promoters Prof. Ana PEREIRA RODERS and Dr. Pirouz NOURIAN, without whose support this dissertation can never be written. Both of my promoters helped me without any reservation and offered me the best from their own fields of expertise. Prof. Ana PEREIRA RODERS guided me throughout the journey, provided me enough freedom as long as I am on the right track, encouraged me to be consistent and systematic, and tolerated and even supported every crazy idea that I had for the research. She is always there, smiling, calmly listening to all sorts of questions, problems, and even troubles, and willing to offer the help with her strongest efforts, no matter how busy she can be. From the first day of my PhD journey, Dr. Pirouz NOURIAN has been enthusiastically suggesting me to pay more attention to the mathematical side of the “wrong” but “useful” models. This view has also shaped me on how I would approach a research question and write a scientific paper. I really appreciate the long days of discussion that we had, side by side or face to face (across laptop screens), working out every line of derivations in the papers.

I would also like to extend my gratitude to the PhD committee members: Prof. Tao CHENG, Prof. Ricardo DA SILVA TORRES, Prof. Laure ITARD, Prof. Steffen NIJHUIS, Prof. Michael TURNER, and Prof. Li ZHANG. Together with Dr. Francesca NOARDO, Prof. Steffen NIJHUIS and Prof. Michael TURNER were also the commmittee members of my Go/No Go Meeting, who made quite a few important and constructive comments that benefited my PhD research significantly.

One biggest attraction for me of my PhD position before coming to the Netherlands, was, of course, the various activities organized by the HERILAND network. As a pan-European research and training network composed of six universities from five countries, we were supposed to meet every half a year in a different place for celebrating our timely achievements, which was unfortunately disrupted by the Covid-19 pandemic after the very first workshops in the Netherlands. Even though many of the on-site get-togethers did not happen, we do manage to form an intimate and robust international and multidisciplinary team with strong professional and personal links. Next to mentioning the funding organization, the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813883, I must first thank the coordinators and management team of the gigantic project, Prof. Gert-Jan BURGERS, Dr. Niels VAN MANEN, and Mr. Ruud VAN OOIJEN, for their efforts of making our project flexible, resilient, and responsible. The second level of gratitude goes to the supervisors from all partner institutions including but not limited to Prof. Francesco CELLINI, Dr. Linde EGBERTS, Mr. Graham FAIRCLOUGH, Dr. Susanne FREDHOLM, Dr. Ron JANSSEN, Dr. María MARGARITA SEGARRA LAGUNES, Dr. Eva LÖFGREN, Prof. John PENDLEBURY, Dr. Roberto ROCCO, Prof. Maggie ROE, Prof. Michael TURNER, Prof. Sam TURNER, Dr. Frank VAN DER HOEVEN, Dr. Els VERBAKEL, Dr. Pieter WAGENAAR, and Prof. Ola WETTERBERG. The supervisors are from very divergent backgrounds touching all aspects of cultural heritage and landscape planning, bringing in precious insights from different perspectives. I am also honoured to get to know Mr. Teun VAN DEN ENDE through the Heriland Network, who introduced me to Ms. Inez WEYERMANS from the City of Amsterdam, and arranged a unique experience for me to talk about my research with a broad audience through a podcast. Similarly, I am glad to have met Ms. Martina TENZER at the Heriland Conference, who shares the same research passion with me. My deepest thanks go to, naturally, all the Early-Stage Researchers (ESRs) who ever joined the team, despite some of them did not continue to the end: Anastasia AKERMAN, Sophia ARBARA, Alana CASTRO DE AZEVEDO, Maitri DORE, Marta DUCCI, Farnaz FARAJI, Maciej JAKUB SWIDERSKI, Ana JAYONE YARZA PÉREZ, Moses KATONTOKA, Marilena MELA, Tinatin MEPARISHVILI, Rusudan MIRZIKASHVILI, Komal POTDAR, Rebecca STAATS, Ana TONK, Georgia TSIANTA, and Maria VALESE. I especially appreciate the days and nights during the Writing Retreat we had in Sweden. In the past years, we see each other grow.

I would like to take the chance to thank my co-authors throughout the years, Nour ABUZAID, Shervin Azadi, Prof. Raoul BUNSCHOTEN, Prof. Tao CHENG, Dr. Bruno DE ANDRADE, Yajing DI, Huichao DING, Marta DUCCI, Dr. Weixin HUANG, Jianan LI, Dr. Renqian LUO, Hongtao MA, Rusudan MIRZIKASHVILI, Dr. Pirouz NOURIAN, Prof. Ana PEREIRA RODERS, Meram-Irina PIENARU, Samaneh REZVANI, Dr. Pei SUN, Roy UIJTENDAAL, Prof. Lu WANG, Anping XIE, Wanting XU, and Wenqia YE. From all of them I have learnt a lot. Special thanks to my high school colleague and friend, Dr. Renqian LUO, for guiding me step by step in publishing in computer science venues.

With the same helpful and kind spirit as our captain Prof. Ana PEREIRA RODERS, the entire HEVA PhD group is like a big caring family, who are always there for help. I am sure I have received much more than I managed to offer. Thank you all, Sophia ARBARA, Clara CABRAL, Mahda FOROUGHI, Manal GINZARLY, Joana GONÇALVES, Ana JAYONE YARZA PÉREZ, Moses KATONTOKA, Muxi LEI, Ji (Leo) LI, Mi (Emeline) LIN, Adva MATAR, Soscha MONTEIRO DE JESUS, Bruna Ribeiro NUNES, Nadia PINTOSSI, Komal POTDAR, Ilaria ROSETTI, Lidwine SPOORMANS, Ana TARRAFA SILVA, Maria VALESE, Loes VELDPAUS, Han ZHANG, and our newest member, Yan ZHOU.

I would love to extend my gratitude to our section Heritage & Architecture, especially Dr. Bruno DE ANDRADE, Dr. Azadeh ARJOMAND KERMANI, Dr. Dan CHEN, Dr. Nicholas CLARK, Prof. Wessel DE JONGE, Mr. Alexander DE RIDDER, Drs. Judith FRAUNE, Dr. Yasuko KAMEI, Dr. Marzia LODDO, Dr. Barbara LUBELLI, Ms. Silvia NALDINI, Dr. Ivan NEVZGODIN, Prof. Uta POTTGIESSER, Dr. Wido QUIST, Dr. Charlotte VAN EMSTEDE, Dr. Marie-Thérèse VAN THOOR, and Dr. Hielkje ZIJLSTRA, with whom I ever chatted about my research. Special thanks to our section leader, Dr. Wido QUIST and our HR officers, Ms. Nathalie BAST and Ms. Julia KREUWEL, for sorting out many managerial complexities for me. Many thanks to our AE+T secretary, especially Daniel PANTOPHLET, Linda VAN RIJSBERGEN, Bo SONG, Carla VAN DER KROGT, Francoise VAN PUFFELEN, Barbara VAN VLIET, Yvonne VREDEVELD, the department manager Dr. Onno DE WIT, and the department head Michiel KREUTZER, for all the help with administrative issues. Thank the PhD mentors from our department, Dr. Regina BOKEL and Dr. Arjen MEIJER. Thank the organization committee of AE+T Bites. Thank the faculty Graduate School, Nouzha CHAMKH, Inge MEULENBERG, and Prof. Paul CHAN, for their constant support. Thank the former data steward of our A+BE faculty, Dr. Yan WANG, and the privacy officer, Ms. Marlou VELOO, for arranging the very first DPIA in our faculty for me. Thank Dr. Frank VAN DER HOEVEN and the LaTeX team, for allowing me to be the very first candidate from our faculty to use the ABE_v3 LaTeX template to write their thesis. Another special thank to our graphic designer Ms. Véro CRICKX, for her thorough support with the book cover and thesis layout.

Thank the faculty and university Graduate School to arrange all the courses, among them I would love to mention a few: Dr. Queena QIAN, Prof. Cor WAGENAAR, and Dr. Juan DURÁN, for forcing me to critically think of my research and prepare for the first proposal; Mr. Sören JOHNSON, for guiding me to write clear and concise journal articles; Dr. Trijsje FRANSSEN, for giving me a chance to discuss the scientific ethics of my own research by writing a science fiction novella; Dr. Aliza GLASBERGEN-PLAS and Dr. Lavinia MARIN, for preparing me for the final versions of my dissertation and propositions; as well as Drs. Gabriël HOEZEN and Drs. Astrid VAN LAAR, for bringing my Dutch to a significant next level. Thank Coursera, edX, MITx, and Stanford Online, for offering so many high-quality Computer Science courses, where I learned many of the theories and techniques used in this thesis.

For my pre-PhD days, I would first love to thank my master supervisor Prof. Lu WANG, who taught me how to comment on design, how to think about research, and how to appreciate life. I would also love to thank my supervisors at different stages of study, Prof. Li ZHANG, Dr. Pei SUN, Prof. Raoul BUNSCHOTEN, and the mentors who instructed me for different lengths of periods (from a few words of encouragement to a few years of interactions), Prof. Michael BATTY, Dr. Anrong DANG, Dr. He HUANG, Dr. Weixin HUANG, Prof. Jun JIA, Prof. Jian KANG, Prof. Ed MANLEY, Dr. Luke LI, Dr. Mingming LI, Dr. Chang LIU, Dr. Chen LIU, Dr. Hailong LIU, Dr. Lun LIU, Dr. Ying LONG, Dr. Dan LUO, Dr. Wei LUO, Prof. Zhou LV, Prof. Kaiping PENG, Mr. Bin QI, Prof. Thekla

Inge SCHULZ-BRIZE, Dr. Kehan SHEN, Dr. Liping WANG, Dr. Wei XIAO, Dr. Peng XUE, Prof. Feng YU, Dr. Dan ZHANG, Dr. Lu ZHENG, Dr. Xiaodi ZHENG, and Prof. Weimin ZHUANG. They have prepared me for conducting research, trained me with various skills, some more professional, some more personal.

A very special acknowledgement for a late family friend and mentor, Dr. Baisheng YE, who has always been my role model, motivating me to learn smarter and work harder.

My internship at the Palace Museum in 2019 was a valuable transition for me between studying as a student and working as a researcher, during which I learnt a lot about the practice of World Heritage management. I would love to thank my instructors Anping XIE, Yajing DI, Peng ZHAO, the director Dr. Xudong WANG, the seniors Mengyu LIU, Renhao LIU, Shiyu PIAO, Zhongjin WANG, Meng WU, Qiong ZHANG, Xiaogu ZHANG, and Meng ZHU, as well as my peers Yuehan AN, Lu BAO, Ruoxi FENG, Yao LU, Qiuhui YANG, Baofang ZHANG, Xincheng ZHANG, and many more, for the experience.

My secondment at the SpaceTimeLab, UCL in 2022-2023 brought me to the field of GIS and GeoAI, which I always wanted to get involved in the past. I would love to thank Prof. Tao CHENG for her kindness and generosity, who provided me with abundant guidance, resources, and inspirations during my stay. I appreciate the group meetings and personal chats with Mustafa CAN OZKAN, Dr. Huanfa CHEN, Xiao FU, Xiaowei GAO, Dr. James HAWORTH, Yihuai HONG, Natchapon JONGWIRIYANURAK, Zhengxiang SHI, Zhihui SONG, Meihui WANG, Xinglei WANG, Zichao ZENG, Xianghui ZHANG, and Fangzhou ZHOU. I also greatly value the opportunities to attend courses offered by people from different institutions including CASA and Alan Turing Institute, and to communicate with prestigious researchers Prof. Elsa ARCAUTE, Prof. Michael BATTY, Dr. Josep GRAU-BOVÉ, Dr. Tian LAN, Dr. Pakhee KUMAR, Dr. Stephen LAW, and Dr. Chen ZHONG, and my peer researchers Jian FENG, Huajie GUO, Deming HU, Yuhan JI, Chenyang LI, Xiuzheng LI, Yiming LIU, Mengdi MAO, Yichang SUN, Yu SHEN, Lei SONG, Maoran SUN, Yushen SUN, Yu WAN, Yahui WANG, Yikang WANG, Xiaoyue XING, Xiaoduo (Amy) XU, Chi ZHANG, and Wenlan ZHANG. I am glad to have the chance to stand out of the box and see how other relevant fields work.

My upcoming Postdoc position at Wageningen University and Research is aligned with my current PhD thesis. I would love to thank Prof. Art DEWULF, Prof. Ricardo DA SILVA TORRES, Prof. Anna FENSEL, and Dr. Tamara METZE for their trust, and Jasper JANSEN, Maarit JUNNIKKALA, and Lennart VERHOEVEN for their support.

I am also honored to attend many conferences and symposiums, either on site or online, where I got to know leading researchers from different disciplines, among which I would love to name a few: Prof. Mario SANTANA QUINTERO from Carleton University, Prof. Leiqing XU from Tongji University, Prof. Yuan LI from Xiamen University, Dr. Jyoti HOSAGRAHAR from World Heritage Centre, Ms. Daniela PALACIOS from DLR, Prof. Jie HE from Harbin Institute of Technology, Shenzhen, Dr. Xinyuan WANG and Dr. Fulong CHEN from HIST, Prof. Ming CAO from Groningen University, Prof. Xiaoming FU from Göttingen University, Dr. Linchuan YANG from SWJTU, Ms. Yan HE from CHCD, ICOMOS China, Dr. Fan ZHANG from the Chinese

University of Hong Kong, Dr. Rui ZHU from University of Bristol, Dr. Yuhao KANG from University of South Carolina, and Prof. Yang YUE from Shenzhen University.

Thank Wen-Yu CHEN, Xingyu FANG, and Chunyu JIN for trusting and allowing me to supervise their Master's theses. They all did great works and inspired me a lot.

I feel lucky to be sitting in the PhD Room 01.West.250 with most of my peers from the AE+T Department, where we share the same joy and help each other with the same PhD-related troubles. I was warmly welcomed and relentlessly helped during the first days by seniors Tiantian DU, Xiao GUO, and Dadi ZHANG, and by the department secretary Bo SONG. I enjoy the days spent together with colleagues and friends from AE+T Department and the A+BE Faculty, including but not limited to: Istiaque AHMED, Miktha Farid ALKADRI, Felipe Bucci ANCAPI, Fatemeh Hedieh ARFA, Tatiana ARMIJOS MOYA, Yuting BAO, Pedro BARRA LUEGMAYER, Prateek BHUSTALI, Tess BLOM, Simone BOMBACI, Ju BU, Stefano CALZATI, Cheng CHANG, Enshan CHEN, Yu CHEN, Yuan CHEN, Yu-Chou CHIANG, Adriana CIARDIELLO, Tianchen DAI, Wei DAI, Er (Erica) DING, Lu DING, Shenglan DU, Yizhao DU, Tamara EGGER, Berk EKICI, Ahmed FELIMBAN, Nima FOROUZANDEH, Yingying GAN, Longmiao GAO, Weixiao GAO, Zoheir HAGHIGHI, Amneh HAMIDA, Hamza HAMIDA, Mohammad HAMIDA, Yu HUAN, Jin HUANG, Nail IBRAHIMLI, Amin JALILZADEH, Lin JIA, Zhuoran JIA, Chi JIN, Ameya KAMAT, Pooyan KAZEMISANGEDEHI, Burcu KÖKEN, Aga KUŚ, Junkai LAN, Baiyi LI, Bo LI, Qingxiang LI, Yinghan LI, Yu LI, Jinsen LIAN, Yang LIANG, Weibin LIN, Mei LIU, Zhengxuan LIU, Chujie LU, Peijun LU, Jiefang MA, Federica MARULO, Meng MENG, Ziead METWALLY, Fillipo MOLAIONI, Fatemeh MOSTAFAVI, Phan Anh NGUYEN, Marco ORTIZ, Ran PAN, Wang PAN, Mauro PARRAVICINI, Alessio PELAGALLI, Dejian PENG, Yuyang PENG, Marije PEUTE, Xiaogeng REN, Federica ROMAGNOLI, Yan SONG, Cynthia SOUAID, Dong SU, Soma SUGANO, Diwen TAN, Tan TAN, Yinhua TAO, Nick TEN CAAT, Lia TRAMONTINI, Tanya TSUI, Prateek WAHI, Biyue WANG, Shiyu WANG, Tong WANG, Tongjin WANG, Wei WANG, Ziao WANG, Jun WEN, Bei WU, Di WU, Yanming WU, Lei XIA, Tian XIA, Juan YAN, Yaqi YAN, Ding YANG, Cehao YU, Cinco YU, Yingwen YU, Gong ZHANG, Haoxiang ZHANG, Shuyu ZHANG, Xiaoxia ZHANG, Yifei ZHAO, Kaiyi ZHU, Penglin ZHU, and many more, to have the academic or leisure activities inside or outside the faculty.

I am also lucky to have met new friends here and there outside of the faculty, and have kept close contact with many old friends and mentors during my PhD journey. From all these people, I would like to take the chance to name a few: Qian BAI, Yunhao BAI, Hantong BAO, Shufan (Fandy) CAI, Wenyan CAI, Zhaochong CAI, Weiyi CAO, Zijian CAO, Shuoyu CHANG, Guanyu CHEN, Hongxi CHEN, Jie CHEN, Nan CHEN, Shu CHEN, Shuangyun CHEN, Ting CHEN, Xuejiao CHEN, Yang CHEN, Yao CHEN, Yufei CHENG, Yanggang DAI, Huishu DENG, Shaobin DENG, Ziqi DENG, Jingliang DU, Qiaochu FAN, Rui FENG, Yuan FENG, Zhenduo FENG, Jie GAO, Junyi GAO, Xiang GAO, Zhong GAO, Ha'en GE'AI, Yu GENG, Li GONG, Xiaolan GONG, Yifan GUO, Haoqing HAN, Qi HAO, Tianhao HE, Wenxiao HOU, Zhe HOU, Zhirong HOU, Rui HU, Xiyang HU, Heye HUANG, Jingyang HUANG, Siqi HUANG, Yetong HUANG, Lu JI, Rujun JIA, Tianlong JIA, Yanwei JIA, Kun JIANG, Chaxuan JIN, Yang JIN, Jiaqi JING, Xueting KANG, Bowen LEI, Bowen LI, Chenyu LI, Hanming LI, Haopeng LI, Jingyu LI, Mingjun LI, Mingxi LI, Mingxin LI, Qiumeng LI, Tianying LI, Weitao LI, Xia LI, Xin LI, Xing LI, Xinyu LI, Yanliang

LI, Yifei LI, Yue’er LI, Yufei LI, Yun LI, Zeyu LI, Zongchen LI, Xiaoxu LIANG, Xuanshuo LIANG, Yang LIANG, Yumin LIN, Chenxi LIU, Gan LIU, Hanqing LIU, Huajian LIU, Lan LIU, Lijuan LIU, Mengyao LIU, Qingzhou LIU, Siyuan LIU, Xiaoyu LIU, Xuanyu LIU, Yu LIU, Xinghao LOU, Dongxu LU, Renqian LUO, Ding LV, Mangting LV, Hongtao MA, Wenting MA, Chun Bon (Calvin) MAN, Suping MENG, Yihang PAN, Lingbo PANG, Chen PENG, Youfang PENG, Ying QI, Li QIAN, Bingjia QIN, Yujia REN, Difei SHAN, Xu SHAN, Xuan SHAO, Canxi SHEN, Lunhao SHEN, Yiqing SHEN, Yiyuan SHI, Wanyu SHUAI, Wenfei SONG, Yu SONG, Yutao SONG, Tianyu SU, Xia SU, Nanbin SUN, Ping SUN, Weijia SUN, Yushen SUN, Maosheng TANG, Jiarui TIAN, Li TIAN, Yixiang TU, Haoran WANG, Jingyang WANG, Ke WANG, Likai WANG, Mingyang WANG, Ruocha WANG, Siyu WANG, Shiqi WANG, Xiaoping WANG, Xuefei WANG, Yaohua WANG, Yicong WANG, Yunyun WANG, Zeyu WANG, Zhenwu WANG, Junhan WEN, Han WU, Quhang WU, Jinlai XIANG, Kechao XIANG, Yuting XIAO, Jun XIE, Shiduo XIN, Xinyu XIU, Jing XU, Jinxi XU, Junzhong XU, Yixin YAN, Haochen YANG, Jie YANG, Junhan YANG, Kejing YANG, Yufeng YANG, Zi YANG, Yu YAO, Zi YE, Tianyinan YIN, Haoxuan YU, Jianwen YU, Qiang YU, Wenbo YU, Sen YUAN, Shuai YUAN, Zibo YUAN, Sijie ZENG, Yuanchen ZENG, Bosen ZHANG, Chengzhang ZHANG, Congyi ZHANG, Haotian ZHANG, Jian ZHANG, Pan ZHANG, Qingrui ZHANG, Wen ZHANG, Wenzhao ZHANG, Yimeng ZHANG, Yingxin ZHANG, Yiyi ZHANG, Yue ZHANG, Yuqi ZHANG, Yuxiang ZHANG, Zihao ZHANG, Zhicheng ZHANG, Rong ZHAO, Shipan ZHAO, Tongzhou ZHAO, Yuguang ZHAO, Xinying ZHENG, Yanggu ZHENG, Yi ZHENG, Yu ZHENG, Yichen ZHONG, Changji ZHOU, Linda ZHOU, Yufan ZHOU, Xinhan ZUO, and many more. They all have made positive and significant contributions to my research or my mental health in some way.

My special thanks to Huishu DENG, Longmiao GAO, Zhuoran JIA, Jiaqi JING, Xiuzheng LI, Weitao LI, Jinsen LIAN, Xu SHAN, Tian XIA, Shiduo XIN, Mingyang WANG, and Zi YE, for the company during the otherwise lonesome days, and their chats with me over the work, the life, and the world. And most specially, to my “partners in crime”, Zibo LIU, Ziao WANG, and Jie YANG, for all the relaxing and busy days that we share, all the ups and downs that we experience, and all the wild tracks that we step on.

Thank TU Delft X, Pathé, Museumkaart, Amazon Books, FanC, West End London, MET Opera New York, Bilibili, Alfred’s Piano, and Nintendo Switch for filling the rainy days.

Last but not least, I would love to thank all my beloved and caring family members who mainly live in Lanzhou and Wuwei, Gansu, China. My cousin Ting LU, and my nephew Zilu HE have been hearing me talk about my research and asking me valuable questions, albeit the latter is currently only ten years old. My parents, Shengfu BAI and Linhong ZHAO, are the ones who will always love me, encourage me, and support me, no matter what I do, without any word of complaints. I love you, and I am lucky and honoured to be your boy.

Thank you all - Dank u wel - Obrigado - Moteshakeram - 感激不尽!

Nan BAI
Lanzhou and Shanghai, August 2023

Contents

List of Tables

List of Figures

List of Acronyms

AI Artificial Intelligence
AMS Data of Amsterdam, the Netherlands
ANOVA Analysis of Variance
API Application Programming Interface
BERT Bidirectional Encoder Representations from Transformers
BoE Bag of Embeddings
CNN Convolutional Neural Network
CPU Central Processing Unit
CV Computer Vision
DBSCAN Density-based Spatial Clustering of Applications with Noise
DL Deep Learning
GAT Graph Attention Network
GCN Graph Convolution Network
GIS Geographic Information System/Science
GNB Gaussian Naive Bayes
GNN Graph Neural Network
GPU Graphical Processing Unit
GRU Gated Recurrent Unit
GSA Graph Sample and Aggregate (GraphSAGE) Models
HA Heritage Attributes
HGSA Heterogeneous GraphSAGE Network
HGT Heterogeneous Graph Transformers
HREs Heritage-Related Events
HUL Historic Urban Landscape
HV Heritage Values
ICOMOS International Council on Monuments and Sites
IoU Intersection over Union
IUCN International Union for Conservation of Nature

KNN K-Nearest Neighbour

LDA Latent Dirichlet Allocation

LS Label Smoothing

MDS Multi-Dimensional Scaling

ML Machine Learning

MLP Multi-layer Perceptron

MML Multi-modal Machine Learning

NLP Natural Language Processing

NMF Non-Negative Matrix Factorization

OUV Outstanding Universal Value

PCA Principal Component Analysis

POI Point of Interest

PyG PyTorch Geometric

QAP Quadratic Assignment Procedure

RDC Random Classifier

RF Random Forest

RNN Recurrent Neural Network

SNA Social Network Analysis

SOUV Statements of OUV

SUZ Data of Suzhou, China

SVM Support Vector Machine

UGC User-Generated Content

ULMFiT Universal Language Model Fine-tuning

UNESCO The United Nations Educational, Scientific and Cultural Organization

VEN Data of Venice, Italy

VEN-XL The extra-large version of Venice data

WH World Heritage

WHL World Heritage List

WoS Web of Science

Summary

This dissertation entitles "Sensing the Cultural Significance with AI for Social Inclusion: A Computational Spatiotemporal Network-based Framework of Heritage Knowledge Documentation using User-Generated Content".

Core Premises
Social Inclusion has been growing as a goal in heritage management in the past decade. Whereas the 2011 UNESCO Recommendation on the Historic Urban Landscape (HUL) called for tools of knowledge documentation, social media already function as a resourceful platform for online communities to actively involve themselves in heritage-related discussions. Such discussions happen both in the baseline scenarios when people calmly share their experience of the cities they live in or travel to, and in the activated scenarios when radical events trigger their emotions. Analyses have been recently performed on User-Generated Content (UGC) from social media platforms to actively collect opinions of the [online] public and to map cultural significance conveyed by various stakeholders in urban environments. Machine learning, deep learning, or more generally, Artificial Intelligence (AI) is shown to be indispensable for organizing, processing, and analysing the unstructured multi-modal massive data from social media efficiently and systematically.

Research Aim
The aim of this research is to explore the use of AI in a methodological framework to include the contribution of a larger and more diverse group of participants and facilitate the knowledge documentation of cultural significance in cities with user-generated social media data. To reach the aim, five parts are used to elaborate the exploration process of the proposed methodological framework. PART A builds up a theoretical BASIS for the dissertation with a general introduction in Chapter 1 and a systematic literature review in Chapter 2. PART B develops the MODELLING process using AI to construct a machine replica of the authoritative view on cultural significance through UNESCO World Heritage Statements of Outstanding Universal Value in Chapter 3. Then the dissertation goes into two directions in PART C and PART D, respectively exploring the two variants of the methodological framework for knowledge documentation. PART C focuses on the CONTEXT of collective opinions in the everyday baseline scenarios with the data collection workflow in Chapter 4 and the mapping process of perceived cultural significance in Chapter 5. PART D focuses on the DYNAMICS of the discussions triggered by radical events by inspecting the spatiotemporal patterns of the content (especially emotions and proposed actions) and intensity of posting behaviours in Chapter 6. PART E concludes the dissertation

in Chapter 7 and discusses on how the proposed methodological framework and the empirical findings can contribute to social INCLUSION in heritage management.

Methods Applied
It is an interdisciplinary study integrating the methods and knowledge from the broad fields of heritage studies, computer science, social sciences, network science, and spatial analysis. State-of-the-art methods from the AI communities were applied, nurtured, and tested within the research. The whole bundle of AI-based methods include ideas and models from Natural Language Processing, Computer Vision, Graph Neural Networks, Semi-Supervised Classification, Multi-modal Machine Learning, Topic Modelling, etc. Datasets of the UGC on social media platforms are collected and structured as networks/graphs, representing the spatial, temporal, and social connections among the posts. AI-based models are employed to help analyse the massive information content to derive the knowledge concerning cultural significance perceived and expressed by the online community in case study cities Venice, Paris, Suzhou, Amsterdam, and Rome. The results are further analysed and visualized with [spatial] statistics and mapping techniques as knowledge documentation.

Main Findings
Cultural significance perceived and conveyed by the online community to the cities was found to be strongly embedded in their spatiotemporal and social contexts. Tobler's First Law of Geography was still shown as relevant for urban heritage on social media. In the baseline scenarios, cultural significance has been perceived and expressed by social media users at a broad variety of locations in cities with urban areas inscribed in the UNESCO World Heritage List, other than the conventional tourist destinations. In the activated scenarios, the triggered discussions reached places far beyond geographical boundaries during the event, forming a temporary global heritage community, where people mainly shared information about the event, expressed their emotions, and proposed or broadcast actions on how to help. Therefore, the AI-based methodological framework is shown to be able to collect information and map the knowledge of the community about the cultural significance of the cities, fulfilling the expectation and requirement of HUL, useful and informative for future socially inclusive heritage management processes.

Limitations and Drawbacks
The use of AI and social media data is never the "eternal solution" for mapping cultural significance, which could potentially create new challenges and opportunities compared to what it managed to solve. The AI models are always biased based on the available data and training methods, which can fall into sub-optimal solutions. Besides data privacy and ethical issues that need to be considered, the use of specific social media platforms as the data source implies that the people being included have been pre-defined, which may also have strong limitations to getting a comprehensive picture that may eventually result in systematic biases. The AI-based approach, therefore, needs to be accompanied by other sorts of qualitative and quantitative studies involving broader stakeholders. Nevertheless, this research makes the first steps to bridging the gaps towards collaborations between AI and heritage experts.

Samenvatting

Deze dissertatie is getiteld “Culturele Betekenis Detecteren met AI voor Sociale Inclusie: Een computationeel, spatiotemporeel, op netwerken gebaseerd kader voor het documenteren van erfgoedkennis met door gebruikers genenereerde inhoud”.

Belangrijke Vooronderstellingen

Sociale inclusie is het voorbije decennium als doel in erfgoedbeheer gegroeid. Terwijl de UNESCO 2011 Recommendation on the Historic Urban Landscape (HUL) oproept tot tools voor kennisdocumentatie, werken sociale media al als vindingrijk podiums voor online gemeenschappen om zichzelf actief te betrekken in erfgoedgerelateerde discussies. Zulke discussies vinden zowel plaats in de basis-scenario’s waar mensen rustig hun ervaringen delen over de steden waarin ze wonen of waarnaartoe ze reizen, als in de geactiveerde scenario’s waar radicale gebeurtenissen hun emoties triggeren. Onlangs zijn er analyses uitgevoerd op door gebruikers gegenereerde inhoud (User-Generated Content, UGC) van sociale media om meningen van het online publiek te verzamelen en de culturele betekenis te karteren die door diverse belanghebbenden in stedelijke omgevingen wordt uitgedragen. Machine learning, deep learning, of in het algemeen Kunstmatige Intelligentie (Artificial Intelligence, AI) blijkt onmisbaar te zijn voor het efficiënt en systematisch organiseren, verwerken en analyseren van de ongestructureerde multimodale massale gegevens uit sociale media.

Doel van het Onderzoek

Het doel van dit onderzoek is het gebruik van AI in een methodologisch kader te ontdekken om de bijdrage van grotere en diversere deelnemers op te nemen en de kennisdocumentatie over culturele betekenis in steden met UGC te faciliteren. Om het doel te bereiken, worden vijf delen gebruikt om de exploratie uit te werken. DEEL A bouwt een theoretische BASIS op voor de dissertatie met een algemene inleiding in Hoofdstuk 1 en een systematisch literatuuronderzoek in Hoofdstuk 2. DEEL B ontwikkelt het MODELLEREN met AI om een machinale replica van de autoritaire visie op de culturele betekenis te construeren via Werelderfgoed Uitzonderlijke Universele Waarde in Hoofdstuk 3. Vervolgens gaat de dissertatie naar twee richtingen in DEEL C en DEEL D, waarin twee varianten van het kader voor kennisdocumentatie worden verkend. DEEL C richt zich op de CONTEXT van collectieve meningen in de alledaagse basis-scenario’s met de workflow voor gegevensverzameling in Hoofdstuk 4 en het karteren van de waargenomen culturele betekenis in Hoofdstuk 5. DEEL D gaat over de DYNAMIEK van de discussies veroorzaakt door radicale gebeurtenissen en inspecteert de spatiotemporele patronen van de inhoud (emoties en acties) en de intensiteit van het postgedrag in Hoofdstuk 6. DEEL E sluit de dissertatie af in hoofdstuk 7 en bespreekt hoe het voorgestelde methodologische kader en de empirische bevindingen kunnen bijdragen tot sociale INCLUSIE in erfgoedbeheer.

Toegepaste Methoden
Het is een interdisciplinair onderzoek dat methoden en kennis integreert uit de brede domeinen van erfgoedstudies, computer-, sociaal-, en netwerk-wetenschappen, en ruimtelijke analyse. Geavanceerde AI methoden werden toegepast, ontwikkeld en getest. De hele bundel AI methoden omvat ideeën en modellen uit Natuurlijke Taalverwerking, Computervisie, Neurale Grafische Netwerken, Semi-Supervised Classificatie, Multi-modaal Machine Learning, Topic Modelling, enz. Datasets met UGC uit sociale media worden verzameld en als netwerken gestructureerd, die de ruimtelijke, temporele en sociale connecties tussen de posts representeren. AI modellen worden gebruikt om de enorme informatie-inhoud te helpen analyseren en daaruit de kennis af te leiden over de culturele betekenis die wordt waargenomen en uitgedrukt door de online gemeenschap in Venetië, Parijs, Suzhou, Amsterdam en Rome. De resultaten worden verder geanalyseerd en gevisualiseerd met [ruimtelijke] statistieken en kaarttechnieken als kennisdocumentatie.

Voornaamste Bevindingen
De door de online gemeenschap waargenomen culturele betekenis bleek sterk ingebed te zijn in hun spatio-temporele en sociale context. De Eerste Geografische Wet van Tobler bleek nog steeds relevant te zijn voor stedelijk erfgoed op sociale media. In de basis-scenario's werd culturele betekenis waargenomen en uitgedrukt door sociale mediagebruikers op diverse locaties in steden gedeeltelijk opgenomen in de Werelderfgoedlijst, anders dan de conventionele toeristische bestemmingen. In de geactiveerde scenario's bereikten de getriggerde discussies tijdens de gebeurtenis plaatsen ver buiten de geografische grenzen en vormden zo een tijdelijke wereldwijde erfgoedgemeenschap, waar mensen voornamelijk informatie deelden, emoties uitten en acties voorstelden of uitzonden over hoe ze konden helpen. Daarom bleek het AI-gebaseerde methodologische kader geschikt te zijn om informatie te verzamelen en de kennis van gemeenschappen over de culturele betekenis te documenteren. Dit voldoet aan de verwachtingen en vereisten van de HUL en is nuttig en informatief voor toekomstige sociaal inclusieve erfgoedbeheerprocessen.

Beperkingen en Nadelen
Het gebruik van AI en sociale media is nooit de "definitieve oplossing" voor het karteren van culturele betekenis, wat mogelijk nieuwe uitdagingen en kansen kan creëren in vergelijking met wat het heeft kunnen oplossen. De AI-modellen zijn altijd bevooroordeeld op basis van de beschikbare gegevens en trainingsmethoden, wat tot suboptimale oplossingen zou kunnen leiden. Naast de dataprivacy en ethische kwesties waarmee rekening moet worden gehouden, impliceert het gebruik van specifieke sociale media als gegevensbron dat de mensen die worden opgenomen vooraf zijn gedefinieerd, wat ook sterke beperkingen kan hebben voor het verkrijgen van een alomvattend beeld, wat uiteindelijk in systematische biases zou kunnen resulteren. De AI-gebaseerde aanpak moet daarom worden gecombineerd met andere soorten kwalitatieve en kwantitatieve studies waarbij belanghebbenden in bredere zin worden betrokken. Desalniettemin zet dit onderzoek de eerste stappen om de kloof tussen AI en erfgoedexperts te overbruggen.

摘要总结

本论文题目可译为“旨在社会包容的文化意义感知——基于人工智能的方法框架”。它以社交媒体上的用户生成内容为数据源，以计算建模、时空分析与网络科学为方法核心，提出了一套利用人工智能对遗产的文化意义进行知识记录的方法框架。

核心前提

在过去十年中，社会包容逐渐成为遗产管理过程中的一个重要目标。联合国教科文组织于2011年通过了《关于城市历史景观的建议书》，其中特别提到应当开发一些工具手段，用于对遗产特征的知识进行记录与绘制，但事实上社交媒体已经作为这样的一个工具发挥着重要且丰富的作用，网络用户群体利用社交媒体主动参与遗产相关问题的讨论，积极地表达着他们对于身边文化与自然遗产的关切。在特定的“应激”情境下，也即当一些重大事件（例如巴黎圣母院的火灾）触发公众情绪时，这类讨论会频繁出现；而在那些更普遍的“基线”情境下，也即当公众对于自己生活或旅行来到的城市进行日常分享时，类似的关于遗产文化意义的讨论也随处可见。在遗产研究领域，近期已经有一系列研究在利用社交媒体平台上的用户生成内容对城市的文化意义进行分析，它们积极地收集网络公众对于遗产的看法，试图将不同的利益攸关者对城市文化意义的理解绘制与纪录下来。为了更加系统且高效地整合、处理与分析社交媒体海量的非结构化数据，以机器学习与深度学习为代表的人工智能也逐渐变得不可或缺。

研究目的

本研究旨在提出并探索一套方法框架，利用社交媒体上的用户生成数据，以人工智能为工具对城市的文化意义进行知识记录与绘制，以此将更丰富、更多元的参与者的贡献纳入到遗产管理过程当中。为了达成这一目标，本论文共利用五个模块对方法框架的提出与探索过程进行拆解。模块A构建了全文的理论基础，由第1章的前言导论与第2章的系统文献综述组成。模块B为后续社交媒体分析进行准备，在第3章中以世界遗产《突出的普遍价值声明》为数据集训练了一组人工智能分类模型，令其能够从专家的视角出发理解遗产文化意义的概念。论文接下来被拆分为两个平行的模块C和D，分别讨论本文提出的知识记录方法框架在不同情境下的两个分支。模块C聚焦于日常基线情境，重点放在公众对于遗产讨论时的背景语境上，第4章提供了一套开源的用于数据收集的流程，第5章则对公众认知到的文化意义的整理绘制过程进行了讨论。模块D则聚焦于受到大事件诱导刺激的应激情境，第6章探索了此情境下公众讨论的内容（特别是情绪与行动）与其发言行为的强度中所体现的时空分布规律。最后，模块E总结了全文的贡献与不足，在第7章中落回到遗产管理过程当中的社会包容概念，重点讨论本文所提出的方法框架与得到的经验结论的可能应用场景。

研究方法

作为一个多视角、跨学科的研究，本文集合了遗产研究、计算机科学、社会科学、网络科学与空间分析等学科的不同知识方法。特别地，人工智能领域的一系列前沿的模型方法在本研究中得到了应用、发展以及测试。本文涉及多个人工智能子领域的理念与模型：自然语言处理、计算机视觉、图神经网络、半监督分类、多模态机器学习、主题模型，等等。社交媒体上的用户生成内容作为数据集被收集下来，并通过网络/图的结构对数据之间的时间、空间以及社会关联进行表征。在威尼斯、巴黎、苏州、阿姆斯特丹与罗马等研究案例中，人工智能模型被用于分析海量的信息内容，从中抽取关于公众所认知到与表达出的城市遗产文化意义的相关知识，其结果进一步通过一系列空间统计学与制图方法得到分析与展示，由此成为一套完整的知识记录工具手段。

主要发现

本研究揭示出在城市当中，公众认知与接受到的遗产文化意义与其时空与社会语境有着紧密的联系。特别地，沃尔多·托布勒所提出的“地理学第一定律”在讨论社交媒体上的城市遗产认知与表达时依然适用。在基线情境下，对于被完全或部分纳入《世界遗产名录》的城市来说，社交媒体用户所认知到与表达出的遗产文化意义往往分布在广泛且多样的城市空间中，而不仅仅局限于传统意义上的旅游景点。而在应激情境下，因为大事件的发生，关于遗产地的激烈讨论拓展到了远超地理边界限制的诸多地区，由此形成了一个临时的全球性遗产社群（共同体），公众在其中积极地传播事件相关信息，表达自己的情绪与观点，并且提议或宣扬各类救助行动。上述发现证明，本文提出的基于人工智能的方法框架可以用于从公众的角度出发，对城市遗产的文化意义进行知识记录与绘制，这是与城市历史景观（也称为历史性城市景观）的预期与要求相吻合的，有望在未来被用于助力更具社会包容性的遗产管理过程。

研究缺陷

利用人工智能与社交媒体数据进行研究并不是对文化意义进行记录与绘制的唯一“最终答案”，相较于它解决的问题，它可能创造出一些新的机遇与挑战。人工智能模型总会因为提供的数据与训练方法的特性而存在偏见，有时也会被困于次优解当中。除去数据隐私与伦理方面的讨论，以特定的社交媒体平台作为数据源也意味着研究能够涵盖的人群已经被预先进行了限制，这对于全面理解社会认知中的遗产文化意义可能是一个障碍，并有可能最终导向具有一定系统误差的结论。因此，这类基于人工智能的方法应当与其他质性与量化研究相结合，让更广泛的利益攸关方参与进来。尽管如此，本研究依然为城市遗产研究架起了一座新的桥梁，希望在不远的将来，人工智能与遗产专家、城市规划师能够有更好的合作。

PART A

The Basics

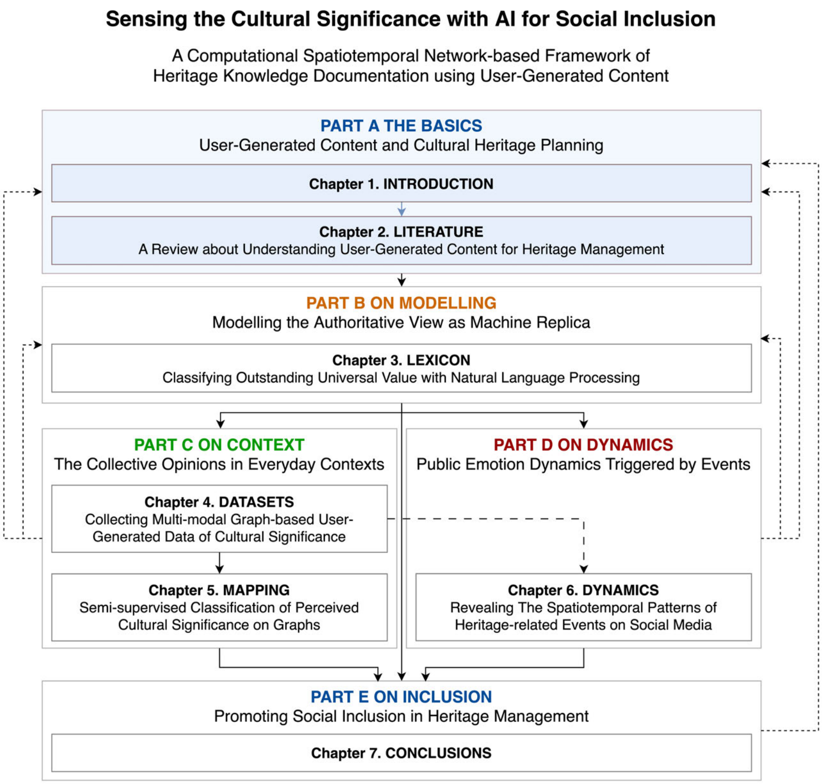
User-Generated Content and Cultural Heritage Planning

This part of dissertation builds up the theoretical and methodological basis of the research. It defines the interdisciplinary scope of the dissertation, introduces the concepts of baseline scenarios and activated scenarios of user-generated social media posting behaviour concerning urban cultural heritage, and sets up the research aim and questions to be answered by this dissertation. A systematic literature review gives an overview of how the content, structure, and context of user-generated content (UGC) are understood computationally in the broad field of heritage management.

Two chapters are included in this part:

Chapter 1 Introduction.

Chapter 2 Literature - A Systematic Review about Understand UGC for Heritage Management.


Sensing the Cultural Significance with AI for Social Inclusion
A Computational Spatiotemporal Network-based Framework of
Heritage Knowledge Documentation using User-Generated Content
PART A THE BASICS
User-Generated Content and Cultural Heritage Planning
Chapter 1. INTRODUCTION
Chapter 2. LITERATURE
A Review about Understanding User-Generated Content for Heritage Management
PART B ON MODELLING
Modelling the Authoritative View as Machine Replica
Chapter 3. LEXICON
Classifying Outstanding Universal Value with Natural Language Processing
PART C ON CONTEXT
The Collective Opinions in Everyday Contexts
Chapter 4. DATASETS
Collecting Multi-modal Graph-based User-
Generated Data of Cultural Significance
Chapter 5. MAPPING
Semi-supervised Classification of Perceived
Cultural Significance on Graphs
PART D ON DYNAMICS
Public Emotion Dynamics Triggered by Events
Chapter 6. DYNAMICS
Revealing The Spatiotemporal Patterns of
Heritage-related Events on Social Media
PART E ON INCLUSION
Promoting Social Inclusion in Heritage Management
Chapter 7. CONCLUSIONS

1 Introduction

Parts of this chapter have been published in Bai et al. (2021b) and Bai et al. (2023b).
Bai N, Nourian P, Pereira Roders A. (2021b). Global Citizens and World Heritage: Social Inclusion of Online Communities in Heritage Planning. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021. p. 23–30.
Bai N, Ducci M, Mirzikashvili R, Nourian P, Pereira Roders, A. (2023b). Mapping Urban Heritage Images with Social Media Data and Artificial Intelligence, A Case Study in Testaccio, Rome. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023. p. 139–146.

1.1 Background

1.1.1 Social Inclusion and Knowledge Documentation

Social inclusion and public participation have been extensively discussed in heritage management in the last decades, both in research and practice. Heritage is diverse in category (natural and cultural, tangible and intangible, etc.), and also in nomination level, ranging from international lists such as UNESCO World Heritage List (WHL), to national, regional, and local levels. The different categories and levels of heritage often overlap in their **attributes** (what to conserve) and **values** (why to conserve) with cultural significance (detailed definition will be provided in Section 1.2.1). Aside from the official listing, heritage often has overlaying meanings and cultural significance conveyed by local citizens, tourists, and experts (Waterton et al., 2006; Australia ICOMOS, 2013; Pereira Roders, 2019). However, the aim of social inclusion is harder to achieve when only societal representatives (public sectors and experts) decide on heritage management. According to Taylor and Gibson (2017), simply providing digitized heritage to stakeholders will not increase the perception of social inclusion. As defined by social psychologists, **Social Inclusion** is

> “the degree to which an individual perceives that the group provides him or her with a sense of belonging and authenticity",

with belonging and authenticity as two major dimensions (Jansen et al., 2014).

This also applies to heritage management and planning. The Recommendation on the Historic Urban Landscape (HUL) adopted by UNESCO in 2011 promotes the participation of a broader variety of stakeholders in heritage management, including actors from local to international, private to public, and experts to communities. The Recommendation also calls for developing "knowledge and planning tools" that enable knowledge documentation (UNESCO, 2011; Bandarin and Van Oers, 2012):

> "**Knowledge and planning tools** should help protect the integrity and authenticity of the attributes of urban heritage. They should also allow for the recognition of cultural significance and diversity, and provide for the monitoring and management of change to improve the quality of life and of urban space. These tools would include **documentation and mapping** of cultural and natural characteristics. Heritage, social and environmental impact assessments should be used to support and facilitate decision-making processes within a framework of sustainable development."

Meanwhile, social media is foreseen to strongly contribute as such a knowledge documentation tool for better social inclusion. With the help of social media, everyone can join the heritage management processes by expressing their opinions and emotions publicly and instantly, even if not directly involved in decision-making. Social media provides the chance to rationally increase and develop the public's input as systematic knowledge into heritage management (Olsson, 2008).

Right after the fire in Notre-Dame de Paris on the 15th of April, 2019, sorrow and shock spread over social media worldwide, growing rapidly in platforms like Facebook, Twitter, Instagram, WeChat, and TripAdvisor. Many posts also pointed out which sector was to blame, and discussed whether or not should Notre-Dame be repaired, restored, or redesigned. This conversation has continued all the way to the recent days - two years after the fire and one year since the Covid-19 pandemic started to spread and paused normal social life. The main emotion to be found on social media comes back to normal states, and people start to integrate Notre-Dame again in their posts sharing their daily lives, expressing how "lovely" Notre-Dame still is though it is still "healing" and "ongoing to rebuild"[1]. In some similar cases like the fire in Notre-Dame, when radical events happened, communities worldwide would use social media platforms as a tool for actively getting involved and therefore included in heritage management. They temporally formulate a group of concerned global citizens and [re]act actively. Their emotions, opinions, and reactions, are also calling the attention of heritage managers and experts to make more responsible planning decisions. Emotions and opinions can be spread through social media in a viral way when such drastic events happen, sometimes even forming a secondary crisis for the heritage managers (Bruns et al., 2012; Schroeder et al., 2013; Lipizzi et al., 2015; Adamic et al., 2016; Zhai et al., 2020).

Social media also function as platforms for expressing collective ideas on the idea of people-centred heritage in an everyday scenario (Ginzarly et al., 2019). By sharing

[1] The terms are induced from Flickr, Instagram, and TripAdvisor posts.

pictures, making comments, leaving tags, and giving rates to places listed as heritage, people are deliberately or unintentionally passing the understanding and perception of the values the places convey to them. By actively expressing the immediate observations on heritage, stakeholders including locals and tourists are involved to co-create the heritage experiences, which in turn could inform heritage management (van Dijck, 2011; Munar, 2012), thus becoming positive “prosumers" in a much more democratic designing procedure (Fischer, 2009; Monti et al., 2018).

1.1.2 Baseline and Activated Scenarios

Borrowing concepts from neuroscience, the aforementioned two different scenarios can be interpreted as **baseline** (everyday) scenarios and **activated** (event-triggered) scenarios, as occurred at Notre-Dame. Both scenarios are crucial in understanding the social inclusion processes and their potential influence on heritage management (Roders and Van Oers, 2011). Figure 1.1 shows the evenly-distributed relative search interests of four major heritage properties between 2015-2020, comparing to the activated scenario caused by the fire in Notre-Dame de Paris on Google Trend search engine[2]. The vertical axes show the relative search interest while the most searched term during the shown period would be counted as 100, and the other points would be scaled accordingly. The extreme focus on Notre-Dame de Paris in April 2019 when the fire happened diminished all the other interests on a relative scale. A further example of the online communities concerning with Notre-Dame globally can also be seen in Figure 1.2. When looking at the dominant searching keyword among the five cultural heritage properties (same as in Figure 1.1), it could be observed that one year before and one year after the fire, the global search interests have been more diverse. And during the outburst of the fire, almost the whole globe focused on Notre-Dame, clearly showing that the world got more densely connected and “smaller" in the activated scenarios (Milgram, 1967; Watts and Strogatz, 1998), which can transcend the geographical and cultural boundaries.

As a more specific definition, this dissertation refers to “activated" scenario when radical events happen concerning with a heritage property, causing a peak in online discussion and search interest for a short period, while the ‘baseline" scenario refers to all other ordinary time. This distinction is shown in Figure 1.1 with the case of the large peak caused by fire in Notre-Dame and the several small peaks with Pantheon. Google Trends Engine could detect such “breakout" events as “rising searches" based on their algorithms[3]. However, such detection would not automatically promise that the outbursts would exactly match and relate to the heritage properties. Additional checks have to be paired as specific interpretations for such detected events. For example, the breakout of searches on “Pantheon" in August 2019 was due to the rework of a character with the same name in the video game “League of Legends", which is weakly relevant to the former Roman temple, though not totally

[2] https://trends.google.com/trends
[3] https://support.google.com/trends/answer/4355000

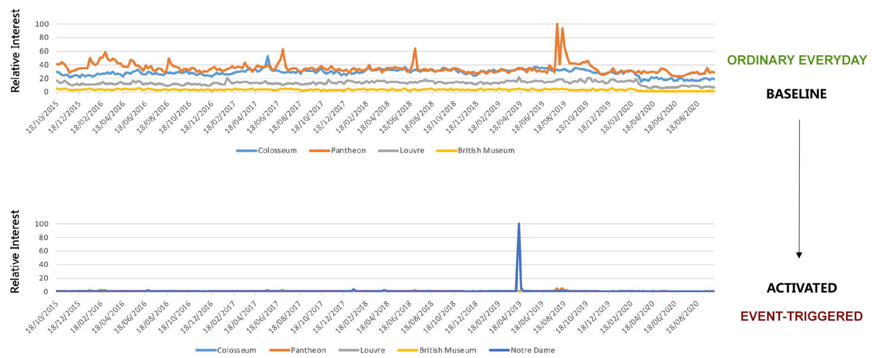


FIG. 1.1 The relative search interest of five heritage properties between 2015 to 2020 on Google Trend search engine. Adding Notre-Dame diminishes the evenly distributed relative search interests. These graphs illustrate both the everyday baseline scenarios (evenly-distributed dates along the two graphs) and activated event-triggered scenarios (the occasional peaks in both graphs).

unrelated. Furthermore, the radical events raising public attention are not necessarily negative. A similar search as Figure 1.1 conducted with five cities from April 2019 to April 2020 is demonstrated in Figure 1.3. The two peaks in Venice were respectively about the exceptional flood in November 2019 and the appearance of dolphins in the canals in March 2020. Both activated cases about the fire in Notre-Dame de Paris and the flood in Venice will be revisited in Chapter 6.

1.1.3 Relevance for Heritage Management

Both baseline and activated scenarios could be relevant to the practice of heritage management and spatial planning (Janssen et al., 2017). According to Couclelis (2005), the function of planning can be operational, managerial, and strategic, corresponding to different time orientations of the past, present, and future, respectively. Planning actions in the heritage context could have different meanings. For the baseline scenario, the planning actions can inform the official narratives towards the heritage attributes and values, which are usually decisions by both the local heritage managers and global organizations, e.g., The United Nations Educational, Scientific and Cultural Organization (UNESCO), International Council on Monuments and Sites (ICOMOS), and International Union for Conservation of Nature (IUCN), on what has to be preserved, what can be changed, and what must be erased (strategic planning including actions to adapt, prepare, shape et al.). For the activated scenario, the planning actions can refer to the official reactions towards the events and their further strategies (operational and managerial planning including actions to react, respond, mitigate, and manage). Both planning approaches correspond with the main steps (which is not necessarily sequential) in HUL (check Appendix A for the full list), i.e., step 2 “to help determine what to protect for the future” and step 4 “to integrate the cultural heritage in city development” for baseline

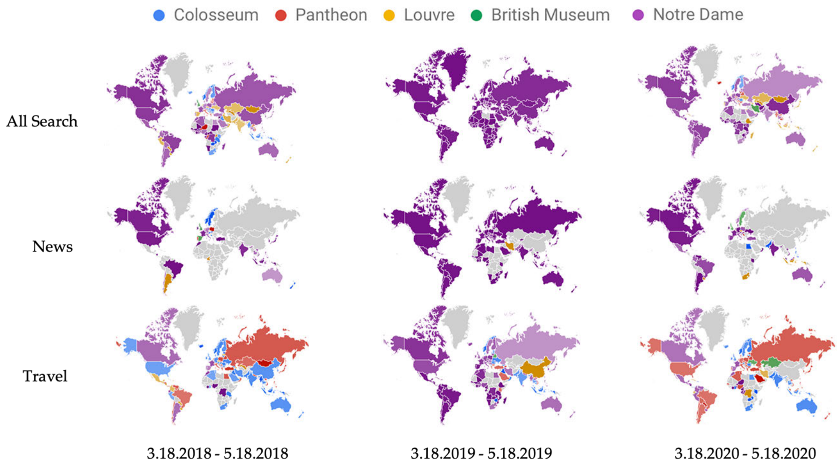


FIG. 1.2 The heatmaps showing relative search interest for countries and regions globally of five heritage properties within two months (March 18 to May 18) in 2018-2020 on Google Trend search engine under the search categories of "All Search", "News", and "Travel", respectively.

scenario, as well as step 5 "to prioritize actions for conservation and development" and step 6 "to develop mechanisms for coordination of the various activities between different actors" for activated scenario (Pereira Roders, 2019). Considering the relationship between the online public and the authorised heritage discourse in both scenarios, social inclusion could be confirmed and further enhanced.

As for the **baseline** scenario, this relationship means how well the attributes identified by the experts conveying values are reflected in daily life for ordinary people. For most listed UNESCO World Heritage (WH) properties, there is a thorough statement defining its OUV, pointing out the unique and exceptional attributes associated with certain values, satisfying one or many of the ten Criteria for Selection (UNESCO, 1972, 2008; IUCN et al., 2010). The OUV shows different aspects of exceptional values of the listed properties, which form a values system, including some of the eight values, such as social, economic, political, historic, aesthetic, scientific, age, and ecological values (Pereira Roders, 2007; Tarrafa Silva and Pereira Roders, 2012). As such, a theoretical framework is provided to describe and justify the cultural significance, as the values play a role in heritage listings and urban conservation. However, for local people or tourists visiting the property, it is not expected that any of them would read the inscription documents and know about the concept of cultural significance. By comparing the official discourse and the expressions on social media, heritage experts can investigate what are the values explicitly understood and perceived by the locals and visitors, which can improve future policy-making (van Dijck, 2011; Ginzarly et al., 2019). Both the values and attributes referenced in Statements of OUV but not broadly expressed by the public, and the ones that are popular in the public yet not listed in global, national, regional, and/or local official documents with cultural significance, are crucial information for heritage management in the constant process of cultural changes (Rochon, 1998;

FIG. 1.3 The relative search interest of five cities with urban areas inscribed in the UNESCO World Heritage List between April 2019 to 2020 on Google Trend search engine. The two main events happened in Venice causing discussion peaks are paired with corresponding news articles.

Pereira Roders, 2019). This relationship can also become a reference for future inscribing and delisting decisions of World Heritage property nominations, by taking more information from the public. In such a way, the selection, maintenance, and management of the heritage can become a dynamic evolution process, which could be more responsible and rational for the whole society as well as future generations (Jokilehto, 2006, 2008).

On the other hand, as for the **activated** scenario, this relationship mainly concerns how decision-makers and heritage managers deal with radical events properly considering the collective reaction from online communities. At the beginning phase of the epidemic spreading of event-related information, the collective emotions (usually anger, sorrow, and happiness if the event is positive) infect the network rapidly by contagious contacts and are easily out of control (Zhai et al., 2020). It is rather strategic for operators and managers to choose a proper moment, a proper way, and some proper sources to broadcast the official reaction (Easley and Kleinberg, 2010; Aggarwal, 2011; Dong et al., 2012; Pentland, 2015). How instant the reactions are, and to how much extent the public concern is reflected in the reactions, can strongly influence the reliability and credibility of the heritage managers in the local, regional, and even global communities. For the later decision-making phase concerning a new policy with the same issue, if no concerns from the previous discussions by the online communities are properly reflected and reacted to, the activation of public opinions can be called up again (Cheng et al., 2016). All such reactions and emotions associated with events could also be reported in documents such as Periodical Reporting, State of Conservation reporting, and Reactive Monitoring about threats, common to World Heritage properties.

Furthermore, as pointed out in Rochon (1998) and ICOMOS (2013), culture and cultural significance are constantly changing over time, place, and group of people. The dynamic and diversity of values render a methodological challenge, that the results from conventional methods in heritage management quickly outdated and easily partial, which happens in both baseline and activated scenarios. As such, a methodological framework is urgently needed to investigate in real-time the complex relationship of both authority-public discourses and baseline-activated scenarios under the common umbrella of "cultural significance". The results therefrom could be applied by researchers, local heritage managers, and global heritage institutions in both baseline and activated scenarios, and become an important source of knowledge documented from the public for future inclusive heritage management processes.

1.2 Problem Fields

Due to the interdisciplinary nature of this dissertation, before going deeper into the research questions, some basic concepts from the four main components of this research will be first briefly introduced, defining its problem fields and building a common ground for the discussions.

1.2.1 Cultural Significance and Heritage Management

Albeit a term frequently used in the context of UNESCO World Heritage, "cultural significance" was never mentioned together in the 1972 "Convention Concerning the Protection of the World Cultural and Natural Heritage" (UNESCO, 1972), and only referred to but not defined in the "Operational Guidelines for the Implementation of the World Heritage Convention" (UNESCO, 2008). Only in the Burra Charter for Places of Cultural Significance (Australia ICOMOS, 2013), a formal definition is given, interchangeable with "cultural heritage significance" and "cultural heritage value":

> "**Cultural significance** means aesthetic, historic, scientific, social or spiritual value for past, present or future generations.
>
> "Cultural significance is embodied in the place itself, its fabric, setting, use, associations, meanings, records, related places and related objects."

This definition completes the definition of one of the most important concepts in World Heritage, i.e., the Outstanding Universal Value (UNESCO, 2008):

> "**Outstanding universal value** means cultural and/or natural significance which is so exceptional as to transcend national boundaries and to be of common importance for present and future generations of all humanity."

The Burra Charter further argues that places may have a range of values for different individuals or groups, that cultural significance may change over time and with use, and that its understanding may change with new information. With this premise, successful policy decision-making and management of heritage properties need to follow the understanding of the cultural significance by a sequence of collecting and analysing information (Australia ICOMOS, 2013). The "Burra Charter Process" is well-aligned with the UNESCO 2011 Recommendation on the Historic Urban Landscape (HUL) (UNESCO, 2011), where the urban area is

> "understood as the result of a historic layering of cultural and natural values and attributes, extending beyond the notion of 'historic centre' or 'ensemble' to include the broader urban context and its geographical setting",

focusing on heritage values (why to conserve) (Turner, 2008; Pereira Roders, 2007; Tarrafa Silva and Pereira Roders, 2012) and heritage attributes (what to conserve) (Veldpaus, 2015) from the perspective of various stakeholders (Pereira Roders, 2019). The HUL approach encourages cities to find their best-fit processes, methods, and tools for managing their heritage with cultural significance, going beyond the List of UNESCO World Heritage. Following the Burra Charter and the Recommendation of HUL, the practice of heritage management, previously dominated by the authorised discourse, is transforming towards a more inclusive, dynamic, and diversified alternative. In this light, recent studies about the cultural significance and heritage management touch upon topics including public and community participation (Li et al., 2020, 2021; Kırmızı and Karaman, 2021; Rosetti et al., 2022; Foroughi et al., 2023), adaptive reuse and redesign as interventions (Pintossi et al., 2019, 2023; Pinto et al., 2020; Gonçalves et al., 2021; Yarza Pérez and Verbakel, 2022; Lin et al., 2023), people-centred heritage (Liu, 2011; Ginzarly et al., 2019; Liu et al., 2019; Spoormans et al., 2023), consensus building (Wilkinson, 2019; Foroughi et al., 2023), integration with spatial planning and policy-making (Veldpaus, 2015; Sánchez et al., 2020; Tarrafa Silva et al., 2023), monitoring the urban dynamic globally (Taubenböck et al., 2012; Verbruggen et al., 2014; Valese et al., 2020), and so on.

1.2.2 Social Media and User-Generated Content

Since technological advances in the internet, especially with Web 2.0, changed the way how data is created and exchanged in the early 2000s, social media as a group of internet-based applications was given an increasingly crucial role in daily life (Kaplan and Haenlein, 2010). Social media, as a collective concept, may refer to a variety of information services used collaboratively by many people, including blogs (e.g., Blogger), micro-blogs (e.g., Twitter, Sina Weibo, RED/Xiaohongshu), opinion mining (e.g., Yelp, TripAdvisor, Dianping), photo and video sharing (e.g., Flickr, YouTube, Instagram, and TikTok), social bookmarking (e.g., Reddit, Pinterest), social networking (e.g., Facebook/Meta, WeChat, WhatsApp, LinkedIn, Snapchat), social news (e.g., Digg), wikis (e.g., Wikipedia), etc (Barbier and Liu, 2011). As of the year 2023, registered and active users of popular social media platforms can easily reach the scale of millions, and in the most extreme cases including Facebook, YouTube, WhatsApp, Instagram, WeChat, and TikTok, the scale of billions (Ruby, 2023). All users distributed globally generate each day massive data with high volume, velocity, variety, value, and veracity (Bello-Orgaz et al., 2016; Liu et al., 2020). Kaplan and Haenlein (2010) defined **User-Generated Content (UGC)** as:

> "the various forms of media content that are publicly available and created by end-users. UGC needs to fulfil three basic requirements:
>
> - Published either on a publicly accessible website or on a social networking site accessible to a selected group of people;
> - Showing a certain amount of creative effort;
> - Created outside of professional routines and practice."

The UGC available on social media platforms, as well as the social media itself as a special form of online social network (a network of interactions or relationships), are being extensively studied in the past decades in the broad fields of media studies, communication science, social sciences, computer science, as well as business and marketing (Bakshy et al., 2011; Zeng and Gerritsen, 2014; Holt, 2016; Monti et al., 2018). Social media platforms usually also provide Application Programming Interface (API) for researchers to obtain data and make sense of the social network structures and user-generated content (de Souza et al., 2004; Aggarwal, 2011). The essential aim of such studies is to discover knowledge and identify novel and actionable patterns (i.e., data mining) in the social media data (Barbier and Liu, 2011). A well-known family of methodological frameworks applied to social media data is the so-called Social Network Analysis (SNA), investigating the interactions between people and determining the important structural patterns in such interactions, albeit SNA as a special perspective of social science and behaviour science and as an application of graph theory has a far longer history than social media (Wasserman and Faust, 1994; Aggarwal, 2011). Analogue to remote sensing, where various sensors are used to collect data describing the physical features of the Earth's surface, **social sensing** treats each social media user as a "sensor" and collectively describes the social dynamics in the society and the socioeconomic features of the physical world, especially when the social media posts are geo-tagged and time-stamped (Liu et al., 2015; Wang et al., 2019; Galesic et al., 2021).

1.2.3 Artificial Intelligence, Machine Learning, and Deep Learning

According to Russell and Norvig (2010), Artificial Intelligence (AI) is a system that has or ideally should have the ability to both "think" and "act" in a way that is both "rationally" and "humanly".

> "**Artificial Intelligence** is the study of agents that receive percepts from the environment and perform actions. Each such agent implements a function that maps percept sequences to actions."

Even though already envisioned in the 1950s by pioneers including Donald Hebb, Marvin Minsky, Alan Turing, Claude Shannon, John McCarthy, and Herbert Simon, AI gradually developed in the late 20th century (Russell and Norvig, 2010; Zhang et al., 2021) and only reached its exploding prosperity in the 2010s, thanks to the availability of massive datasets and the computing power offered by Graphical Processing Unit (GPU), in addition to the traditional computation on Central Processing Unit (CPU). Two other terminologies stand closely with AI and are often used interchangeably in research and media reporting, i.e., Machine Learning (ML) and Deep Learning (DL) (Bishop and Nasrabadi, 2006; LeCun et al., 2015; Goodfellow et al., 2016; Zhang et al., 2021; Zhou, 2021). As pointed out and illustrated by Goodfellow et al. (2016), Machine Learning is a subset of Artificial Intelligence, and Deep Learning is a subset of Machine Learning, where:

> "**Machine Learning** is a form of applied statistics with increased emphasis on the use of computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions."
>
> And "**Deep Learning** methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level. It particularly makes use of deep neural network models with many layers."

Here representation generally refers to high-dimensional vectors containing feature information on the data point. Despite the hierarchical relations of the concepts AI, ML, and DL, in this dissertation, Deep Learning is referred to as the models employing either deep neural networks or Multi-layer Perceptron (MLP), whereas Machine Learning is referred to as the conventional statistical methods such as Support Vector Machine (SVM), Density-based Spatial Clustering of Applications with Noise (DBSCAN), and Random Forest (RF) without the use of neural networks, and Artificial Intelligence is used as the general term including the other two.

ML and DL can be categorized differently based on different perspectives (Zhang et al., 2021; Zhou, 2021). Depending on the inputs and outputs, they can be distinguished as:

- "supervised learning", where both features and labels are given, and models need to learn the mapping functions from features to labels, which can be separated to:
- "regression", if the labels are numerical variables;
- "classification", if the labels are categorical;
- "unsupervised learning", where only features but no labels are given, and models need to find out some intrinsic patterns within the data, which may include:
- "clustering", if the task is to group similar data points;
- "dimensionality reduction", if the task is to reduce the number of features;
- "semi-supervised learning", where only a small proportion of labels are provided during the training process;

- "reinforcement learning", where a group of agents learn how to act optimally based on the rewards given.

Depending on the type of tasks and the nature of input features, there are two popular families of ML and DL models:

- Computer Vision (CV) or Image Recognition in specific, aiming "to build autonomous systems which could perform some of the tasks which the human visual system can perform (and even surpass it in many cases)" (Huang, 1996), which has witnessed great advances since the dataset of ImageNet was proposed in the early 2010s (Deng et al., 2009; Russakovsky et al., 2015) and the development of ResNet models in 2016 (He et al., 2016);
- Natural Language Processing (NLP) or sometimes also referred to as Text Mining, Computer Linguistic, or Information Retrieval (Manning and Schutze, 1999; Bird et al., 2009; Manning, 2009; Collobert et al., 2011), concerning interactions between computers and humans that use natural language, many of which are based on language models that define a probability distribution over sequences of words, characters, or bytes in a natural language (Bird et al., 2009; Goodfellow et al., 2016; Zhang et al., 2021). It has been revolutionized since the introduction of the Attention mechanism and Transformers in 2017 (Vaswani et al., 2017; Devlin et al., 2019; Rao and McMahan, 2019) and large language models such as GPT-3 (Generative Pre-trained Transformer 3) (Floridi and Chiriatti, 2020; Brown et al., 2020).

Both families of tasks contain various sub-tasks, forming a huge community of computer science researchers. However, it must be noted that albeit currently dominantly being approached with Deep Learning models, both tasks of CV and NLP have a longer research history than DL. Recently, a new branch of research called Multi-modal Machine Learning (MML) has been discovering the combination of different modalities including images, texts, audio, videos, etc., in order to make better reasoning, more accurate inference, and higher-quality generation (Baltrusaitis et al., 2019; Kim et al., 2021; Bubeck et al., 2023).

Depending on the architecture of DL models, Convolutional Neural Network (CNN) (LeCun et al., 1989; Szegedy et al., 2015) and Recurrent Neural Network (RNN) (Rumelhart et al., 1986; Lipton et al., 2015) might be the two most important backbone structures that have been used for different purposes. Even though CNN models have been conventionally used in CV tasks, and RNN models and their variants LSTM (Long short-term memory) (Hochreiter and Schmidhuber, 1997) and Gated Recurrent Unit (GRU) (Cho et al., 2014) have been popular solutions in NLP tasks, they are also being used interchangeably. In the cases where a network/graph structure of the data points is also available as training input, a family of Graph Neural Network (GNN) might come into play (Zhang and Cheng, 2020; Ma and Tang, 2021; Wu et al., 2022).

For AI systems and DL models to be used broadly in application fields other than computer science algorithm development, the concept of "**transfer learning**" is crucial (Pan and Yang, 2010; Kang et al., 2021), which refers to the process to

> "extract the knowledge from one or more source tasks [learned previously] and applies the knowledge to a [novel] target task."

As such, users can take the full benefit of pre-trained large models: by conducting fine-tuning or prompt-tuning with an unseen dataset on the thoroughly trained models (Liu et al., 2022), the knowledge can be transferred to tasks in a new domain, without the need for collecting massive data and repeating the training process which usually costs resources (both time and computation power) unaffordable for normal users. Moreover, the ready-to-use Python-based ML and DL libraries and frameworks including Scikit-learn (Pedregosa et al., 2011), TensorFlow (Abadi et al., 2016), and PyTorch (Paszke et al., 2019), as well as the open-sharing culture of the AI community further pushes its applications in all fields forward in a revolutionary way.

1.2.4 Network Science and Spatiotemporal Analysis

As a relatively new field of research aiming at describing the networks (relationships among a collection of items) within the complex systems that are omnipresent in the world, network science is strongly based on the mathematical field of graph theory that originated already in 1735 with the well-known "Seven Bridges of Königsberg" problem (Newman, 2010; Easley and Kleinberg, 2010; Batty, 2013; Barabási et al., 2016; Latora et al., 2017). Two sets of terminologies exist in Graph Theory and Network Science that are basically used interchangeably:

- in Graph Theory, the collection of items is called "vertex/vertices", the collection of relationships is called "edges", and the entire system is called a "graph";
- in Network Science, the terms are respectively called "nodes", "links", and "network".

Nevertheless, both can be universally represented with the mathematical expression:

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}), \mathcal{V} = \{v_i\}, \mathcal{E} = \{(v_i, v_j) | v_i, v_j \in \mathcal{V}\} \subseteq \mathcal{V} \times \mathcal{V}, \tag{1.1}$$

where $\mathcal{G}$ is the graph/network, $\mathcal{V}$ is the set of vertices/nodes, and $\mathcal{E}$ is the set of pairs of vertices/nodes represented as edges/links. The relationship (connectivity) of the nodes is typically represented with an adjacency matrix $\boldsymbol{A}_{|\mathcal{V}|\times|\mathcal{V}|}$. Different types of graphs can be named depending on the nature of their components. A graph is:

- "undirected" if all edges have two directions between vertices;
- "directed" if some edges only have one direction from one vertex to another;
- "weighted" if the adjacency matrix is not binary, giving a weight to each edge;

- "attributed" if the vertices and/or edges have an accompanying numerical or categorical value, vector, or matrix describing their features;
- "dynamic" if the vertices and/or edges have an accompanying timestamp indicating the temporal development of the graph;
- "bipartite" if the vertices are composed of two separate groups where connections only exist between two different groups but not within the same groups;
- "heterogeneous" if the vertices are composed of items with different natures;
- "multi-dimensional" if the edges are composed of connections with different natures that can co-exist among two vertices.
- ......

The different settings of the graph/network are used to model various real-world phenomena with the statistical features and computable metrics of both the graph/network (such as density, diameter, clustering coefficient, etc.) and the vertices/nodes (such as degrees, closeness centrality, betweenness centrality, eigenvector centrality, PageRank, etc.) (Katz, 1953; Watts and Strogatz, 1998; Page et al., 1999; Newman, 2010; Batty, 2013; Nourian et al., 2016). This approach has been proved with numerous examples to be powerful in describing, explaining, predicting, and even improving the phenomena, functionality, and mechanisms of complex systems in many application domains including social communication (Bingham-Hall and Law, 2015; Luo and Cheng, 2015), transportation and mobility (Gonzalez et al., 2008), economics and marketing (Easley and Kleinberg, 2010), medicine and chemistry (Barabasi and Oltvai, 2004), urban and regional planning (Arcaute et al., 2016; Bai et al., 2023c), architecture design (Nourian, 2016; Jia et al., 2023), disease epidemics (Manríquez et al., 2021), sport science (Stival et al., 2023), and many more (Barabási et al., 2016; Latora et al., 2017). Specifically, when the graph vertices or network nodes are representing humans or human-generated posts, the field of Social Network Analysis introduced in Section 1.2.2 came into the spotlight; and when the graph vertices are urban streets or street intersections, then the studies of Space Syntax and spatial network analysis (Hillier and Hanson, 1989; Ratti, 2004; Turner, 2007) are created.

On a parallel line, when the subject of interest is humans living in cities and moving between cities, which is the main consideration for social sciences, urban studies, and human geography, many data (including the UGC on social media) are unavoidably geo-tagged and time-stamped, giving them a specific spatiotemporal context. To study the spatial dependency and divergence, as well as the temporal periodicity and dynamics of the geographic phenomena, methods, models, and theories have been extensively developed for the purposes of description, explanation, prediction, visualization, and simulation in the separate fields of Spatial Analysis (Batty, 1976; Goodchild and Longley, 1999; Anselin, 2003; Rogerson, 2021) and Time-Series Analysis (Wooldridge, 2013; Hamilton, 2020), as well as in the integrated field of Spatiotemporal Analysis (Batty et al., 1999; Cheng and Wang, 2009; Cheng et al., 2012).

One of the most important claims in the field of spatiotemporal analysis is the so-called “The First Law of Geography” by Tobler (1970):

> “Everything is related to everything else, but near things are more related than distant things.”

It has been tested and/or challenged with examples in many geographical and social systems (Nourian, 2016; Rogerson, 2021). Also worth noting is that spatiotemporal analyses are often conducted on spatial networks, indicating the close relationship between the two aspects (Batty, 1976; Cheng et al., 2012; Rogerson, 2021).

1.2.5 State-of-the-Art in Bridging the Problem Fields

Using social media data to facilitate heritage management is not uncommon. At an early stage of social media when the geolocation service was still not precise, Monteiro et al. (2014) already explored the possibility to understand the relationship of people with world heritage using Twitter data. The temporal evolution of the opinion on Twitter was associated with heritage-related events in activation scenarios such as the possible delisting of world heritage property. In a more recent study (Ginzarly et al., 2019), ways of mapping the HUL values and attributes extracted from the social media platform (Flickr, to be specific) in baseline scenarios have been explored. The posted pictures of the users were clustered into tangible and intangible aspects, and for each of the aspects, several topics have been discovered. The tags given by the users and the geolocation of that user posting the pictures have also been transformed into a spatial map to explore the similarities and differences between local people and visitors. However, the clustering and classification were mainly conducted manually as an expert-based process, which is rather time-consuming and case-specific, thus hard to be generalized to other cultural heritage properties. Nevertheless, the recent development in computer science, or AI, to be specific, offers some powerful tools to transform this process into an automated or semi-automated learning task for computers (Pan and Yang, 2010).

Recently, much research has been conducted to integrate AI techniques to analyse UGC on social media platforms to understand the opinions and behaviours of people in the urban environment. By associating the semantic meaning of the UGC (images and comments) with their geolocations, the most representative images for different places can be identified (Miah et al., 2017; Gomez et al., 2019; Zhang et al., 2019). Similarly, analysis of UGC can help to automate the process of characterizing the urban mobility pattern of locals and tourists along with the important places (Gabrielli et al., 2014; van der Zee and Bertocchi, 2018; Clemente et al., 2019), and to distinguish the exact opinions or sentiments of people on different topics within the places (Afzaal et al., 2019; Taecharungroj and Mathayomchan, 2019). All those mentioned studies can be interpreted as indications of heritage values from a bottom-up approach. However, none of them explicitly referred to such a connection, as they are mostly from different disciplines such as tourism, urban planning, management and marketing, ecosystem service, and computer science. While the research from different fields may have different focuses, they are all related to heritage management in some way.

This research intends to link the evidence-based and data-driven spatiotemporal and social characteristics in the urban environment to the perceived and expressed cultural significance, with the help of artificial intelligence as automation tools. With the proposed systematic methodological workflows of making classification from the user-generated content and generating maps related to cultural significance, tested on case study cities with urban areas inscribed in the UNESCO WHL, this research could be regarded as a pioneer knowledge documentation tool, as called for by the Recommendation on the Historic Urban Landscape (UNESCO, 2011).

As such, all the components in the title of this dissertation, i.e., "Sensing the Cultural Significance with AI for Social Inclusion, A Computational Spatiotemporal Network-based Framework of Heritage Knowledge Documentation using User-Generated Content" have been covered in this Section. The eventual aim and goal "Social Inclusion" is defined in Section 1.1.1; the objects "cultural significance" and "heritage knowledge documentation" are reflected with Section 1.2.1; the data "user-generated content" is introduced in Section 1.2.2; the tools "AI" and "computational" are discussed in Section 1.2.3; and the contexts "spatiotemporal" and "network-based" are covered in Section 1.2.4. Specifically, the word "sensing" both refers to the concept of "social sensing" brought up in Section 1.2.2, and implies that this research revisits the approach of Monteiro et al. (2014), proposing a tool for "sensing World Heritage".

1.3 Research Framework

1.3.1 Research Questions

The aim of this research is to:

> **explore the use of AI in a methodological framework to include the contribution of a larger and more diverse group of participants and facilitate the knowledge documentation of cultural significance in cities with user-generated social media data**.

Four main topics will be considered, respectively: the mathematical modeling of the social media networks, the contexts of opinions about heritage values and attributes, the dynamics of emotions about radical events, and the inclusion in the planning process, as visualized in Figure 1.4. The research questions that the framework is designed to answer are specified as:

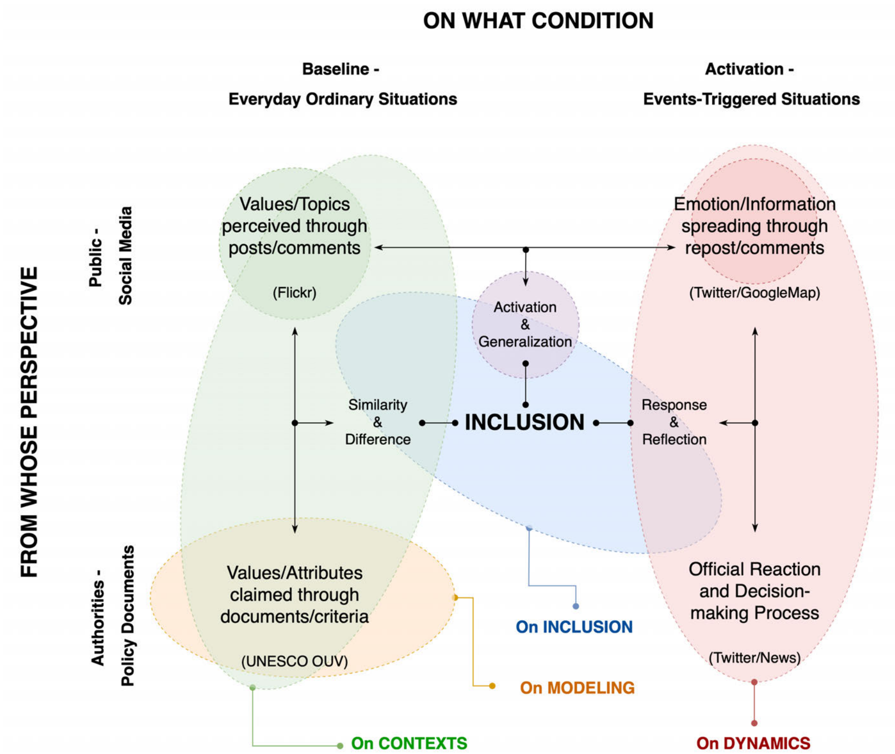


FIG. 1.4 The main research topics and work packages included in this dissertation.

1 How can mathematical and/or computational **MODELLING** help to construct a machine replica of the authoritative view of the cultural significance of UNESCO World Heritage properties as the basis for analyzing User-Generated Content?

2 As for a baseline scenario, how can a computational method help to map the spatiotemporal and social **CONTEXTS** of the public opinions about the cultural significance in a normal everyday setting?

3 As for an activated scenario, how can the **DYNAMICS** and mechanism of the emotion/information spreading on social media platforms be described when some radical events happen about a heritage property?

4 How can the evidence-based research findings improve the power and degree of social **INCLUSION** in future heritage management in broader cases?

The first three questions can be specified further in terms of identifying the [generalized] structure of communication networks on social media from given signal responses in baseline and activated scenarios (Adams and MacKay, 2007). In other words, it would be preferred to find an abstract network model as a graph that could function the same way as the real communication network consisting of many individuals on social media, whose nodes would represent the cliquish communities of global citizens who care enough about the heritage to a degree that they would express their opinions and emotions on social media platforms (Katz, 1953; Wasserman and Faust, 1994; Lazer et al., 2009; Pentland, 2015). Moreover, the results of future studies in this direction will have the potential to be extended to other domains of application such as Participatory Value Evaluation to inform policymakers on policy choices through Civic Engagement, which will be addressed in the fourth question (Bond and Messing, 2015; Calder et al., 2018).

1.3.2 Overall Methodology

This is an interdisciplinary research project that utilizes and combines various methods from conventional quantitative and qualitative to state-of-the-art mathematical/computational methods utilizing artificial intelligence. A set of workflows are proposed as a systematic and reproducible toolbox for collecting, processing, structuring, analysing, and mapping the information about the cultural significance of urban cultural heritage on social media platforms. The methods applied include Natural Language Processing, Image Recognition, Multi-modal Machine Learning, and Graph Neural Networks from the disciplinary of computer science, Graph Theory and Optimization from mathematics, Social Network Analysis and Statistical Hypothesis Testing from social sciences, Spatial Statistics and Spatiotemporal Modelling from Human Geography and Geographical Information Science, as well as Document Inspection and Case Studies from the interdisciplinary field of Social Sciences and Heritage Studies.

This research starts with a systematic literature review concerning the usage of User-Generated Content and Artificial Intelligence in the broad field of urban heritage management, in order to recognize the most prominent models, algorithms, and data sources in the existing literature. Then the official document of Statements of Outstanding Universal Value by the UNESCO World Heritage Centre is fed into artificial intelligence classifiers to train a machine replica for recognizing the relevant OUV selection criteria mentioned in a generic sentence, which is to be used throughout the dissertation as a key proxy for the concept of cultural significance. Furthermore, the sub-questions of the baseline and the activated scenarios both start with the collection and pre-processing of unstructured social media User-Generated Content data (raw images and texts) using state-of-the-art models and algorithms from either Natural Language Processing or Image Recognition, depending on the data format. The processed data are used to construct structured datasets in the form of networks with spatial, temporal, and social contexts, which are then analysed with Social Network Analysis, Graph Theory, and Spatial Statistics. The results of the analyses are visualized as maps and embedded in the discussion arena of heritage studies, eventually leading to planning recommendations in seek of inclusive heritage management processes.

The main datasets collected and/or applied in this dissertation include:

- WHOSe Heritage (Bai et al., 2021a), collected from UNESCO World Heritage Centre based on the Statements of Outstanding Universal Value, structured as a hierarchical multi-class single-label text classification dataset, introduced and analysed in Chapter 3, further applied in Chapters 4 till 6;
- Tripoli HUL (Ginzarly et al., 2019), adapted from a previous study collected from Flickr in the city of Tripoli, Lebanon, structured as a multi-class single-label image classification dataset, introduced and analysed in Chapter 4;
- Heri-Graphs (Bai et al., 2022), collected from Flickr in cities of Amsterdam, Suzhou, Venice (and additionally also Testaccio area of Rome), structured as a graph-based multi-modal multi-class multi-task node classification dataset, introduced in Chapter 4 and further analysed in Chapter 5;
- HREs (Bai et al., 2023a), collected from Twitter for the events of the 2019 fire in Notre-Dame de Paris and the 2019 flood in Venice, structured as an unlabelled attributed graph dataset with textual features and spatiotemporal contexts, introduced and analysed in Chapter 6;
- OSMnx (Boeing, 2017), a python-based framework for collecting crowd-sourcing OpenStreetMap datasets (Haklay and Weber, 2008) about the complex graph structure of urban street networks, applied in Chapters 4 till 6.

Other task-specific datasets used in part of the research pipelines within a single chapter will be introduced accordingly.

1.3.3 Overview of Case Selections

Due to the theoretical nature of proposing instantiable methodological workflows, this dissertation is not a conventional case-study-based research in the fields of heritage studies and urban studies. However, according to Ragin and Becker (1992),

> “every study is a case study because it is an analysis of social phenomena specific to time and place”.

This dissertation can also be interpreted as a Case Study (Johansson, 2007), defined by Ragin and Becker (1992) as:

"a phenomenon of some sort occurring in bounded [spatiotemporal] context".

Within this research, the social phenomenon is the perception of cultural significance, which is represented by the opinions and emotions on social media platforms, bounded by the time of the posting period after the existence of social media, and by the place of the selected city containing UNESCO World Heritage properties.

Venice has been chosen as the main study case (spatial bounds) showcasing the proposed methodological workflows throughout the dissertation from Chapters 3 to 6, while Amsterdam, Suzhou, Rome, Notre-Dame de Paris, as well as the entire UNESCO World Heritage List, are respectively analysed as test cases in different chapters. Most of the cases are built heritage properties concerning buildings, urban spaces, and cities. Note the cases used in this dissertation are mainly for illustrative and demonstrative purposes. An overview of all the case studies can be seen in Table 1.1. The detailed official Statement of OUV for the UNESCO World Heritage properties in the study cases can be found in Appendix A.

Specifically, "Venice and its Lagoon" is one of the only three UNESCO World Heritage properties up to 2023 that are justified with all six cultural OUV selection criteria (together with Mount Taishan and Mogao Caves in China). Taking Venice as the case study could cover a broad variety of cultural significance, especially as there are strong indications of natural elements in the city, though not explicitly justified with natural OUV selection criteria. Furthermore, whereas Mount Taishan is a huge natural landscape with cultural elements as scenic spots (mixed heritage) and the Mogao Caves are a group of small-scale caves containing statues and wall paintings, Venice [and its Lagoon] is the only property at an urban scale among the three. Since an important theoretical basis of this dissertation is the HUL, the necessity of analysing Venice across the Chapters can be further rationalised.

TABLE 1.1 A brief overview of the case studies in this dissertation listed in alphabetical order.

Study Case	State Party	OUV Selection Criteria	Data Source	Data Type	Scenarios	Chapters
Amsterdam	the Netherlands	(i)(ii)(iv)	Flickr	Texts, Images, Contexts	Baseline	Chapter 4
Paris*	France	(i)(ii)(iv)	Twitter	Texts, Contexts	Activated	Chapter 6
Rome**	Italy	(i)(ii)(iii)(iv)(vi)	Flickr	Texts, Images, Contexts	Baseline	Chapter 4
Suzhou	China	(i)(ii)(iii)(iv)(v)	Flickr	Texts, Images, Contexts	Baseline	Chapter 4
Venice***	Italy	(i)(ii)(iii)(iv)(v)(vi)	UNESCO, Flickr, Twitter	Texts, Images, Contexts	Baseline, Activated	Chapters 3, 4, 5, 6
World Heritage List	-	(i)(ii)(iii)(iv)(v)(vi)(vii)(viii)(ix)(x)	UNESCO	Texts	Baseline	Chapter 3

*Activated case about the fire in Notre Dame de Paris, April 2019.
**Baseline case about the Testaccio neighbourhood in Rome.
***Activated case about the flood in Venice, November 2019.

The selection of the social media platform in this research will mainly be based on the criteria of:

- the accessibility of the data and the privacy policy;
- the popularity of usage in general public for generating cultural-heritage-related content;
- a considerable even distribution of locals and visitors in the user group;
- the choice of previous researchers for a similar topic in the literature;
- the data needs to be well-structured, ideally having a timestamp, a spatial geo-tag, and a user profile along with the photos and/or comments.

Flickr and Twitter have been selected under the above-mentioned criteria for the baseline and activation scenarios, respectively.

1.3.4 Data Management and Research Ethics

As this dissertation deals with user-generated content from different social media platforms, it has been made sure that this research respected and followed the data privacy policy in Europe, the research integrity of TU Delft and HERILAND College of heritage planning, as well as the usage legal restrictions of the APIs from the social media platforms being used. Specifically, the data from UNESCO World Heritage Centre (Chapter 3), Flickr (Chapter 4), and Twitter (Chapter 6) are used at different stages of this dissertation. The data management plan of this research has been verified and approved by the data steward in the Faculty Architecture and Built Environment, TU Delft. In the meanwhile, as this research is related to data generated by humans and posted online, this research has gone through a Data Protection Impact Assessment (DPIA) process, which has been supported by the Privacy Officer from TU Delft. Part of the research data and computational workflows (from Chapter 3 and 4) have been open-sourced on the code-sharing platform GitHub. The research data of the entire dissertation will be shared with the 4TU Centre for Research Data, with the condition that all the personal trackable data are deleted and made anonymous and/or pseudonymous.

1.3.5 Research Limitations

The following subjects can be more or less related to the subject as can be also seen from the literature review, nevertheless, they fall out of the direct scope of this research, and could be discussed in future studies:

- proposing or stimulating the usage of social media platforms by heritage managers and the general public as a tool for cultural heritage participatory planning process and/or for place branding and social media marketing;
- comprehensively seeking for participation and social inclusion of all the possible stakeholders in heritage management;
- actively gathering stakeholders for some participatory planning projects, in order to reveal conflicts and seek consensus;
- comparing the different behavioural patterns of User-Generated Content on different social media platforms of various natures and target groups;
- figuring out the exact proportion of social media users in the whole society and ensuring the representativeness [with respect to sample size/diversity] of the social media users against the whole society;
- exploring the travel behaviour and preference of human beings with the help of their visit routes and trajectories;
- developing a new virtual platform or offline exhibition to collect the user-generated content and showcase the research outcomes;
- making use of digital technology to transform urban heritage into a [playable] digital twin and including people for the cyberspace documentation with crowd-sourcing and citizen engagement;
- exhaustively collecting as much information as possible for a specific study case from multiple data sources and creating an in-detailed cultural heritage information system;
- legitimating the research workflow using social media platforms for social inclusion in official documents and policies;
- explicitly applying the research outcome in practices as real-world cultural heritage management, spatial planning, adaptive reuse, and/or strategic urban design project;
- investigating the methodology into indoor museum exhibitions and social life scenarios in search of the human preference for movable cultural relics and intangible heritage;
- Improving the current selection criteria, inscription procedure, and Outstanding Universal Value system by UNESCO World Heritage Committee with the empirical results and/or creating a new heritage values and attributes standard.

1.4 Thesis Structure

Expanding the framework of Figure 1.4 and the discussion in Section 1.3.2, the structure of this dissertation is visualized in Figure 1.5. The dissertation is composed of 7 chapters and 5 parts in a hierarchical structure.

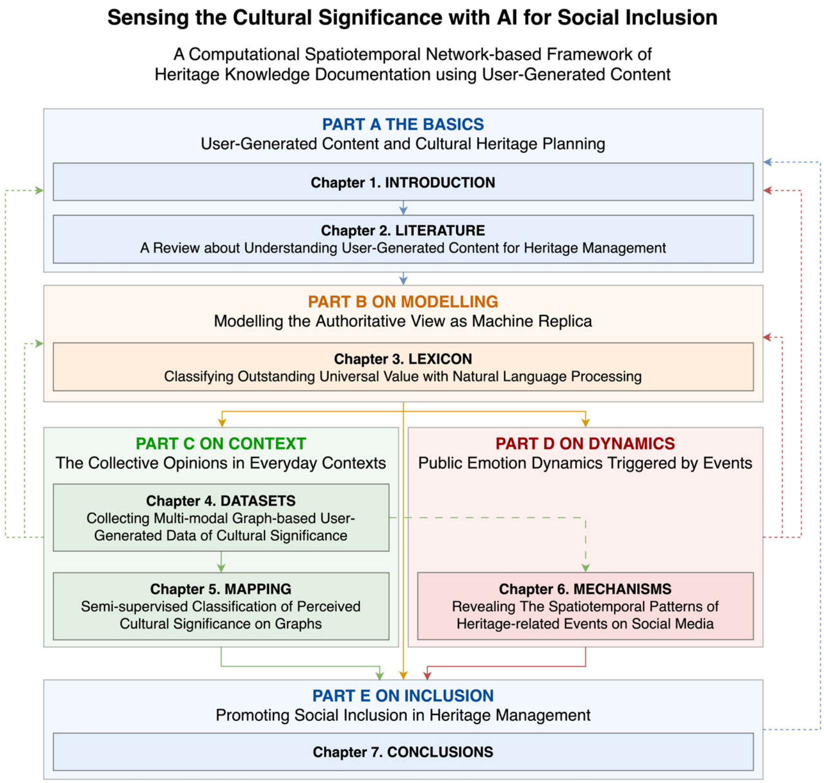


FIG. 1.5 The structure of the thesis. Arrows with solid lines demonstrate direct feed-forwards of research outcomes, and arrows with dashed lines entail potential feedback mechanisms.

The dissertation starts with a theoretical **BASIS** (PART A) of user-generated content and cultural heritage management, with the building-up of the methodological framework in Chapter 1 and a systematic literature review in Chapter 2. Then the dissertation continues to develop the mathematical/computational **MODELLING** (PART B) of the authoritative view on the cultural significance as a machine replica, through training natural language processing models on Outstanding Universal Value and obtaining an OUV-related lexicon in Chapter 3. From there the dissertation will be split into two parallel lines: the first one focuses on the **CONTEXT** (PART C) of the collective opinions in the everyday baseline scenarios, through constructing a baseline dataset in Chapter 4 and mapping the perceived cultural significance in Chapter 5; the second one focuses on the **DYNAMICS** (PART D) of the public emotions triggered by radical events in activated scenarios, by inspecting the mechanisms and the spatiotemporal patterns of public discussion in Chapter 6. Finally, the dissertation comes back in an integral discussion about applying the proposed workflows to promote social **INCLUSION** (PART E) in future heritage management, with the conclusions in Chapter 7. Hypothetically, results can emerge from later parts of the dissertation, which can be informative or even revolutionary for the knowledge obtained in earlier parts. With further iterations and updates in the future, it entails a combined feed-forward and feedback loop visualized with arrows in Figure 1.5.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In Osdi, volume 16, pages 265–283. Savannah, GA, USA.

Adamic, L. A., Lento, T. M., Adar, E., and Ng, P. C. (2016). Information evolution in social networks. WSDM 2016 - Proceedings of the 9th ACM International Conference on Web Search and Data Mining, pages 473–482.

Adams, R. P. and MacKay, D. J. (2007). Bayesian online changepoint detection. arXiv preprint arXiv:0710.3742.

Afzaal, M., Usman, M., Fong, A. C., and Fong, S. (2019). Multiaspect-based opinion classification model for tourist reviews. Expert Systems, 36(2):e12371.

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In Aggarwal, C. C., editor, Social Network Data Analytics, chapter 1, pages 1–15. SPRINGER.

Anselin, L. (2003). An introduction to spatial autocorrelation analysis with geoda. Spatial Analysis Laboratory, University of Illinois, Champagne-Urbana, Illinois.

Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., and Batty, M. (2016). Cities and regions in britain through hierarchical percolation. Royal Society open science, 3(4):150691.

Australia ICOMOS (2013). The Burra Charter: The Australia ICOMOS Charter for Places of Cultural Significance (1999). Technical report, Australia ICOMOS.

Bai, N., Cheng, T., Nourian, P., and Pereira Roders, A. (2023a). An exploratory data analysis of the spatiotemporal patterns of heritage-related events on twitter. The 30th International Conference on Geoinformatics (CPGIS 2023), London, UK.

Bai, N., Ducci, M., Mirzikashvili, R., Nourian, P., and Pereira Roders, A. (2023b). Mapping urban heritage images with social media data and artificial intelligence, a case study in testaccio, rome. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023:139–146.

Bai, N., Luo, R., Nourian, P., and Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage statements of ”Outstanding Universal Value” with soft labels. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 366–384, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2022). Heri-graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. ISPRS International Journal of Geo-Information, 11(9).

Bai, N., Nourian, P., and Pereira Roders, A. (2021b). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.

Bai, N., Nourian, P., Pereira Roders, A., Bunschoten, R., Huang, W., and Wang, L. (2023c). Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data. Environment and Planning B: Urban Analytics and City Science, 50(1):94–116.

Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Everyone’s an influencer: quantifying influence on twitter. In Proceedings of the fourth ACM international conference on Web search and data mining, pages 65–74.

Baltrusaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2):423–443.

Bandarin, F. and Van Oers, R. (2012). The historic urban landscape: managing heritage in an urban century. John Wiley & Sons.

Barabási, A.-L. et al. (2016). Network Science. Cambridge University Press.

Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell’s functional organization. Nature reviews genetics, 5(2):101–113.

Barbier, G. and Liu, H. (2011). Data mining in social media. Social network data analytics, pages 327–352.

Batty, M. (1976). Urban modelling. Cambridge University Press Cambridge.

Batty, M. (2013). The new science of cities. MIT press.

Batty, M., Xie, Y., and Sun, Z. (1999). Modeling urban dynamics through gis-based cellular automata. Computers, environment and urban systems, 23(3):205–233.

Bello-Orgaz, G., Jung, J. J., and Camacho, D. (2016). Social big data: Recent achievements and new challenges. Information Fusion, 28:45–59.

Bingham-Hall, J. and Law, S. (2015). Connected or informed?: Local twitter networking in a london neighbourhood. Big Data & Society, 2(2):2053951715597457.

Bird, S., Klein, E., and Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

Bishop, C. M. and Nasrabadi, N. M. (2006). Pattern recognition and machine learning, volume 4. Springer.

Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65:126–139.

Bond, R. and Messing, S. (2015). Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook. American Political Science Review, 109(1):62–78.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901.

Bruns, A., Burgess, J. E., Crawford, K., and Shaw, F. (2012). # qldfloods and QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods. Technical Report Cci, ARC Centre of Excellence for Creative Industries and Innovation.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712.

Calder, M., Craig, C., Culley, D., de Cani, R., Donnelly, C. A., Douglas, R., Edmonds, B., Gascoigne, J., Gilbert, N., Hargrove, C., Hinds, D., Lane, D. C., Mitchell, D., Pavey, G., Robertson, D., Rosewell, B., Sherwin, S., Walport, M., and Wilson, A. (2018). Computational modelling for decision-making: Where, why, what, who and how. Royal Society Open Science, 5(6).

Cheng, J., Adamic, L. A., Kleinberg, J., and Leskovec, J. (2016). Do cascades recur? 25th International World Wide Web Conference, WWW 2016, pages 671–681.

Cheng, T., Haworth, J., and Wang, J. (2012). Spatio-temporal autocorrelation of road network data. Journal of Geographical Systems, 14:389–413.

Cheng, T. and Wang, J. (2009). Accommodating spatial associations in drnn for space–time analysis. Computers, Environment and Urban Systems, 33(6):409–418.

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. Syntax, Semantics and Structure in Statistical Translation, page 103.

Clemente, P., Calvache, M., Antunes, P., Santos, R., Cerdeira, J. O., and Martins, M. J. (2019). Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a natural park in portugal. Ecological indicators, 96:59–68.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of machine learning research, 12(ARTICLE):2493–2537.

Couclelis, H. (2005). "Where has the future gone?" Rethinking the role of integrated land-use models in spatial planning. Environment and Planning A, 37(8):1353–1371.

de Souza, C. R., Redmiles, D., Cheng, L.-T., Millen, D., and Patterson, J. (2004). Sometimes you need to see through walls: a field study of application programming interfaces. In Proceedings of the 2004 ACM conference on Computer supported cooperative work, pages 63–71.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dong, W., Heller, K., and Pentland, A. (2012). Modeling infection with multi-agent dynamics. In Social Computing, Behavioral-Cultural Modeling and Prediction: 5th International Conference, SBP 2012, College Park, MD, USA, April 3-5, 2012. Proceedings 5, pages 172–179. Springer.

Easley, D. and Kleinberg, J. (2010). Networks, Crowds and Markets: Reasoning about a Highly Connected World. Cambridge University Press, Cambridge.

Fischer, G. (2009). Democratizing design: New challenges and opportunities for computer-supported collaborative learning. Computer Supported Collaborative Learning Practices, CSCL 2009 Conference Proceedings - 9th International Conference, pages 282–286.

Floridi, L. and Chiriatti, M. (2020). Gpt-3: Its nature, scope, limits, and consequences. Minds and Machines, 30:681–694.

Foroughi, M., de Andrade, B., Roders, A. P., and Wang, T. (2023). Public participation and consensus-building in urban planning from the lens of heritage planning: A systematic literature review. Cities, 135:104235.

Gabrielli, L., Rinzivillo, S., Ronzano, F., and Villatoro, D. (2014). From Tweets to Semantic Trajectories: Mining Anomalous Urban Mobility Patterns. In Nin, J and Villatoro, D., editor, CITIZEN IN SENSOR NETWORKS, volume 8313 of Lecture Notes in Artificial Intelligence, pages 26–35, HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY. SPRINGER-VERLAG BERLIN.

Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S. L., Kreuter, F., Olsson, H., Prelec, D., Stein, D. L., and van Der Does, T. (2021). Human social sensing is an untapped resource for computational social science. Nature, 595(7866):214–222.

Ginzarly, M., Pereira Roders, A., and Teller, J. (2019). Mapping historic urban landscape values through social media. Journal of Cultural Heritage, 36:1–11.

Gomez, R., Gomez, L., Gibert, J., and Karatzas, D. (2019). Learning from #barcelona instagram data what locals and tourists post about its neighbourhoods. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11134 LNCS:530–544.

Gonçalves, J., Mateus, R., Silvestre, J. D., Roders, A. P., and Bragança, L. (2021). Attitudes matter: Measuring the intention-behaviour gap in built heritage conservation. Sustainable Cities and Society, 70:102913.

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. nature, 453(7196):779–782.

Goodchild, M. and Longley, P. (1999). The future of gis and spatial analysis. In Geographical information systems: principles, techniques, management and applications, volume 1, pages 567–580. John Wiley New York.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MIT press.

Haklay, M. and Weber, P. (2008). Openstreetmap: User-generated street maps. IEEE Pervasive computing, 7(4):12–18.

Hamilton, J. D. (2020). Time series analysis. Princeton university press.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.

Hillier, B. and Hanson, J. (1989). The Social Logic of Space. Cambridge University Press.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8):1735–1780.

Holt, D. (2016). Branding in the age of social media. Harvard business review, 94(3):40–50.

Huang, T. (1996). Computer vision: Evolution and promise. 1996 CERN SCHOOL OF COMPUTING, page 21.

ICOMOS, A. (2013). The Burra Charter: The Australia ICOMOS charter for places of cultural significance 2013. Australia ICOMOS Incorporated.

IUCN, ICOMOS, ICROM, and WHC (2010). Guidance on the preparation of retrospective Statements of Outstanding Universal Value for World Heritage Properties. Technical report, IUCN, ICOMOS, ICROM and WHC.

Jansen, W. S., Otten, S., van der Zee, K. I., and Jans, L. (2014). Inclusion: Conceptualization and measurement. European journal of social psychology, 44(4):370–385.

Janssen, J., Luiten, E., Renes, H., and Stegmeijer, E. (2017). Heritage as sector, factor and vector: conceptualizing the shifting relationship between heritage management and spatial planning. European Planning Studies, 25(9):1654–1672.

Jia, Z., Nourian, P., Luscuere, P., and Wagenaar, C. (2023). Spatial decision support systems for hospital layout design: A review. Journal of Building Engineering, page 106042.

Johansson, R. (2007). On case study methodology. Open house international, 32(3):48–54.

Jokilehto, J. (2006). World Heritage: Defining the Outstanding Universal Value. City & Time, 2(2):1–10.

Jokilehto, J. (2008). What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties. Technical report, ICOMOS, ICOMOS Berlin.

Kang, Y., Cho, N., Yoon, J., Park, S., and Kim, J. (2021). Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos. ISPRS International Journal of Geo-Information, 10(3):137.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. Business horizons, 53(1):59–68.

Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika, 18(1):39–43.

Kim, W., Son, B., and Kim, I. (2021). Vilt: Vision-and-language transformer without convolution or region supervision. In International Conference on Machine Learning, pages 5583–5594. PMLR.

Kırmızı, Ö. and Karaman, A. (2021). A participatory planning model in the context of historic urban landscape: The case of kyrenia's historic port area. Land use policy, 102:105130.

Latora, V., Nicosia, V., and Russo, G. (2017). Complex networks: principles, methods and applications. Cambridge University Press.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Social science. computational social science. Science (New York, NY), 323(5915):721–723.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature, 521(7553):436–444.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4):541–551.

Li, J., Krishnamurthy, S., Roders, A. P., and Van Wesemael, P. (2020). Community participation in cultural heritage management: A systematic literature review comparing chinese and international practices. Cities, 96:102476.

Li, J., Krishnamurthy, S., Roders, A. P., and van Wesemael, P. (2021). Imagine the old town of lijiang: Contextualising community participation for urban heritage management in china. Habitat International, 108:102321.
Lin, M., Pereira Roders, A., Nevzgodin, I., and de Jonge, W. (2023). Values and interventions: dynamic relationships in international doctrines. Journal of Cultural Heritage Management and Sustainable Development, ahead-of-print(ahead-of-print). Publisher: Emerald Publishing Limited.
Lipizzi, C., Iandoli, L., and Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using twitter streams. International Journal of Information Management, 35(4):490–503.
Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.
Liu, J., Li, T., Xie, P., Du, S., Teng, F., and Yang, X. (2020). Urban big data fusion based on deep learning: An overview. Information Fusion, 53:123–133.
Liu, S. B. (2011). Grassroots heritage: A multi-method investigation of how social media sustain the living heritage of historic crises. PhD thesis, University of Colorado at Boulder.
Liu, T., Butler, R. J., and Zhang, C. (2019). Evaluation of public perceptions of authenticity of urban heritage under the conservation paradigm of historic urban landscape—a case study of the five avenues historic district in tianjin, china. Journal of architectural conservation, 25(3):228–251.
Liu, X., Ji, K., Fu, Y., Tam, W., Du, Z., Yang, Z., and Tang, J. (2022). P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 61–68.
Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., and Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. Annals of the Association of American Geographers, 105(3):512–530.
Luo, J.-D. and Cheng, M.-Y. (2015). Guanxi circles' effect on organizational trust: Bringing power and vertical social exchanges into intraorganizational network analysis. American Behavioral Scientist, 59(8):1024–1037.
Ma, Y. and Tang, J. (2021). Deep learning on graphs. Cambridge University Press.
Manning, C. and Schutze, H. (1999). Foundations of statistical natural language processing. MIT press.
Manning, C. D. (2009). An introduction to information retrieval. Cambridge university press.
Manríquez, R., Guerrero-Nancuante, C., Martínez, F., and Taramasco, C. (2021). Spread of epidemic disease on edge-weighted graphs from a database: A case study of covid-19. International Journal of Environmental Research and Public Health, 18(9):4432.
Miah, S. J., Vu, H. Q., Gammack, J., and McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. Information & Management, 54(6):771–785.
Milgram, S. (1967). The small world problem. Psychology today, 2(1):60–67.
Monteiro, V., Henriques, R., Painho, M., and Vaz, E. (2014). Sensing World Heritage An Exploratory Study of Twitter as a Tool for Assessing Reputation. In Murgante, B and Misra, S and Rocha, AMAC and Torre, C and Rocha, JG and Falcao, MI and Taniar, D and Apduhan, BO and Gervasi, O., editor, COMPUTATIONAL SCIENCE AND ITS APPLICATIONS - ICCSA 2014, PT II, volume 8580 of Lecture Notes in Computer Science, pages 404–419. Univ Minho; Univ Perugia; Univ Basilicata; Monash Univ; Kyushu Sangyo Univ; Assoc Portuguesa Investigacao Operac.
Monti, L., Delnevo, G., Mirri, S., Salomoni, P., and Callegati, F. (2018). Digital invasions within cultural heritage: Social media and crowdsourcing. In Smart Objects and Technologies for Social Good: Third International Conference, GOODTECHS 2017, Pisa, Italy, November 29-30, 2017, Proceedings 3, pages 102–111. Springer.
Munar, A. M. (2012). Social Media Strategies and Destination Management. SCANDINAVIAN JOURNAL OF HOSPITALITY AND TOURISM, 12(2):101–120.
Newman, M. (2010). Networks: An Introduction. Oxford University Press.
Nourian, P. (2016). Configraphics: Graph Theoretical Methods for Design and Analysis of Spatial Configurations. TU Delft.
Nourian, P., Rezvani, S., Sariyildiz, I., and van der Hoeven, F. (2016). Spectral modelling for spatial network analysis. In Proceedings of the Symposium on Simulation for Architecture and Urban Design (simAUD 2016), pages 103–110. SimAUD.
Olsson, K. (2008). Citizen input in urban heritage management and planning: A quantitative approach to citizen participation. Town Planning Review, 79(4):371–395.
Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.
Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Pentland, A. (2015). Social Physics: How social networks can make us smarter. Penguin.
Pereira Roders, A. (2007). Re-architecture: lifespan rehabilitation of built heritage. PhD thesis, Technische Universiteit Eindhoven.
Pereira Roders, A. (2019). The Historic Urban Landscape Approach in Action: Eight Years Later. In Reshaping Urban Conservation, pages 21–54. Springer.
Pinto, M. R., Viola, S., Fabbricatti, K., and Pacifico, M. G. (2020). Adaptive reuse process of the historic urban landscape post-covid-19. the potential of the inner areas for a "new normal". VITRUVIO-International Journal of Architectural Technology and Sustainability, 5(2):87–105.
Pintossi, N., Ikiz Kaya, D., and Pereira Roders, A. (2019). Adaptive reuse of cultural heritage in amsterdam: Identifying challenges and solutions through the historic urban landscape approach. In Proceedings of the LDE Heritage Conference on Heritage and the Sustainable Development Goals: Proceedings, pages 304–314.
Pintossi, N., Kaya, D. I., van Wesemael, P., and Roders, A. P. (2023). Challenges of cultural heritage adaptive reuse: A stakeholders-based comparative study in three european cities. Habitat International, 136:102807.
Ragin, C. C. and Becker, H. S. (1992). What is a case?: exploring the foundations of social inquiry. Cambridge university press.
Rao, D. and McMahan, B. (2019). Natural Language Processing with PyTorch - Build Intelligent Language Applications Using Deep Learning. O'Reilly Media, Inc.
Ratti, C. (2004). Space syntax: Some inconsistencies. Environment and Planning B: Planning and Design, 31(4):487–499.
Rochon, T. R. (1998). Culture moves: Ideas, activism, and changing values. Princeton University Press.
Roders, A. P. and Van Oers, R. (2011). World heritage cities management. Facilities, 29(7/8):276–285.
Rogerson, P. A. (2021). Spatial Statistical Methods for Geography. SAGE Publications Ltd.
Rosetti, I., Bertrand Cabral, C., Pereira Roders, A., Jacobs, M., and Albuquerque, R. (2022). Heritage and sustainability: Regulating participation. Sustainability, 14(3):1674.
Ruby, D. (2023). Social media users in the world — (2023 demographics). demandsage. (version: 2023-03-20) (access date: 2023-05-27).
Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088):533–536.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115:211–252.
Russell, S. J. and Norvig, P. (2010). Artificial Intelligence: A modern approach. Pearson Education, Inc.
Sánchez, M. L., Cabrera, A. T., and Del Pulgar, M. L. G. (2020). Guidelines from the heritage field for the integration of landscape and heritage planning: A systematic literature review. Landscape and Urban Planning, 204:103931.
Schroeder, A., Pennington-Gray, L., Donohoe, H., and Kiousis, S. (2013). Using Social Media in Times of Crisis. Journal of Travel and Tourism Marketing, 30(1-2):126–143.
Spoormans, L., Czischke, D., Pereira Roders, A., and de Jonge, W. (2023). "do i see what you see?"—differentiation of stakeholders in assessing heritage significance of neighbourhood attributes. Land, 12(3):712.
Stival, L., Pinto, A., Andrade, F. d. S. P. d., Santiago, P. R. P., Biermann, H., Torres, R. d. S., and Dias, U. (2023). Using machine learning pipeline to predict entry into the attack zone in football. PloS one, 18(1):e0265372.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9.
Taecharungroj, V. and Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tourism Management, 75(July):550–568.
Tarrafa Silva, A. and Pereira Roders, A. (2012). Cultural Heritage Management and Heritage (Impact) Assessments. In Joint CIB W070, W092 & TG72 International Conferenceon Facilities Management, Procurement Systems and Public Private Partnership.
Tarrafa Silva, A., Pereira Roders, A., Cunha Ferreira, T., and Nevzgodin, I. (2023). Critical analysis of policy integration degrees between heritage conservation and spatial planning in amsterdam and ballarat. Land, 12(5):1040.
Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., and Dech, S. (2012). Monitoring urbanization in mega cities from space. Remote sensing of Environment, 117:162–176.
Taylor, J. and Gibson, L. K. (2017). Digitisation, digital interaction and social media: embedded barriers to democratic heritage. International Journal of Heritage Studies, 23(5):408–420.
Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. Economic geography, 46(sup1):234–240.
Turner, A. (2007). From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis. Environment and Planning B: planning and Design, 34(3):539–555.

Turner, M. (2008). Values of heritage in great religious and cultural areas: From existentialism to historicism-a view of the holy land and the sites of jesus and the apostles. In Values of Heritage in Great Religious and Cultural Areas, pages 1000–1007. Polistampa.

UNESCO (1972). Convention Concerning the Protection of the World Cultural and Natural Heritage. Technical Report november, UNESCO, Paris.

UNESCO (2008). Operational guidelines for the implementation of the world heritage convention. Technical Report July, UNESCO World Heritage Centre.

UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.

Valese, M., Noardo, F., and Pereira Roders, A. (2020). World heritage mapping in a standard-based structured geographical information system. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4-2020:81–88.

van der Zee, E. and Bertocchi, D. (2018). Finding patterns in urban tourist behaviour: A social network analysis approach based on tripadvisor reviews. Information Technology & Tourism, 20(1-4):153–180.

van Dijck, J. (2011). Flickr and the culture of connectivity: Sharing views, experiences, memories. Memory Studies, 4(4):401–415.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 6000–6010.

Veldpaus, L. (2015). Historic urban landscapes: framing the integration of urban and heritage planning in multilevel governance. PhD thesis, Technische Universiteit Eindhoven.

Verbruggen, R., Pereira Roders, A., Stash, N., Leony, D., and De Bra, P. (2014). Protected urban planet: monitoring the evolution of protected urban areas worldwide. In Special Session'Real Spaces and Cyber Spaces: New Challenges in Regional Science'within ERSA 54th congress Regional Development & Globalisation: Best practices, Saint Petersburg, Russia.

Wang, D., Szymanski, B. K., Abdelzaher, T., Ji, H., and Kaplan, L. (2019). The age of social sensing. Computer, 52(1):36–45.

Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.

Waterton, E., Smith, L., and Campbell, G. (2006). The utility of discourse analysis to heritage studies: The burra charter and social inclusion. International journal of heritage studies, 12(4):339–355.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. nature, 393(6684):440–442.

Wilkinson, A. (2019). The historic urban landscape approach in edinburgh's old and new towns: Implementation of projects on the ground in a living capital city. Reshaping Urban Conservation: The Historic Urban Landscape Approach in Action, pages 223–233.

Wooldridge, J. M. (2013). Introductory Econometrics: A Modern Approach. South-Western Cengage Learning, 5 edition.

Wu, L., Cui, P., Pei, J., and Zhao, L. (2022). Graph Neural Networks: Foundations, Frontiers, and Applications. Springer Singapore, Singapore.

Yarza Pérez, A. J. and Verbakel, E. (2022). The role of adaptive reuse in historic urban landscapes towards cities of inclusion. the case of acre. Journal of Cultural Heritage Management and Sustainable Development.

Zeng, B. and Gerritsen, R. (2014). What do we know about social media in tourism? a review. Tourism management perspectives, 10:27–36.

Zhai, X., Luo, Q., and Wang, L. (2020). Why tourists engage in online collective actions in times of crisis: Exploring the role of group relative deprivation. Journal of Destination Marketing and Management, 16(August 2019).

Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2021). Dive into deep learning. arXiv preprint arXiv:2106.11342.

Zhang, F., Zhou, B., Ratti, C., and Liu, Y. (2019). Discovering place-informative scenes and objects using social media photos. Royal Society open science, 6(3):181375.

Zhang, Y. and Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Computers, Environment and Urban Systems, 79:101403.

Zhou, Z.-H. (2021). Machine learning. Springer Nature.

2 Literature

A Review about Understanding User-Generated Content for Heritage Management

Parts of this chapter have been published in Bai et al. (2021b) and Bai et al. (2023a) and will be submitted as Bai, et al. (2024a).

Bai N, Nourian P, Pereira Roders A. (2021b). Global Citizens and World Heritage: Social Inclusion of Online Communities in Heritage Planning. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021. p. 23–30.

Bai N, Ducci M, Mirzikashvili R, Nourian P, Pereira Roders, A. (2023a). Mapping Urban Heritage Images with Social Media Data and Artificial Intelligence, A Case Study in Testaccio, Rome. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023. p. 139–146.

Bai, N, Nourian P, Pereira Roders, A. (2024a). Mapping the User-Generated Content for Researching Cultural Heritage - A Systematic Literature Review. (Under Preparation).

ABSTRACT The UNESCO 2011 Recommendation on the Historic Urban Landscape promotes mapping the cultural significance of urban heritage from the perspectives of the general public in pursuit of social inclusion in heritage management. The user-generated content already available on social media platforms in the form of images, comments, and ratings can be considered a rich source for collecting data concerning the tourists' image of destinations and their collective perception of urban cultural heritage. Considering the amount of unstructured data on a large scale, artificial intelligence can construct structured feature vectors therefrom and significantly aid the analysis and collation processes compared to the traditional manual approach for mapping public perception of cultural heritage. This chapter presents a systematic literature review on the usage of user-generated content on social media platforms in the specific field of heritage management. A total of 431 records including research articles, conference proceedings, and book chapters that are available at SCOPUS and Web of Science were collected in 2020. After abstract screening and full-paper reviewing, 73 studies are included for qualitative synthesis to reflect the main issue of "how is User-Generated Content from social media platforms [computationally] understood for researching cultural heritage properties". Information about the geographical distribution of the studies, the most frequently

used social media platform, the span and size of collected and analysed user-generated content, the research objectives, focuses, methods, computational algorithms, and models, as well as the distinction between baseline and activated scenarios, has been identified, categorized, and/or summarized. These key aspects coded for the systematic literature review process are eventually mapped on a 2D space using the Multi-Dimensional Scaling algorithm to indicate the co-occurring associations among the aspects. It is also found that although only a small proportion of the included studies declared explicitly that their study case is a cultural heritage property, a majority of them were conducted with at least one study case in a city where one or more World Heritage properties are located. In other words, those studies were concerned with cultural heritage planning or management issues without explicit recognition. This chapter becomes the theoretical basis for the rest of the dissertation.

KEYWORDS Systematic Literature Review, User-Generated Content, Social Media, Cultural Heritage, Multi-Dimensional Scaling

2.1 Introduction

Conventionally, the listing of cultural heritage, especially that of UNESCO World Heritage is determined by authorities and justified by experts, thus mainly a top-down view of the cultural significance (UNESCO, 2008). However, ordinary people including local residents and visiting tourists usually also have their own experiences and opinions on the tangible and intangible resources that they truly value in the place they live, work, or leisure, providing an alternative bottom-up view (Janssen et al., 2017; Bai et al., 2021b). The “image” of a place can be informative for experts during the spatial planning and heritage management decision-making processes, as it adds a potentially more inclusive layer of information concerning the emotional attachment and sense of belonging in a “lived place” (Lynch, 1964; Lefebvre, 2014), which might not be directly and/or necessarily recognised as heritage according to the conventional procedure but does positively contribute to the collective memory (Assmann and Czaplicka, 1995; Bai et al., 2023b). Since the adoption of the 2011 Recommendation on the Historic Urban Landscape (HUL) by UNESCO, mapping the cultural significance of urban heritage as its valued attributes (Veldpaus, 2015) from the perspectives of a broader range of stakeholders including the general public is being recommended (UNESCO, 2011), where tools for knowledge documentation and civic engagement are also actively called for. As argued by Bai et al. (2021b), social media platforms and other online digital applications already partially function as a critical resource for constructing such tools to promote inclusive planning, both as an active medium for crowd-sourcing and participatory design (Watkins, 2007; Estellés-Arolas and González-Ladrón-de

Guevara, 2012; Boy and Uitermark, 2017; Ducci et al., 2023), and as a pre-existing database for social sensing, information mapping, and pattern mining (Schich et al., 2014; Ginzarly et al., 2019; Kumar, 2020; Galesic et al., 2021).

The geo-tagged and time-stamped user-generated information already available on social media platforms in the form of images, comments, and ratings is considered an effective source for collecting data concerning the tourist image of destinations (Kang et al., 2021; Cho et al., 2022), the most representative characteristics of urban scenes (Lai et al., 2017; Zhang et al., 2019; Liu and De Sabbata, 2021), and more specifically, the collective perception of urban cultural heritage (Monteiro et al., 2014; Ginzarly et al., 2019; Bai et al., 2022), all of which could be interpreted as major components of the urban heritage images perceived and expressed by the social media users. However, the amount of user-generated data on social media platforms is usually at a large scale with thousands and even millions of samples that easily exceed the capacity of manual analyses, even if only a general overview is desired.

The current advancements in the field of artificial intelligence (AI), or specifically, the developments of pre-trained machine learning and deep learning models have made semi-automatic analyses of unstructured multi-modal data at scales not only possible but also effective (Deng et al., 2009; LeCun et al., 2015; Vaswani et al., 2017; Baltrusaitis et al., 2019). With large image and language models thoroughly pre-trained on a massive amount of data, structured feature vectors can be easily constructed and/or extracted as effective representations of the raw information, enabling various types of downstream tasks in the application ends at a relatively low cost (Pan and Yang, 2010; Kang et al., 2021). As a vivid example, ChatGPT by OpenAI[1] already showcased how AI models could interact with different use cases including urban, history, and heritage studies, creating revolutionary possibilities, even though also raising moral concerns (Batty, 2023; Fostikov, 2023; Thorp, 2023). Nevertheless, the ethical discussions of applying this emerging technology are out of the scope of this chapter, which only considers AI as one of the possible tools that can aid the analysis and collation processes compared to the traditional manual approach in heritage studies.

This chapter presents a systematic literature review, indenting to answer the question “how is User-Generated Content from social media platforms [computationally] understood for researching cultural heritage properties". The answers to this question would be approached by inspecting, categorizing, and summarizing the existing literature, which could effectively become a source of inspiration for this dissertation and other future studies with similar objectives.

A few other literature reviews or bibliometric studies touched upon similar issues of understanding User-Generated Content in the fields of Tourism (Leung et al., 2013; Lu and Stepchenkova, 2015; Pickering et al., 2018; Alaei et al., 2019), Hospitality (Leung et al., 2013), Marketing (Avila-Robinson and Wakabayashi, 2018), Cultural Studies (van Dijck, 2011), and most recently, in Heritage Studies (Alviz-Meza et al., 2022). However, as far as the author knows, this chapter is the first one to bring

[1] An official description can be found at https://openai.com/blog/chatgpt#OpenAI

together all three aspects of the research context (i.e., geographical distribution, temporal span, data source), research content (i.e., research topics, case studies, heritage-specific categorization), and research methodology (i.e., models, algorithms), which not only result in qualitative and quantitative descriptions but also as a 2D Multi-Dimensional Scaling (MDS) plot showing the empirical and theoretical associations among the aspects.

2.2 Methodology

2.2.1 Searching Strategy

Following the principles suggested by Boland et al. (2017), keyword searches were performed on SCOPUS and Web of Science (WoS) on 24th and 25th March 2020, respectively. The searches included the title, abstract, and keywords of journal articles, conference papers, and book chapters. The search string was finalized as “(Heritage OR UNESCO OR Touris* OR HUL OR ‘Historic Urban Landscape’) AND (‘social media’) AND {(‘Machine Learning’ OR ‘Deep Learning’ OR ‘Information Retriev*’ OR ‘Text Mining’) OR (‘Graph Theory’ OR ‘Social Network’ OR ‘Complex Network’ OR ‘Network Science’) OR [(Negotia* OR Inclusi* OR Democra* OR ’Decision-making’) AND (Planning OR Management)]}", as visualized in Figure 2.1.

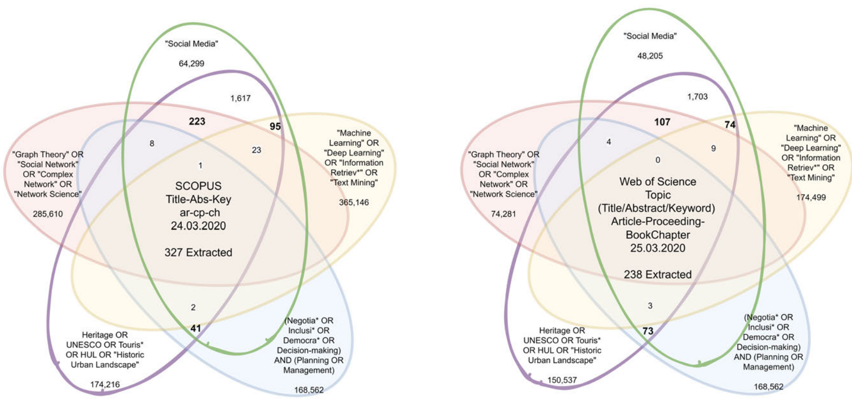


FIG. 2.1 Keyword searching on SCOPUS and Web of Science following the Systematic Literature Review.

The broadest topic of this research which defines the research field is cultural heritage. To keep it general, "(Heritage OR UNESCO OR Touris* OR HUL OR 'Historic Urban Landscape')" is used to catch all the possible related aspects when scholars talk about cultural heritage. All the later searches must also include this set of keywords, as this is the basic research discipline of this dissertation. There are in total 174,216 publications in SCOPUS and 150,637 in WoS available at the moment of search, implying the size of the cultural heritage research community.

Another essential topic of this research is social media, as this is the research object and the data source to answer all the research questions. By combining the term "social media" with the cultural heritage terms, the search results are significantly reduced to 1617 in SCOPUS and 1703 in WoS.

The next level of topics includes three equally important, yet well-separated fields:

1. Concerning the primary methods to deal with the raw user-generated data gathered from social media or other similar data sources, terms related to the computational methods are combined, as "('Machine Learning' OR 'Deep Learning' OR 'Information Retriev*' OR 'Text Mining')";
2. Concerning the secondary methods of network science to structure and analyze the gathered within its spatiotemporal and socioeconomic context, which also reveals the intrinsic topology of social media, terms related to network science are combined, as "('social media') AND {('Machine Learning' OR 'Deep Learning' OR 'Information Retriev*' OR 'Text Mining')";
3. Concerning the ultimate goal of this research, i.e., to enhance social inclusion, public participation, and democratization in heritage management, terms related to these processes are combined, as "(Negotia* OR Inclusi* OR Democra* OR 'Decision-making') AND (Planning OR Management)".

Looking at the combination of the three sub-topics, the quantity of research objects drops critically again. The intersection of machine learning and network science in the field of cultural heritage contributes to 23 articles in SCOPUS and 9 in WoS. Similarly, the intersection of machine learning and inclusive planning contributes to 2 articles in SCOPUS and 3 in WoS. And the quantity for the intersection of network science and inclusive planning is 8 and 4, respectively. This huge contrast between the grand size of the individual communities and the lack of publications when combining terminologies together implies that there exists a huge research gap to understand social media computationally in the field of cultural heritage planning and management, which points out the necessity and significance of this dissertation.

For the further process of systematic review in the next step, all the relevant records that consist of all the essential topics (cultural heritage and social media) and at least one of the technical topics (machine learning, network science, and/or inclusive planning) are considered. The search intended to extract publications related to the use of social media in heritage studies with specific methodological focuses on machine learning, network analyses, and/or inclusive planning, as they were most relevant to the proposed research framework shown in Chapter 1. Initially, 327 publications were extracted from SCOPUS and 238 from Web of Science, making up a total of 431 publications for screening and reading after merging and removing the redundant ones. As a note, terms concerning both scenarios “baseline/everyday" and “activated/event-triggered" were not used explicitly in the search, since it is not desirable to refine the results to only focus on the scenarios. Ideally, both scenarios would be automatically included in the extracted publications since the classification framework is assumed to cover most heritage-related empirical studies using social media data. The same argument is valid for the terms related to the level of analysis, i.e., “spatial", “temporal", or “spatiotemporal".

2.2.2 Inclusion/Exclusion Criteria

Two sets of inclusion/exclusion criteria were applied respectively for the two stages of title/abstract screening and full-text reading (Boland et al., 2017) to filter out the articles weakly related to the proposed framework.

For the first stage, publications that only focused on the hospitality industry including hotel, transportation, and/or gastronomy in the tourism sector (96 records), museum exhibitions on-site or online (17 records), human mobility and destination recommendation systems (26 records), social media marketing strategy (113 records), developing computation algorithms in seek of better performance (76 records), and those that were not openly accessible (11 records) were excluded, yielding 92 publications for the second stage of full-text reading.

For the second stage, 19 publications were further excluded from the qualitative synthesis since they did not include empirical studies (9 records), were not related to any aspect of User-Generated Content on social media (5 records), were literature review papers (4 records), or did not have an English version available (1 record).

As a result, 73 publications were included and analyzed with quantitative description, qualitative synthesis, and statistical tests. The selection and screening of the literature are presented with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) standard (Liberati et al., 2009; Moher et al., 2009; Boland et al., 2017), as shown in Figure 2.2.

2.2.3 Analytical Strategies

The systematic review mainly answers the following questions:

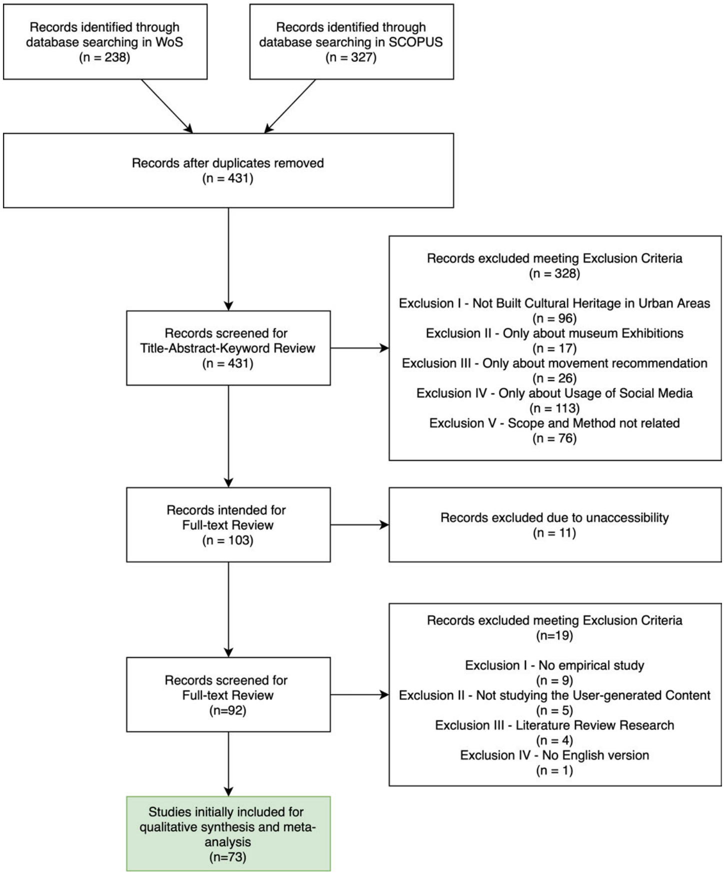


FIG. 2.2 The systematic literature review protocol.

- **Research Context - Spatial** Where have the studies been conducted both globally and in Europe? How were the study cases selected and what was the geographical distribution of the research institute and their study cases?
- **Research Context - Temporal** What were the most frequently used social media platforms? What was the time span of the data collection period and how large were the collected datasets of User-Generated Contents, especially in both baseline/everyday and activated scenarios?
- **Research Content** What information was mostly collected and analysed? What was the focus of the studies, in terms of the main objectives, the focus group, as well as the analytical approach?
- **Research Methodology** What were the most frequently used methods, models, and algorithms for analysing and understanding User-Generated Content, and what was the trend of the methodological development and usage?

During full-text reading, a systematic qualitative coding scheme was used to extract relevant information from the 73 included publications. In this scheme, there are a total of 154 variables under the thematic topics of the research context (geographical distribution, case study category, data collection), research content (main objectives, focus group, analytical approach), research methodology (methods, models, algorithms), and research presentation, not judging from any political standpoint. Among the coding variables, 110 are binary (i.e., denoting if the publication satisfies a certain aspect or not), while the others are numerical, categorical, or textual. These binary variables are not mutually exclusive, meaning multiple aspects can be given a positive value within the same category. On the one hand, those variables are the categorical attributes to describe the included publications; on the other hand, the publication records also give weight to each of the aspects, the co-occurrence of which may suggest the similarity of the aspects under different thematic topics. Under such consideration, a MDS analysis is conducted in Python with sklearn library on the matrix of the records and aspects, in order to show the similarities among all the concept aspects, as well as the clusters composed of the highly related concepts (Kruskal, 1964). This approach is different from most bibliometric studies (van Eck and Waltman, 2014; Alviz-Meza et al., 2022) since the latter most frequently deals with the uni-partite co-occurrence of publication records, while the former makes use of the bipartite record-aspect matrices/networks. A comprehensive list of all the variables in the coding scheme can be found in Appendix B.

Specifically, publications were classified as “baseline/ everyday", “activated", or “both", corresponding with the scenarios defined in Section 1.1. If one study explicitly declared an event as the main focus for the case study, for example, an international exposition, a natural disaster, and/or the crisis reaction for a destination, it was labeled with “activated”; otherwise, if a study focused on the ordinary status of the case, it was labeled with “everyday”; for the special cases where both scenarios were emphasized and compared explicitly, they were labeled as “both".

2.3 Results

2.3.1 Geographical Distribution of the Studies

The included 73 records have a wide distribution globally in terms of the locations of the corresponding research institutions and the study cases. All geographical regions are present as study cases in the literature, though Europe (41 records) and Asia (22 records) are the most studied regions, followed by North America (3 records), Latin America (1 record), Oceania (1 record), and Africa (1 record). Furthermore, 3 studies use the globe as their study case without emphasizing any single region (Monteiro et al., 2014; Paldino et al., 2015; Zhang et al., 2019).

In order to understand the geographical distribution of the studies, a directed graph can be constructed, where each research record can be interpreted as one or more edges, connecting the research institutions as sources, and study cases as targets. Almost half of the records (36) have multiple research institutions participating in the research locally or globally, contributing to edges with multiple sources pointing to single targets. 16 records have multiple study cases, contributing to edges with single sources pointing to multiple targets. 9 records have both multiple research institutes and multiple study cases, resulting in fully connected edges. Edges connecting the same nodes as source and target are also valid, referring to research conducted with a case that is the same as where the research institute is located. The graph can be mapped geographically on a global scale and in a European scale, respectively, as shown in 2.3 and 2.4.

FIG. 2.3 The geographical distribution of research institutes and study cases on a global scale, with a distribution histogram of the distance of edges on a log scale. The size of the point for cases shows the in-degree of the city.

From the geographical distributions, it can be seen clearly that Europe (centred with Italy) and Eastern Asia (especially China) are the two hotspots for research on understanding User-Generated Content on Social Media about cultural heritage properties and tourist destinations. From the histograms of edge distance, it can be observed that short-distance cooperation (less than 2500km globally and less than 1000km in Europe) and local studies where the research institute and study case are based in the same city are still the mainstream. Still, studies with wider cooperation from a longer distance keep on appearing in the past decade (Miah et al., 2017; van Weerdenburg et al., 2019; Thakuriah et al., 2020).

As the term “cultural heritage” is not a restriction for the literature search, a lot of records screened are not originally conducted by researchers from the field of Heritage Management. Rather, the majority of the records come from the fields of Tourism (25 records), Computer Science (18 records), and spatial planning (11 records). Interestingly, as only 22 records (30.1%) declare explicitly that their study case is a cultural heritage property, e.g., in Bellens et al. (2016); Chianese et al. (2016); Micera and Crispino (2017); Campillo-Alhama and Martinez-Sala (2019);

Ginzarly et al. (2019), 76.7% of the included studies (56 records) were conducted with at least one study case in a city with urban areas inscribed in the UNESCO WHL.

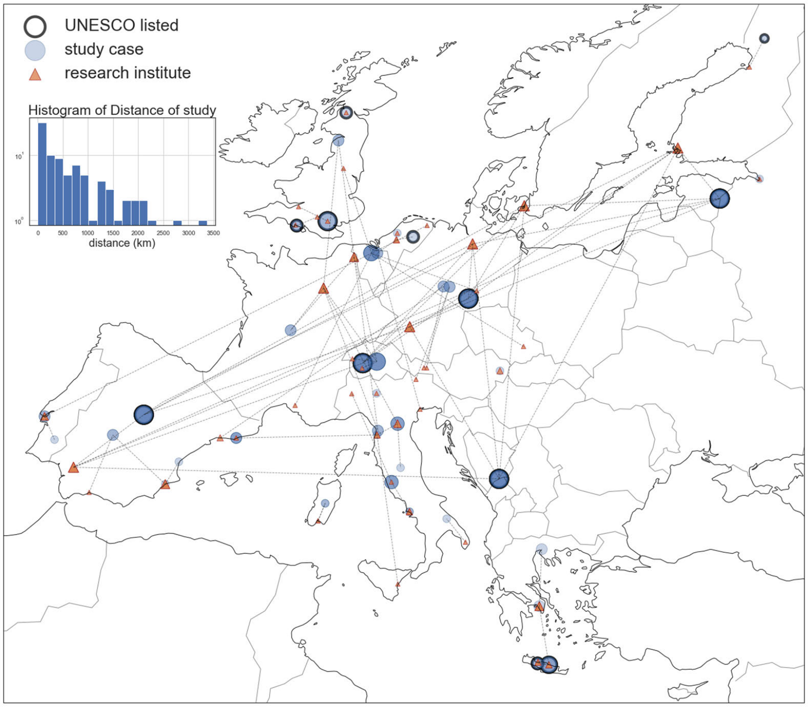


FIG. 2.4 The geographical distribution of research institutes and study cases on a European scale, with a distribution histogram of the distance of edges on a log scale. The size of the point for cases shows the in-degree of the city. UNESCO World Heritage Properties are highlighted with a ring.

2.3.2 Studied Social Media Platforms and User-Generated Content

For the included research, 54 used a single social media platform to collect the data, 8 used two platforms to compare, and 11 used more than two platforms. The study which includes the most various data sources (Martí et al., 2021) used five different social media platforms (Instasights, Foursquare, Twitter, Google Places, and Airbnb). The most frequently used social media platforms are Twitter (22 times) and Flickr (19 times), followed by TripAdvisor (9 times), generic destination websites (7 times), Facebook (7 times), and Instagram (6 times). This ranking does not mimic directly the popularity of social media platforms in daily life, where a report in Australia (Burgess et al., 2015) shows the most used social media sites are Facebook, YouTube, Instagram, Snapchat, LinkedIn, Twitter, and Pinterest, in descending order[2]. The most possible reasons for the inconsistency of the popularity and the choice of Twitter and Flickr as data sources include their accessibility through free and open APIs (Miah et al., 2017; Wu et al., 2018; Clemente et al., 2019), the policy for sharing, viewing, and researching with privacy consideration (Thakuriah et al., 2020; van Dijck, 2011), as well as the quality of data structure (Gabrielli et al., 2014; Monteiro et al., 2014; Clemente et al., 2019; Gosal et al., 2019). Furthermore, Twitter is especially popular with the activated condition (78.6%), due to its timeliness and low time lag for updates (Williams et al., 2017).

According to the convention, the main research topics of social media and social network analysis can be clustered as **Structural** Analysis, which focuses on the relational data of the network, the community, the node and edge properties, as well as the dynamic of network evolution; and **Content**-based Mining, which focuses on the textual, pictorial and multimedia data (Aggarwal, 2011). As the geospatial aspect of social media is neither directly related to the structure, nor to the content, yet very important in the field of tourism, spatial planning, and cultural heritage studies, a third dimension "**Context**" is added as the social network analysis aspect. Whether a record focuses on the structure, content, or context aspect of the social media data is individually coded among all the records, meaning one single record can have multiple focuses. Content-based mining is the most common topic in the included records (76.7%) (Hashida et al., 2018; Feizollah et al., 2019; Giglio et al., 2019; Salur et al., 2019), followed by geographical context mapping (64.4%) (Floris and Zoppi, 2015; Guo et al., 2018; Clemente et al., 2019; Giglio et al., 2020). The structure, dynamics, and mechanism of the social media data are not very commonly studied in the reviewed articles (34.2%) (Barbagallo et al., 2012; Gabrielli et al., 2014; Bellens et al., 2016; Junker et al., 2017). Only two records discussed on all three aspects of structure, content, and context (Sun et al., 2017; Qi et al., 2018).

The type of user-generated content data collected and analyzed from social media is also recorded. Textual data (comments, reviews, tags, captions, titles, et al.) is the most studied data type (75.3%), followed by geolocation (57.5%), timestamp (53.4%), and visual/pictorial data (46.6%). The user interaction (retweet, share, like, mention et al.) (27.4%) (Barbagallo et al., 2012; Campillo-Alhama and Martinez-Sala, 2019; Park et al., 2019; McMullen, 2020), the ratings and scores (11.0%) (Dickinger and Lalicic, 2016; Dickinger et al., 2017; van der Zee and Bertocchi, 2018; Taecharungroj and Mathayomchan, 2019), and video data (4.1%) (Mariani et al., 2016; Micera and Crispino, 2017; Del Vecchio et al., 2018) are not very commonly included. It is also worth noting that although 42.5% of the records collect user information including user ID, name, nationality, origin, etc., only one record (Thakuriah et al., 2020) declared explicitly the data privacy issue and explains the technique they use to anonymize or pseudonymize the data.

[2]Yellow Social Media Report 2018 - Consumer, which is available at https://www.yellow.com.au/wp-content/uploads/2018/06/Yellow-Social-Media-Report-2018-Consumer.pdf

2.3.3 Baseline/Everyday and Activated Scenarios

Among the 73 included publications, 9 were about the “activated" scenario while 59 were about the “baseline" scenario, and only 5 were about both. A summary of the publications with the label of either “activated" or “both" could be seen in Table 2.1. The majority of publications mainly discussed the use of social media for heritage planning without mentioning any special events. The trend of such distinctions could be seen in Figure 2.5. In the past decade, though the research about normal everyday scenarios has kept growing, especially after 2016, studies explicitly about the activated scenario dealing with event-related heritage management issues remained scarce, let alone studies combining and comparing the two.

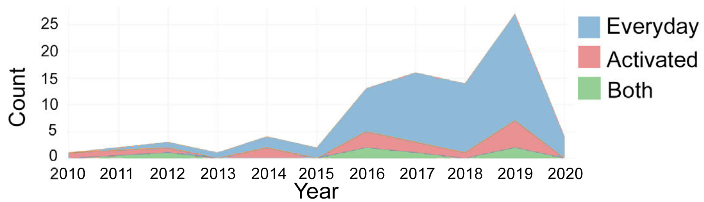


FIG. 2.5 The number of publications with label of “everyday", “activated", and “both" from 2010 to 2020.

Like the example of fire and flood as radical events shown in Section 1.1, plenty of activation came as consequences of natural disasters (Fukui and Ohe, 2019; Park et al., 2019; Taecharungroj and Mathayomchan, 2019). However, activated public engagement on social media could also happen after political events (Claster et al., 2010; Monteiro et al., 2014; Chaabani et al., 2018), large-scale cultural activities (Gabrielli et al., 2014; Amato et al., 2016; Williams et al., 2017; Vassakis et al., 2019), or even general daily events (Barbagallo et al., 2012; Battiato et al., 2016; Chianese et al., 2016; Campillo-Alhama and Martinez-Sala, 2019), therefore not necessarily negative, nor radical.

Among the studies, the majority focused on the regional- or national-level voices from either local residents and/or tourists as a concerned community, while Monteiro et al. (2014) brought together the discussion about events such as the possible delisting of a UNESCO World Heritage property in Australia into the global context, showing the local and global sensitivities regarding World Heritage based on the spatiotemporal evolution of related tweets. Researchers mainly used content-based information (e.g., words, pictures), network structure (e.g., user interaction, connectivity, temporal dynamics), and contextual aspects (e.g., geo-location) from the social media platforms to draw their conclusions in activated scenarios (Aggarwal, 2011), mainly from Twitter due to its timeliness and low time lag for updates (Williams et al., 2017). Natural Language Processing tools such as sentiment analysis and topic models have been applied to mine the public opinions of heritage properties triggered by events (Claster et al., 2010; Gabrielli et al., 2014; Monteiro et al., 2014; Chaabani et al., 2018; Taecharungroj and Mathayomchan, 2019; Fukui and Ohe, 2019), and graphs/networks were constructed to find out the community structures (Barbagallo et al., 2012; Williams et al., 2017), critical influencers

(Barbagallo et al., 2012; Campillo-Alhama and Martinez-Sala, 2019), popular destinations (Gabrielli et al., 2014), and to make personalized recommendations (Amato et al., 2016; Battiato et al., 2016). However, none of the presented studies in Table 2.1 have applied or developed heritage-specific tools targeted at revealing cultural significance, i.e., values and attributes of heritage properties, which should become an important initial step for the proposed framework (Bai et al., 2021a).

TABLE 2.1 A brief overview of the investigated publications in the systematic literature review classified as either "activated" or "both".

Study	Scenario	Data Source	Case Study	Event	Polarity	Collection Duration	Data Type
Amato et al. (2016)	Activated	Twitter	Naples, Italy*	Assumptive guided tour for masterpieces of Caravaggio	Positive	-	Content & Context
Barbagallo et al. (2012)	Both	Twitter	Milan, Italy*	General negative comments in tourism and culture domain	Negative	Jan-Apr 2011	Structure
Battiato et al. (2016)	Both	The Social Picture	Pisa, Italy*	Cultural-related public events	Positive	-	Content & Context
Campillo-Alhama and Martinez-Sala (2019)	Activated	Facebook & Twitter	40 Spanish properties*	Heritage-property-related public events	Positive	Jan-Dec 2017	Structure & Context
Chaabani et al. (2018)	Activated	Twitter	Tunis, Tunisia*	Arab Spring Revolution	Negative	10th-17th July 2016	Context
Chianese et al. (2016)	Both	Twitter	Naples, Bari, Venice & Rome, Italy*	Heritage-property-related public event	Neutral	Dec 2014 - May 2015	Content & Context
Claster et al. (2010)	Activated	Twitter	Bangkok & Phuket, Thailand	Red Shirt Demonstration	Negative	Nov 2009 - May 2010	Context
Fukui and Ohe (2019)	Activated	Twitter	Iwate, Japan*	Earthquake and Tsunami	Negative	2010-2019	Content & Context
Gabrielli et al. (2014)	Activated	Twitter & Foursquare	Barcelona, Spain*	Mobile World Congress 2012	Positive	Feb-Mar 2012	Content & Structure
Monteiro et al. (2014)	Activated	Twitter	The Globe*	Cases including possible delisting of Tasmanian Wilderness from World Heritage	Negative	Dec 2013 - Jan 2014	Content & Context
Park et al. (2019)	Activated	Facebook	Florida	Landfall of Hurrican Irma	Negative	Aug-Sep 2017	Structure & Context
Taecharungroj and Mathayomchan (2019)	Both	TripAdvisor	Phuket, Thailand	Wave-hit on tour boat	Negative	-	Content & Context
Vassakis et al. (2019)	Both	Instagram, Facebook, Foursquare, & Twitter	Heraklion & Chania, Greece	Video shooting of a popular singer	Positive	Nov-Dec 2017	Content & Context
Williams et al. (2017)	Activated	Twitter	Bournemouth, UK	Bournemouth Air Festival	Positive	2011-2015	Structure & Context

*The case study contains at least one UNESCO World Heritage property.

Furthermore, Figure 2.6 shows information about data collection for all the included publications, which contained the collection period, duration, as well as total size of collected data, when related information has been explicitly provided by the authors. It reveals that for the investigated studies within both scenarios, the data collection duration varied significantly, ranging from 3 weeks (Gabrielli et al., 2014) to 12 years (Junker et al., 2017; Ginzarly et al., 2019; Barros et al., 2020). Moreover, a shorter collection duration does not necessarily mean a smaller data quantity. Meanwhile, as the popularity of big data has been growing in the past decade, not all recent studies are processing "bigger data" than before. Remarkably, the two studies with the largest data on the scale of 10^8 were both conducted more than 5 years ago (Claster et al., 2010; Paldino et al., 2015). Furthermore, t-tests showed that the studies focusing on the activated scenarios have a significantly shorter data collection duration than the studies merely focusing on everyday scenarios $(t = -3.22, p < .01)$, while there is no significant difference found with the data quantity for different scenarios $(t = 1.50, p = .14)$. This again suggests that specific tools and algorithms to handle the large datasets on social media to obtain potentially useful information for heritage practitioners and researchers are urgently needed.

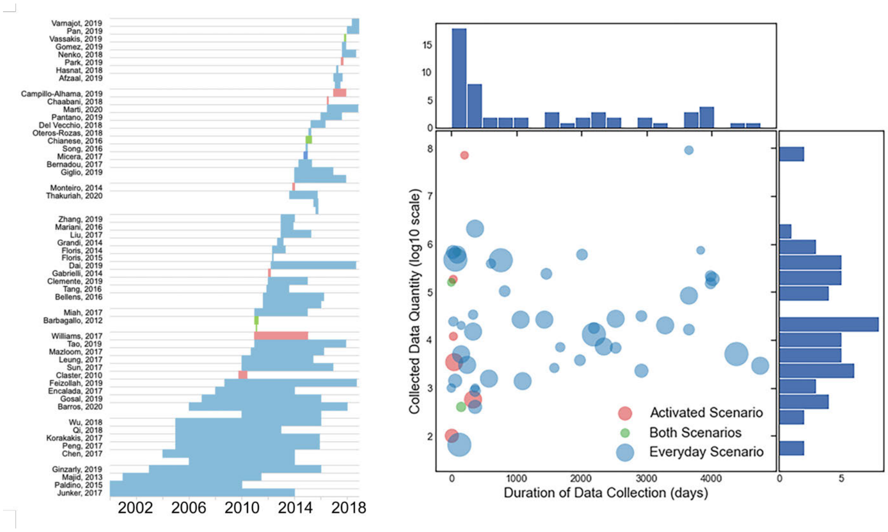


FIG. 2.6 Left: the data collection duration (start time-end time) of the reviewed research; Right: the relationship between data collection duration and data quantity. The sizes of points show the proximity of the research to now, meaning that the later the record is published, the larger the point. The distributions of the duration and quantity are shown with histograms, at the top and right, respectively. The colors in both graphs distinguish literature focused on "everyday", "activated" or "both" scenarios.

2.3.4 Research Content and Focuses

Three main aspects are coded under the thematic topic of research content, i.e., research objective, focus group, and analytical approach. Again, all binary variables under each thematic topic allow multi-label coding and are not mutually exclusive.

For the research objective, different stages of the research-application cycle are reached. As a field of study in human attitude, cognition, perception, and behavior, it is commonly believed that there are four main objectives as a sequence in Psychology – describe, explain, predict, and control (Gerrig et al., 2015). The same procedure can be valid for urban studies and heritage studies when it concerns humans. In the included records, the majority (84.9%) reached the goal of describing the existing property (Monteiro et al., 2014; Miah et al., 2017; Hasnat and Hasan, 2018; Barros et al., 2020), whereas fewer records (21.9%) tried to explain the mechanism of the current phenomena (Gabrielli et al., 2014; Schirpke et al., 2018; Clemente et al., 2019; Ginzarly et al., 2019), and only one record intended to predict future performance through analyses, modelings, or simulations (Qi et al., 2018). However, a slightly higher percentage (30.1%) tried to suggest management policies and design principles as a way of controlling and improving the current situation (Mariani et al., 2016; Song and Kim, 2016; Miah et al., 2017; Wu et al., 2018; Taecharungroj and Mathayomchan, 2019). The unbalance and the jump over the steps in the sequence can be due to the difference in fields, yet also suggest where the research gap exists. A large proportion of the records (76.7%) used their research as an exploration to propose a new workflow (Paldino et al., 2015; Peng and Huang, 2017; Oteros-Rozas et al., 2018; Afzaal et al., 2019), as the usage of social media user-generated content is generally new in the field. Those studies from the field of computer science usually also had the objectives of proposing a new algorithm or improving an existing one (37.0%) (Chen et al., 2017; Junker et al., 2017; Pan et al., 2019; Ramanathan and Meyyappan, 2019), creating a new platform either for collecting user-generated content or for displaying the analytical results to the managers and tourists (11.0%) (Battiato et al., 2016; Liu et al., 2017; Sansonetti et al., 2019; Thakuriah et al., 2020), and developing a better recommendation system for promoting destinations as products (11.0%) (Majid et al., 2013; Mazloom et al., 2017; Korakakis et al., 2017; Figueredo et al., 2018). Besides, a few records (15%) also explored the usage of the social media platform by managers and tourists (Dickinger and Lalicic, 2016; Del Vecchio et al., 2018; Nenko and Petrova, 2018; Varnajot, 2019; McMullen, 2020).

For the focus group, three types of stakeholders are mostly involved: visitors, locals, and officials. The discussion of “whose heritage" has since long been on board in the field of cultural heritage management (Rakic and Chambers, 2008; Taylor and Gibson, 2017; Oteros-Rozas et al., 2018). The Recommendation on the HUL states explicitly that a diverse cross-section of stakeholders should be involved and empowered to identify the key values in their urban areas (UNESCO, 2011; Bandarin and Van Oers, 2012). In the involved studies, the locals start to get the research attention (65.8%) (Chianese et al., 2016; Williams et al., 2017; Nenko and Petrova, 2018; Dai et al., 2019; Ginzarly et al., 2019), but are still under the dominant focus of tourists (95.9%), as the majority of the included records are from the field of tourism. 32.9% records claimed explicitly that there might be a perceptual and behavioural difference between stakeholders (Mariani et al., 2016; Bernadou, 2017; Leung et al., 2017; Campillo-Alhama and Martinez-Sala, 2019; Park et al., 2019). Interestingly, knowing the existence of differences between locals/officials and visitors, 6 studies took measures to exclude the data generated by either locals (Grandi and Neri, 2014; Encalada et al., 2017; Liu et al., 2017; Martínez-Sala et al., 2018) or officials (Hasnat and Hasan, 2018; Tao et al., 2019) from the dataset in order to understand the visitors more thoroughly. While the research from different fields may target on different focus groups, e.g., tourism studies may focus more on the tourists as demanders (Dickinger and Lalicic, 2016; Miah et al., 2017; van der Zee and Bertocchi, 2018; Martí et al., 2021), marketing and management studies may focus more on managers as suppliers (Mariani et al., 2016; Miah et al., 2017; Pantano and Dennis, 2019; McMullen, 2020), spatial planning studies may contribute more to government oriented to future development (Floris et al., 2014; Tang and Li, 2016; Thakuriah et al., 2020; Martí et al., 2021), destination management research may care more about the destination or property itself (Song and Kim, 2016; Sun et al., 2017; Nenko and Petrova, 2018; Taecharungroj and Mathayomchan, 2019), they are all connected to and concerned with heritage management in some way.

For the analytic approach, the majority of the included records (65.8%) used computational algorithms, models, and/or tools to facilitate the analytical understanding of the user-generated content (Abeysinghe et al., 2018; Afzaal et al., 2019; Sansonetti et al., 2019; van Weerdenburg et al., 2019). Half of the records (50.7%) used either descriptive statistics or statistical tests to aid the story-telling and to prove the hypothesis (Majid et al., 2013; Miah et al., 2017; Clemente et al., 2019; Ginzarly et al., 2019). Spatial analysis with the help of Geographic Information System/Science (GIS) and spatial mapping (34.2%) (Monteiro et al., 2014; Tang and Li, 2016; Schirpke et al., 2018; Dai et al., 2019) and qualitative analysis with in-detailed observation and description (31.5%) (Mariani et al., 2016; Song and Kim, 2016; Bernadou, 2017; Oteros-Rozas et al., 2018) were also applied with regard to the research fields of the authors. Graph theory approach modeling the relational structure and dynamics of the social networks (19.2%) (Majid et al., 2013; Bellens et al., 2016; Williams et al., 2017; Sansonetti et al., 2019) and mathematical formulations with some degrees of abstraction (12.3%) (Majid et al., 2013; Chianese et al., 2016; Mazloom et al., 2017; Pan et al., 2019) were the least touched areas. Concerning the general analytical approach and the format of the collected user-generated data, detailed methodologies of one of the five topics, i.e., Natural Language Understanding (67.1%), Image Recognition (28.8%), Machine Learning (63.0%), Spatial Mapping (46.6%), and Graph Analysis (30.1%), were also recorded and coded, which will be further discussed in detail in Section 2.3.5. It is worth noting that though NLP and CV are recently the terminologies from artificial intelligence, the terms were interpreted literally during coding, allowing for more traditional ways of manually reading, tagging, processing, and understanding information in the unstructured texts and images (Albers and James, 1988; McMullen, 2020). This also explains why the proportion of studies with an analytical approach to natural language understanding can be even larger than those with a computational one.

2.3.5 Analytical Approach

As already briefly mentioned in Section 2.3.4, the models, methods, algorithms, tools used, and other methodological aspects are coded with respect to the five broad categories of Graph Theory, Spatial Mapping, Machine Learning, Natural Language Understanding, and Image Recognition.

For the 22 included records which involved **Graph Theory** or [social] network analysis, they mainly constructed networks or graphs about terminologies, concepts and sentiments (knowledge graph, 8 records) (Grandi and Neri, 2014; Monteiro et al., 2014; Afzaal et al., 2019; Tao et al., 2019), visitation patterns and occurrence associations (spatial network, 10 records) (Paldino et al., 2015; Korakakis et al., 2017; Qi et al., 2018; van der Zee and Bertocchi, 2018), and social media interactions (social network, 4 records) (Barbagallo et al., 2012; Bellens et al., 2016; Williams et al., 2017; Abeysinghe et al., 2018). The three types of graphs are useful information for the content, context, and structure aspects of social media UGC, respectively. The knowledge graph usually exists as a by-product for text-mining, as it explains the relationship between topic keywords (Monteiro et al., 2014) and the association between sentiment and its aspect (Afzaal et al., 2019); or as a by-product for creating a recommendation system, as it connects terms, places, et al., to the user (Sansonetti et al., 2019). The spatial graph consists of three types:

1. bipartite or multipartite graphs associating the concepts and/or visitors to their spatial context (Majid et al., 2013; Mazloom et al., 2017);
2. monopartite directed graphs recording the consecutive travel paths of the visitors (Gabrielli et al., 2014; Wu et al., 2018);
3. monopartite undirected graphs recording the co-visiting and cooccurrence pattern (Qi et al., 2018; van der Zee and Bertocchi, 2018).

The social networks keep records of the retweeting, mentioning, and replying relationship in the unit of either user or post (Barbagallo et al., 2012; Williams et al., 2017). The most frequently used network analysis metrics include degree centrality (8 times), density (5 times), degree distribution (4 times), and betweenness centrality, clustering coefficient, and core-periphery structure (3 times). Metrics about the dynamic property of the graph such as the spreading speed of the information are only mentioned once in Barbagallo et al. (2012). Most of the concepts here have been introduced in Section 1.2.4.

For the 34 included records which involve **Spatial Mapping**, the majority (28 records) visualizes the quantity of the collected UGC in the form of heatmap (Battiato et al., 2016; Nenko and Petrova, 2018; Clemente et al., 2019; Ginzarly et al., 2019) or scatter plots (Bellens et al., 2016; Liu et al., 2017; Giglio et al., 2019; Tao et al., 2019). 14 records also included the results of spatial statistical analysis, e.g. the clusters based on DBSCAN (Majid et al., 2013; Miah et al., 2017; Peng and Huang, 2017; Giglio et al., 2019), the local Moran Index (Floris and Zoppi, 2015; Encalada et al., 2017), the spatial similarity and heterogeneity (Oteros-Rozas et al., 2018; Zhang et al., 2019), etc.. 10 records mapped the relevant topics and/or sentiments revealed in the UGC with respect to their corresponding locations (Clemente et al., 2019; Ginzarly et al., 2019). Only 2 records also included the temporal aspects in the mapping by comparing the maps of different seasons (Schirpke et al., 2018) or event-related time periods (Monteiro et al., 2014).

46 records were coded as **Machine Learning** research since they included processes of classification, clustering and/or regression, all of which were introduced in Section 1.2.3. Although coded as Machine Learning research in this review, a small number of records also applied semi-automatic (6 records) (Monteiro et al., 2014; Mariani et al., 2016; Hasnat and Hasan, 2018; Clemente et al., 2019; Ginzarly et al., 2019; Park et al., 2019) or manual (3 records) (Qi et al., 2018; Wu et al., 2018; Dai et al., 2019) processes to achieve the classification/clustering/regression tasks.

- 24 records involved the classification task, which is a supervised learning task for categorical data. Among the classification studies, the most researched aspect (9 records) is to categorize sentiments (Amato et al., 2016; Hashida et al., 2018; Afzaal et al., 2019; Salur et al., 2019). Other notable categorizations include Cultural Ecosystem Services (Clemente et al., 2019; Dai et al., 2019), whether a depicted scene is tangible or intangible heritage (Ginzarly et al., 2019), and to identify the city where the photos were taken (Zhang et al., 2019).
- 19 records involved the clustering task, which is an unsupervised learning task for categorical/numerical data, among which the most researched aspect (7 records) is to identify the popular locations geographically (Majid et al., 2013; Leung et al., 2017; Miah et al., 2017; Giglio et al., 2020). Other application scenarios included the detection of user communities (Williams et al., 2017; Thakuriah et al., 2020) and discussion topics (Park et al., 2019; Taecharungroj and Mathayomchan, 2019).
- 7 records involved the regression task, which is a supervised learning task for numerical data, predicting variables such as the tourists' preference (Floris and Zoppi, 2015; Schirpke et al., 2018), number of visitors (Fukui and Ohe, 2019), and levels of user engagement (Mariani et al., 2016).

The majority (29 records) applied traditional machine learning models and algorithms without using variants of deep neural networks, such as SVM (Support Vector Machine) (Dickinger et al., 2017; Hasnat and Hasan, 2018; van Weerdenburg et al., 2019), NB (Naïve Bayes) (Chaabani et al., 2018; Ramanathan and Meyyappan, 2019; Taecharungroj and Mathayomchan, 2019), and variants of DBSCAN (Majid et al., 2013; Miah et al., 2017; Korakakis et al., 2017; Giglio et al., 2020) for the tasks. As a later development within the framework of machine learning, Deep Learning trains neural network models with hidden layers through massive data, which triggers the development and boosts the performances of both research applications of Computer Vision and Natural Language Processing. A smaller number (9 records) of the included records made use of deep learning models, such as CNN (Convolution Neural Network) (Battiato et al., 2016; Hashida et al., 2018; Gomez et al., 2019; Zhang et al., 2019), general RNN (Recurrent Neural Network) (Amato et al., 2016;

Abeysinghe et al., 2018; Feizollah et al., 2019), and an RNN variant called LSTM (Long Short Term Memory) (Abeysinghe et al., 2018; Feizollah et al., 2019).

Among the 21 records involving the procedure of **Image Recognition**, the majority (10 records) of them still applied manual coding and visual content analysis for the collected image data (Song and Kim, 2016; Bernadou, 2017; Oteros-Rozas et al., 2018; Ginzarly et al., 2019; McMullen, 2020), followed by deep learning (7 records) (Battiato et al., 2016; Mazloom et al., 2017; Figueredo et al., 2018; Giglio et al., 2019; Gomez et al., 2019; Zhang et al., 2019; Thakuriah et al., 2020), traditional machine learning (3 records) (Miah et al., 2017; Guo et al., 2018; Gosal et al., 2019), and hybrid methods (1 record) combining manual work with computational models (Martí et al., 2021). Although deep learning already took up the dominant position of Computer Vision research in 2010 and became quite successful for image recognition tasks (LeCun et al., 2015; Goodfellow et al., 2016; Zhang et al., 2020), its application is still not as popular as it should be within the included research about heritage and tourism studies in the past 5 years. The included records concerned with the scenes or contexts where the images were taken (15 records) (Bernadou, 2017; Guo et al., 2018; Gomez et al., 2019; McMullen, 2020), the main topic of the depicted scenes (14 records) (Peng and Huang, 2017; Figueredo et al., 2018; Ginzarly et al., 2019; Martí et al., 2021), as well as the object appearing in the images (11 records) (Giglio et al., 2019; Gosal et al., 2019; Zhang et al., 2019; Thakuriah et al., 2020).

According to a review paper about the use of sentiment analysis in the field of tourism (Alaei et al., 2019), the methods used in the literature can be clustered into three groups: machine learning, rule-/dictionary-based, and hybrid approaches. The machine learning approach, either supervised or unsupervised, applies methods such as SVM, NB, K-means to train a classifier or to enable clustering tasks. The rule-/dictionary-based approach, however, first builds a context-related lexicon and uses such lexicon and fine-tuned linguistic rules to understand the sentiment orientation of the texts. And the hybrid approach combines the other two approaches in parallel. Although the category **Natural Language Understanding** coded here does not necessarily mean sentiment analysis, which is only one possible application of the former concept, a similar clustering of methodological approaches can be found in the 49 included records of this review. Moreover, as a matter of consistency, "deep learning" could be separately listed from the machine learning approach due to its popularity and enormous contribution to NLP. Before 2017, the dictionary-based approach (14 records) was dominant in the literature (Claster et al., 2010; Majid et al., 2013; Chen et al., 2017; Liu et al., 2017), since which the traditional machine learning (10 records) (Del Vecchio et al., 2018; Muangon et al., 2018; Taecharungroj and Mathayomchan, 2019; van Weerdenburg et al., 2019), deep learning (8 records) (Sun et al., 2017; Mazloom et al., 2017; Gomez et al., 2019; Thakuriah et al., 2020), and hybrid (8 records) (Peng and Huang, 2017; Chaabani et al., 2018; Ramanathan and Meyyappan, 2019; Vassakis et al., 2019) methods have been growing. Similar to image recognition, manual analysis (9 records) for the text qualitatively was still a popular approach for text understanding in heritage and tourism fields (Mariani et al., 2016; Korakakis et al., 2017; Williams et al., 2017; Campillo-Alhama and Martinez-Sala, 2019). As argued above, even though sentiment analysis is only one specific application of Natural Language Understanding, it still took up the majority of the included records (23 records) (Barbagallo et al., 2012; Amato et al., 2016; Salur et al., 2019; Tao et al., 2019), together with two other popular concerns – context (or aspect) of the text showing the general categories such as food, price, service, etc. (21 records) (Chianese et al., 2016; Miah et al., 2017; Dai et al., 2019; Thakuriah et al., 2020), and the detailed topic of the speech (21 records) (Floris and Zoppi, 2015; Mariani et al., 2016; Liu et al., 2017; Varnajot, 2019). The aspect analysis was often combined with sentiment analysis, making up a specific approach called "aspect-based sentiment analysis" (12 records), distinguishing the different sentiment orientations of users when they are discussing in different contexts (Abeysinghe et al., 2018; Afzaal et al., 2019; Ramanathan and Meyyappan, 2019; Tao et al., 2019). Other frequently researched topics included the association between word and word, words and context or topic, and/or words and place (17 records) (Monteiro et al., 2014; Miah et al., 2017; Ginzarly et al., 2019; Gomez et al., 2019), entity extraction retrieving the names of places, objects, and/or heritage properties (14 records) (Gabrielli et al., 2014; Monteiro et al., 2014; Nenko and Petrova, 2018; Park et al., 2019), and emotions of the text, which is one level more detailed than sentiment where only the direction and valence (positive, negative or neutral) is being considered (5 records) (Moreno et al., 2015; Dickinger and Lalicic, 2016; Abeysinghe et al., 2018; Nenko and Petrova, 2018; Pan et al., 2019).

2.3.6 Associations Between Contexts and Contents

The results of the MDS plotting of all binary variables mentioned in the Sections 2.3.2, 2.3.3, and 2.3.4 are shown in Figure 2.7.

As a method to visualize the level of similarity of individual concepts, the two axes do not have an explicit meaning. Rather, it is the distance of each pair of points on the map that matters: the closer the points, the more similar or more related they are. Many pairs of concepts are consistent with common sense, for example "graph analysis" and "graph theory", "spatial analysis" and "spatial planning", "computational", "technology", and "machine learning", etc. Besides those, some interesting patterns can also be observed:

- Twitter, TripAdvisor, Facebook and Instagram are all closer to the activated scenario than to the everyday/baseline scenario, while Flickr has a similar distance to both the scenarios, suggesting the common and popular choices of social media platforms for different research aims and applicational scenarios.
- Natural Language Understanding is close to computational approach, while Image Recognition is closer to qualitative approach, confirming that many included studies used traditional methods rather than computational models to recognize the content and topic shown on images.
- The higher-level research objectives (explain/ control) are closer to the focus group of officials and government, and to the qualitative and spatial analytical approach.

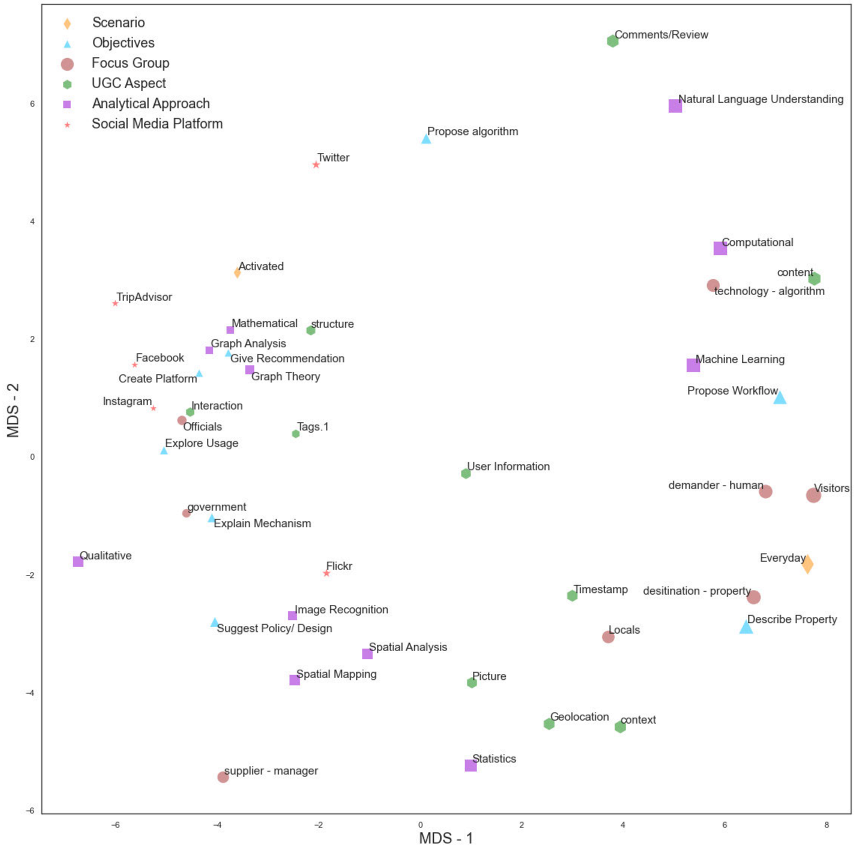


FIG. 2.7 Multi-Dimensional Scaling Plot of all the binary aspects with research content, the research scenario (everyday/activaed), and the key social media platforms based on the screened records. The color and shape of the points show the categories the concepts relate to; the size of the points shows the frequency of the concept appearing in the records.

This can be due to the nature of the direct application of those studies.

– The three dimensions of social media analysis, namely structure, content, and context are broadly distributed in the MDS space. The “structure" dimension is closer to mathematical and graph theory aspects, the activated scenario, and the interaction data from the social media; the “content" dimension is closer to computational aspect, the textual data, and natural language understanding; while the “context" dimension is closer to spatiotemporal data and statistical approach. This confirms the necessity of adding the third dimension of context as argued in Section 2.3.2.

2.3.7 Models, Methods, Algorithms

For all the records, the models, algorithms, and external databases other than the social media platform were also coded during the review process. 33 models, algorithms, and databases were applied more than once in the included records. There is no dominant model, algorithm, or database that appears in more than 10 times or 15% of the records, due to the multi-disciplinary essence of this review. The co-occurrence patterns of the 33 items are visualized in Figure 2.8 as an undirected graph. The size of the nodes and the labels show the number of mentions and applications (degree), and the width of the links shows the times of co-occurrence between the two connected items. Furthermore, an overview of all 33 items concerning algorithms, models, and datasets can be found in Table 2.2.

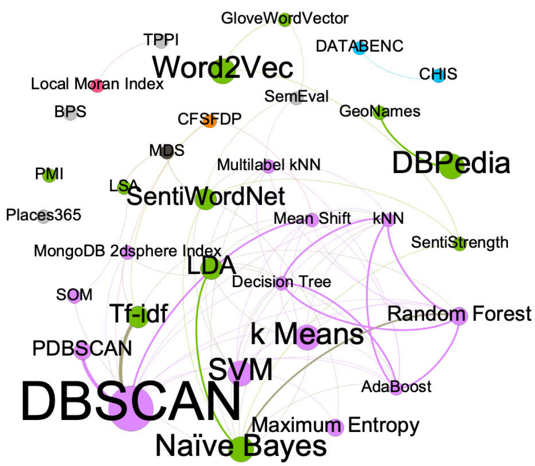


FIG. 2.8 A network (undirected graph) showing the popularity of models, algorithms, and external datasets, and their co-occurrence relationship within the included records. Explanations of the acronyms in the items can be found in Table 2.2. The colors of the nodes correspond to the main application fields of research, such as Machine Learning, Computer Vision, Natural Language Processing, Statistics, Spatial Analysis, Cultural Heritage, and Tourism

The most applied method is DBSCAN, with a strong connection to its variation P-DBSCAN (Ester et al., 1996; Kisilevich et al., 2010; Schubert et al., 2017), which appears 10 times in total. It is a powerful and popular density-based clustering algorithm broadly used for spatial analysis using a very simple parameter selection process. The included records used the geo-locations of social media posts as input data to clusters of Point of Interest (POI), AOIs (areas of interest), or top attractions at the country level (Mendieta et al., 2016), the region level (Hasnat and Hasan, 2018), the city level (Majid et al., 2013; Korakakis et al., 2017; Leung et al., 2017;

Giglio et al., 2019), and the district or even street level (Miah et al., 2017; Peng and Huang, 2017).

TABLE 2.2 An overview of the algorithms, models, and external databases that were applied more than once in the included studies. All the items can be found in Figure 2.8 and are ranked in alphabetic order. Some records excluded after the full-text review step are also counted in the table. The fields are respectively: CH(Cultural Heritage), CV(Computer Vision), ML([general] Machine Learning), NLP(Natural Language Processing), SA(Spatial/Spatiotemporal Analysis), ST(Statistics), and TM(Tourism Management).

Name	Full Name	Type	Referred Paper	Referring Records	Field
AdaBoost		Algorithm	Freund et al. (1996)	Hasnat and Hasan (2018); van Weerdenburg et al. (2019)	ML
BPS	Brand Personality Scale	Database	Aaker (1997)	Dickinger and Lalicic (2016); Moreno et al. (2015)	TM
CFSFDP	Clustering by Fast Search and Find of Density Peaks	Algorithm	Rodriguez and Laio (2014)	Peng and Huang (2017); Wu et al. (2018)	SA
CHIS	Cultural Heritage Information System	Database		Chianese et al. (2016); Castiglione et al. (2018)	CH
DATABENC	Distretto ad Alta Tecnologia per i Beni Culturali*	Database	Bifulco et al. (2016)	Chianese et al. (2016); Castiglione et al. (2018)	CH
DBpedia	Database created with Wikipedia	Database	Auer et al. (2007)	Gabrielli et al. (2014); Chianese et al. (2016); De Angelis et al. (2017); Liu et al. (2017); Figueredo et al. (2018); Sansonetti et al. (2019); Zhang et al. (2019)	NLP
DBSCAN	Density-based spatial clustering of applications with noise	Algorithm	Ester et al. (1996)	Majid et al. (2013); Deeksha et al. (2015); Mendieta et al. (2016); Miah et al. (2017); Korakakis et al. (2017); Leung et al. (2017); Peng and Huang (2017); Hasnat and Hasan (2018); Al-Sultany and Abd Al-Ameer (2019); Giglio et al. (2019)	ML
Decision Tree		Algorithm	Safavian and Landgrebe (1990)	Hasnat and Hasan (2018); van Weerdenburg et al. (2019)	ML
GeoNames		Database	Berman et al. (2012)	De Angelis et al. (2017); Liu et al. (2017)	NLP
GloVe	Global Vectors Word Representation	Model	Pennington et al. (2014)	Abeysinghe et al. (2018); Feizollah et al. (2019)	NLP
k Means		Algorithm	Kanungo et al. (2002)	Williams et al. (2017); Hasnat and Hasan (2018); Schirpke et al. (2018); Taecharungroj and Mathayomchan (2019); Thakuriah et al. (2020)	ML
KNN	k Nearest Neighbours	Algorithm	Manning (2009)	Hasnat and Hasan (2018); van Weerdenburg et al. (2019)	ML
LDA	Latent Dirichlet Allocation	Algorithm	Colace et al. (2014)	Amato et al. (2016); Dickinger et al. (2017); Taecharungroj and Mathayomchan (2019); van Weerdenburg et al. (2019)	NLP
Local Moran I		Algorithm	Anselin (1995)	Floris and Zoppi (2015); Encalada et al. (2017)	SA
LSA	Latent Semantic Association	Algorithm	Landauer et al. (1998)	Ginzarly et al. (2019); Gosal et al. (2019)	NLP
Maximum Entropy		Algorithm	Nigam et al. (1999)	Chaabani et al. (2018); Alaei et al. (2019); Clemente et al. (2019)	ML

TABLE 2.2 Cont.

Name	Full Name	Type	Referred Paper	Referring Records	Field
MDS	Multi-dimensional Scaling	Algorithm	Borg and Groenen (2005)	Ginzarly et al. (2019); Miah et al. (2017)	ST
Mean Shift		Algorithm	Comaniciu and Meer (2002)	Peng and Huang (2017); Hasnat and Hasan (2018)	ML
MongoDB	MongoDB 2dsphere Index	Database		Deeksha et al. (2015); Amato et al. (2016)	ML
ML-KNN	Multilabel k Nearest Neighbours	Algorithm	(Zhang and Zhou, 2007)	Afzaal et al. (2019); van Weerdenburg et al. (2019)	ML
NB	Naïve Bayes	Model	Jindal and Liu (2006)	Claster et al. (2010); Ramanathan and Meyyappan (2019); Salur et al. (2019); Taecharungroj and Mathayomchan (2019); van Weerdenburg et al. (2019)	NLP
PDBSCAN	DBSCAN with Photos	Algorithm	Kisilevich et al. (2010)	Majid et al. (2013); Leung et al. (2017); Miah et al. (2017)	ML
Places365	Images from 365 scene categories	Database, Model	Zhou et al. (2017)	Battiato et al. (2016); Figueredo et al. (2018); Zhang et al. (2019)	CV
PMI	Point-wise Mutual Information	Algorithm	Church and Hanks (1990)	Chen et al. (2017); Pan et al. (2019)	ML
Random Forest		Model	Chan and Paelinckx (2008)	Hasnat and Hasan (2018); Salur et al. (2019); van Weerdenburg et al. (2019)	ML
SemEval	Semantic Evaluation Workshops	Database	Wagner et al. (2014)	Abeysinghe et al. (2018); Afzaal et al. (2019)	NLP
Senti Strength	Detection for Sentiment Strength	Algorithm	Thelwall et al. (2010)	Mazloom et al. (2017); Ramanathan and Meyyappan (2019)	NLP
Senti WordNet	Sentiment Lexicon based on WordNet	Model	Sebastiani and Esuli (2006)	Chaabani et al. (2018); Afzaal et al. (2019); Miah et al. (2017); Ramanathan and Meyyappan (2019)	NLP
SOM	Self Organizing Map	Algorithm	Honkela (1997)	Claster et al. (2010); Gosal et al. (2019)	ML
SVM	Support Vector Machine	Algorithm	Cristianini et al. (2000)	Dickinger et al. (2017); Chaabani et al. (2018); Del Vecchio et al. (2018); Hasnat and Hasan (2018); van Weerdenburg et al. (2019)	ML
tf-idf	Term Frequency - Inverse Document Frequency	Algorithm	Kennedy et al. (2007)	Majid et al. (2013); Deeksha et al. (2015); Peng and Huang (2017); Muangon et al. (2018)	NLP
TPPI	Tourist Positive Preferences Incidence	Index		Floris et al. (2014); Floris and Zoppi (2015)	TM
Word2Vec	Word to Vector	Model	Mikolov et al. (2013b)	Mazloom et al. (2017); Sun et al. (2017); Hashida et al. (2018); Feizollah et al. (2019); Gomez et al. (2019)	NLP

*The High Technology District for Cultural Heritage management of the Campania Region.

DBpedia is the most used external database, mentioned by 4 records. It is a content ontology based on the community contribution of the Wikipedia links, which involves as much as 1.95 million concepts (Auer et al., 2007). Seven major categories in DBpedia are related to culture: art, artwork, artist, sculptor, museum, monument, and humanist, which are also closely related to the tangible and intangible aspects of cultural heritage. DBpedia is used as a tool for the extraction of places and other culture-related categories, in order to further contribute to context-aware recommendation systems (Liu et al., 2017; Sansonetti et al., 2019); to distinguish the locals and tourists based on their registered home location (Gabrielli et al., 2014); and to match the UGC to the DBpedia categories to distinguish the posts related to cultural heritage (Chianese et al., 2016).

Word2Vec is a popular and effective pre-trained word embedding model to transfer textual data into a high-dimensional vector space, which can be further used for machine learning, especially during neural network training (Mikolov et al., 2013a,b). It has been revolutionary for NLP research in the early 2010s and was broadly applied to represent the similarity of words, even though gradually exceeded by context-aware Transformer-based models in the late 2010s (Vaswani et al., 2017; Devlin et al., 2019). However, due to the information gap between the research fields, Transformers have not been broadly applied in the included studies from heritage management and tourism before 2020. Within the included records, the Word2Vec is combined with other deep learning neural-network models such as CNN, RNN, and LSTM, for further sentiment analysis (Sun et al., 2017; Hashida et al., 2018; Feizollah et al., 2019), and for learning the association of verbal information with the visual content and spatial context (Mazloom et al., 2017; Gomez et al., 2019).

SVM, k-Means, Naïve Bayes, Random Forest, and Maximum Entropy being popular machine learning algorithms, tf-idf, Latent Dirichlet Allocation (LDA), and SentiWordNet being popular NLP algorithms or lexicons, they are also mentioned more than 3 times within the included records and are repeatedly mentioned together with each other and other ML and NLP algorithms.

The database of Places365 (Zhou et al., 2017) and its predecessor Places205 (Zhou et al., 2014) are also worth mentioning, as they were applied twice as datasets twice (Figueredo et al., 2018; Zhang et al., 2019) and once as a pre-trained model (Battiato et al., 2016) within the included records. Places365 is a dataset containing more than 10 million images labeled with indoor and outdoor scenes structured in semantic categories, among which there is a specific category of “cultural or historical building/place”, strongly related to the domain of tourism and heritage planning. Several pre-trained CNN models on the Places365 dataset are available and have a considerable level of classification performance.

Moreover, it can also be observed from the edges of the graph in Figure 2.8, that the algorithms, models, and datasets from the same category are more likely to appear together in the research, suggesting the existence of some disciplinary preferences.

2.4 Discussion

As assumed in Section 1.3.2 and 2.2.1, the systematic literature review in this Chapter revealed a highly interdisciplinary research field of studying User-Generated Content from social media platforms for heritage management, combining the knowledge from disciplines including but not limited to computer science, social science, heritage studies, tourism, spatial planning, spatiotemporal analysis, and management. The results from the sections concerning research context (Sections 2.3.1 to 2.3.3), research content (Sections 2.3.4 to 2.3.6), and research methodology (Section 2.3.7) all indicated a complex interplay of different approaches, which is especially clear in Figures 2.7 and 2.8. The finding in Section 2.3.1 showing that most included studies are conducted in cities with urban areas inscribed in the UNESCO WHL is thrilling. In other words, the majority of studies do concern or have the potential to concern cultural heritage planning or management issues without explicitly recognizing and/or declaring them. This could be further compared and embedded in a more general discussion about the role of heritage management processes in the entire planning policy (Janssen et al., 2017). Methods from other relevant yet divergent disciplines could become a great pool of inspiration and references for solving the intended problem of this dissertation as mentioned in Section 1.3.1.

The theoretical framework proposed in Chapter 1 and in Bai et al. (2021b), which distinguishes the social interaction of the global online public for heritage properties as “everyday/baseline" and “activated/event-triggered" scenarios, has been supported with pieces of evidence from the systematic literature review in this chapter as contextualization and validation. According to the review results, some critical gaps are identified for future studies. Studies are needed:

1. to develop heritage-specific tools to deal with large-scale data from social media,
2. to construct proper spatiotemporal and social networks for both scenarios to capture useful information on the opinions and emotions of the online communities,
3. to apply a variety of case studies in a global context to validate the generalizability of the methods, and
4. to link back to real-world heritage management and planning actions to facilitate decision-making processes.

This confirms the need of conducting this dissertation, as to fill in some of the gaps (mostly No.1-3, and partly No.4) and link the evidence-based spatial characteristics in the urban environment to the heritage attributes and values, with the help of machine learning as an automation tool.

Many popular methods identified from the systematic review will constantly reappear in the following Chapters, embedding this dissertation in a broader State of the Art.

For example, choosing Flickr (Chapters 4 and 5) and Twitter (Chapter 6) as social media platforms; involving the content, context (both in Chapters 4 to 6), and structure (Chapter 6) of social media data; discussing baseline (Chapters 4 and 5) and activated scenarios (Chapter 6); including text, images, timestamps, geolocations, and user-information as the data format (in all of Chapters 4 to 6); and applying MDS (Chapter 2), GloVe, Naïve Bayes (both in Chapter 3), Places365 (Chapter 4), Local Moran I (Chapter 5), as well as DBSCAN, LDA, and GeoNames (all in Chapter 6). This suggests that while reviewing the dissertation Chapters and corresponding publications with the same coding scheme, they would hypothetically appear in a central position connecting different sub-fields in this study.

This systematic literature review was originally conducted in 2020, so it unavoidably excluded the highly-related studies published thereafter, such as Bigne et al. (2021); Kang et al. (2021); Ginzarly et al. (2022); Kim and Kang (2022); Tenzer (2022) about everyday baseline scenario, and Kumar et al. (2020); Kumar (2020); Garduño Freeman and Gonzalez Zarandona (2021); Lorini et al. (2022) about the activated event-triggered scenario. Moreover, throughout the research journey, a few other highly relevant studies published before 2020 in the fields of social sciences, digital humanity, geography, and urban studies emerged, which were also not initially included in this systematic literature review (Lee et al., 2011; Lansley and Longley, 2016; Boy and Uitermark, 2017; Lai et al., 2017), probably because they did not include any keywords from the broadest search string restricting the scope (i.e., Heritage OR UNESCO OR Touris* OR HUL OR 'Historic Urban Landscape'). Depending on the eventual purpose, future studies can also consider adding more terms to the search string, allowing for the inclusion of studies from an even broader scope. Another iteration of the same process of literature searching on WoS and SCOPUS could be conducted, together with snowballing methods to incorporate 1) the development of the field ever since, and 2) the neglected yet highly relevant studies due to searching and exclusion strategies. Nevertheless, since the coding scheme, the analytical framework, and the generation of visualizations in this Chapter are all highly structured and modulated, facilitated with a Python-based program, no additional changes are needed when adding new research records. If new dimensions are discovered during later iterations of reviewing process, however, earlier records will also need to be revisited for the coding of specific topics.

2.5 Conclusions

This chapter presented a systematic literature review conducted to answer the questions about how User-Generated Content on social media platforms is collected, processed, analysed, and discussed in the broad field of heritage planning and tourism management. A complex and multi-/inter-disciplinary pattern has been found within the research context, research content, and research methodology of all the reviewed studies. While only a small percentage of reviewed articles explicitly referred to heritage in their writing, the majority of them actually took place in cities with urban areas inscribed in the UNESCO WHL. The methods, algorithms, models, and analytical approaches summarized in this literature review can be inspiring and beneficial for systematically analyzing social media User-Generated Content related to heritage management at scale. Future studies, including the following chapters in this dissertation, need to reflect on this complexity and incorporate inspiring methods from different disciplines.

References

Aaker, J. L. (1997). Dimensions of brand personality. Journal of marketing research, 34(3):347–356.

Abeysinghe, S., Manchanayake, I., Samarajeewa, C., Rathnayaka, P., Walpola, M. J., Nawaratne, R., Bandaragoda, T., and Alahakoon, D. (2018). Enhancing decision making capacity in tourism domain using social media analytics. In 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), pages 369–375. IEEE.

Afzaal, M., Usman, M., Fong, A. C., and Fong, S. (2019). Multiaspect-based opinion classification model for tourist reviews. Expert Systems, 36(2):e12371.

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In Aggarwal, C. C., editor, Social Network Data Analytics, chapter 1, pages 1–15. SPRINGER.

Al-Sultany, G. A. and Abd Al-Ameer, A. A. (2019). Geotagged photos clustering using adapted density-based spatial clustering of applications with noise algorithm. Journal of Computational and Theoretical Nanoscience, 16(3):1056–1061.

Alaei, A. R., Becken, S., and Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. Journal of travel research, 58(2):175–191.

Albers, P. C. and James, W. R. (1988). Travel photography: A methodological approach. Annals of tourism research, 15(1):134–158.

Alviz-Meza, A., Vásquez-Coronado, M. H., Delgado-Caramutti, J. G., and Blanco-Victorio, D. J. (2022). Bibliometric analysis of fourth industrial revolution applied to heritage studies based on web of science and scopus databases from 2016 to 2021. Heritage Science, 10(1):189.

Amato, F., Cozzolino, G., Di Martino, S., Mazzeo, A., Moscato, V., Picariello, A., Romano, S., and Sperlí, G. (2016). Opinions analysis in social networks for cultural heritage applications. Smart Innovation, Systems and Technologies, 55:577–586.

Anselin, L. (1995). Local indicators of spatial association—lisa. Geographical analysis, 27(2):93–115.

Assmann, J. and Czaplicka, J. (1995). Collective memory and cultural identity. New german critique, pages 125–133.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007+ ASWC 2007, Busan, Korea, November 11-15, 2007. Proceedings, pages 722–735. Springer.

Avila-Robinson, A. and Wakabayashi, N. (2018). Changes in the structures and directions of destination management and marketing research: A bibliometric mapping study, 2005–2016. Journal of Destination Marketing & Management, 10:101–111.

Bai, N., Ducci, M., Mirzikashvili, R., Nourian, P., and Pereira Roders, A. (2023a). Mapping urban heritage images with social media data and artificial intelligence, a case study in testaccio, rome. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023:139–146.

Bai, N., Luo, R., Nourian, P., and Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage statements of "Outstanding Universal Value" with soft labels. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 366–384, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2022). Heri-graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. ISPRS International Journal of Geo-Information, 11(9).

Bai, N., Nourian, P., and Pereira Roders, A. (2021b). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.

Bai, N., Nourian, P., Pereira Roders, A., Bunschoten, R., Huang, W., and Wang, L. (2023b). Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data. Environment and Planning B: Urban Analytics and City Science, 50(1):94–116.

Baltrusaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2):423–443.

Bandarin, F. and Van Oers, R. (2012). The historic urban landscape: managing heritage in an urban century. John Wiley & Sons.

Barbagallo, D., Bruni, L., Francalanci, C., and Giacomazzi, P. (2012). An Empirical Study on the Relationship between Twitter Sentiment and Influence in the Tourism Domain. In Fuchs, M and Ricci, F and Cantoni, L., editor, INFORMATION AND COMMUNICATION TECHNOLOGIES IN TOURISM 2012, pages 506–516, SACHSENPLATZ 4-6, A-1201 VIENNA, AUSTRIA. SPRINGER-VERLAG WIEN.

Barros, C., Moya-Gómez, B., and Gutiérrez, J. (2020). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. Current Issues in Tourism, 23(10):1291–1310.

Battiato, S., Farinella, G. M., Milotta, F. L., Ortis, A., Addesso, L., Casella, A., D'Amico, V., and Torrisi, G. (2016). The social picture. In Proceedings of the 2016 ACM on international conference on multimedia retrieval, pages 397–400.

Batty, M. (2023). A new kind of search. Environment and Planning B: Urban Analytics and City Science, 50(3):575–578.

Bellens, A., Banc, N. V. L., Eloire, F., Grabar, N., Kergosien, E., and Severo, M. (2016). Social media and european cultural routes: Instagram networks on the via francigena. In Proceedings of the 8th International Conference on Management of Digital EcoSystems, pages 122–128.

Berman, M. L., Åhlfeldt, J., and Wick, M. (2012). Historical gazetteer system integration: Chgis, regnum francorum, and geonames. Placing names: enriching and integrating gazetteers. Indiana University Press, Bloomington and Indianapolis.

Bernadou, D. (2017). Construire l'image touristique d'une région à travers les réseaux sociaux : le cas de l'Émilie-romagne en italie. Cybergeo, 2017.

Bifulco, F., RUSSO SPENA, T., et al. (2016). The databenc experience: a smart innovation habitat. In Managing Cultural heritage, pages 46–63. McGraw-Hill.

Bigne, E., Ruiz, C., Cuenca, A., Perez, C., and Garcia, A. (2021). What drives the helpfulness of online reviews? a deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. Journal of Destination Marketing & Management, 20:100570.

Boland, A., Cherry, M. G., and Dickson, R. (2017). Doing a Systematic Review: A Student's Guide. SAGE PUBLICATIONS INC.

Borg, I. and Groenen, P. J. (2005). Modern multidimensional scaling: Theory and applications. Springer Science & Business Media.

Boy, J. D. and Uitermark, J. (2017). Reassembling the city through instagram. Transactions of the Institute of British Geographers, 42(4):612–624.

Burgess, S., Sellitto, C., Buultjens, J., and Cox, C. (2015). How australian smes engage with social media. In Proceedings of the European conference on e-learning, pages 45–51.

Campillo-Alhama, C. and Martinez-Sala, A.-M. (2019). Events 2.0 in the transmedia branding strategy of World Cultural Heritage Sites. PROFESIONAL DE LA INFORMACION, 28(5).

Castiglione, A., Colace, F., Moscato, V., and Palmieri, F. (2018). Chis: A big data infrastructure to manage digital cultural items. Future Generation Computer Systems, 86:1134–1145.

Chaabani, Y., Toujani, R., and Akaichi, J. (2018). Sentiment analysis method for tracking touristics reviews in social media network. Smart Innovation, Systems and Technologies, 76:299–310.

Chan, J. C.-W. and Paelinckx, D. (2008). Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 112(6):2999–3011.

Chen, F.-W., Guevara Plaza, A., and Alarcón Urbistondo, P. (2017). Automatically extracting tourism-related opinion from chinese social media. Current Issues in Tourism, 20(10):1070–1087.

Chianese, A., Marulli, F., and Piccialli, F. (2016). Cultural Heritage and Social Pulse: A Semantic Approach for CH Sensitivity Discovery in Social Media Data. In Proceedings - 2016 IEEE 10th International Conference on Semantic Computing, ICSC 2016, pages 459–464. Institute of Electrical and Electronics Engineers Inc.

Cho, N., Kang, Y., Yoon, J., Park, S., and Kim, J. (2022). Classifying tourists' photos and exploring tourism destination image using a deep learning model. Journal of Quality Assurance in Hospitality & Tourism, pages 1–29.

Church, K. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. Computational linguistics, 16(1):22–29.

Claster, W. B., Cooper, M., and Sallis, P. (2010). Thailand–tourism and conflict: Modeling sentiment from twitter tweets using naïve bayes and unsupervised artificial neural nets. In 2010 second international conference on computational intelligence, Modelling and Simulation, pages 89–94. IEEE.

Clemente, P., Calvache, M., Antunes, P., Santos, R., Cerdeira, J. O., and Martins, M. J. (2019). Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a natural park in portugal. Ecological indicators, 96:59–68.

Colace, F., De Santo, M., Greco, L., Amato, F., Moscato, V., and Picariello, A. (2014). Terminological ontology learning and population using latent dirichlet allocation. Journal of Visual Languages & Computing, 25(6):818–826.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. IEEE Transactions on pattern analysis and machine intelligence, 24(5):603–619.

Cristianini, N., Shawe-Taylor, J., et al. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

Dai, P., Zhang, S., Chen, Z., Gong, Y., and Hou, H. (2019). Perceptions of cultural ecosystem services in urban parks based on social network data. Sustainability, 11(19):5386.

De Angelis, A., Gasparetti, F., Micarelli, A., and Sansonetti, G. (2017). A social cultural recommender based on linked open data. In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, pages 329–332.

Deeksha, S., Ashrith, H., Bansode, R., and Kamath, S. (2015). A spatial clustering approach for efficient landmark discovery using geo-tagged photos. In 2015 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pages 1–6. IEEE.

Del Vecchio, P., Mele, G., Ndou, V., and Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. Information Processing & Management, 54(5):847–860.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dickinger, A. and Lalicic, L. (2016). An analysis of destination brand personality and emotions: a comparison study. Information Technology & Tourism, 15:317–340.

Dickinger, A., Lalicic, L., and Mazanec, J. (2017). Exploring the generalizability of discriminant word items and latent topics in online tourist reviews. International Journal of Contemporary Hospitality Management, 29(2):803–816.

Ducci, M., Janssen, R., Burgers, G.-J., and Rotondo, F. (2023). Mapping local perceptions for the planning of cultural landscapes. International Journal of E-Planning Research (IJEPR), 12(1):1–27.

Encalada, L., Boavida-Portugal, I., Cardoso Ferreira, C., and Rocha, J. (2017). Identifying tourist places of interest based on digital imprints: Towards a sustainable smart city. Sustainability, 9(12):2317.

Estellés-Arolas, E. and González-Ladrón-de Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information science, 38(2):189–200.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, page 226–231. AAAI Press.

Feizollah, A., Ainin, S., Anuar, N. B., Abdullah, N. A. B., and Hazim, M. (2019). Halal products on twitter: Data extraction and sentiment analysis using stack of deep learning algorithms. IEEE Access, 7:83354–83362.

Figueredo, M., Ribeiro, J., Cacho, N., Thome, A., Cacho, A., Lopes, F., and Araujo, V. (2018). From photos to travel itinerary: A tourism recommender system for smart tourism destination. In 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), pages 85–92. IEEE.

Floris, R., Campagna, M., et al. (2014). Social media geographic information in tourism planning. TEMA, 257:417–430.

Floris, R. and Zoppi, C. (2015). Social media-related geographic information in the context of strategic environmental assessment of municipal masterplans: A case study concerning sardinia (italy). Future Internet, 7(3):276–293.

Fostikov, A. (2023). First impressions on using ai powered chatbots, tools and search engines: Chatgpt, perplexity and other–possibilities and usage problems. Review of the National Center for Digitization 2023 (preprint).

Freund, Y., Schapire, R. E., et al. (1996). Experiments with a new boosting algorithm. In icml, volume 96, pages 148–156. Citeseer.

Fukui, M. and Ohe, Y. (2019). Assessing the role of social media in tourism recovery in tsunami-hit coastal areas in Tohoku, Japan. Tourism Economics.

Gabrielli, L., Rinzivillo, S., Ronzano, F., and Villatoro, D. (2014). From Tweets to Semantic Trajectories: Mining Anomalous Urban Mobility Patterns. In Nin, J and Villatoro, D., editor, CITIZEN IN SENSOR NETWORKS, volume 8313 of Lecture Notes in Artificial Intelligence, pages 26–35, HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY. SPRINGER-VERLAG BERLIN.

Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S. L., Kreuter, F., Olsson, H., Prelec, D., Stein, D. L., and van Der Does, T. (2021). Human social sensing is an untapped resource for computational social science. Nature, 595(7866):214–222.
Garduño Freeman, C. and Gonzalez Zarandona, J. A. (2021). Digital spectres: the notre-dame effect. International Journal of Heritage Studies, 27(12):1264–1277.
Gerrig, R. J., Zimbardo, P. G., Campbell, A. J., Cumming, S. R., and Wilkes, F. J. (2015). Psychology and life. Pearson Higher Education AU.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2019). Using social media to identify tourism attractiveness in six italian cities. Tourism management, 72:306–312.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2020). Machine learning and points of interest: typical tourist italian cities. Current Issues in Tourism, 23(13):1646–1658.
Ginzarly, M., Pereira Roders, A., and Teller, J. (2019). Mapping historic urban landscape values through social media. Journal of Cultural Heritage, 36:1–11.
Ginzarly, M., Srour, F. J., and Roders, A. P. (2022). The interplay of context, experience, and emotion at world heritage sites: a qualitative and machine learning approach. Tourism Culture & Communication, 22(4):321–340.
Gomez, R., Gomez, L., Gibert, J., and Karatzas, D. (2019). Learning from #barcelona instagram data what locals and tourists post about its neighbourhoods. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11134 LNCS:530–544.
Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MIT press.
Gosal, A. S., Geijzendorffer, I. R., Václavík, T., Poulin, B., and Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosystem Services, 38:100958.
Grandi, R. and Neri, F. (2014). Sentiment analysis and city branding. In New Trends in Databases and Information Systems: 17th East European Conference on Advances in Databases and Information Systems, pages 339–349. Springer.
Guo, T., Guo, B., Ouyang, Y., Yu, Z., Lam, J. C., and Li, V. O. (2018). Crowdtravel: scenic spot profiling by using heterogeneous crowdsourced data. Journal of Ambient Intelligence and Humanized Computing, 9:2051–2060.
Hashida, S., Tamura, K., and Sakai, T. (2018). Classifying sightseeing tweets using convolutional neural networks with multi-channel distributed representation. In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 178–183. IEEE.
Hasnat, M. M. and Hasan, S. (2018). Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. Transportation Research Part C: Emerging Technologies, 96:38–54.
Honkela, T. (1997). Self-organizing maps in natural language processing. PhD thesis, Citeseer.
Janssen, J., Luiten, E., Renes, H., and Stegmeijer, E. (2017). Heritage as sector, factor and vector: conceptualizing the shifting relationship between heritage management and spatial planning. European Planning Studies, 25(9):1654–1672.
Jindal, N. and Liu, B. (2006). Identifying comparative sentences in text documents. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pages 244–251.
Junker, C., Akbar, Z., and Cuquet, M. (2017). The network structure of visited locations according to geotagged social media photos. arXiv, pages 1–8.
Kang, Y., Cho, N., Yoon, J., Park, S., and Kim, J. (2021). Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos. ISPRS International Journal of Geo-Information, 10(3):137.
Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. IEEE transactions on pattern analysis and machine intelligence, 24(7):881–892.
Kennedy, L., Naaman, M., Ahern, S., Nair, R., and Rattenbury, T. (2007). How flickr helps us make sense of the world: context and content in community-contributed media collections. In Proceedings of the 15th ACM international conference on Multimedia, pages 631–640.
Kim, J. and Kang, Y. (2022). Automatic classification of photos by tourist attractions using deep learning model and image feature vector clustering. ISPRS International Journal of Geo-Information, 11(4):245.
Kisilevich, S., Mansmann, F., and Keim, D. (2010). P-dbscan: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. In Proceedings of the 1st international conference and exhibition on computing for geospatial research & application, pages 1–4.
Korakakis, M., Spyrou, E., Mylonas, P., and Perantonis, S. J. (2017). Exploiting social media information toward a context-aware recommendation system. Social Network Analysis and Mining, 7:1–20.
Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29(1):1–27.
Kumar, P. (2020). Twitter, disasters and cultural heritage: A case study of the 2015 nepal earthquake. Journal of Contingencies and Crisis Management, 28(4):453–465.
Kumar, P., Ofli, F., Imran, M., and Castillo, C. (2020). Detection of disaster-affected cultural heritage sites from social media images using deep learning techniques. Journal on Computing and Cultural Heritage (JOCCH), 13(3):1–31.

Lai, J., Cheng, T., and Lansley, G. (2017). Improved targeted outdoor advertising based on geotagged social media data. Annals of GIS, 23(4):237–250.

Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. Discourse processes, 25(2-3):259–284.

Lansley, G. and Longley, P. A. (2016). The geography of twitter topics in london. Computers, Environment and Urban Systems, 58:85–96.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature, 521(7553):436–444.

Lee, R., Wakamiya, S., and Sumiya, K. (2011). Discovery of unusual regional social activities using geo-tagged microblogs. World Wide Web, 14:321–349.

Lefebvre, H. (2014). The production of space (1991). In The people, place, and space reader, pages 323–327. Routledge.

Leung, D., Law, R., Van Hoof, H., and Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. Journal of travel & tourism marketing, 30(1-2):3–22.

Leung, R., Vu, H. Q., and Rong, J. (2017). Understanding tourists' photo sharing and visit pattern at non-first tier attractions via geotagged photos. Information Technology & Tourism, 17:55–74.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., and Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of clinical epidemiology, 62(10):e1–e34.

Liu, P. and De Sabbata, S. (2021). A graph-based semi-supervised approach to classification learning in digital geographies. Computers, Environment and Urban Systems, 86:101583.

Liu, Z., Shan, J., Glassey Balet, N., and Fang, G. (2017). Semantic social media analysis of chinese tourists in switzerland. Information technology & tourism, 17:183–202.

Lorini, V., Rufolo, P., and Castillo, C. (2022). Venice was flooding... one tweet at a time. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2):1–16.

Lu, W. and Stepchenkova, S. (2015). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. Journal of Hospitality Marketing & Management, 24(2):119–154.

Lynch, K. (1964). The image of the city. MIT press.

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., and Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science, 27(4):662–684.

Manning, C. D. (2009). An introduction to information retrieval. Cambridge university press.

Mariani, M. M., Di Felice, M., and Mura, M. (2016). Facebook as a destination marketing tool: Evidence from italian regional destination management organizations. Tourism management, 54:321–343.

Martí, P., García-Mayor, C., and Serrano-Estrada, L. (2021). Taking the urban tourist activity pulse through digital footprints. Current Issues in Tourism, 24(2):157–176.

Martínez-Sala, A.-M., Albeza, R., and Martínez Cano, F. J. (2018). Social networks of tourist destination marketing organizations as potential sources of ewom. Observatorio, 12:246–271.

Mazloom, M., Hendriks, B., and Worring, M. (2017). Multimodal context-aware recommender for post popularity prediction in social media. In Proceedings of the on Thematic Workshops of ACM Multimedia 2017, pages 236–244.

McMullen, M. (2020). 'pinning'tourist photographs: Analyzing the photographs shared on pinterest of heritage tourist destinations. Current Issues in Tourism, 23(3):376–387.

Mendieta, J., Suárez, S., Vaca, C., Ochoa, D., and Vergara, C. (2016). Geo-localized social media data to improve characterization of international travelers. In 2016 Third International Conference on eDemocracy & eGovernment (ICEDEG), pages 126–132. IEEE.

Miah, S. J., Vu, H. Q., Gammack, J., and McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. Information & Management, 54(6):771–785.

Micera, R. and Crispino, R. (2017). Destination web reputation as "smart tool" for image building: the case analysis of naples city-destination. International Journal of Tourism Cities.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. In Bengio, Y. and LeCun, Y., editors, 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26:3111–3119.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Grp, P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). Physical Therapy, 89(9):873–880.

Monteiro, V., Henriques, R., Painho, M., and Vaz, E. (2014). Sensing World Heritage An Exploratory Study of Twitter as a Tool for Assessing Reputation. In Murgante, B and Misra, S and Rocha, AMAC and Torre, C and Rocha, JG and Falcao, MI and Taniar, D and Apduhan, BO and Gervasi, O., editor, COMPUTATIONAL SCIENCE AND ITS APPLICATIONS - ICCSA 2014, PT II, volume 8580 of Lecture Notes in Computer Science, pages 404–419. Univ Minho; Univ Perugia; Univ Basilicata; Monash Univ; Kyushu Sangyo Univ; Assoc Portuguesa Investigacao Operac.

Moreno, A., Jabreel, M., and Huertas, A. (2015). Automatic analysis of the communication of tourist destination brands through social networks. In 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), pages 546–553. IEEE.

Muangon, W., Muangprathub, J., saelee, J., Soonklang, T., Pongpinigpinyo, S., and Sitdhisanguan, K. (2018). An information retrieval system on thailand tourism community websites. In Proceedings of the 2018 10th International Conference on Information Management and Engineering, pages 101–105.

Nenko, A. and Petrova, M. (2018). Emotional geography of st. petersburg: detecting emotional perception of the city space. In Digital Transformation and Global Society: Third International Conference, DTGS 2018, St. Petersburg, Russia, May 30–June 2, 2018, Revised Selected Papers, Part II 3, pages 95–110. Springer.

Nigam, K., Lafferty, J., and McCallum, A. (1999). Using maximum entropy for text classification. In IJCAI-99 workshop on machine learning for information filtering, volume 1, pages 61–67. Stockholom, Sweden.

Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., and Plieninger, T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five european sites. Ecological Indicators, 94:74–86.

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., and González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. EPJ Data Science, 4(1):1–17.

Pan, J., Mou, N., and Liu, W. (2019). Emotion analysis of tourists based on domain ontology. In Proceedings of the 2019 International Conference on Data Mining and Machine Learning, pages 146–150.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359.

Pantano, E. and Dennis, C. (2019). Store buildings as tourist attractions: Mining retail meaning of store building pictures through a machine learning approach. Journal of Retailing and Consumer Services, 51:304–310.

Park, D., Kim, W. G., and Choi, S. (2019). Application of social media analytics in tourism crisis communication. Current Issues in Tourism, 22(15):1810–1824.

Peng, X. and Huang, Z. (2017). A novel popular tourist attraction discovering approach based on geo-tagged social media big data. ISPRS International Journal of Geo-Information, 6(7):216.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In Moschitti, A., Pang, B., and Daelemans, W., editors, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1532–1543. ACL.

Pickering, C., Rossi, S. D., Hernando, A., and Barros, A. (2018). Current knowledge and future research directions for the monitoring and management of visitors in recreational and protected areas. Journal of Outdoor Recreation and Tourism, 21(November 2017):10–18.

Qi, S., Wong, C. U. I., Chen, N., Rong, J., and Du, J. (2018). Profiling macau cultural tourists by using user-generated content from online social media. Information Technology & Tourism, 20:217–236.

Rakic, T. and Chambers, D. (2008). World Heritage: Exploring the Tension Between the National and the 'Universal'. Journal of Heritage Tourism, 2(3):145–155.

Ramanathan, V. and Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about oman tourism. In 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), pages 1–5. IEEE.

Rodriguez, A. and Laio, A. (2014). Clustering by fast search and find of density peaks. science, 344(6191):1492–1496.

Safavian, S. R. and Landgrebe, D. (1990). Topics in inference and decision-making with partial knowledge. Technical report, Purdue University.

Salur, M. U., Aydin, İ., and Alghrsi, S. A. (2019). Smartsenti: A twitter-based sentiment analysis system for the smart tourism in turkey. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), pages 1–5. IEEE.

Sansonetti, G., Gasparetti, F., Micarelli, A., Cena, F., and Gena, C. (2019). Enhancing cultural recommendations through social and linked open data. User Modeling and User-Adapted Interaction, 29:121–159.

Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L., and Helbing, D. (2014). A network framework of cultural history. science, 345(6196):558–562.

Schirpke, U., Meisch, C., Marsoner, T., and Tappeiner, U. (2018). Revealing spatial and temporal patterns of outdoor recreation in the european alps and their surroundings. Ecosystem services, 31:336–350.

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., and Xu, X. (2017). Dbscan revisited, revisited: why and how you should (still) use dbscan. ACM Transactions on Database Systems (TODS), 42(3):1–21.

Sebastiani, F. and Esuli, A. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of the 5th international conference on language resources and evaluation, pages 417–422. European Language Resources Association (ELRA) Genoa, Italy.

Song, S.-G. and Kim, D.-Y. (2016). A pictorial analysis of destination images on pinterest: The case of tokyo, kyoto, and osaka, japan. Journal of Travel & Tourism Marketing, 33(5):687–701.

Sun, Y., Ma, H., and Chan, E. H. (2017). A model to measure tourist preference toward scenic spots based on social media data: A case of dapeng in china. Sustainability, 10(1):43.

Taecharungroj, V. and Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tourism Management, 75(July):550–568.

Tang, J. and Li, J. (2016). Spatial network of urban tourist flow in xi'an based on microblog big data. Journal of China Tourism Research, 12(1):5–23.
Tao, Y., Zhang, F., Shi, C., and Chen, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. Sustainability, 11(18):5070.
Taylor, J. and Gibson, L. K. (2017). Digitisation, digital interaction and social media: embedded barriers to democratic heritage. International Journal of Heritage Studies, 23(5):408–420.
Tenzer, M. (2022). Tweets in the peak: Twitter analysis-the impact of covid-19 on cultural landscapes. Internet Archaeology, 59.
Thakuriah, P. V., Sila-Nowicka, K., Hong, J., Boididou, C., Osborne, M., Lido, C., and McHugh, A. (2020). Integrated multimedia city data (imcd): a composite survey and sensing approach to understanding urban living and mobility. Computers, Environment and Urban Systems, 80:101427.
Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American society for information science and technology, 61(12):2544–2558.
Thorp, H. H. (2023). Chatgpt is fun, but not an author.
UNESCO (2008). Operational guidelines for the implementation of the world heritage convention. Technical Report July, UNESCO World Heritage Centre.
UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.
van der Zee, E. and Bertocchi, D. (2018). Finding patterns in urban tourist behaviour: A social network analysis approach based on tripadvisor reviews. Information Technology & Tourism, 20(1-4):153–180.
van Dijck, J. (2011). Flickr and the culture of connectivity: Sharing views, experiences, memories. Memory Studies, 4(4):401–415.
van Eck, N. J. and Waltman, L. (2014). Visualizing bibliometric networks. Measuring scholarly impact: Methods and practice, pages 285–320.
van Weerdenburg, D., Scheider, S., Adams, B., Spierings, B., and van der Zee, E. (2019). Where to go and what to do: Extracting leisure activity potentials from web data on urban space. Computers, Environment and Urban Systems, 73:143–156.
Varnajot, A. (2019). Digital rovaniemi: contemporary and future arctic tourist experiences. Journal of Tourism Futures, 6(1):6–23.
Vassakis, K., Petrakis, E., Kopanakis, I., Makridis, J., and Mastorakis, G. (2019). Location-based social network data for tourism destinations. Springer Singapore.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 6000–6010.
Veldpaus, L. (2015). Historic urban landscapes: framing the integration of urban and heritage planning in multilevel governance. PhD thesis, Technische Universiteit Eindhoven.
Wagner, J., Arora, P., Vaíllo, S. C., Barman, U., Bogdanova, D., Foster, J., and Tounsi, L. (2014). Dcu: Aspect-based polarity classification for semeval task 4. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 223–229.
Watkins, J. (2007). Social media, participatory design and cultural engagement. In Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces, pages 161–166.
Williams, N. L., Inversini, A., Ferdinand, N., and Buhalis, D. (2017). Destination eWOM: A macro and meso network approach? Annals of Tourism Research, 64:87–101.
Wu, X., Huang, Z., Peng, X., Chen, Y., and Liu, Y. (2018). Building a spatially-embedded network of tourism hotspots from geotagged social media data. IEEE Access, 6:21945–21955.
Zhang, F., Zhou, B., Ratti, C., and Liu, Y. (2019). Discovering place-informative scenes and objects using social media photos. Royal Society open science, 6(3):181375.
Zhang, M.-L. and Zhou, Z.-H. (2007). Ml-knn: A lazy learning approach to multi-label learning. Pattern recognition, 40(7):2038–2048.
Zhang, Z., Cui, P., and Zhu, W. (2020). Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering.
Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6):1452–1464.
Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using places database. Advances in neural information processing systems, 27.

PART B

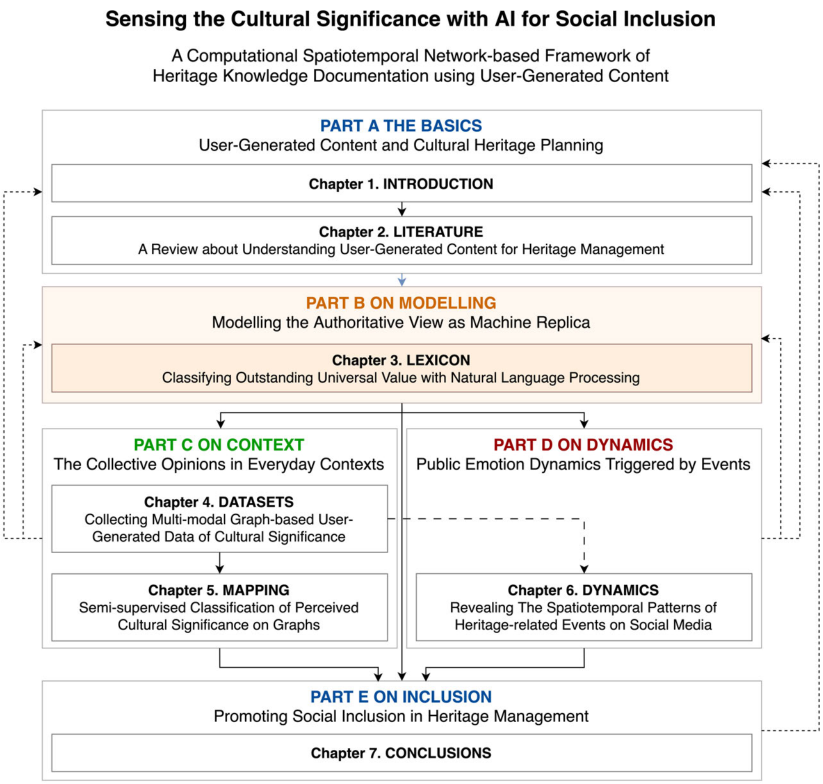
On Modelling

Modelling the Authoritative View as Machine Replica

This part of dissertation models the authoritative view on UNESCO World Heritage Outstanding Universal Value (OUV). It trains a few Natural Language Processing models to replicate the classification and justification of OUV selection criteria based on the semantic information of a generic sentence. The associations among the natural and cultural selection criteria are revealed. This part is based on the knowledge acquired from PART A and prepares for social media analyses in the following PART C and PART D, since a heritage-specific analytic tool that is both reproducible and scalable is needed in response to the massive amount of user-generated social media data.

One chapter is included in this part:

Chapter 3 Lexicon - Classifying Outstanding Universal Value with Natural Language Processing.


Sensing the Cultural Significance with AI for Social Inclusion
A Computational Spatiotemporal Network-based Framework of
Heritage Knowledge Documentation using User-Generated Content
PART A THE BASICS
User-Generated Content and Cultural Heritage Planning
Chapter 1. INTRODUCTION
Chapter 2. LITERATURE
A Review about Understanding User-Generated Content for Heritage Management
PART B ON MODELLING
Modelling the Authoritative View as Machine Replica
Chapter 3. LEXICON
Classifying Outstanding Universal Value with Natural Language Processing
PART C ON CONTEXT
The Collective Opinions in Everyday Contexts
Chapter 4. DATASETS
Collecting Multi-modal Graph-based User-Generated Data of Cultural Significance
Chapter 5. MAPPING
Semi-supervised Classification of Perceived Cultural Significance on Graphs
PART D ON DYNAMICS
Public Emotion Dynamics Triggered by Events
Chapter 6. DYNAMICS
Revealing The Spatiotemporal Patterns of Heritage-related Events on Social Media
PART E ON INCLUSION
Promoting Social Inclusion in Heritage Management
Chapter 7. CONCLUSIONS

3 Lexicon

Classifying Outstanding Universal Value with Natural Language Processing

Parts of this chapter have been published in Bai et al. (2021a,b).
Bai N. Luo R, Nourian P, Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage" Outstanding Universal Value" Documents with Soft Labels. In Findings of the Association for Computational Linguistics: EMNLP 2021. p. 366-384. Association for Computational Linguistics.

Bai N, Nourian P, Luo R, Pereira Roders A. (2021b). "What is OUV" Revisited: A Computational Interpretation on the Statements of Outstanding Universal Value. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-M-1-2021. p. 25–32.

ABSTRACT Evaluating and justifying the Outstanding Universal Value (OUV) is essential for each site inscribed in the UNESCO World Heritage List, and yet a complex task, even for experts, since the selection criteria of OUV are not mutually exclusive. The 2008 ICOMOS Report "What is OUV" has been a successful example of interpreting OUV as an integral concept by inspecting the associations of the selection criteria in all inscribed properties. Furthermore, manual annotation of heritage values and attributes from multi-source textual data, which is currently dominant in heritage studies, is knowledge-demanding and time-consuming, impeding systematic analysis of such authoritative documents in terms of their implications on heritage management. This chapter applies state-of-the-art Natural Language Processing models to build a classifier on Statements of OUV, seeking an explainable and scalable automation tool to facilitate the nomination, evaluation, research, and monitoring processes of World Heritage sites. Label smoothing is innovatively adapted to improve the model performance by adding prior inter-class relationship knowledge to generate soft labels. The study shows that the best models can reach 94.3% top-3 accuracy, that the lexicon derived from computational techniques can capture the essential concepts of OUV, and that the selection criteria are consistently associated with each other in different similarity metrics. A human study with an expert evaluation of the model prediction shows that the models are sufficiently generalizable. This study provides a quantitative and qualitative interpretation of the Statements of OUV and the associations of selection criteria, which can be seen as an

elaborated computational extension of the 2008 Report, useful for future inscription and evaluation process of World Heritage nominations. Code and data for this project are available at https://github.com/zzbn12345/WHOSe_Heritage.

KEYWORDS UNESCO World Heritage, Outstanding Universal Value, Natural Language Processing, Label Smoothing, Text Classification

3.1 Introduction

Since the World Heritage Convention was adopted in 1972, 1121 sites have been inscribed worldwide in the World Heritage List (WHL) of UNESCO up to 2019, aiming at the collective protection of the cultural and natural heritage of Outstanding Universal Value (OUV) for mankind as a whole (UNESCO, 1972; von Droste, 2011; Pereira Roders and van Oers, 2011). First proposed in 1976, OUV, meaning the

> "cultural and/or natural significance which is so exceptional as to transcend national boundaries and to be of common importance for present and future generations of all humanity",

has been operationalized and formalized into an administrative requirement instead of an independent qualification for new inscriptions on the WHL since the adoption of the Operational Guidelines in 2005 (Jokilehto, 2006, 2008; UNESCO, 2008). Ten selection criteria exist as the core of OUV, among which criteria (i) - (vi) generally refer to cultural values, and (vii) - (x) to natural ones. At least one of the ten criteria must be fulfilled by any nomination. Further details are available in Appendix A.

Since 2007, complete **Statements of OUV (SOUV)** need to be submitted and approved for new World Heritage (WH) nominations, which should include, among others, a section of "justification for criteria", giving a short paragraph to explain why a site (also known as property) satisfies each of the criteria it is inscribed under. These statements are to be drafted by the State Parties after scientific research for any tentative nominations, further reviewed and revised by the Advisory Bodies from ICOMOS and/or IUCN and eventually approved and adopted by the World Heritage Committee for inscription. Similarly, Retrospective Statements of OUV were also prepared during the Second Cycle of Periodic Reporting (2008-2015) by 812 properties[1] inscribed before 2006, to revise or refill the section of justification for criteria if it was incomplete or not agreed on at the time of inscription (IUCN et al.,

[1] this number is calculated based on the data provided in the Reports of each region available at http://whc.unesco.org/en/pr-questionnaire/

2010). However, the evaluation of SOUV can be ambiguous in the sense that: 1) the selection criteria are not mutually exclusive and contain common information about historical and aesthetic/artistic values as an integral part (Jokilehto, 2008); 2) the key stakeholders to evaluate the SOUV for a nomination occasionally disagree with each other at early stages, leading to recursive reviews and revisions, though all are considered to be domain experts (Jokilehto, 2008; Tarrafa Silva and Pereira Roders, 2010; von Droste, 2011). A tool to check the accuracy, objectivity, consistency, and coherence of such statements can significantly benefit the inscription process involving thousands of experts worldwide each year.

Not only for new nominations, but the SOUV is also an essential reference point for monitoring and interpreting inscribed heritage sites (IUCN et al., 2010). Researchers and practitioners actively and regularly check if the justified criteria are still relevant for the sites, so as to decide on further planning and managerial actions. Moreover, these same statements are also used in support of legal court cases, should WH sites be endangered by human development (Pereira Roders, 2010; von Droste, 2011). Under the support of the Recommendation of Historic Urban Landscape and the recent Our World Heritage campaign, multiple data sources (e.g., news articles, policy documents, social media posts) are encouraged in such analyses of identifying and mapping OUV (UNESCO, 2011; Bandarin and Van Oers, 2012; Ginzarly et al., 2019). The traditional method of manually annotating heritage values and attributes by experts can be time-consuming and knowledge-demanding for analysing massive social media posts by people in cities with urban areas inscribed in the UNESCO WHL to find OUV-related statements, albeit dominantly applied in practice (Tarrafa Silva and Pereira Roders, 2010, 2012; Abdel Tawab, 2019).

Investigating OUV and comparing it to the selection criteria and justifications applied to the listed WH properties is not uncommon. Most research, however, focuses on a single case or a few cases for comparative study, thus mainly concerning a small number of SOUV (Tarrafa Silva and Pereira Roders, 2010; Shah, 2015; Abdel Tawab, 2019; Ruffino et al., 2019). Whereas the 2007 International Conference on Values and Criteria in Heritage Conservation explicitly organized sessions to discover the definition and evolution of OUV as an integral concept, discussing the terms used in the current (by then) WH justifications and proposing possible enhancement to clarify the concepts (Fejérdy, 2007; Jokilehto, 2007; Petzet, 2007). The whole discussion of this conference resulted in the well-known ICOMOS report "**What is OUV**, Defining the Outstanding Universal Value of Cultural World Heritage Properties", published in 2008. The report described the evolution of OUV since first proposed, summarized the essential focuses of each cultural selection criterion, and matched the criteria to the main themes in existing WH properties (Jokilehto, 2008). In that report, the concepts of OUV are illustrated from both a deductive perspective by interpreting the definitions in Operational Guidelines, and an inductive perspective by giving examples from justification texts of WH properties. Keywords in the justifications are highlighted to indicate why this piece of text reflects the selection criterion it describes. Furthermore, the report suggests that the criteria are strongly associated with each other, since the

> "historical value is an integral part of the majority of... criteria (i)-(vii)",

and that

> "the aesthetic /artistic value also plays a role in several OUV criteria".

Such associations have been further investigated in the report by looking at how often a specific criterion is used together with others. This line of interpreting OUV and the selection criteria is rather effective and contributes to a better understanding of the concepts. However, such processes of keyword highlighting are heavily dependent on expert knowledge, which may not be easily applicable and intelligible for the general public, let alone being prone to inevitable personal and disciplinary biases.

To approximate both ultimate goals of this study: 1) aiding the inscription process by checking the coherence and consistency of SOUV, and 2) identifying heritage values from multiple data sources (e.g., social media posts), a computational solution rooted in SOUV is desired. By training Natural Language Processing (NLP) models with the officially written and approved SOUV, a machine replica of the collective authoritative view could be obtained. This machine replica will not be employed at this stage to justify OUV for new nominations from scratch. Rather, it will assess the written SOUV of WH sites (either existing or new) and classify OUV-related texts with the learned collective authoritative view. Furthermore, it can investigate the existing SOUV from the bottom up and capture the subtle intrinsic associations within the statements and among the corresponding selection criteria (Bai et al., 2021b). This yields a new perspective on interpreting the WHL, which would give insights for furthering amending the concept of OUV and selection criteria to be better discernible.

Therefore, this study aims at training an explainable and scalable classifier that can reveal the intrinsic associations of World Heritage OUV selection criteria, which can be feasible to apply in real-world analyses by researchers and practitioners. As outcome, this chapter presents the classifier of UNESCO **W**orld **H**eritage Statements of **O**UV with **S**oft Lab**e**ls (**WHOSe** Heritage).

The contributions of this chapter can be summarized as follows:

1. A text classification dataset is presented, concerning a domain-specific task about Outstanding Universal Value for UNESCO World Heritage sites;
2. Innovative variants of label smoothing are applied to introduce the prior knowledge of label association into training as soft labels, which turned out effective to improve performance in most investigated popular models as baselines in this task;
3. Several classifiers are trained and compared on the Statements of OUV classification task as initial benchmarks, supplemented with explorations on their explainability and generalizability using expert evaluation;
4. An OUV-related lexicon is provided from the trained classifiers, which can be used to highlight keywords in a generic text on relevant selection criteria;
5. Three types of matrix-based similarity metrics (i.e., co-occurrence matrix in the WHL, confusion matrix by the classifiers, and similarity matrix of the lexicon) are proposed

from different sources to represent the pair-wise associations of selection criteria, which are analysed quantitatively and qualitatively, giving insights to more clearly defining OUV in future practice.

3.2 Related Work

Text classification
In the past decades, numerous models have been proposed from shallow to deep learning models for text classification tasks. In shallow learning models, the raw input text is pre-processed to extract features of the text, which are then fed into machine learning classifiers, e.g., Naive Bayes (Maron, 1961) and Support Vector Machine (SVM) (Joachims, 1998) for prediction. In deep learning models, deep neural networks are leveraged to extract information from the input data, such as Convolutional Neural Network (CNN) (Kim, 2014; Johnson and Zhang, 2017), Recurrent Neural Network (RNN) (Cho et al., 2014; Tai et al., 2015), attention networks (Yang et al., 2016) and Transformers (Devlin et al., 2019). Multi-class and multi-label tasks are two extensions of the simplest binary classification, where every sample can belong to one or more classes within a class list (Aly, 2005; Tsoumakas and Katakis, 2007), where the labels may also be correlated (Pal et al., 2020). This work explores the combined application of some popular shallow and deep learning models for a multi-class classification task.

Label Smoothing
Label Smoothing (LS) is originally proposed as a regularization technique to alleviate overfitting in training deep neural networks (Szegedy et al., 2016; Müller et al., 2019). It assigns a noise distribution on all the labels to prevent the model from predicting too confidently on ‘ground-truth’ labels. It is widely used in computer vision (Szegedy et al., 2016), speech (Chorowski and Jaitly, 2017) and natural language processing (Vaswani et al., 2017) tasks. Originally the distribution is uniform across the labels, which is data independent. Recently, other variants of LS are also proposed that are able to incorporate the interrelation information from the data into the distribution (Zhong et al., 2016; Krothapalli and Abbott, 2020; Zhang et al., 2020). In this work, the technique is applied to generate soft labels with a distribution derived from domain knowledge since the classes in this task are clearly interrelated with each other.

Transfer Learning in NLP
In many real-world applications, labelled data are limited and expensive to collect. Training models with limited data from scratch affects the performance. Transfer

learning (Pan and Yang, 2010) is widely used to solve this by using word embeddings that are pretrained on massive corpus and fine-tuning them on target tasks. Earlier works (Mikolov et al., 2013; Pennington et al., 2014) provide static word embeddings that ignore the contextual information in the sentences. More recent works, e.g., Universal Language Model Fine-tuning (ULMFiT) (Howard and Ruder, 2018) and Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019), take the context into account and generate dynamic contextualized word vectors, showing excellent performance, which also proves to be sufficiently generalizable across many tasks. This task, with a relatively small data size, employs the idea of transfer learning and applies both embedding methods.

3.3 Data and Materials

3.3.1 Case Studies: UNESCO World Heritage List

Since this study aims to train a general model that is useful for heritage at a global level, the entire UNESCO World Heritage List is selected as the case study for this chapter. UNESCO World Heritage Centre openly releases a syndication dataset of the sites in XLS format[2], which includes information of the inscribed World Heritage sites such as ID, name, short description, justification of criteria et al. Among them, the field of justification provides a paragraph for each selection criterion the site fulfills[3], contributing as the input data for this task. In total, 1052 out of 1121 WH sites contain the justification data[4], while the remaining 69 await the Retrospective SOUV to be approved as introduced in Section 3.1. As an example, in Venice and Its Lagoon, the paragraph on **criterion (i)** shows:

> ...The lagoon of Venice also has one of the highest concentrations of masterpieces in the world: from Torcello's Cathedral to the church of Santa Maria della Salute. The years of the Republic's extraordinary Golden Age are represented by monuments of incomparable beauty...[5]

[2] http://whc.unesco.org/en/syndication.

[3] This field is not complete in the original XLS dataset. The WH Centre website is walked through to fill in the missing values.

[4] The statistics are up to the 44th session of the World Heritage Committee held in Fuzhou, China in July 2021, after which the total number of WH sites grew to 1154.

[5] https://whc.unesco.org/en/list/394

3.3.2 Data Collection and Pre-processing

For any inscribed WH site $p_i \in \mathcal{P}$, where $\mathcal{P}$ is the set of all the sites, it may fulfill one or more of the ten selection criteria. By checking if each criterion is justified for the site p_i, a non-negative vector $\boldsymbol{\gamma}_i := [\gamma_{i,k}]_{\kappa\times 1},\ k \in [1, \kappa], \kappa = 10$ can be formed as the "parental" label for the site:

$$\gamma_{i,k} = \begin{cases} 1, & \text{if } p_i \text{ meets the } k_{\text{th}} \text{ criterion,} \\ 0, & \text{otherwise.} \end{cases} \tag{3.1}$$

Meanwhile, the paragraphs $\boldsymbol{X}_i$ in the justification field of p_i, describing all criteria that p_i has, are split into sentences. For the j_{th} sentence $\boldsymbol{x}_{i,j,k}$ describing the criterion k possessed by the site p_i, a non-negative one-hot vector $\boldsymbol{y}_{i,j,k}$ can be formed as the "ground-truth" label for this single sentence:

$$\boldsymbol{y}_{i,j,k} = \boldsymbol{e}_k \in \{0, 1\}^{\kappa\times 1}. \tag{3.2}$$

Each sentence $\boldsymbol{x}_{i,j,k} \in \boldsymbol{X}_i$ is treated as a sample, with two labels: a one-hot "ground-truth label" $\boldsymbol{y}_{i,j,k}$ for the particular sentence, and a multi-class "parental label" $\boldsymbol{\gamma}_i$ for all sentences that belong to the site p_i. The sentence-level setup is desirable here since paragraphs may contain overwhelming information on multiple OUV criteria, as will be shown in Section 3.3.3. As such, a more specific indication of OUV tendencies in each part of the texts could be differentiated. Complementarily, the fine-grained sentence-level prediction vectors could still be aggregated into paragraph/text levels without losing lower-level details, which will be demonstrated in Figure 3.3. As the sentences were written, revised, and approved by various domain experts at local and global levels during the inscription process, the labels can be considered as having a good "inter-annotator agreement" (Jokilehto, 2008; Nowak and Rüger, 2010).

The following data pre-processing techniques are applied to construct the final dataset used for training: 1) all letters are turned into lower-case; 2) the umlauts and accents are normalized; 3) numbers are replaced with a special <NUM> token; 4) only sentences with a length between 8 and 64 words are kept, based on the dataset distribution; 5) the sentences are randomly split into train/validation/test sets with a proportion of 8:1:1. Additionally, the official definition sentences of selection criteria[6] as given in Table A.1 of Appendix A are respectively appended into the train split with the same one-hot sentence and parental labels for each criterion. Stop-words are not removed since BERT and ULMFiT to be applied generally prefer natural texts with context information. Furthermore, an additional 11_{th} class "Others" is introduced by appending an arbitrary noise of $\gamma_{i,\kappa+1} = 0.2$ to all parental labels $\boldsymbol{\gamma}_i$, and a 0 to all "ground-truth" labels $\boldsymbol{y}_{i,j,k}$, so that the models are not forced to give predictions only to the ten criteria even when the relevance to all of them is weak. For each sentence, the 11_{th} "Others" class and the complement sets of its parental labels could be regarded as the negative classes for classification since the site this sentence

[6] http://whc.unesco.org/en/criteria/

describes is not justified with those values. An exemplary pre-processed data sample is shown in Table 3.1. On average, 27.97 ± 11.04 words appear in each sentence. A summary of the number of samples in sentence level in each split for each criterion is presented in the first three rows of Table 3.2.

TABLE 3.1 An example of data sample concerning the WH property "Kalwaria Zebrzydowska: the Mannerist Architectural and Park Landscape Complex and Pilgrimage Park" in Poland, with the attributes of text data $\boldsymbol{x}_{i,j,k}$, sentence label as discrete index k, sentence label as one-hot vector $\boldsymbol{y}_{i,j,k}$ (appended with 0 in the end for the negative class "Others"), parental label as vector $\boldsymbol{\gamma}_i$ (appended with 0.2 in the end), sample length $|\boldsymbol{x}_{i,j,k}|$, index of parental WH property i, and the data split.

Attribute	Notation	Data
data	$\boldsymbol{x}_{i,j,k}$	the counter reformation of the late $<$ NUM $>$ th century led to a flowering in the creation of calvaries in europe
single label	k	Criterion (iv)
sentence label	$\boldsymbol{y}_{i,j,k}$	[0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0]
parental label	$\boldsymbol{\gamma}_i$	[0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0.2]
length	$\|\boldsymbol{x}_{i,j,k}\|$	18 (tokens)
property ID	i	905
data split		train

Similarly, the paragraphs $\boldsymbol{S}_i$ in the field short description of WH site p_i, giving a general introduction of the site, which are not originally written to describe any specific OUV selection criterion, are pre-processed into an additional independent test dataset Short Description (SD) to evaluate the generalizability of the classifiers on unseen data that comes from a slightly different distribution. For those sentences $\boldsymbol{s}_{i,o} \in \boldsymbol{S}_i$, both ground-truth and parental labels are the same as $\boldsymbol{\gamma_i}$ for the site they describe. The total number of samples that contain each criterion in Short Description (SD) dataset is shown in the fourth row of Table 3.2.

TABLE 3.2 The number of samples in sentence level that contain each criterion as a label, annotated with C1 to C6 for cultural values and N7 to N10 for natural values. The first three rows show the data split using the field justification; the fourth row shows a new dataset only for testing using the field short description (SD); the last row shows the potential samples the models can see for each criterion after introducing label smoothing (LS).

Split	C1	C2	C3	C4	C5	C6	N7	N8	N9	N10	Sum
train	333	631	651	774	209	327	386	261	370	572	4514
valid	40	71	83	89	28	49	43	42	42	76	563
test	41	79	72	92	35	47	45	32	50	71	564
test in SD	815	1563	1647	2049	554	876	510	334	465	548	9361
seen w LS	1077	1747	1832	2131	609	1063	1130	630	1047	1251	12517

3.3.3 Association between Classes

Jokilehto (2008) summarized the selection criteria with their main focuses by inspecting the official definitions and the justification texts of WH sites. Details about the definitions of the criteria could be found in Appendix A. However, as stated in Section 3.1, the criteria are not mutually exclusive. The **criterion (i)** justification of Venice in Section 3.3.2 will be again used as an example. Judging as a domain expert, it clearly describes criterion (i) as labelled, since it explicitly uses the terms “masterpieces” and “monuments of incomparable beauty”. However, traces can still be found on other values: 1) as it describes the “Cathedral”, “church”, and “monuments”, it also concerns the criterion (iv) about architectural typology; 2) as it talks about the “Golden Age”, it also points to criterion (ii) about influence and criterion (iii) about testimony. In fact, Venice is also justified with criteria (ii), (iii), and (iv). Pragmatically speaking, for sites fulfilling more than one OUV selection criteria, it is hard to avoid talking about the other criteria while isolating one criterion alone (Pereira Roders, 2010).

Furthermore, the association between each pair of criteria can be different. The distinction between criteria is generally larger when the pair comes from a different category (cultural v.s. natural). For a pair of criteria from the same category, the association level can also vary. For example, Jokilehto (2008) pointed out that

> “criteria (i) and (ii) can reinforce each other while (iv) is often used as an alternative”.

This complex association pattern can also be seen in the co-occurrence matrix $\boldsymbol{A} := [A_{k,l}]_{\kappa\times\kappa},\ k,l \in [1,\kappa]$ of the criteria in all the inscribed sites $\mathcal{P}$, where the diagonal entries record the number of cases when each criterion is used alone, and the off-diagonal entries $A_{k,l}, k \neq l$ are the number of properties that satisfy both criteria k and l (shown in Figure 3.6a):

$$A_{k,l} = \begin{cases} \sum_i \left(\gamma_{i,k}\gamma_{i,l}\right), & \text{if } k \neq l, \\ \sum_i \left\lfloor \frac{\gamma_{i,k}}{\sum_{j\in[1,\kappa]} \gamma_{i,j}} \right\rfloor, & \text{otherwise.} \end{cases} \tag{3.3}$$

TABLE 3.3 The distribution of the total number of selection criteria $\sum_{k=1}^{\kappa} \gamma_{i,k}$ a property is justified with.

N	Count	Proportion	Example
1	188	16.75%	Sydney Opera House
2	468	41.71%	Babylon
3	304	27,09%	City of Bath
4	103	9.18%	Yellowstone National Park
5	34	3.0%	Acropolis, Athens
6	4	0.36%	Venice and its Lagoon
7	2	0.18%	Mount Taishan

Among all the 1121 properties inscribed in the World Heritage List up to 2019, only

188 are justified with merely one criterion. The distribution of the total number of criteria justified for each property (i.e. $\sum_{k=1}^{\kappa} \gamma_{i,k}$) is shown in Table 3.3. This is an indication of the extent of how the problem characterizes a multi-label classification nature. Note 85.5% of properties are justified with no more than 3 criteria. The criteria from the same category are co-justified more often, while criteria (ii-iv), (iii-iv), and (ii-iii) are the most frequently co-occurred pairs.

This intrinsic association implied by the co-occurrence pattern is to be used as the prior knowledge for the classification task.

3.4 Experiments

3.4.1 Soft Labels Generation

Section 3.3.3 argues that the selection criteria are not mutually exclusive, and that co-justified criteria of a WH site that have a stronger association may be reflected in the sentences describing a specific criterion. In other words, classifying such sentences is not purely a single-label multi-class classification task. Rather, it also has a multi-label characteristic considering the “parental labels” of the sites.

To leverage the problem between the two sorts of tasks and to prevent the models from being over-confident at the only “ground-truth" labels, this paper proposes to apply the label smoothing (LS) technique with two novel variants to combine the “ground-truth” sentence label $\boldsymbol{y}_{i,j,k}$ and the parental document label $\boldsymbol{\gamma}_i$ into a single vector $\widetilde{\boldsymbol{y}}_{i,j,k}$ as soft labels for training process. This is similar to the hierarchical LS approach proposed by Zhong et al. (2016) to reflect the prior label similarity distribution. We propose three variants: **vanilla** that assigns identical “noises” to all classes, which will be proved equivalent to the original LS in Appendix B from Equations (B.1) to (B.7); **uniform** that treats all co-justified associated criteria in the parental label equally; and **prior** that weights the co-justified criteria based on the frequency that the pair co-occurs in matrix $\boldsymbol{A}$:

$$\widetilde{\boldsymbol{y}}_{i,j,k} = \begin{cases} \mathbf{f}(\boldsymbol{y}_{i,j,k} + \alpha\mathbf{1}), & \text{if } \textbf{vanilla}, \\ \mathbf{f}(\boldsymbol{y}_{i,j,k} + \alpha\boldsymbol{\gamma}_i), & \text{if } \textbf{uniform}, \\ \mathbf{f}(\boldsymbol{y}_{i,j,k} + \alpha\boldsymbol{\mu}_k \odot \boldsymbol{\gamma}_i), & \text{if } \textbf{prior}. \end{cases} \tag{3.4}$$

Here $\mathbf{f} : \mathbb{R}_+^d \to [0,1]^d$ is a variant of the original softmax function so that it maps a

$d-$dimensional vector of non-negative numbers to a distribution that sums up to 1:

$$\mathbf{f}(\boldsymbol{z})_t = \frac{e^{z_t} - 1}{\sum_{l=0}^{d} e^{z_l} - d}, \text{ or } \mathbf{f}(\boldsymbol{z}) = \frac{e^{\boldsymbol{z}} - \mathbf{1}}{e^{\boldsymbol{z}^\intercal}\mathbf{1} - d}, \tag{3.5}$$

$$\text{for } t \in [0, d), \mathbf{1} := [1]_{d\times 1} \text{ and } \boldsymbol{z} := [z_t]_{d\times 1} \in \mathbb{R}_+^d;$$

α is a scalar that leverages the effect of LS; $\boldsymbol{\mu}_k := [\mu_{l,k}]_{(\kappa+1)\times 1}$ is a criterion-specific non-negative vector showing the inter-criteria associations:

$$\mu_{l,k} = \frac{A_{l,k}}{\sum_i A_{i,k}}, \ l \in [1, \kappa + 1], \tag{3.6}$$

and $\odot$ represents the element-wise Hadamard-Schur product of vectors. This variant of the softmax function introduced in Equation (3.5) is preferable since it transforms the combined non-negative labels-vectors in Equation (3.4) to a "probability" distribution while keeping non-related labels still as 0. For example, a combined vector $[2, 0, 1, 0]^\intercal$ becomes $[.62, .08, .22, .08]^\intercal$ with normal softmax, and $[.79, 0, .21, 0]^\intercal$ with this variant.

All three variants are considered as options during training, and tuned as hyperparameters together with the scalar $\alpha \in \{0, 0.01, 0.05, 0.1, 0.2, 0.5, 1\}$. For all variants, the problem is purely multi-class when $\alpha = 0$, and approaches multi-label when α gets larger, giving parental labels larger weights.

The following benefits can be achieved with the use of proposed LS variants:

- The knowledge of the actual association of classes (selection criteria) are introduced into the training in both uniform and prior variants, giving the model chances to learn these intrinsic associations with soft labels;
- The freedom on the design decision of whether the problem should be multi-class or multi-label is provided for the model training process;
- The models can potentially see more instances for each class during training with LS variants, as shown in the last row of Table 3.2;
- The computed soft label vector $\widetilde{\boldsymbol{y}}_{i,j,k}$ is mathematically more similar to the prediction vector $\widehat{\boldsymbol{y}}_{i,j,k}$ than one-hot vectors, both of which are discrete "probability" distributions, pushing the use of Cross-entropy Loss closer to its original definition (Rubinstein and Kroese, 2013).

3.4.2 Natural Language Processing Models

Five models $\mathcal{M} = \{\mathfrak{m}_m | m = [0, 5)\}$ are selected as baselines: 1) N-gram (Cavnar and Trenkle, 1994) embedding followed by Multi-layer Perceptron (MLP); 2) Bag of Embeddings (BoE) using GloVe (Pennington et al., 2014); 3) Gated Recurrent Unit (GRU) (Cho et al., 2014) with Attention (Bahdanau et al., 2014; Yang et al., 2016) (denoted as GRU+Attn); 4) Pretrained ULMFiT language model (Howard and Ruder, 2018) further fine-tuned on the full WHL domain dataset; and 5) uncased base BERT model (Devlin et al., 2019). The former three models are trained mostly from scratch (where BoE and GRU+Attn used the GloVe-6B-300d vectors as initial embeddings), while the latter two are extensively pretrained and fine-tuned on this specific classification task. The model implementation details and the hyperparameter configurations are shown in Appendix B.

3.4.3 Evaluation Metrics for Model Training

For the training process, **Cross-Entropy** is used as the loss-function for two soft label vectors, while three metrics are used to evaluate the model performance as a multi-class classification task: 1) **Top-1 Accuracy** which counts the instances when the predicted class with the highest output value matches the ground-truth sentence label; 2) **Top-k Accuracy** which counts the instances when the ground-truth sentence label is among the top k predicted classes with the highest output values; 3) **Macro-averaged F1** which calculates the overall cross-label performance. **Per-class Metrics** (i.e., top-1 precision, recall, and F1) for each selection criteria are also calculated for evaluation purposes.

For the independent Short Description (SD) test set, two metrics are defined here to evaluate the model performance as a multi-label classification task: 1) **Top-1 Match** which counts the instances when at least one of the parental labels matches the predicted class; 2) **Top-k Match** which counts the instances when at least one parental label is among the top k predicted classes. Arguably, the top-1 and top-k matches are more tolerant extensions of top-1 and top-k accuracy into multi-label classification scenarios.

For all evaluation metrics, k is chosen to be 3 following the rationale introduced in Section 3.3.3.

Moreover, for model $\mathfrak{m}_m$, three confusion matrices $\boldsymbol{C}^{(m,s)} = [C_{k,l}^{(m,s)}]_{\kappa\times\kappa},\ k,l \in [0,\kappa)$, $s \in \{\text{train, val, test}\}$ were computed, where the entries $C_{k,l}^{(m,s)}$ represent the total number of data samples with a true label of criterion k being classified as criterion l by model $\mathfrak{m}_m$ in the s set (train, validation, or test). An example of the confusion matrix $\boldsymbol{C}^{(4,\text{test})}$ of $\mathfrak{m}_4$'s (ULMFiT) performance on test dataset is shown in Figure 3.6b.

3.4.4 Experiment Setup for Model Training

The experiment consists of three successive steps for each baseline:

1. Grid search within a small range is performed to tune the hyperparameters with a single random seed, and the best configuration is selected according to the top-k accuracy on the validation split;
2. LS with different α values under all three conditions (vanilla, uniform, and prior) is tested using the configuration from step 1, repeated with 10 different random seeds, treated as another round of hyperparameter tuning, saving the best LS configuration according to the performance mean and variance over the seeds;
3. The best LS configuration in step 2 is applied to save a model with the same random seed used in step 1 and evaluated together with the baseline model without LS, both on validation/test splits and on Short Description (SD) test set;

Early-stopping is applied during all training processes based on the top-k accuracy on the validation split. The models are implemented in PyTorch (Rao and McMahan, 2019) and experiments are performed on NVIDIA Tesla P100 GPU and Intel Core i7-8850H CPU, respectively. The inference is performed entirely on a CPU to test the models' feasibility in more general application scenarios when GPU can be unavailable for end-users. More details of the model configuration, training resource utilization, model size, and inference time are shown in Appendix B.

3.5 Ablation Studies

3.5.1 Expert Evaluation of Trained Models

Eight heritage researchers with rich experience in identifying heritage values and attributes were invited for a human study adapted from Nguyen (2018), Schuff (2020), and He et al. (2022), to test the models' reliability and generalizability. They were presented with 56 sentences about Venice harvested from "Justification" (14) and "Brief Synthesis" (13) in SOUV and Social Media platforms (29). Each sentence was given three positive classes as top-1 and top-3 criteria predictions from BERT and ULMFiT models, and one negative class as another random cultural criterion. Not knowing that the criteria are predictions by computer models, the experts were asked to rate the relevance of the sentences and each criterion on a 5-point Likert scale.

Materials

The materials about the WH property "Venice and Its Lagoon" for expert evaluation were harvested from three data sources: 1) all 14 sentences from Justification for Criteria section of SOUV, where each sentence has one "ground-truth" sentence label and a parental property label of Venice, which is also within the data $\boldsymbol{X}_i$ used during

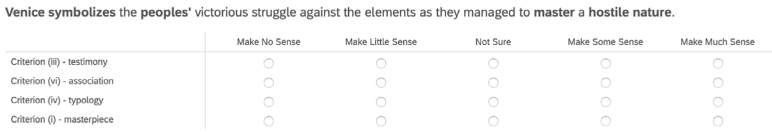


FIG. 3.1 The evaluation interface on Qualtrics.

model training and testing; 2) all 13 sentences from Brief Synthesis section of SOUV, where sentences only have a same multi-label parental label of Venice, which is similar with the Short Description (SD) test data $\boldsymbol{S}_i$ used for generalization test; 3) Social Media data sampled from a total of 1687 social media posts where a textual description is written, collected from Flickr in the region of Venice with a resolution of 5km using Flickr API[7]. Among the 1687 social media posts, there are 820 unique textual descriptions in English. By splitting the unique posts into sentences, removing html symbols, and filtering out the texts about camera parameters, image formats, and advertisements, 1132 sentences were obtained. The 1132 sentences were fed into the trained BERT and ULMFiT models. The sentences were further filtered based on the predictions: 1) the total confidence scores of top-3 predictions need to be larger than .8 by both models; 2) the Intersection over Union (IoU) of top-3 predictions by two models needs to be larger than .5 (i.e. maximum one different predicted class). 388 Social Media sentences that potentially convey OUV-related information were obtained. Furthermore, 29 sentences were randomly sampled from those 388 for the expert evaluation.

Survey Design

Each of the 56 sentences was fed into BERT and ULMFiT models to obtain the predictions and confidence scores. The predicted selection criteria with the highest confidence scores by both models were considered as the top-1 predictions. Two other criteria within the top-3 classes predicted by the both models with relatively high confidence scores were considered as the top-3 predictions for the survey. Another random cultural criterion that was not predicted by any model to be top-3 classes was considered as the negative class for each sentence. Natural criteria were not sampled as negative classes as they are not easily confused with positive ones. As a result, each sentence got four criteria to be evaluated. All four criteria were presented in a random order for each sentence, asking for an evaluation of the relevance of the sentence conveying the criterion on a 5-point Likert scale (from “5: make much sense”, to “1: make no sense”). The “important” words with higher attention weights in the GRU model were highlighted in bold. An example of such evaluation on the Qualtrics platform is shown in Figure 3.1. The sentences from the three data sources were grouped in four separate sessions, while the social media data were split into two sessions. The session of “justification for criteria” was always presented first during evaluation, also as a practice for the experts. The other three sessions were presented in a randomized order to prevent systematic errors caused by impatience or tiredness. Additional questions about the familiarity with heritage value identification, their familiarity with Venice, their confidence in evaluation, the

[7] https://pypi.org/project/flickrapi/

usefulness of highlighted words, and overall enjoyment and difficulty of the exercise were respectively raised before and after the evaluation, also with a 5-point Likert scale.

Since the expert evaluations are in ordinal scales, non-parametric statistical tests including Kruskal-Wallis H test, which is analogous to Analysis of Variance (ANOVA), and Mann-Whitney U test, which is analogous to $t-$ test, are conducted. The statistic analyses are performed with Scipy[8] and Statsmodels[9] libraries.

3.5.2 Computation of a Keyword Lexicon

A total of 2353 phrases composed of 1- to 5-Gram features (phrases with 1 to 5 consequent words) that appeared more than 15 times and less than 600 times in the SOUV were fed to each model $\mathfrak{m}_m$ mentioned above, predicting the scores of each phrase belonging to each criterion $k, k \in [0, \kappa+1)$, where the 11_{th} criterion referred to an additional negative class of "Others" related to none of the criteria. A series of ordered sets $\mathcal{W}_k^{(m)} = \{(\text{phrase } w, \text{rank } r)\}, |\mathcal{W}_k^{(m)}| = 50, r \in [1, 50]$ of phrases was obtained to contain the ranked top-50 keywords for criterion k predicted by the model $\mathfrak{m}_m$. The initial vocabulary can be composed of all the phrases as $\mathcal{V}^{(0)} = \bigcup_{k=0}^{\kappa+1} \bigcup_{m=0}^{5} \{w | (w, *) \in \mathcal{W}_k^{(m)}\}, |\mathcal{V}^{(0)}| = 1782$. A three-dimensional array $\boldsymbol{\Upsilon} = [\upsilon_{n,k,m}]_{|\mathcal{V}^{(0)}| \times (\kappa+1) \times 5}$ can be constructed for the j_{th} phrase w_n in the vocabulary $\mathcal{V}^{(0)}$ pertaining to its rank r in the criterion k predicted by model $\mathfrak{m}_m$, such that:

$$\upsilon_{n,k,m} = \begin{cases} r, & \text{if } (w_n, r) \in \mathcal{W}_k^{(m)}, \\ 0, & \text{otherwise.} \end{cases} \tag{3.7}$$

Lexicon, literally defined as

> “all the words and phrases used in a particular language or subject"[10]

was originally a linguistic concept, which requires some “morpholexical rules" to specify whether words should be members of some classes (Lieber, 1980). However, in modern NLP literature, the term “lexicon" is frequently referred to as a list of words that “carry particularly strong cues" of certain word senses, usually sentiment (Faruqui et al., 2015; Jurafsky and Martin, 2020). One of the most popularly used lexicons is the SentiWordNet, where each word is given scores for its tendency of being positive, negative, and objective (Esuli and Sebastiani, 2006). Such lexicons can be constructed by manual annotation, semi-supervised induction, and/or supervised learning. The initial entire vocabulary $\mathcal{V}^{(0)}$ has the following problems to

[8] https://docs.scipy.org/doc/scipy/reference/stats.html
[9] https://github.com/statsmodels/statsmodels
[10] Oxford Learner’s Dictionary

be considered as a lexicon, which needs to be revised and filtered: 1) some terms only appear in a limited number of models (especially in the worse performing models such as $\mathfrak{m}_1$ N-Gram model), which may be caused by the randomness of the models (e.g., “foot" was predicted with a high rank by $\mathfrak{m}_1$); 2) some terms always have lower confidence scores (lower ranks) in all models, which may suggest that they are not strongly relevant to the topic; 3) some terms are redundant since the longer N-Gram features may be accompanied by their subsets, for example “directly and tangibly associated" appears together with “directly and tangibly", “and tangibly associated", etc.; 4) stop-words such as prepositions and articles differentiate the word senses in their contexts (Devlin et al., 2019), but may not introduce additional semantic meanings when considered as keywords (e.g., “art of", “art in", and “art and" are all about the concept “art").

To improve these aspects, keywords are aggregated by taking advantage of the ensemble of models. Since the performance of the model may suggest the general reliability of predicted keywords, a model-related weight vector $\boldsymbol{\omega} := [\omega_m]_{5\times 1} = [1, 1, 1, \lambda_0, \lambda_0]^\intercal, \lambda_0 \geq 1 \in \mathbb{R}^+$ is arbitrarily formed to give the predictions by the latter two models a higher weight. Similarly, keywords predicted with higher confidence scores (higher ranks) may suggest that they are more related to the topic. Therefore, a rank-related weight vector $\boldsymbol{\zeta} := [\zeta_r]_{51\times 1} = [0, \lambda_1^2, ..., \lambda_1^2, \lambda_1, ..., \lambda_1, 1, ..., 1]^\intercal$, $\lambda_1 \geq 1 \in \mathbb{R}^+$ is also arbitrarily constructed to give higher-ranked keywords more importance, where the top-10 are amplified by the scalar λ_1^2, the $11_{\text{th}} - 25_{\text{th}}$ ranked phrases are amplified by λ_1, the $26_{\text{th}} - 50_{\text{th}}$ are kept the same, and those not ranked are omitted. The three-dimensional array $\boldsymbol{\Upsilon}$ in equation 3.7 can be therefore flattened on the model axis m to a matrix $\boldsymbol{\Upsilon}' := [\upsilon'_{n,k}]_{|\mathcal{V}_0|\times(\kappa+1)}$, such that:

$$\upsilon'_{n,k} = \sum_{m=0}^{4} \zeta[\upsilon_{n,k,m}]\omega[m]. \tag{3.8}$$

With a threshold $\lambda_2 \in \mathbb{R}^+$ to filter the computed weights in the matrix $\boldsymbol{\Upsilon}'$, a group of aggregated keyword sets $\mathcal{W}'_k$ can be obtained for each criterion k, such that:

$$\mathcal{W}'_k = \{(w_n, \upsilon'_{n,k}) | \upsilon'_{n,k} \geq \lambda_2\}. \tag{3.9}$$

Finding a properly filtered group of sets $\mathcal{W}'_k$ can be formulated as the following optimization problem, where $\mathcal{W}'_k$ is effectively a function of the three variables $\lambda_0, \lambda_1, \lambda_2$:

$$\max_{\lambda_0,\lambda_1,\lambda_2} \quad \frac{|\bigcup\limits_{\substack{k,l=0\\k\neq l}}^{\kappa+1} (\{w|(w,*) \in \mathcal{W}'_k\} \cap \{w|(w,*) \in \mathcal{W}'_l\})|}{|\bigcup\limits_{k=0}^{\kappa+1} \{w|(w,*) \in \mathcal{W}'_k\}| \times \sigma_{|\mathcal{W}'_k|} + \epsilon}, \tag{3.10a}$$

$$\text{subject to} \quad |\bigcup_{k=0}^{\kappa+1} \{w|(w,*) \in \mathcal{W}'_k\}| \leq N_0 = 800 \tag{3.10b}$$

$$\lambda_0, \lambda_1, \lambda_2 \in \{1.0, 1.1, 1.2, \ldots, 4.9\}. \tag{3.10c}$$

Where $\sigma_{|\mathcal{W}'_k|}$ denotes the standard deviation of the sizes of sets $\mathcal{W}'_k$, and ϵ is a small number to avoid zero division. This optimization ensures that: 1) there are enough phrases that fulfill more than one criterion (ensured by the nominator of equation (3.10a)); 2) the total size of the vocabulary is concise (ensured by N_0 in equation (3.10b)); 3) the sizes of keyword sets are evenly distributed across the criteria (ensured by $\sigma_{|\mathcal{W}'_k|}$ in the denominator of equation (3.10a)); and 4) the weights are in reasonable ranges for the filtering computation (ensured by equation (3.10c)).

Using a brute-force search for solving this optimization from a total of $|\lambda_0||\lambda_1||\lambda_2| = 64000$ configuration possibilities of discretized $\lambda_0, \lambda_1, \lambda_2$, a configuration of $\lambda_0 = 2.2, \lambda_1 = 1.2, \lambda_2 = 2.6$ yields the best filtering with a total vocabulary size of $|\mathcal{V}^{(1)}| = |\bigcup_{k=0}^{\kappa+1}\{w|(w,*) \in \mathcal{W}'_k\}| = 552$, among which 78 occur in more than one selection criteria. For the new vocabulary $\mathcal{V}^{(1)}$, `Stop-words` and `WordNet Lemmatizer` tools in the NLTK package (Miller, 1995; Loper and Bird, 2002) are used to further normalize and merge the keywords (as with the example of "art"). Furthermore, phrases composed of more than 2 words are merged to their longest N-Gram features (as with the example of "directly and tangibly associated").

After merging, a final lexicon as sets $\mathcal{W}_k$ is obtained, yielding a vocabulary size of $|\mathcal{V}| = |\bigcup_{k=0}^{\kappa+1}\{w|(w,*) \in \mathcal{W}_k\}| = 354$, among which 77 occur in more than one selection criteria.

3.5.3 Construction of Similarity Matrices

Co-occurrence matrix $\boldsymbol{A}$ of the selection criteria, as introduced in Section 3.3.3, shows how often two criteria are justified together, i.e. marked as relevant, for a WH property. The more often two criteria are fulfilled simultaneously, the more similar and associated they arguably are with one another. The term "similarity" here is from a **structural** viewpoint on the dataset. By normalizing matrix $\boldsymbol{A}$, the upper triangular entries can be "unrolled" and form a long vector $\boldsymbol{\alpha} = [\alpha_t]_{\frac{\kappa(\kappa-1)}{2}\times 1}, t \in [0, \frac{\kappa(\kappa-1)}{2})$, indexed with the ordered pair $(k,l), k < l$, representing the pair-wise similarity of the criteria, such that:

$$\{\alpha_t\} = \left\{ \frac{\kappa A_{k,l}}{\sum_{k_0}\sum_{l_0} A_{k_0,l_0}} | k,l \in [0,\kappa), k < l \right\}. \tag{3.11}$$

On the other hand, the confusion matrices $\boldsymbol{C}^{(m,s)}$ of the models during training and testing processes mentioned in Section 3.4.3 reveal how easily different selection criteria are to be misclassified as each other. Suppose the models are properly trained and represent certain degrees of truth, two criteria shall be more similar to one another as the models literally "confuse" them more often (Zhang et al., 2019). The term "similarity" here is an **experimental** viewpoint on the data concerning the NLP models' performances. However, before arguing that the confusion matrices reflect some intrinsic similarity, one must first prove that the models behave in a consistent manner, i.e., different models have difficulties at the same criteria pairs by easily confusing them. For each combination of the performance of model $\mathfrak{m}_m$ on either validation or test set s (training set performances are disregarded since the other two are supposed to better represent the prediction power of models), a similar construction as equation (3.11) can be applied to obtain long vectors $\boldsymbol{\beta}^{(m,s)} = [\beta_t^{(m,s)}]_{\frac{\kappa(\kappa-1)}{2}\times 1}, t \in [0, \frac{\kappa(\kappa-1)}{2})$ from the confusion matrices $\boldsymbol{C}^{(m,s)}$ following (Zhang et al., 2019), such that:

$$\{\beta_t^{(m,s)}\} = \left\{ \frac{C_{k,l}^{(m,s)}}{\sum_{k_0} C_{k_0,l}^{(m,s)}} + \frac{C_{l,k}^{(m,s)}}{\sum_{l_0} C_{l_0,k}^{(m,s)}} | k, l \in [0, \kappa), k < l \right\}. \tag{3.12}$$

Since the co-occurrence matrix $\boldsymbol{A}$ is symmetrical, the summation in Equation (3.12) is desirable as it transforms the generally asymmetrical confusion matrices into symmetric ones. The long vectors $\boldsymbol{\beta}^{(m,s)}$ are first compared to each other using Spearman's Rank Correlation to check the consistency of the models' performances. However, the null hypotheses in normal correlation analyses on such vectors can be easily refuted falsely because of the auto-correlated structures in matrices, making the normal significance tests invalid. A method called Quadratic Assignment Procedure (QAP) has been proposed to solve this problem (Krackhardt, 1988; Liu, 2007). By repeating the process of simultaneously permuting the rows and columns of one of the matrices before unrolling it to a vector for correlation computation, a theoretical distribution of the correlation coefficients can be obtained as a simulation outcome. The percentile of the original correlation coefficient (the one calculated without permutation) in this theoretical distribution can instead estimate the significance level of the correlation analyses effectively. The vectors are then fed to Principal Component Analysis (PCA) and Non-Negative Matrix Factorization (NMF) algorithms in Scikit-learn to perform dimensionality reduction and obtain the aggregated vector $\boldsymbol{\beta} = [\beta_t]_{\frac{\kappa(\kappa-1)}{2}\times 1}, t \in [0, \frac{\kappa(\kappa-1)}{2})$, representing the pair-wise confusion of the selection criteria (Févotte and Idier, 2011).

Furthermore, the final lexicon $\mathcal{V} = \bigcup_{k=0}^{\kappa+1}\{w|(w,*) \in \mathcal{W}_k\}$ discussed in section 3.5.2 can provide another level of interpretation on the criteria similarity. As suggested by the NLP literature (Wallach, 2006; Mikolov et al., 2013; Pennington et al., 2014), the pre-computed word embedding vectors provide good semantic meanings of the phrases, which can be further aggregated to represent the document topics composed of the ensemble of words. Therefore, another matrix $\boldsymbol{H} = [H_{k,l}]_{\kappa\times\kappa}$, $k, l \in [0, \kappa)$ showing the **semantic** similarity of the criteria can be constructed by computing the pair-wise cosine similarities of the averaged embedding vectors $\boldsymbol{f}_k$ of phrases in $\mathcal{W}_k$ for each criterion k, such that:

$$\boldsymbol{f}_k = \frac{\sum_{j=0}^{|V|} \mathbf{g}(w_n)}{|\mathcal{W}_k|} |(w_n, v'_{n,k}) \in \mathcal{W}_k. \tag{3.13}$$

Where $\mathbf{g}(w_n)$ is a function to look up the 300-dimensional GloVe embedding vectors of all the words in the phrase w_n and take the sum of the vectors. Similar to equation (3.11), another long vector $\boldsymbol{\gamma} = [\gamma_t]_{\frac{\kappa(\kappa-1)}{2}\times 1}, t \in [0, \frac{\kappa(\kappa-1)}{2})$ can be obtained to represent the pair-wise semantic similarities of the criteria.

$$\{\gamma_t\} = \left\{ H_{k,l} = \frac{\boldsymbol{f}_k^\top \boldsymbol{f}_l}{||\boldsymbol{f}_k||_2 ||\boldsymbol{f}_l||_2} | k, l \in [0, \kappa), k < l \right\}. \tag{3.14}$$

The three vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ are further compared to each other using Spearman's Rank Correlation (as they have different value distributions) to check the relationship and consistency of different similarity definitions based on QAP significance level.

3.5.4 Visualization

The vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ representing the pair-wise similarity of the selection criteria can be also interpreted as the edge weights of three undirected weighted unipartite graphs $\mathcal{G}_\alpha, \mathcal{G}_\beta, \mathcal{G}_\gamma$, where each node represents a specific criterion k. The graphs are visualized in Gephi using the Force Atlas algorithm based on the edge weights (Bastian et al., 2009; Jacomy et al., 2014). Since those graphs are (almost) complete with significantly divergent edge weights, different thresholds $\xi_\alpha, \xi_\beta, \xi_\gamma$ are applied to show only the edges whose weights are larger than the threshold based on the weight distributions, in order to give clearer structural information of the associations between the criteria.

Furthermore, the lexicon, i.e., the ensemble of sets $\bigcup_{k=0}^{\kappa+1} \mathcal{W}_k = \{(w_n, v'_{n,k})\}$ can also be interpreted as the edge table of an undirected weighted bipartite graph $\mathcal{B}_w$, where the two sets of nodes are respectively the vocabulary $\mathcal{V}$ and all the selection criteria. Moreover, as introduced in section 3.5.2, some phrases may belong to more than one criteria, and edge weights of such phrases can also vary across criteria. For example, the term "architectural" belongs to both Criterion (iv) with a weight of 5.70 and Criterion (i) with a weight of 4.75. In such cases, the degree of nodes representing the phrases will be the sum of weights from all edges connected to them. The lexicon as a bipartite graph is also visualized in Gephi using the Force Atlas algorithm based on the edge weights (Bastian et al., 2009; Jacomy et al., 2014).

3.6 Results

3.6.1 Experiment Results for Model Training

The averaged top-k accuracies of experiments conducted with 10 random seeds are shown in Figure 3.2. In most cases (except for BoE), the models with proposed LS variants (uniform or prior) either strictly or weakly outperform the baselines (without LS or with vanilla LS) based on multiple experiments. Furthermore, the proposed LS variants seem to make the models more robust to over-fitting and catastrophic forgetting problems, especially with the cases of BERT and ULMFiT. The uniform variant of LS with different α values appears in most models. A possible explanation is that uniform LS introduces the prior knowledge from the parental labels as "noise" in a simple way during the training, balancing yet not challenging the "ground-truth" sentence labels (Müller et al., 2019). Yet, the complex effect of LS on different baselines invites further investigation.

Table 3.4 shows the performance of the models with and without LS on the validation split, test split, and Short Description (SD) test set. Except for BoE, introducing LS increased the performance of most baselines in most metrics. Generally speaking, the pretrained models dominate the performance, and the highest score for all the metrics occurs in either ULMFiT or BERT, mostly with LS. Still, top-1 accuracy only reaches 71% in the best models, while top-k accuracy manages to reach 94%, suggesting that it would be more reliable to look at the top 3 predictions during application in this task. The models perform remarkably well in the SD test set, though given a relatively simpler task than in training, indicating the generalizability of the classifiers.

TABLE 3.4 The performance of models with and without LS on validation split, test split (top-1 accuracy, top-k accuracy, and averaged macro F1), and independent SD test set (top-1 match and top-k match), where k=3. The best score for each metric is highlighted in bold, and underlined if the best score occurs in models with LS in either variant of uniform (uni) or prior (pri). The effect of adding LS to each baseline is marked with background colors: blue indicates a rise in performance, red indicates a drop, while grey indicates a tie. The darker background color indicates a larger variation in performance.

Model	Config	val 1	val k	val F1	test 1	test k	test F1	SD 1	SD k
N-gram	w/o LS	67.38	90.82	63.11	59.96	88.87	58.87	70.49	95.13
	uni 0.1	67.19	91.21	62.11	59.57	89.65	58.24	71.12	95.26
BoE	w/o LS	64.84	91.99	63.11	62.11	91.60	61.93	68.80	94.53
	pri 0.01	64.26	91.60	62.48	62.70	91.41	62.14	66.15	94.14
GRU	w/o LS	64.26	91.60	60.83	60.55	91.41	59.28	64.27	92.71
+Attn	uni 0.2	64.26	91.80	61.36	61.52	90.23	61.06	66.35	94.06
ULMFiT	w/o LS	69.34	93.95	68.40	66.41	92.38	66.09	70.21	96.15
	pri 0.1	70.12	**<u>94.34</u>**	68.83	67.19	93.16	66.97	70.65	**<u>96.22</u>**
BERT	w/o LS	70.31	**94.34**	69.60	**67.58**	93.55	67.15	**71.56**	95.96
	uni 0.2	**<u>71.68</u>**	93.95	**<u>70.42</u>**	66.99	**<u>94.53</u>**	**<u>67.34</u>**	71.51	96.15

TABLE 3.5 The average per-class metrics over all models on validation and test splits with LS, and the main focus of each criterion adapted from Jokilehto (2008).

OUV	Focus	Prec	Recall	F1	OUV	Focus	Prec	Recall	F1
C1	Masterpiece	46.68	71.52	56.18	C6	Associations	58.28	67.89	61.27
C2	Values/Influences	69.19	66.34	67.56	N7	Natural Beauty	78.94	70.89	74.35
C3	Testimony	63.96	58.60	61.01	N8	Geological Process	66.92	80.42	72.39
C4	Typology	61.10	54.23	57.24	N9	Ecological Process	60.16	67.23	63.45
C5	Land-Use	40.98	52.30	45.01	N10	Bio-diversity	86.89	78.54	82.48

The per-class top-1 metrics of the best models in each baseline on the validation and

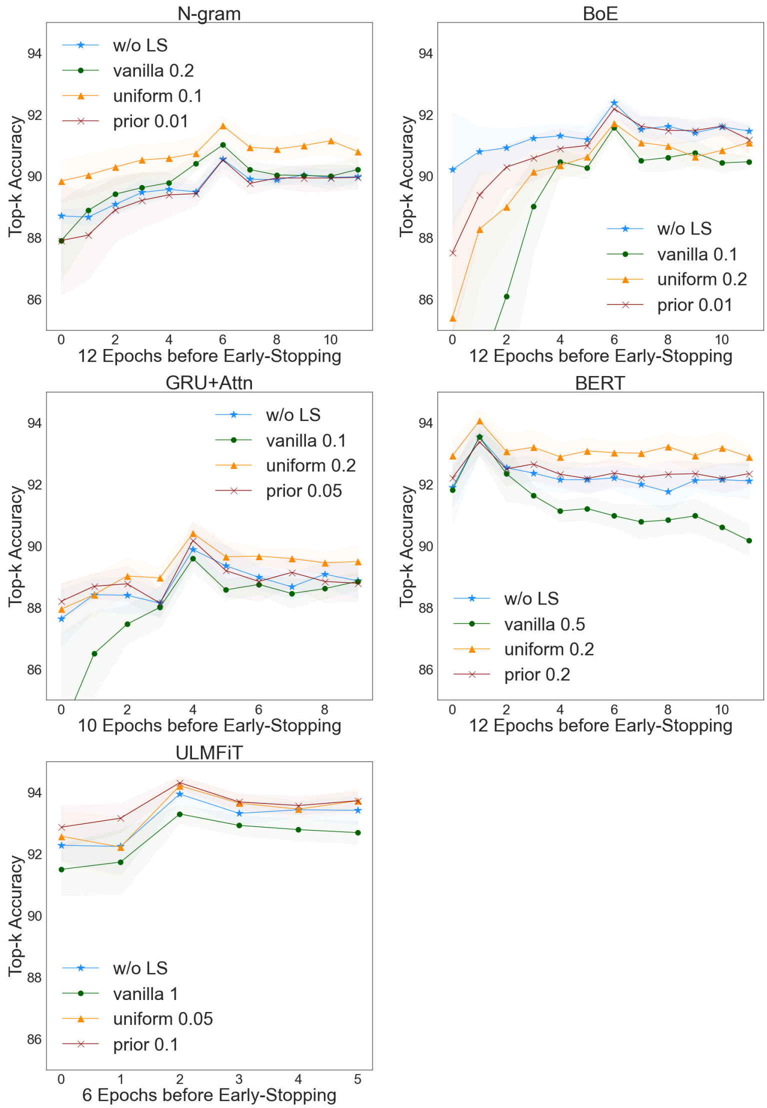


FIG. 3.2 The average training curve of best-performing models in experiments under 10 random seeds for each baseline on validation split. The x-axes show several epochs before the early-stopping happened. The numbers of epochs are different for each baseline as described in Appendix B. Orange curves with triangles show the top-k (k=3) accuracy with uniform LS, red curves with crosses the performance of prior LS, green curves with circles for vanilla LS, and blue curves with stars show the performance without LS. 95 % confidence intervals of the performance based on the 10 random seeds are shown in shades.

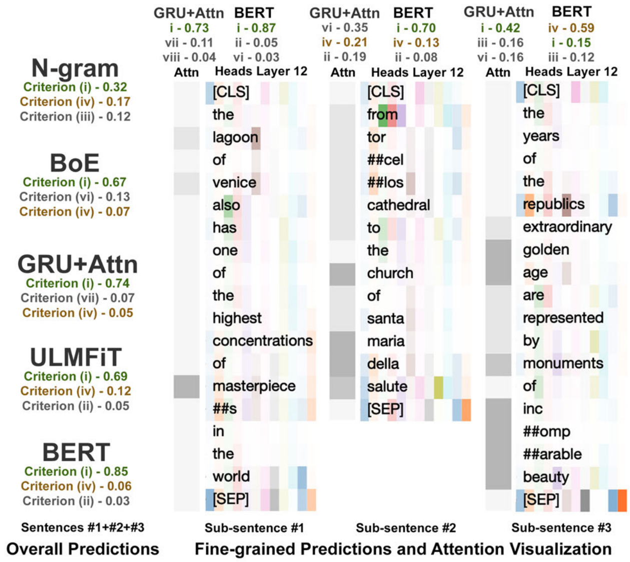


FIG. 3.3 The overall and fine-grained top-3 predictions of models, and attention weights of GRU+Attn and BERT models on the exemplary sub-sentences concerning criterion (i) in Venice. The left part of the image reports the top-3 predictions of all 5 models when the models take the aggregated paragraph as input. The top part reports the fine-grained top-3 predictions of two models on each sub-sentence. The rest of the image visualizes the attention weights. Attention weights of GRU+Attn is visualized in grey-scale, and that of BERT is illustrated using BertViz as coloured bars.

test split (Table 3.5) make it evident that the difficulty of classifying each selection criterion varies. T-test shows that the F1 score is significantly different between the cultural and natural criteria ($T = 8.20, p < .001$), suggesting that natural criteria are probably more clearly defined, while cultural ones might be closely intertwined. The poor performance on criterion (v) is consistent with its smallest sample size (as shown in Table 3.2); meanwhile, the models perform reasonably well for criterion (viii) with the second smallest sample size. This suggests that except for sample size, the strong associations between the classes can also influence the difficulty for NLP models (and probably also for human experts) to distinguish the nuance of criteria. Criterion (i) has a far poorer precision than recall, suggesting that samples from other criteria, especially from criterion (iv) based on the confusion matrices shown in Figure 3.6c, are easily mistaken as this one. This is also comprehensible since criterion (i), emphasizing that a site is a masterpiece, can be easily mentioned “unintentionally” in the description of criterion (iv) that regards the value of some specific architectural typology.

3.6.2 Error Analysis and Explainability

Although sometimes challenged (Serrano and Smith, 2020), attention mechanisms are believed to be effective for visualizing NLP model performance in an explainable manner (Yang et al., 2016; Vaswani et al., 2017; Tang et al., 2019; Sun and Lu, 2020). The same example on OUV selection criterion (i) in Venice as in Section 3.3.2 and 3.3.3 will be demonstrated here using the trained models from the attention-enabled GRU+Attn and BERT, as shown in Figure 3.3, with the help of BertViz library (Vaswani et al., 2018; Vig, 2019). GRU+Attn employs a single universal attention mechanism to all inputs, while BERT has 12 attention heads for the [CLS] token on its last layer, both of which manage to capture the meaningful keywords and phrases such as masterpiece, church, golden age, monuments, and incomparable beauty in the sentences. As a note, Clark et al. (2019) used probing to find out that some BERT attention heads correspond to certain linguistic phenomena. In this study, the attention heads from the last layer also seem to focus on different semantic information of OUV. This observation invites further studies.

Figure 3.3 also shows the top-3 predictions of the models on the exemplary sentences. In the overall predictions taking the sentences as a paragraph for input, all models manage to give the ground-truth label criterion (i) the highest predicted value (from 0.32 in N-gram to 0.85 in BERT). Remarkably, all models also include criterion (iv) in the top-3 predictions (from 0.05 in GRU+Attn to 0.17 in N-gram), suggesting that the sentences might also be related to criterion (iv). The fine-grained predictions taking each sub-sentence as input, however, show a different pattern. Although criterion (i) is almost always present in the top-3 predictions, criterion (iv) shows to take a higher place in the second sentence by GRU+Attn, and in the third sentence by BERT. This behaviour is not necessarily an error per se in prediction. Rather, considering the arguments in Section 3.3.3, those sub-sentences could be indeed relevant to other criteria (in this case, criterion iv) based on the association pattern, indicating why criterion (iv) is always included in the overall predictions.

3.6.3 Expert Evaluation Results

The expert evaluation mentioned in Section 3.5.1 took 55.10 ± 20.74 minutes to finish. The eight experts are all very familiar with the concept of OUV (4.38 ± 0.70) and the heritage values and attributes identification (4.75 ± 0.43), while not all are familiar with OUV justification (3.00 ± 1.50), nor with the cultural heritage in Venice (3.00 ± 1.41). The experts agree that the exercise in the evaluation was very hard (4.13 ± 0.93) and not so enjoyable (2.63 ± 1.32). They are more confident with identifying irrelevant sentence-criterion pairs (3.88 ± 0.78) than evaluating the relevant ones (3.00 ± 1.12). These show that the results of the expert evaluation are sufficiently reliable, that the heritage experts are cautious and critical of the process, that OUV justification is a difficult task even for experts as it is time-consuming and knowledge-demanding, and that a computational model is urgently needed to automate the classification if to be applied with massive social media data. The experts are not fully convinced that the highlighted words helped them with the justification process (2.88 ± 1.05), since the words provide both relevant information (3.13 ± 1.27) and irrelevant information (4.38 ± 0.70). This suggests that the explainability using the GRU attention mechanism needs further development.

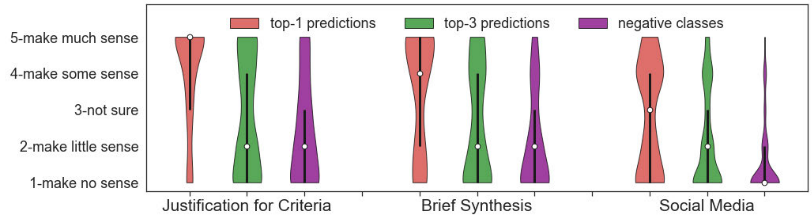


FIG. 3.4 The distribution as violin plots of expert evaluations given to the relevance of selection criteria and sample sentences about Venice from three sources. The scores for top-1 and top-3 classes and the negative class predicted by the models are plotted separately. The 25%, 75% percentiles and the medians are shown.

The distributions of all the ratings are shown in Figure 3.4. Kruskal-Wallis H tests show significant differences among the three types of criteria labels for all data sources, including for "justification of criteria" [$H(2) = 68.412, p < .001$], for "brief synthesis" [$H(2) = 40.351, p < .001$], and for "social media" [$H(2) = 102.321$, $p < .001$]. Post-hoc Mann-Whitney tests were used to compare all pairs of groups, as is shown in Table 3.6. The all-significant results of U tests show that the human experts gave significantly higher ratings to top-1 predictions than top-3 predictions, and to top-3 predictions than negative classes. The average ratings of experts for each sentence-criterion pair show a strong correlation with the average confidence scores of models ($r_p = .618, p < .001$). In other words, the human experts and computer models are consistently similar in differentiating the positive and negative criteria for the sentences concerning their relevance.

TABLE 3.6 The results of post-hoc Mann-Whitney U tests for the three types of labels within each data source. The medians (M) and counts (n) of each type are given together with the statistics from U tests.

Data Source	Type-1	Type-2	M_1	M_2	n_1	n_2	U value	p value
Justification	top-1 prediction	top-3 prediction	5	2	120	240	8157.0***	<.001
of Criteria	top-1 prediction	negative class	5	2	120	120	3161.0***	<.001
	top-3 prediction	negative class	2	2	240	120	12638.0*	.026
Brief	top-1 prediction	top-3 prediction	4	2	96	192	6256.0***	<.001
Synthesis	top-1 prediction	negative class	4	2	96	96	2401.5***	<.001
	top-3 prediction	negative class	2	2	192	96	7603.5**	.006
Social	top-1 prediction	top-3 prediction	3	2	232	464	40629.0***	<.001
Media	top-1 prediction	negative class	2	1	232	232	13784.5***	<.001
	top-3 prediction	negative class	2	1	464	232	39284.5***	<.001

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Some exemplary ratings of the experts and model predictions are given in Table 3.7. Some heritage experts seem to be rather cautious and reserved to assess informal texts as “culturally significant” without further historical contexts and comparative studies. For example, the third sentence in Table 3.7 from social media,

> “In 1952, the station was finalized on a design by the architect Paul Perilli”

with a predicted label of criterion (i) got extremely divergent expert scores. For some experts, it is clearly related to criterion (i) about masterpiece based on the semantic content. However, for the experts who rated a low score, merely declaring that some building is designed by a certain architect does not automatically entail that it is a masterpiece. Further investigations have to be made to fully convince them. Although such an example shows disagreement amongst the experts and between the experts and the computer models, it does not limit the machine’s ability to differentiate between positive and negative classes. The expert evaluation proves that the models are sufficiently reliable and capable of identifying OUV-related statements even from the less formal social media data, useful for the ultimate motivations of this study discussed in Section 3.1.

TABLE 3.7 Some example ratings on sentence-criterion relevance by human experts. The confidence scores by the computer models BERT and ULMFiT are also given.

Text	Criteria	Source	Type	BERT	ULMFiT	Ratings
With the unusualness of an archaeological site which still breathes life, Venice bears testimony unto itself.	iii	justification	top-1	0.744	0.825	5,5,5,3 5,5,4,5
Human interventions show high technical and creative skills in the realization of the hydraulic and architectural works in the lagoon area.	i	synthesis	top-1	0.607	0.590	4,5,5,1 4,4,2,5
In 1952, the station was finalized on a design by the architect Paul Perilli.	i	social media	top-1	0.757	0.529	5,4,1,1 1,3,1,1

3.6.4 OUV-related Lexicon of Selection Criteria

The visualized lexicon as bipartite graph $\mathcal{B}_w$ containing all phrases in $\mathcal{V}$ and their relationship with the selection criteria (including the negative class “Others") are shown in Figure 3.5. Generally, the essential topics of the criteria also appear to have the largest weights as the prediction from computational models. This is obvious in the cases of Criterion (i) with the phrase “masterpiece" and “human creative genius", (ii) with “influence" and “development", (iii) with “bear exceptional testimony", (iv) with “outstanding example" and “building", (v) with “traditional human settlement", (vi) with “directly and tangibly associated", (vii) with “exceptional natural beauty", (viii) with “geological process", (ix) with “ecological", and (x) with “species".

For each criterion, not only adjectives and verb phrases describing the **values**, but

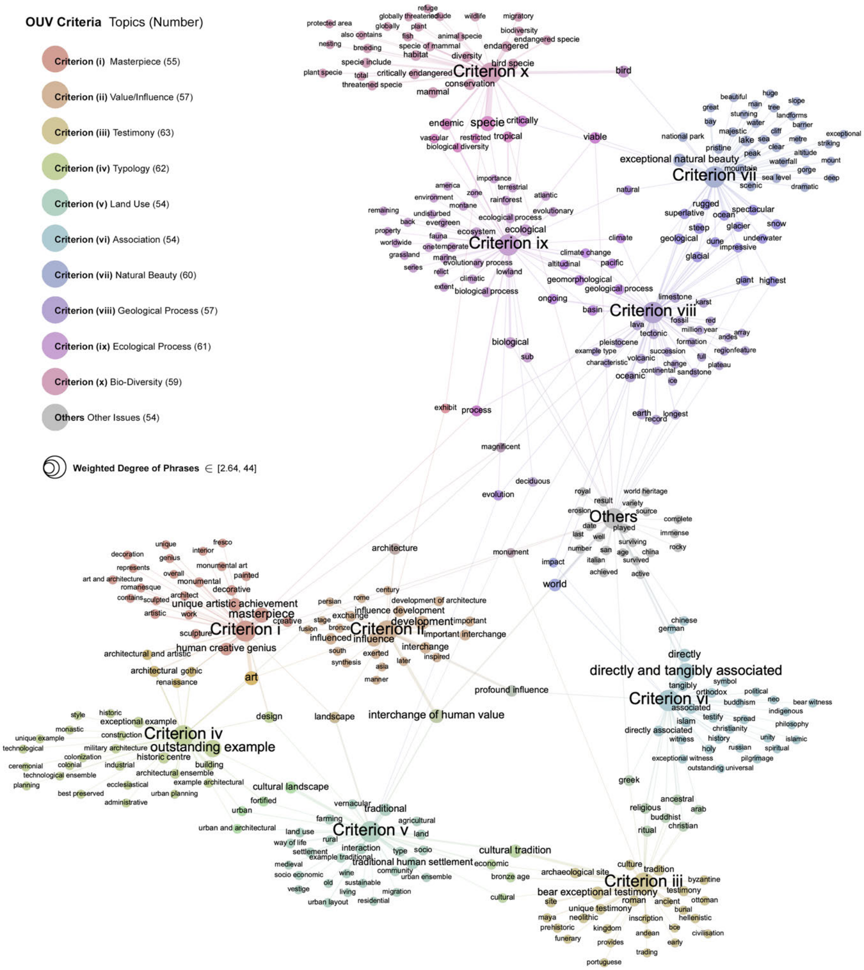


FIG. 3.5 The lexicon of selection criteria, i.e., the bipartite graph $\mathcal{B}_w$, visualized as a word network based on the Force Atlas algorithm in Gephi. Thicker edges indicate higher weights of the phrases in vocabulary $\mathcal{V}$ regarding a specific criterion. Nodes with higher edge weights are placed closer to each other in the visualization. Larger nodes and font sizes indicate larger total weighted degrees of the phrases. The colors of the phrase nodes are rendered the same as the criterion they belong to. The nodes of phrases belonging to two or more criteria are placed between the criteria clusters, and the colors of the nodes are also the mixture of the criteria colors. The general topics of criteria according to the ICOMOS report (Jokilehto, 2008) and the total number of keywords belonging to each criterion, i.e., $|\mathcal{W}_k|$ are demonstrated in the legend. This graph (lexicon) could be used to locate specific words regarding their relations with different selection criteria, and to observe and select the most relevant words while drafting and/or evaluating the Statements of OUV. Detailed interpretations of the lexicon are presented in Section 3.6.4.

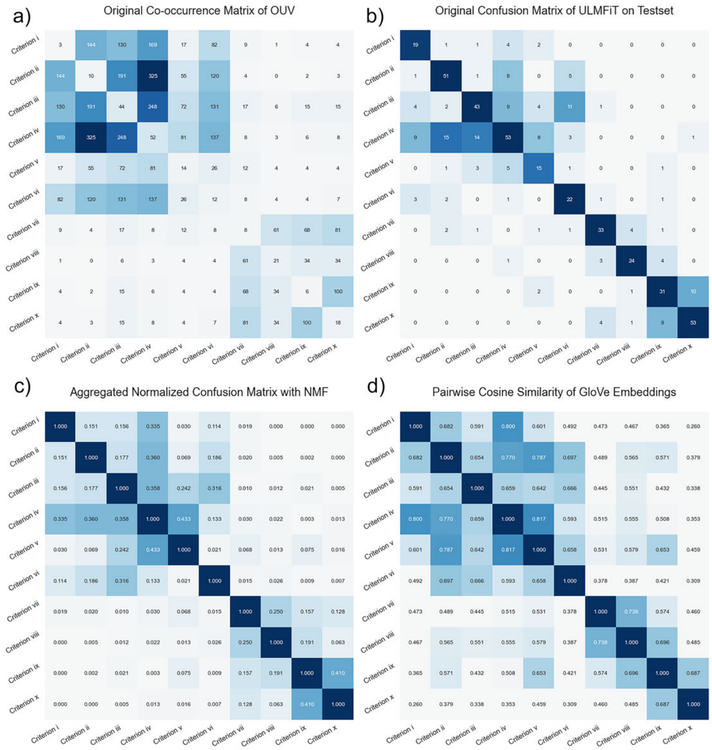


FIG. 3.6 The matrices representing the pairwise similarity and associations between selection criteria. a) the original (unnormalized) co-occurrence matrix $\boldsymbol{A}$; b) the original (unnormalized) confusion matrix $\boldsymbol{C}^{(4,\text{test})}$ by $\mathfrak{m}_4$ ULMFiT; c) the aggregated normalized confusion constructed from the NMF vector $\boldsymbol{\beta}$; d) the semantic similarity matrix $\boldsymbol{H}$ of the pairwise cosine similarity of GloVe embeddings for each criterion.

also nouns and noun phrases showing the critical **attributes** can be found. Take Criterion (i) as an example, phrases such as “unique artistic achievement, creative, genius, artistic, monumental" highlight the main artistic, aesthetic, and historic values associated with this criterion. Meanwhile representative attributes such as “fresco, sculpture, interior, decoration, art and architecture" demonstrate where those values are applied to.

Inspecting the phrases associated with more criteria can provide some insights into interpreting the common justifications of OUV. The terms “art" and “design" connect Criteria (i)(ii)(iv), while “landscape" connects Criteria (i)(ii)(v), and “cultural landscape" connects Criteria (iv)(v), showing the common stand-points and nuances in the focuses of those criteria. Moreover, the groups of phrases related to religions connecting Criteria (iii) and (vi), phrases about architectural art connecting (i) and (iv), about urban form connecting (iv) and (v), about natural phenomena between (vii) and (viii), as well as phrases about bio-creatures between Criteria (ix) and (x), etc., all imply some common characteristics within the OUV concept.

3.6.5 Associations and Similarities of Selection Criteria

All vector pairs from $\boldsymbol{\beta}^{(m,s)}$ have a high Spearman's Rank Correlation coefficient from $.713$ to $.933$, while all correlations are significant with $p < .001$ based on QAP simulation. This suggests that all the investigated confusion matrices perform consistently across models and datasets. Though models such as BERT and ULMFiT generally have better prediction accuracy, they are similarly confused at the same criteria pairs as the worse-performing models. Therefore, it is appropriate to aggregate the vectors $\boldsymbol{\beta}^{(m,s)}$ into $\boldsymbol{\beta}$ to represent the overall confusion patterns of the models. The first PCA component of the vectors manages to explain 89.7% of the variance in $\boldsymbol{\beta}^{(m,s)}$. However, due to the nature of PCA, some elements in its component are unavoidably negative, which can be hard to interpret as a similarity metric. Alternatively, the first component computed from NMF is non-negative, and has a Pearson Correlation of $r_p = 1.0$, $p < .001$ with the first PCA component. Therefore, the first NMF component from $\boldsymbol{\beta}^{(m,s)}$ is used as $\boldsymbol{\beta}$ for later analysis. This vector effectively makes a single matrix representative of the 10 possible variants of the $\mathcal{G}_\beta$, thus making this graph comparable to the other two graphs.

The values of the vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ are reflected in Figure 3.6 (a), (c), and (d), respectively. The matrix heatmaps generally illustrate a consistent visual pattern: 1) the top left corner indicating the cultural criteria associations and the bottom right corner indicating the natural criteria associations are stronger and create two relatively dense sub-matrices; 2) the off-diagonal entries highlight similar places, such as the entries representing the relation between Criteria (ii)(iv) and between Criteria (ix)(x). These patterns are further proved with correlation analysis. The Spearman's Rank Correlation of the vectors representing the similarities between selection criteria is shown in Table 3.8. All three pairs are significantly correlated with a high coefficient between $.615$ and $.838$, proving that the three proposed similarity matrices representing the structural (as co-occurrence matrix), experimental (as aggregated confusion matrix), and semantic (as cosine similarity matrix of GloVe embedding) information of the criteria are consistent with each other, though each one of the three may capture different aspects of the pair-wise associations. These aspects will be discussed extensively in Sections 3.3.3 and 3.7. The p values from QAP simulations out of 1000 random permutations indicate that such high correlations are significant, i.e. not caused by randomness.

TABLE 3.8 The Spearman's Rank Correlation ρ of three long vectors from the three matrices. The significance level p is computed based on QAP simulation.

Vector 1	Vector 2	ρ value	p value
$\boldsymbol{\alpha}$ (Structural from co-occurrence matrix $\boldsymbol{A}$)	$\boldsymbol{\beta}$ (Experimental from confusion matrix $\boldsymbol{C}$)	0.838*	$<.001$
$\boldsymbol{\alpha}$ (Structural from co-occurrence matrix $\boldsymbol{A}$	$\boldsymbol{\gamma}$ (Semantic from similarity matrix $\boldsymbol{H}$)	0.615*	$<.001$
$\boldsymbol{\beta}$ (Experimental from confusion matrix $\boldsymbol{C}$)	$\boldsymbol{\gamma}$ (Semantic from similarity matrix $\boldsymbol{H}$)	0.793*	$<.001$

*$p < .001$ with QAP simulation of 1000 permutations.

The similarity matrices showing the associations of selection criteria are further visualized in 2D as weighted graphs $\mathcal{G}_\alpha, \mathcal{G}_\beta, \mathcal{G}_\gamma$ in Figure 3.7, where the nodes representing more similar criteria are placed closer to each other. The graphs on the top are complete graphs showing all edge weights, while the graphs on the bottom are filtered graphs only showing the edges whose weights are equal or higher than the first two cross-domain edges linking cultural (i-vi) and natural (vii-x) criteria. The thresholds $\xi_\alpha, \xi_\beta, \xi_\gamma$ for conducting the filtering are also plotted on the histograms of the edge weights. It can be observed from the histograms that the edge weights in $\mathcal{G}_\alpha$ and $\mathcal{G}_\beta$ are more divergent, while in $\mathcal{G}_\gamma$, the edge weights are more homogeneous. As a consequence, $\mathcal{G}_\gamma$ is also visually more different from the other two similarity graphs.

By inspecting the visualization in Figure 3.7, consistent association and similarity patterns of the criteria can be observed from the graphs: 1) the in-domain edges generally have a larger weight than cross-domain edges, thus creating two sub-graph clusters for cultural and natural criteria in all graphs, suggesting that cultural and natural criteria are relatively independent of each other; 2) the first several cross-domain edges connecting cultural and natural criteria always involve either Criterion (v) about Land-Use or Criterion (iii) about Testimony, suggesting that these two cultural criteria also have a natural aspect; 3) the cultural criteria are generally more connected and interrelated than the natural ones, suggesting that the cultural criteria are probably more similarly defined and associated with each other than the natural criteria; 4) the edges between Criteria (ii) and (iv), and between Criteria (i) and (iv) are always among the top-5 weights in all three graphs (see the lists of Top 5 edges in Figure 3.7b/e/h), proving the strong association of Architectural Typology with both Masterpiece and cultural Influences; 5) the edge between Criteria (iv) and (v) appears to be the top-1 weight of both $\mathcal{G}_\beta$ and $\mathcal{G}_\gamma$, but is only the 13_{th} in $\mathcal{G}_\alpha$, showing that the association of Architectural heritage and Urban heritage might be stronger than indicated by the actual co-justification in WHL; 6) Contrarily, the edges between Criteria (iii) and (iv), and between Criteria (ii) and (iii) are ranked top-3 in graph $\mathcal{G}_\alpha$, yet respectively rank as 11_{th} in $\mathcal{G}_\gamma$ and $\mathcal{G}_\beta$, showing that although these criteria are usually co-justified in WH properties, they may not be that semantically similar or empirically confusing.

Remarkably, the strong associations indicated by the graphs in Figure 3.7 are also clearly illustrated with many common phrases (lexicon) in Figure 3.5, though the two figures are derived from different data sources and resolutions. The bipartite lexicon graph $\mathcal{B}_w$ in Figure 3.5 can be interpreted more as a zoomed-in view on the selection criteria composed of phrases, while the graphs $\mathcal{G}_\alpha, \mathcal{G}_\beta, \mathcal{G}_\gamma$ in Figure 3.7 arguably reflect a zoomed-out view on the characteristics of criteria themselves.

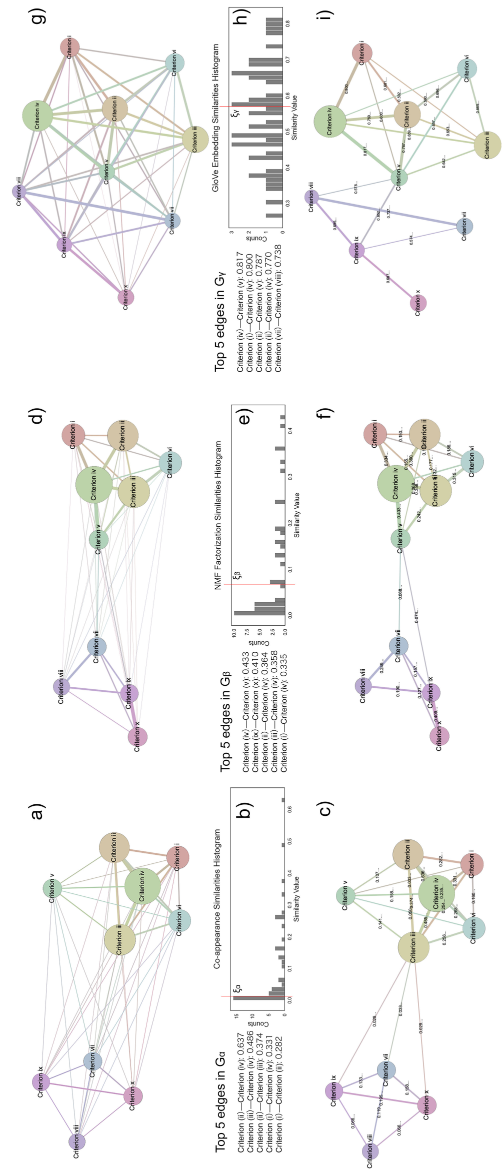


FIG. 3.7 The graph visualizations of the similarity matrices represented by $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ as edge weights using the Force Atlas algorithm in Gephi. a-c) Co-occurrence graph $\mathcal{G}_\alpha$; d-f) Confusion graph $\mathcal{G}_\beta$; g-i) Semantic similarity $\mathcal{G}_\gamma$; a/d/g) Complete graphs with all edge weights visualized; c/f/i) Filtered graphs that only show edges whose weights are higher than the first two cross-domain cultural-natural criteria pair; b/e/h) Histogram of edge weights and the threshold $\xi_\alpha, \xi_\beta, \xi_\gamma$ during filtering, the top-5 edges being listed with their weights. Node size represents the total World Heritage properties justified with this selection criterion.

3.7 Discussion

3.7.1 Application Scenarios and Broader Impact

This research of training a machine replica of the authoritative view with NLP models could help the identification and justification of heritage values across the world for various stakeholders, including both heritage experts and lay-persons, through text classification, as is pointed out in Section 3.1 and 3.7. It can lead to a better understanding of the OUV criteria and the association among them. This work is intended to aid, but not replace the workload of human stakeholders: for State Parties to identify OUV-related statements through documentation, for Advisory Bodies and WH Committee to review and revise the yearly nomination proposals, for researchers to investigate massive official discourse and user-generated content, and for the public to visually understand the values of their World Heritage around them. Therefore, this work **WHOSe Heritage** can be another milestone for the digital transformation of World Heritage Studies, aiming at a more socially inclusive future practice. Nevertheless, the interpretation of the classification result needs to be carefully conducted by researchers and practitioners, especially during policy decision-making on World Heritage for the social benefit of the entire human species. WH inscription and OUV justification are far more complicated than only reading written texts and identifying the described values. Rather, it is a systematic thematic study based on scientific research and always rooted in a comparative study across the globe (Jokilehto, 2008). The actual decisions of including new nominations into the WHL have to be made by humans with heritage investigations. This is also evident in the results of expert evaluation and during the open discussion about the exercise with invited experts. As stated in the example shown in Section 3.6.2, thorough heritage investigations are always needed to determine if a site truly justifies certain OUV selection criteria. Such investigations, however, would be out of the scope of an NLP study investigating the semantic and syntactic content of written official documents. Therefore, a human has to be involved in the loop during application.

The dataset used in this work is collected by the author(s) from the public website of UNESCO World Heritage Centre via XLS syndication respecting the terms of use and copyrights. The description of the dataset is sufficiently revealed in section 3.3.2. All labels used are based on the official OUV justification given by local and global heritage experts and involve no crowd workers or other new annotators. The dataset and the methods used in the chapter do not contain demographic/identity characteristics. Once deployed, the model does not learn from user inputs, and it generates no harmful output to users. The expert evaluation involving human study was totally voluntary, did not collect any personal information, and the privacy of the experts was fully protected. Though initially unaware of the true purpose of the evaluation to reduce bias, the experts were explained with the study afterwards. BERT and ULMFiT with LS proved to perform best in all investigated metrics. However, there is a trade-off to consider for real-world applications. As claimed in Appendix B and Section 3.6.2, ULMFiT has a relatively shorter inference time compared to BERT, while BERT is potentially more explainable due to the attention mechanism. Both models might work optimally for different application scenarios.

The lexicon presented in Figure 3.5 could become a tool for researchers and practitioners to automatically highlight the keywords in a sentence about World Heritage properties and indicate the best matching selection criteria, which also has the potential to facilitate the drafting and revising of SOUV, useful to support new WH nominations and their evaluation by the Advisory Body Evaluation parties, ICOMOS and IUCN. Since the computational models were trained with the authoritative context of WH properties, the lexicon derived from this study provides a chance to empirically investigate the patterns frequently appeared in SOUV which are captured and learned by the NLP models, while they can be easily neglected or undervalued with traditional methods. For example, Criterion (i) is officially defined as “to represent a masterpiece of human creative genius" in the Operational Guidelines and summarized as “masterpiece" by the 2008 report (Jokilehto, 2008; UNESCO, 2008). However, the term “unique artistic achievement" is boldly stressed by the computational models and the lexicon shown in Figure 3.5, suggesting that artistic value is also expected to be of high importance for the WH properties justified with Criterion (i). Similarly, though Jokilehto stressed more on the “value/influence" dimension of Criterion (ii), the terms related to “development" and “interchange" in its definition also seem to have alike importance. As the next step, the lexicon could be further updated with additional human engineering such as expert-based rating, as the current version is the outcome of a semi-automated procedure.

Some visual similarities can already be observed in Figure 3.6, as the heatmaps seem to highlight matrix entries in a similar pattern. This was also probably the assumption in ICOMOS 2008 report about the OUV associations, as argued in Section 3.1. Yet these similarities would be hard to prove and falsify without a quantitative methodology, such as the one presented in this paper. The correlation coefficients shown in Section 3.6.5 and the graphs $\mathcal{G}_\alpha, \mathcal{G}_\beta, \mathcal{G}_\gamma$ in Figure 3.7 confirm this intuitive assumption based on observations. Furthermore, while graph $\mathcal{G}_\alpha$ based on the co-occurrence pattern of the OUV criteria may vary radically due to the change of interest or focus of the WH Committee during the nomination procedure, the other two graphs might be more static along the time. The 2008 report argued that

> “[Criteria] (i) and (ii) can reinforce each other, while (iv) is often used as an alternative"

based on the co-occurrence pattern at that time, when cases co-justifying Criteria (i) and (ii) were almost twice as many as the cases with Criteria (i) and (iv) (Jokilehto, 2008). This observation is no longer true for the situation in 2019, when the latter, i.e. cases with Criteria (i) and (iv), appears even more frequently than the former. However, both associations are observed in the 4_{th} finding presented in Section 3.3.3. As graph $\mathcal{G}_\beta$ and $\mathcal{G}_\gamma$ are both based on the written texts and terms collectively used in the entire Statements of OUV, they may be more robust to new nominations unless very unusual terms are to be systematically introduced. It can also be informative in future studies to investigate the changing dynamic of presented graphs over time.

The qualitative and quantitative analyses show that the selection criteria pairs have different association strengths. For a thoroughly trained expert (either human or computer), nuances between pairs such as Criteria (i) and (iv) can already be rather hard to distinguish, let alone someone from the general public. To make the World Heritage management more socially inclusive, the concept of OUV more intelligible, and the future inscription process more effective, extra efforts may need to be made to further sharpen and clarify the definitions of criteria, and to make sure the OUV statements written by future practitioners and researchers are sufficiently consistent and coherent.

3.7.2 Limitations

Label Smoothing parameters proposed in this chapter were not tuned together with other hyperparameters during the training. Yet, it still showed an improvement in most baselines. However, the complex effect of LS on different baselines needs more investigation. The top-1 accuracy is limited even on the best models, which is not uncommon in the literature for non-binary multi-class classification when the labels are not sufficiently distinct (Sun et al., 2019). Applying data augmentation and training supplemental binary classifiers may improve the performance on difficult classes. The choice of replacing all numbers into `<NUM>` tokens might introduce both advantages and drawbacks in terms of semantic context and generalizability when historical dates might be crucial information, which invites more investigations. Moreover, more studies on the generalizability and reliability of the models on data from different distributions (e.g., from policy documents or news articles) are needed before further application. This work would support a series of follow-up studies respectively exploring the intrinsic associations of OUV based on the models' behaviour (Bai et al., 2021b), application of the proposed methods in social media mining in Venice (Bai et al., 2021c), and generalizability in case studies worldwide.

This study and the obtained NLP models are inherently less biased than manual annotation by a single expert in the sense that they avoid adding too much implicit personal experience into the written texts, and that the trained models represent the collective views of many human experts in the past. This can also be seen in some divergent evaluation outcomes by the eight invited experts, as demonstrated in Section 3.5.1: though one specific expert may be more cautious and critical at a certain sample, the overall trend of all experts can consistently differentiate the positive and negative classes. However, the computational models trained on SOUV can also be a double-edged sword in the sense that they are highly dependent on the existing descriptions, which may contain historical unfairness.

Researchers and practitioners, especially those outside of the Computer Science field, need to be explicitly informed and even warned before usage on the limitations of such models, to avoid automation bias, which shows that people favour the results automatically generated from systems for decision-making (Parasuraman and Manzey, 2010). Wrongly under-judging the value of a WH nomination merely based on text classification results and consequently deferring or even refusing the inscription can cause a great loss to human culture in the worst scenario, as it can hamper its access to the available heritage management and conservation programs. Therefore, this work functions as a supplemental tool and reference for the understanding/evaluating of World Heritage OUV implied in text descriptions, which will and shall not replace the human effort and/or deviate the expert knowledge in WH decision-making process. Instead, it has two ultimate goals as use-cases: 1) aiding inscription processes by checking the coherence and/or consistency of OUV statements; 2) mining heritage-values-related texts from multiple data sources (e.g., social media).

Although filtering as described in Section 3.5.2 has been applied, not every phrase in the lexicon makes sense. Some failure examples include the term “one" and “back" within Criterion (ix), “total" within Criterion (x), and “overall" within Criterion (i). Those terms should have been rather neutral, but probably the consistent writing style and word usage preference in Statements of OUV give some phrases a misleading score. Furthermore, the lexicon can be used as initial “seed words" in future studies to construct a more comprehensive and concrete World Heritage OUV-related lexicon by incorporating other larger and maturer semantic lexicons such as WordNet (Miller, 1995; Jurafsky and Martin, 2020).

3.8 Conclusions

This chapter presents a new text classification benchmark from a real-world problem about UNESCO World Heritage Statements of Outstanding Universal Value (OUV). The problem is essentially a multi-class single-label classification task, while the classes are not necessarily mutually exclusive. The prior knowledge of the class association is added to the training process as soft labels through novel variants of label smoothing (LS). The study shows that introducing LS improved the performance on most baselines, reaching a top-3 accuracy of 94.3%. The models also performed reasonably well in an independent test dataset and received positive outcomes in a human study with domain experts, suggesting that the classifiers have the potential to be further developed and applied in the World Heritage research and practice.

This chapter also presents the computational interpretation of the associations of OUV selection criteria conveyed by the properties, as an evolution of the ICOMOS report “What is OUV" published in 2008, applying a novel methodology integrating state-of-the-art technology. It provides an OUV-related lexicon showing relevant phrases of each selection criterion, proposes three similarity graphs using different data sources to show various aspects of the criteria associations, and conducts quantitative and qualitative analyses on the lexicon and similarity graphs to make sense of the observations. This study may give some insights into further evolutions and improvements of the concept of both World Heritage and OUV, as is also regularly revised by the World Heritage Committee[11].

References

Abdel Tawab (2019). The Assessment of Historic Towns' Outstanding Universal Value Based on the Interchange of Human Values They Exhibit. Heritage, 2(3):1874–1891.

Aly, M. (2005). Survey on multiclass classification methods. Neural Netw, 19:1–9.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

Bai, N., Luo, R., Nourian, P., and Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage statements of "Outstanding Universal Value" with soft labels. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 366–384, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2021b). "What is OUV" revisited: A computational interpretation on the statements of Outstanding Universal Value. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-M-1-2021:25–32.

Bai, N., Nourian, P., and Pereira Roders, A. (2021c). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.

Bandarin, F. and Van Oers, R. (2012). The historic urban landscape: managing heritage in an urban century. John Wiley & Sons.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In Proceedings of the International AAAI Conference on Web and Social Media, volume 3, pages 361–362.

Cavnar, W. B. and Trenkle, J. M. (1994). N-gram-based text categorization. In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, pages 161–175.

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. Syntax, Semantics and Structure in Statistical Translation, page 103.

Chorowski, J. and Jaitly, N. (2017). Towards better decoding and language model integration in sequence to sequence models. In Proc. Interspeech 2017, pages 523–527.

Clark, K., Khandelwal, U., Levy, O., and Manning, C. D. (2019). What does BERT look at? an analysis of BERT's attention. In Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pages 276–286, Florence, Italy. Association for Computational Linguistics.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In LREC, volume 6, pages 417–422.

Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., and Smith, N. A. (2015). Retrofitting word vectors to semantic lexicons. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1606–1615.

Fejérdy, T. (2007). Evolution and possible enhancement of the concept of ouv. In Values and Criteria in Heritage Conservation, pages 323–328. Polistampa.

[11] http://whc.unesco.org/en/criteria/

Févotte, C. and Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. Neural computation, 23(9):2421–2456.
Ginzarly, M., Pereira Roders, A., and Teller, J. (2019). Mapping historic urban landscape values through social media. Journal of Cultural Heritage, 36:1–11.
He, Z., Deng, N., Li, X., and Gu, H. (2022). How to “read” a destination from images? machine learning and network methods for dmos’ image projection and photo evaluation. Journal of Travel Research, 61(3):597–619.
Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. In Gurevych, I. and Miyao, Y., editors, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers, pages 328–339. Association for Computational Linguistics.
IUCN, ICOMOS, ICROM, and WHC (2010). Guidance on the preparation of retrospective Statements of Outstanding Universal Value for World Heritage Properties. Technical report, IUCN, ICOMOS, ICROM and WHC.
Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. PloS one, 9(6):e98679.
Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In Nedellec, C. and Rouveirol, C., editors, Machine Learning: ECML-98, 10th European Conference on Machine Learning, Chemnitz, Germany, April 21-23, 1998, Proceedings, volume 1398 of Lecture Notes in Computer Science, pages 137–142. Springer.
Johnson, R. and Zhang, T. (2017). Deep pyramid convolutional neural networks for text categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 562–570, Vancouver, Canada. Association for Computational Linguistics.
Jokilehto, J. (2006). World Heritage: Defining the Outstanding Universal Value. City & Time, 2(2):1–10.
Jokilehto, J. (2007). Aesthetics in the world heritage context. In Values and Criteria in Heritage Conservation, pages 183–194. Polistampa.
Jokilehto, J. (2008). What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties. Technical report, ICOMOS, ICOMOS Berlin.
Jurafsky, D. and Martin, J. H. (2020). Speech and language processing An Intoduction to Natural Language Processing, Computational Liguistics, and Speech Recognition. Third Edition Draft.
Kim, Y. (2014). Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.
Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. Social networks, 10(4):359–381.
Krothapalli, U. and Abbott, A. L. (2020). Adaptive label smoothing. arXiv preprint arXiv:2009.06432.
Lieber, R. (1980). On the organization of the lexicon. PhD thesis, Massachusetts Institute of Technology.
Liu, J. (2007). Qap: A unique method of measuring "relationships" in relational data. Chinese Journal of Sociology(in Chinese Version), 27(4):164–174.
Loper, E. and Bird, S. (2002). Nltk: the natural language toolkit. In Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1, pages 63–70.
Maron, M. E. (1961). Automatic indexing: An experimental inquiry. J. ACM, 8(3):404–417.
Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. In Bengio, Y. and LeCun, Y., editors, 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings.
Miller, G. A. (1995). Wordnet: a lexical database for english. Communications of the ACM, 38(11):39–41.
Müller, R., Kornblith, S., and Hinton, G. (2019). When does label smoothing help? In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Red Hook, NY, USA. Curran Associates Inc.
Nguyen, D. (2018). Comparing automatic and human evaluation of local explanations for text classification. In Walker, M. A., Ji, H., and Stent, A., editors, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers), pages 1069–1078. Association for Computational Linguistics.
Nowak, S. and Rüger, S. (2010). How reliable are annotations via crowdsourcing. In Proceedings of the international conference on Multimedia information retrieval, pages 557–566.
Pal, A., Selvakumar, M., and Sankarasubbu, M. (2020). MAGNET: multi-label text classification using attention-based graph neural network. In Rocha, A. P., Steels, L., and van den Herik, H. J., editors, Proceedings of the 12th International Conference on Agents and Artificial Intelligence, ICAART 2020, Volume 2, Valletta, Malta, February 22-24, 2020, pages 494–505. SCITEPRESS.
Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359.
Parasuraman, R. and Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3):381–410.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In Moschitti, A., Pang, B., and Daelemans, W., editors, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1532–1543. ACL.

Pereira Roders, A. (2010). Revealing the World Heritage cities and their varied natures. In Heritage 2010: Heritage and Sustainable Development, Vols 1 and 2, chapter Heritage a, pages 245–253. Green Lines Institute.

Pereira Roders, A. and van Oers, R. (2011). World Heritage cities management. Facilities, 29(7):276–285.

Petzet, M. (2007). What is outstanding universal value. In Values and Criteria in Heritage Conservation, pages 315–322. Polistampa.

Rao, D. and McMahan, B. (2019). Natural Language Processing with PyTorch - Build Intelligent Language Applications Using Deep Learning. O'Reilly Media, Inc.

Rubinstein, R. Y. and Kroese, D. P. (2013). The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning. Springer Science & Business Media.

Ruffino, P., Permadi, D., Gandino, E., Haron, A., Osello, A., and Wong, C. (2019). Digital technologies for inclusive cultural heritage: The case study of serralunga d'alba castle. In ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, volume 4, pages 141–147.

Schuff, H. (2020). Explainable question answering beyond f1: metrics, models and human evaluation. Master's thesis, Universitaet Stuttgart.

Serrano, S. and Smith, N. A. (2020). Is attention interpretable? ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, pages 2931–2951.

Shah, K. (2015). Documentation and cultural heritage inventories case of the historic city of ahmadabad. In ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, volume 2, pages 271–278.

Sun, C., Qiu, X., Xu, Y., and Huang, X. (2019). How to Fine-Tune BERT for Text Classification? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11856 LNAI(2):194–206.

Sun, X. and Lu, W. (2020). Understanding Attention for Text Classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3418–3428.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826.

Tai, K. S., Socher, R., and Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers, pages 1556–1566. The Association for Computer Linguistics.

Tang, M., Gandhi, P., Kabir, M. A., Zou, C., Blakey, J., and Luo, X. (2019). Progress notes classification and keyword extraction using attention based deep learning models with BERT. arXiv.

Tarrafa Silva, A. and Pereira Roders, A. (2010). The cultural significance of World Heritage cities : Portugal as case study. In Heritage and Sustainable Development, pages 255–263, Évora, Portugal.

Tarrafa Silva, A. and Pereira Roders, A. (2012). Cultural Heritage Management and Heritage (Impact) Assessments. In Joint CIB W070, W092 & TG72 International Conferenceon Facilities Management, Procurement Systems and Public Private Partnership.

Tsoumakas, G. and Katakis, I. (2007). Multi-label classification: An overview. International Journal of Data Warehousing and Mining (IJDWM), 3(3):1–13.

UNESCO (1972). Convention Concerning the Protection of the World Cultural and Natural Heritage. Technical Report november, UNESCO, Paris.

UNESCO (2008). Operational guidelines for the implementation of the world heritage convention. Technical Report July, UNESCO World Heritage Centre.

UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.

Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A., Gouws, S., Jones, L., Kaiser, Ł., Kalchbrenner, N., Parmar, N., et al. (2018). Tensor2tensor for neural machine translation. In Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track), pages 193–199.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 6000–6010.

Vig, J. (2019). A multiscale visualization of attention in the transformer model. In Costa-jussà, M. R. and Alfonseca, E., editors, Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 3: System Demonstrations, pages 37–42. Association for Computational Linguistics.

von Droste, B. (2011). The concept of outstanding universal value and its application: "From the seven wonders of the ancient world to the 1,000 world heritage places today". Journal of Cultural Heritage Management and Sustainable Development, 1(1):26–41.

Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. In Proceedings of the 23rd international conference on Machine learning, pages 977–984.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. J., and Hovy, E. H. (2016). Hierarchical attention networks for document classification. In Knight, K., Nenkova, A., and Rambow, O., editors, NAACL HLT 2016, pages 1480–1489.

Zhang, C.-B., Jiang, P.-T., Hou, Q., Wei, Y., Han, Q., Li, Z., and Cheng, M.-M. (2020). Delving deep into label smoothing. arXiv preprint arXiv:2011.12562.

Zhang, F., Zhou, B., Ratti, C., and Liu, Y. (2019). Discovering place-informative scenes and objects using social media photos. Royal Society open science, 6(3):181375.

Zhong, Q., Li, C., Zhang, Y., Sun, H., Yang, S., Xie, D., and Pu, S. (2016). Towards good practices for recognition & detection. In CVPR workshops, volume 1.

PART C

On Context

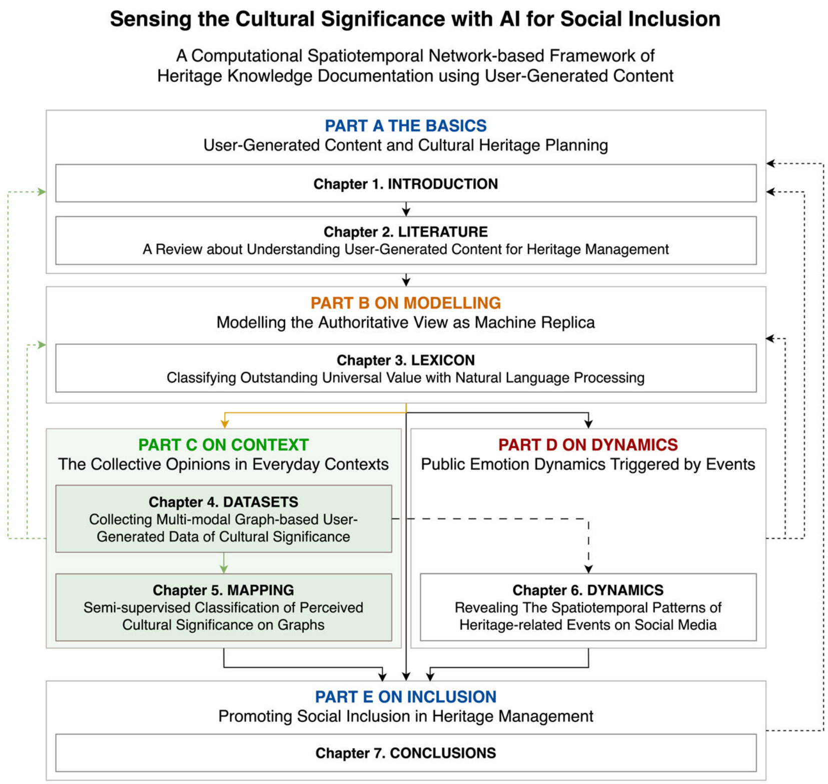
The Collective Opinions in Everyday Contexts

This part of dissertation focuses on the everyday baseline scenario when the collective opinions are shared on social media about the cities people visit or live in. A methodological framework is proposed to map the cultural significance conveyed to people by the collection, process, analysis, and summary of user-generated multi-modal social media data. The unstructured images and texts are converted to structured vectors understandable by computers with the aid of pre-trained deep learning models including the machine replica developed in PART B. Following the knowledge base of PART A, this part treats the spatiotemporal and social contexts of social media data as a crucial component of information in addition to the content, connecting them in network structures. The information extracted from social media posts are projected to a series of spatial maps, supposedly comprehensible by designers, planners, and decision-makers.

Two chapters are included in this part:

Chapter 4 Datasets - Collecting Multi-modal Graph-based User-Generated Data of Cultural Significance.

Chapter 5 Mapping - Semi-supervised Classification of Perceived Cultural Significance on Graphs.


Sensing the Cultural Significance with AI for Social Inclusion
A Computational Spatiotemporal Network-based Framework of
Heritage Knowledge Documentation using User-Generated Content
PART A THE BASICS
User-Generated Content and Cultural Heritage Planning
Chapter 1. INTRODUCTION
Chapter 2. LITERATURE
A Review about Understanding User-Generated Content for Heritage Management
PART B ON MODELLING
Modelling the Authoritative View as Machine Replica
Chapter 3. LEXICON
Classifying Outstanding Universal Value with Natural Language Processing
PART C ON CONTEXT
The Collective Opinions in Everyday Contexts
Chapter 4. DATASETS
Collecting Multi-modal Graph-based User-Generated Data of Cultural Significance
Chapter 5. MAPPING
Semi-supervised Classification of Perceived Cultural Significance on Graphs
PART D ON DYNAMICS
Public Emotion Dynamics Triggered by Events
Chapter 6. DYNAMICS
Revealing The Spatiotemporal Patterns of Heritage-related Events on Social Media
PART E ON INCLUSION
Promoting Social Inclusion in Heritage Management
Chapter 7. CONCLUSIONS

4 Datasets

Collecting Multi-modal Graph-based User-Generated Data of Cultural Significance

Parts of this chapter have been published in Bai et al. (2022) and Bai et al. (2023).
Bai N, Nourian P, Luo R, Pereira Roders A. (2022). Heri-Graphs: A Dataset Creation Framework for Multi-Modal Machine Learning on Graphs of Heritage Values and Attributes with Social Media. ISPRS International Journal of Geo-Information. 11(9): 469.
Bai N, Ducci M, Mirzikashvili R, Nourian P, Pereira Roders, A. (2023). Mapping Urban Heritage Images with Social Media Data and Artificial Intelligence, A Case Study in Testaccio, Rome. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023. p. 139–146.

ABSTRACT Values (why to conserve) and Attributes (what to conserve) are essential concepts of cultural heritage. Recent studies have been using social media to map values and attributes conveyed by public to cultural heritage. However, it is rare to connect heterogeneous modalities of images, texts, geo-locations, timestamps, and social network structures to mine the semantic and structural characteristics therein. This study presents a methodological framework for constructing such multi-modal datasets using posts and images on Flickr for graph-based machine learning (ML) tasks concerning heritage values and attributes. After data pre-processing using pre-trained ML models, the multi-modal information of visual contents and textual semantics are modelled as node features and labels, while their social relationships and spatiotemporal contexts are modelled as links in Multi-Graphs. The framework is tested in three cities with urban areas inscribed in the UNESCO WHL - Amsterdam, Suzhou, and Venice, which yielded datasets with high consistency for semi-supervised learning tasks. The entire process is formally described with mathematical notations, ready to be applied in provisional tasks both as ML problems with technical relevance and as urban/heritage study questions with societal interests. This study could also benefit the understanding and mapping of heritage values and attributes for future research in global cases, aiming at inclusive heritage management practices. Moreover, the proposed framework could be summarized as creating attributed graphs from unstructured social media data sources, ready to be applied in a wide range of use cases.

KEYWORDS UNESCO World Heritage, Flickr, Multi-modal Dataset, Graph Construction, Machine and Deep Learning.

4.1 Introduction

In the context of UNESCO World Heritage (WH) Convention, "values" (why to conserve) and "attributes" (what to conserve) have been used extensively to detail the cultural significance of heritage (UNESCO, 1972, 2008). Meanwhile, researchers have provided categories and taxonomies for heritage values and attributes, respectively (Pereira Roders, 2007; Tarrafa Silva and Pereira Roders, 2010; Veldpaus, 2015). Both concepts are essential for understanding the significance and meaning of cultural and natural heritage, and for making more comprehensive management plans (Veldpaus, 2015). However, the heritage values and attributes are not only to define the significance of Outstanding Universal Value (OUV) in the particular context of World Heritage List (WHL), but all kinds of significance, ranging from listed to unlisted, natural to cultural, tangible to intangible, and from global to national, regional and local (Rakic and Chambers, 2008; Tarrafa Silva and Pereira Roders, 2010; Bonci et al., 2018; Pereira Roders, 2019; Bai et al., 2021c). Moreover, the 2011 UNESCO Recommendation on the Historic Urban Landscape (HUL) stressed that heritage should also be recognized through the lens of local citizens, tourists and experts, calling for tools for civic engagement and knowledge documentation (UNESCO, 2011; Pereira Roders, 2019; Bai et al., 2021c).

Thereafter, in the past decade, analyses have been performed on User-Generated Content (UGC) from social media platforms to actively collect opinions of the [online] public, and to map heritage values and attributes conveyed by various stakeholders in urban environments (Lu and Stepchenkova, 2015; Pickering et al., 2018). In Machine Learning (ML) literature, a modality is defined as

> "the way in which something happens or is experienced",

which can include natural language, visual contents, vocal signals, etc. (Baltrusaitis et al., 2019). Most of studies mapping heritage values and attributes from UGC focused only on a few isolated modalities, such as textual topics of comments and/or tags (Marine-Roig and Anton Clavé, 2015; Amato et al., 2016; Lee and Kang, 2021), visual contents of depicted scenes (Giglio et al., 2019b; Zhang et al., 2019), social interactions (Liew, 2014; Williams et al., 2017; Campillo-Alhama and Martinez-Sala, 2019), and geographical distribution of the posts (Gabrielli et al., 2014; Giglio et al., 2019a).

However, the heterogeneous multi-modal information from social media can enrich the understanding of posts, as textual and visual contents, temporal and geographical contexts, and underlined social network structures could show both complementary and contradictory messages (Aggarwal, 2011; Bai et al., 2021c). A few studies have analysed different modalities to reveal the discussed topics and depicted scenes about cultural heritage (Monteiro et al., 2014; Ginzarly et al., 2019). However, since they (mostly) adapted analogue approaches during analyses and the multi-modal information was not explicitly paired, linked, and analysed together, these studies could not yet be classified as Multi-modal Machine Learning (MML), aiming to

> "build models that can process and relate information from multiple modalities"

to enrich the conclusions that could not be easily achieved with isolated modalities (Baltrusaitis et al., 2019). On the other hand, Crandall et al. (2009) proposed a global dataset collected from Flickr with visual and textual features, as well as geographical locations. Graphs were constructed with multi-modal information to map, cluster, and retrieve the most representative landmark images for major global cities. Gomez et al. (2019) trained multi-modal representation models of images, captions, and neighbourhoods with Instagram data within Barcelona, able to retrieve the most relevant photos and topics for each municipal district, being used to interpret the urban characteristics of different neighbourhoods. More recently, the continuous research line demonstrated in Kang et al. (2021) and Cho et al. (2022) applied transfer learning (Pan and Yang, 2010) techniques to classify geo-tagged images into hierarchical scene categories and connected the depicted tourist activities to the urban environments that these cultural activities took place. Although not all of them explicitly referred to heritage, these studies could provide useful information for scholars and practitioners to gather knowledge from the public about their perceived heritage values and attributes in urban settings, as suggested by HUL (UNESCO, 2011; Bai et al., 2021c). Among the five main MML challenges summarized by Baltrusaitis et al. (2019), representation (to present and summarize multi-modal data in a joint or coordinated space) and fusion (to join information for prediction) can be the most relevant for heritage and urban studies, as to 1) retrieving visual and/or textual information related to certain heritage values and attributes, and 2) aggregating individual posts in different geographic and administrative levels as the collective summarized knowledge of a place.

Furthermore, according to the First Law of Geography (Tobler, 1970),

> "everything is related to everything else, but near things are more related than distant things".

This argument can also be assumed to be valid in other distance measures other than geographical ones where a random walk could be performed (Pearson, 1905), such as in a topological space abstracted from spatial structure (Batty, 2013; Nourian, 2016; Ren et al., 2019; Zhang and Cheng, 2020) or a social network constructed based on common interests (Wasserman and Faust, 1994; Lazer et al., 2009; Barabási, 2013; Pentland, 2015). In this light, it would be beneficial to construct graphs of UGC posts where Social Network Analysis (SNA) could be performed, showing the socio-economic and spatio-temporal context among them, reflecting the inter-related dependent nature of the posts (Cheng and Wicks, 2014). Such a problem definition could help with both the classification and the aggregation tasks mentioned above, as has been demonstrated as effective and powerful by applications in the emerging field of Machine and Deep Learning on Graphs (Zhang et al., 2020; Ma and Tang, 2021).

This paper describes the methodological framework of creating multi-modal graph-based datasets about heritage values and attributes using unstructured social media data. The core question from generating such datasets could be formulated as: while heritage values and attributes have been historically inspected from site visiting and document reviewing by experts, can computational methods and/or artificial intelligence aid the process of knowledge documentation and comparative studies by mapping and mining multi-modal social media data? Even if the acceleration of the processes is not a priority, the provision of such a framework is aimed to encourage consistency and inclusion of communities in the discourse of cherishing, protecting, and preserving cultural heritage. In other words, the machine can eventually represent the voice of the community (Bai et al., 2021c). The main contributions of this manuscript could be summarized as:

1. Domain-specific multi-modal attributed graph datasets about heritage values and attributes (or more precisely, the values and attributes conveyed by public to urban cultural heritage) are collected and structured with the User-Generated Content from the social media platform Flickr in three cities (Amsterdam, Suzhou, and Venice) with urban areas inscribed in the UNESCO WHL, which could benefit the knowledge documentation and mapping for heritage and urban studies, aiming at a more inclusive heritage management process;
2. Several pre-trained machine learning and deep learning models have been extensively applied and tested for generating multi-modal features and [pseudo-]labels with full mathematical formulations as its problem definition, providing a reproducible methodological framework that could also be tested in other cases worldwide;
3. Multi-graphs have been constructed to reflect the temporal, spatial, and social relationships among the data samples of collected User-Generated Content, ready to be further tested on several provisional tasks with both scientific relevances for Graph-based Multi-modal Machine Learning and Social Network research, and societal interests for Urban Studies, Urban Data Science, and Heritage Studies.

4.2 General Framework

Before zooming into the domain-specific case studies with technological details, this section first describes the general process of creating multi-modal datasets as attributed graphs from unstructured volunteered information contents harvested from social media. These graphs would encode connections between posts of content publishers on social media; connections that can be established by virtue of similarities or proximities in spatial, temporal, or social domains. The whole process consists of five core components - data acquisition and cleaning (Section 4.3.2), multi-modal feature representation (Section 4.4), [pseudo-] label generation (Section 4.5), contextual graph construction (Section 4.6), and qualitative inspection and validation (Section 4.7), as visualized in Figure 4.1.

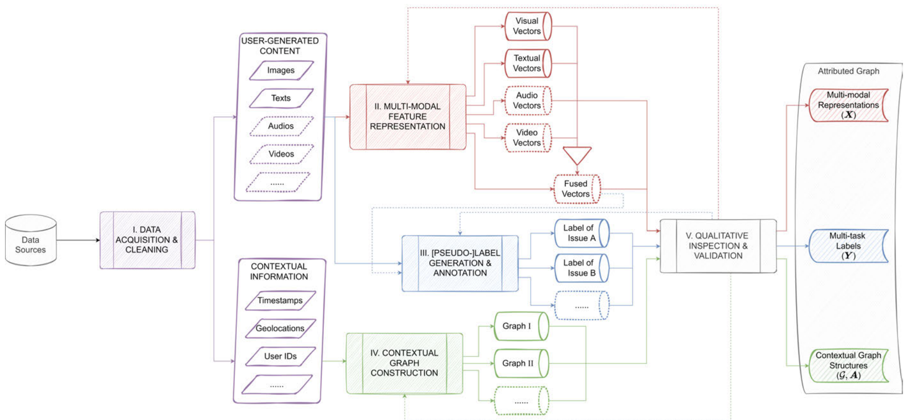


FIG. 4.1 The framework to create multi-modal machine learning datasets as attributed graphs from unstructured data sources.

As argued in Aggarwal (2011) and Bai et al. (2021c), the analyses on social media (or social network data) could be categorized as studying its content (traditionally texts and images, possibly also audio and video), structure (social linkages among users entailing interactions), and context (spatiotemporal and socio-economic manifolds). While the former is mainly about constituent data points themselves, the latter two (both are contextual information under different scenarios) provide explicit data about the potential linkage between the data points. For any data source (social media platform) of interest, the proposed framework suggests acquiring both content and contextual information for a rigid understanding of the social network. After data acquisition and cleaning, the input data would be highly unstructured and non-standard and thus challenging to feed into data science workflows, which need to be transformed as machine-readable formats - presumably vectors - using certain feature representation techniques. For different modalities, various techniques could be employed: from hand-engineered features, to pre-trained embeddings, and to other end-to-end techniques such as auto-encoders. Moreover, the fusion of different modalities could happen in various forms, from the most simple concatenation, to more complex techniques using neural networks (Baltrusaitis et al., 2019). Even though unsupervised learning applications of spatial clustering and auto-correlation are not uncommon, it is still preferable to have semantic labels concerning various issues of interest to make more sense out of the data points. In situations where human annotation can be expensive and challenging, semi-automatic labeling with transfer learning, pseudo-label generation and/or active learning using either the raw data or the generated multi-modal features could be applied to efficiently circumvent this bottle-neck process (Prince, 2004; Pan and Yang, 2010; Nowak and Rüger, 2010; Zhou and Li, 2010; Settles, 2011; Lee et al., 2013). Furthermore, the graph construction process makes use of the proximity or similarity of the contextual information to link the data points as [multi-] graphs. Contextualization of the data points and creating a coherent picture of the datasets are necessary tasks, without which the task of data analysis would remain at the level of dealing with a bag containing powder-like data points. Graph datasets can be of essential value in interpolation and extrapolation tasks, simply put for diffusing or transferring information from the neighbours of a data point to it. In cases where some graph attribute is missing on a data point, a graph representation can help in creating consistency and coherence. This is especially important for semi-supervised learning scenarios on social media data, where missing features could be very common (Kipf and Welling, 2016). Before storing the results as valid attributed graph datasets with graph structures ($\mathcal{G}, \boldsymbol{A}$) and node features ($\boldsymbol{X}, \boldsymbol{Y}$), a bundle of processes for qualitatively and quantitatively inspecting the quality, consistency, and validity of the generated results is necessary. This could also possibly include humans in the loop.

The rest of this manuscript will explain each component in detail with specific instances tailored for the use case of mapping heritage values and attributes as demonstration (such as the selection of the three cities in Section 4.3.1, the choice of Flickr as data source in Section 4.3.2, and the decisions of pre-trained ML models in Sections 4.4 and 4.5). However, in principle, the case study to be instantiated and technology to be employed could be specified, enhanced, and updated based on the actual use cases within a wide range of scenarios, taking advantage of the most suitable tools and the most current technological developments. This will be further discussed in Section 4.8.2.

4.3 Data and Materials

4.3.1 Case Studies: Venice, Amsterdam, and Suzhou

Without loss of generality, this research selected three cities in Europe and China that are related to UNESCO WH and HUL as case studies: Amsterdam (AMS), the Netherlands; Suzhou (SUZ), China; and Venice (VEN), Italy. All three cities are with urban areas entirely or partially inscribed in the UNESCO WHL, such as Venice and its Lagoon[1] and Seventeenth-Century Canal Ring Area of Amsterdam inside the Singelgracht[2], or contain WHL in multiple spots of the city, such as Classical Gardens of Suzhou[3], showcasing different spatial typologies of cultural heritage in relation to its urban context (Pereira Roders, 2010; Valese et al., 2020).

TABLE 4.1 The case studies and their World Heritage status.

City	Geo-location	WHL Name	OUV Criteria	Property Area	Inscription Date
Amsterdam (AMS)	52.365000N 4.887778E	Seventeenth-Century Canal Ring Area of Amsterdam inside the Singelgracht	(i)(ii)(iv)	198.2 ha	2010
Suzhou (SUZ)	31.302300N 120.631300E	Classical Gardens of Suzhou	(i)(ii)(iii)(iv)(v)	11.9 ha	2000
Venice (VEN)	45.438759N 12.327145E	Venice and its Lagoon	(i)(ii)(iii)(iv)(v)(vi)	70,176.2 ha	1987

As shown in Table 4.1, the three cases have very different scales, yet all strongly demonstrate the relationship between urban fabric and the water system. Interestingly, Amsterdam and Suzhou have been, respectively, referred to as “the Venice of the North” and “the Venice of the East” by the media and public. Moreover, the concept of OUV introduced in Section 4.1 reveals the core cultural significance of WH properties. The OUV of a property would be justified with ten selection criteria, where criteria (i)–(vi) reflect various cultural values, and criteria (vii)–(x) natural ones (Jokilehto, 2007, 2008; UNESCO, 2008; Bai et al., 2021b), as explained in Appendix Table A.1. The three selected cases include a broad variety of all cultural heritage OUV selection criteria, implying the representativeness of the datasets constructed in this study. Full documents of SOUV for the three cases can be found in Appendix A.

[1] https://whc.unesco.org/en/list/394
[2] https://whc.unesco.org/en/list/1349
[3] http://whc.unesco.org/en/list/813

4.3.2 Data Collection and Pre-processing

Numerous studies have collected, annotated, and distributed open-source datasets from the image-sharing social media platform Flickr owing to its high-quality image data, searchable metadata, and convenient Application Programming Interface (API), although its possible drawbacks include relatively low popularity, limited social and geographical coverage of users, and unbalanced information quantities of images and texts (van Dijck, 2011; Lin et al., 2014; Tenkanen et al., 2017; Li et al., 2018). A collection of Flickr-based datasets could include MirFlickr-1M (Huiskes and Lew, 2008), NUS-WIDE (Chua et al., 2009), Flickr (Tang and Liu, 2009), ImageNet (Deng et al., 2009; Krizhevsky et al., 2012), Microsoft Common Object in COntext (MS COCO) (Lin et al., 2014), Flickr30k (Plummer et al., 2015), SinoGrids (Zhou and Long, 2016), and GRAPH Saint (Zeng et al., 2019), etc. These datasets containing one or more of the visual, semantic, social, and/or geographical information of UGC are widely used, tested, but also sometimes challenged by different ML communities including Computer Vision, Multi-modal Machine Learning, and Machine Learning on Graphs. However, they are more or less suitable for bench-marking general ML tasks and testing computational algorithms, which are not necessarily tailor-made for heritage and urban studies. On the other hand, the motivation of data collection in this research is to provide datasets that could be both directly applicable for ML communities as a test-bed, and theoretically informative for heritage and urban scholars to draw conclusions on for informing the decision-making process. Therefore, instead of adapting the existing datasets that can be weakly related to the problems of interest in this study, new data are directly collected and processed from Flickr as an instance of the proposed framework in Section 4.2. Further possibilities of merging other existing datasets and data from other sources in response to the limitations of Flickr will be briefly addressed in Section 4.8.2.

FlickrAPI python library[4] was used to access the `photo.search` API method provided by Flickr[5], using the Geo-locations in Table 4.1 as the centroids to search a maximum of 5000 IDs of geo-tagged images within a fixed radius covering the major urban area (5km for Venice and Suzhou, and 2km for Amsterdam), to form comparable and compatible datasets from the three cities, since only 4229 IDs were found in Suzhou during the time of data collection, reflecting the relatively scarse use of Flickr in China. Only images with a `candownload` flag indicated by the owner were further queried, respecting the privacy and copyrights of Flickr users. Those images are further queried through `photo.getInfo` and `photo.getSizes` API methods to retrieve the following information: owner's ID; owner's registered location on Flickr; the title, description, and tags provided by user; geo-tag of the image; timestamp marking when the image was taken, and URLs to download the `Large Square` (150×150 px) and `Small 320` (320×240 px) versions of the original image. The images that have the user tag of "erotic" were excluded from the query. Then all the images of both sizes were saved and transformed into RGB format as raw visual data.

The retrieved raw textual fields of `description`, `title`, and `tags` could all provide

[4] https://stuvel.eu/software/flickrapi/
[5] https://www.flickr.com/services/api/

useful information, yet not all posts have these fields, and not all posts are necessarily written to express thoughts and share knowledge about the place (considered as valid in the context of this study). A stop-word list has been used to remove the HTML (HyperText Markup Language) symbols and other formatting elements from the texts and to filter out textual data that were mainly 1) a description of the camera used, 2) a default image name generated by the camera, 3) an advertisement or a promotion. The textual fields of the posts were cleaned, translated, and merged into a `Revised Text` field as raw English textual data, after recording the detected original language of posts on sentence level using Google Translator API from the Deep Translator python library[6]. Moreover, many posts shared by the same user were uploaded at once, thus having the same duplicated textual fields for all of them. To handle such redundancy, a separate dataset of all the unique processed textual data on sentence level was saved for each city, while the original post ID of each sentence was marked and could easily be traced back.

Furthermore, the public friend and subscription lists of all the retrieved owners were queried through the `people.getPublicGroups` and the `contacts.getPublicList` API methods, while all personal information was only considered as a [semi-] anonymous ID with respect to the privacy policy.

To test the scalability of the methodological workflow, another larger dataset without the limit of maximum 5000 IDs has also been collected in Venice (VEN-XL). The API of Flickr has a limitation at the scale of queries, which would return occasional errors while the server gets in burden. This requires a different strategy during data collection of the larger dataset. In this study, a 20×20 grid was tiled in the area of Venice (from 45.420855N 12.291054E to 45.448286N 12.369234E) to collect the post IDs from the centroid of each tile with a radius of 0.3km, which were later aggregated by removing the duplicated IDs collected by multiple tiles to form the entire large dataset, similar to Bekker (2020). The further steps of data cleaning and pre-processing remained the same with the smaller datasets.

The data collection procedure took place from 28 December 2020–10 January 2021 and 10 February 2022–25 February 2022, respectively. The earliest captured photos collected date back to 1946 in Data of Amsterdam, the Netherlands (AMS), 2007 in Data of Suzhou, China (SUZ), 1954 in Data of Venice, Italy (VEN), and 1875 in The extra-large version of Venice data (VEN-XL), and for all cities, the most recent photos were taken in 2021–2022.

Table 4.2 shows the number of data samples (posts) and owners (users) for the three case study cities at each stage. Note the numbers of posting owners are relatively unbalanced in different cities. Intuitionally, a larger number of owners could suggest a better coverage of social groups and provide better representativeness for the datasets. However, since the unit of data points in this study is a single post, not a unique social media user (content publisher), it could be assumed that the latter only provides sufficient [social] contextual information for the former.

[6] https://deep-translator.readthedocs.io/en/latest/

TABLE 4.2 The number of data samples collected at each stage, the bold numbers mark the sample size of the final datasets.

City	AMS	SUZ	VEN	VEN-XL
IDs Collected	5000	4229	5000	116,675
Is Downloadable	3727	3137	2952	80,964
Downloaded Posts	**3727**	**3137**	**2951**	**80,963**
Has Textual Data *	3404	2692	2801	77,644
Has Unique Texts **	3130	1963	1952	59,396
Unique Sentences	2247	361	3249	61,253
Original Posts **	2904	754	1761	49,823
Posting Owners	195	95	330	6077

* At least one of `Description`, `Title` and `Tag` fields is not empty.
** The two rows of numbers are different because of posts without any valid sentences.

4.3.3 Formal Description of the Dataset

To formally describe the data, define the problem, and propose a generalizable workflow as a methodological framework, mathematical notations are used in the rest of this manuscript. Since the same process is valid for all three cities (and probably also for other unselected cases worldwide) and has been repeated exactly three times, no distinctions would be made among the cities, except for the cardinality of sets reflecting sample sizes.

Let i be the index of a generic sample of the dataset for one city, then its raw data could be denoted as a tuple $\boldsymbol{\mathfrak{d}}_i = (\boldsymbol{\mathfrak{I}}_i, \mathcal{S}_i, \mathfrak{u}_i, \mathfrak{t}_i, \mathfrak{l}_i)$, $\boldsymbol{\mathfrak{d}}_i \in \mathfrak{D} = \{\boldsymbol{\mathfrak{d}}_1, \boldsymbol{\mathfrak{d}}_2, ..., \boldsymbol{\mathfrak{d}}_K\}$, where K is the sample size of the dataset in a city (as shown in Table 4.2), $\boldsymbol{\mathfrak{I}}_i$ is a three-dimensional tensor of the image size with three RGB channels, $\mathcal{S}_i = \{s_i^{(1)}, s_i^{(2)}, ..., s_i^{(|\mathcal{S}_i|)}\}$ or $\mathcal{S}_i = \varnothing$ is a set of revised English sentences that can also be an empty set for samples without any valid textual data, $\mathfrak{u}_i \in \mathcal{U}$ is a user ID that is one instance from the user set $\mathcal{U} = \{\mu_1, \mu_2, ..., \mu_{|\mathcal{U}|}\}$, $\mathfrak{t}_i \in \mathcal{T}$ is a timestamp that is one instance from the ordered set of all the unique timestamps $\mathcal{T} = \{\tau_1, \tau_2, ..., \tau_{|\mathcal{T}|}\}$ from the dataset at the level of weeks, and $\mathfrak{l}_i = (\mathfrak{x}_i, \mathfrak{y}_i)$ is a geographical coordinate of latitude $(\mathfrak{y}_i)$ and longitude $(\mathfrak{x}_i)$ marking the geo-location of the post. A complete nomenclature of all notations used in this chapter can be found in the Appendix Tables B.3 and B.4.

Figure 4.2 demonstrates the data flow of one sample post in Venice, which will be formally explained in the following sections.

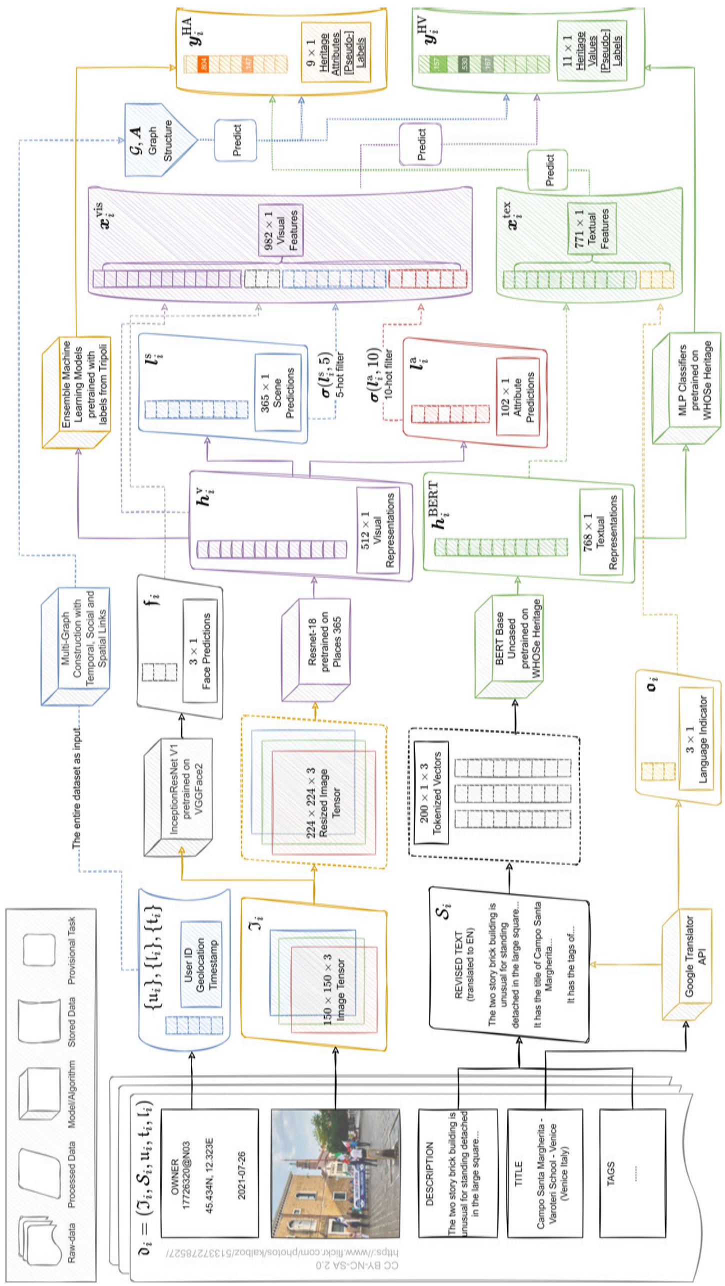


FIG. 4.2 Data flow of the multi-modal feature generation process of one sample post in Venice, while graph construction requires all data points of the dataset. The original post owned by user 17726320@N03 is under CC BY-NC-SA 2.0 license.

4.4 Multi-Modal Feature Generation

4.4.1 Visual Features

Places365 is a dataset containing 1.8 million images from 365 scene categories, which includes a relatively comprehensive collection of indoor and outdoor places (Zhou et al., 2014, 2017). The categories can be informative for urban and heritage studies to identify depicted scenes of images and to further infer heritage attributes (Veldpaus, 2015; Ginzarly et al., 2019). A few Convolutional Neural Network (CNN) models were pre-trained by Zhou et al. (2017) using state-of-the-art backbones to predict the depicted scenes in images, reaching a top-1 accuracy of around 55% and top-5 accuracy of around 85%. Furthermore, the same set of pre-trained models have been used to predict 102 discriminative scene attributes based on SUN Attribute dataset (Patterson and Hays, 2012; Patterson et al., 2014), reaching top-1 accuracy of around 92% (Zhou et al., 2017). These scene attributes are conceptually different from heritage attributes, as the former are mostly adjectives and present participles describing the scene and activities taking place. Therefore, both heritage values and attributes could be effectively inferred therefrom.

This study used the open ResNet-18 model (He et al., 2016) pre-trained on Places365 with PyTorch[7]. This model was adjusted to effectively yield three output vectors: 1) the last softmax layer of the model $\boldsymbol{l}^{\mathsf{s}}_{365\times 1}$ as logits over all scene categories; 2) the last hidden layer $\boldsymbol{h}^{\mathsf{v}}_{512\times 1}$ of the model; 3) a vector $\boldsymbol{l}^{\mathsf{a}}_{102\times 1}$ as logits over all scene attributes. Such a process for any image input $\mathfrak{I}_i$ could be described as:

$$\boldsymbol{l}^{\mathsf{s}}_i, \boldsymbol{l}^{\mathsf{a}}_i, \boldsymbol{h}^{\mathsf{v}}_i = \boldsymbol{f}_{\text{ResNet-18}}(\mathfrak{I}_i|\boldsymbol{\Theta}_{\text{ResNet-18}}), \tag{4.1}$$

or preferably in a vectorized format:

$$\boldsymbol{L}^{\mathsf{s}}, \boldsymbol{L}^{\mathsf{a}}, \boldsymbol{H}^{\mathsf{v}} = \boldsymbol{f}_{\text{ResNet-18}}([\mathfrak{I}_1, \mathfrak{I}_2, ..., \mathfrak{I}_K]|\boldsymbol{\Theta}_{\text{ResNet-18}}), \tag{4.2}$$

where

$$\boldsymbol{L}^{\mathsf{s}} := [\boldsymbol{l}^{\mathsf{s}}_i]_{365\times K}, \boldsymbol{L}^{\mathsf{a}} := [\boldsymbol{l}^{\mathsf{a}}_i]_{102\times K}, \boldsymbol{H}^{\mathsf{v}} := [\boldsymbol{h}^{\mathsf{v}}_i]_{512\times K}. \tag{4.3}$$

Considering that the models demonstrate reasonable performance in top-n accuracy, to keep the visual features explainable, a n-hot soft activation filter $\boldsymbol{\sigma}^{(n)}$ is performed on both logit outputs, to keep the top-n prediction entries active, while smoothing all the others based on the confidence of top-n predictions ($n = 5$ for scene categories $\boldsymbol{L}^{\mathsf{s}}$ and $n = 10$ for scene attributes $\boldsymbol{L}^{\mathsf{a}}$). Let $\max(\boldsymbol{l}, n)$ denote the n_{th} maximum element of a d-dimensional logit vector $\boldsymbol{l}$ (the sum of all d entries of $\boldsymbol{l}$ equals 1), then the activation filter $\boldsymbol{\sigma}^{(n)}$ could be described as:

$$\boldsymbol{\sigma}^{(n)}(\boldsymbol{l}_{d\times 1}) = \boldsymbol{l} \odot \boldsymbol{m} + \frac{1 - \boldsymbol{l}^{\mathsf{T}}\boldsymbol{m}}{d - n}(\boldsymbol{1}_{d\times 1} - \boldsymbol{m}), \tag{4.4}$$

$$\boldsymbol{m} := [m_\iota]_{d\times 1}, m_\iota = \begin{cases} 1 & \text{if } l_\iota \geq \max(\boldsymbol{l}, n) \\ 0 & \text{otherwise} \end{cases}, \tag{4.5}$$

[7] https://github.com/CSAILVision/places365

where $\boldsymbol{m}$ is a mask vector indicating the positions of top-n entries, and $\boldsymbol{l}^{\mathsf{T}}\boldsymbol{m}$ is effectively the total confidence of the model for top-n predictions. Note that this function could also take a matrix as input and process it as several column vectors to be concatenated back.

Furthermore, as the Places365 dataset is tailor-made for scene detection tasks rather than facial recognition (Zhou et al., 2017), the models pre-trained on it may become confused when a new image is mainly composed of faces as "typical tourism pictures" and self-taken photos, which is not uncommon in the case studies as popular tourism destinations. As the ultimate aim of constructing such datasets is not to precisely predict the scene each image depicts, but to help infer heritage values and attributes, it would be unfair to simply exclude those images containing a significant proportion of faces. Rather, the existence of humans in images showing their activities would be a strong cue of intangible dimension of heritage properties. Under such consideration, an Inception ResNet-V1 model[8] pre-trained on the VGGFace2 Dataset (Schroff et al., 2015; Cao et al., 2018) has been used to generate features about depicted faces in the images. A three-dimensional vector $\mathfrak{f}_i$ was obtained for any image input $\mathfrak{I}_i$, where the non-negative first entry $\mathfrak{f}_{1,i} \in \mathbb{N}$ counts the number of faces detected in the image, the second entry $\mathfrak{f}_{2,i} \in [0,1]$ records the confidence of the model for face detection, and the third entry $\mathfrak{f}_{3,i} \in [0,1]$ calculates the proportion of area of all the bounding boxes of detected faces to the total area of the image. Similarly, the vectorized format could be written as $\boldsymbol{F} := [\mathfrak{f}_i]_{3\times K}$ over the entire dataset.

Finally, all obtained visual features were concatenated vertically to generate the final visual feature $\boldsymbol{X}^{\mathsf{vis}}_{982\times K}$:

$$\boldsymbol{X}^{\mathsf{vis}}_{982\times K} = \left[\boldsymbol{H}^{\mathsf{v}\mathsf{T}}, \boldsymbol{F}^{\mathsf{T}}, \boldsymbol{\sigma}^{(5)}(\boldsymbol{L}^{\mathsf{s}})^{\mathsf{T}}, \boldsymbol{\sigma}^{(10)}(\boldsymbol{L}^{\mathsf{a}})^{\mathsf{T}}\right]^{\mathsf{T}}, \tag{4.6}$$

where $[\cdot,\cdot]$ denotes the horizontal concatenation of matrices.

This final matrix is to be used in future MML tasks as the vectorized descriptor of the uni-modal visual contents of the posts, with both more abstract hidden features only to be understood by machines, and more specific information about predicted categories interpretable by humans, which is common practice in MML literature (Baltrusaitis et al., 2019). All models are tested on both 150×150 and 320×240 px images to compare the consistency of generated features. The workflow of generating visual features is illustrated in the top part of Figure 4.2.

4.4.2 Textual Features

In the last decade, attention- and Transformer-based pre-trained models have taken over the field of Natural Language Processing (NLP), increasing the performance of

[8] https://github.com/timesler/facenet-pytorch

models in both general machine learning tasks, and domain-specific transfer learning scenarios (Vaswani et al., 2017). As an early version, the pre-trained Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) is still regarded as a powerful base model to be fine-tuned on specific downstream datasets and for various NLP tasks. Specifically, the output on the `[CLS]` token of BERT models is regarded as an effective representation of the entire input sentence, being used extensively for classification tasks (Clark et al., 2019; Sun et al., 2019). In the heritage studies domain, Bai et al. (2021a) fine-tuned BERT on the dataset WHOSe Heritage that they constructed from the UNESCO World Heritage inscription document, followed by a Multi-Layer Perceptron (MLP) classifier to predict the OUV selection criteria that a sentence is concerned with, showing top-1 accuracy of around 71% and top-3 accuracy of around 94%, which has been shown in Chapter 3.

This study used the openly-released BERT model fine-tuned on WHOSe Heritage with PyTorch[9]. The BERT model took both the entire sentence sets $\mathcal{S}_i$ and individual sentences of the sets $\{s_i^{(1)}, s_i^{(2)}, ..., s_i^{(|\mathcal{S}_i|)}\}$ as paragraph- and sentence-level inputs, respectively, for the comparison of consistency on predicted outputs of this new dataset. Furthermore, taking the entire sentence sets $\mathcal{S}_i$ as input, the 768-dimensional output vector $\boldsymbol{h}_{768\times 1}^{\text{BERT}}$ of the `[CLS]` token was retrieved on samples that have valid textual data:

$$\boldsymbol{h}_i^{\text{BERT}} = \boldsymbol{f}_{\text{BERT}}(\mathcal{S}_i|\boldsymbol{\Theta}_{\text{BERT}}), \text{ where } \boldsymbol{f}_{\text{BERT}}(\varnothing|\boldsymbol{\Theta}_{\text{BERT}}) = \mathbf{0}_{768\times 1} \tag{4.7}$$

or preferably in a vectorized format:

$$\boldsymbol{H}^{\text{B}} = \boldsymbol{f}_{\text{BERT}}([\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_K]|\boldsymbol{\Theta}_{\text{BERT}}), \text{where } \boldsymbol{H}^{\text{B}} := [\boldsymbol{h}_i^{\text{BERT}}]_{768\times K}. \tag{4.8}$$

Moreover, the original language of each sentence may provide additional information to the verbal context of posts, informative to effectively identify and compare locals and tourists. A three-dimensional vector $\mathfrak{o}_i \in \{0,1\}^3$ was obtained with Google Translator API. The three entries, respectively, marked whether there were sentences in English, local languages (Dutch, Chinese, or Italian, respectively), and other languages in the set $\mathcal{S}_i$. The elements of vector $\mathfrak{o}_i$ or the matrix form $\boldsymbol{O} := [\mathfrak{o}_i]_{3\times K}$ could be in a range from all zeros (when there were no textual data at all) to all ones (where the post comprised different languages in separate sentences).

Similar to visual features, final textual features $\boldsymbol{X}_{771\times K}^{\text{tex}}$ could be obtained as:

$$\boldsymbol{X}_{771\times K}^{\text{tex}} = \left[\boldsymbol{H}^{\text{B}\mathsf{T}}, \boldsymbol{O}^{\mathsf{T}}\right]^{\mathsf{T}}. \tag{4.9}$$

The workflow of generating textual features is illustrated in the bottom part of Figure 4.2.

[9]https://github.com/zzbn12345/WHOSe_Heritage

4.4.3 Contextual Features

As mentioned in Section 4.3.2, the user ID $\mathfrak{u}_i$ and timestamp $\mathfrak{t}_i$ of a post are both instances from their respective set $\mathcal{U}$ and $\mathcal{T}$, since multiple posts could be posted by the same user, and multiple images could be taken during the same week. To help formulate and generalize the problem under the practice of relational database (Reiter, 1989), the information from both can be transformed as one-hot embeddings $\boldsymbol{U} := [u_{j,i}]_{|\mathcal{U}|\times K} \in \{0,1\}^{|\mathcal{U}|\times K}$ and $\boldsymbol{T} := [t_{k,i}]_{|\mathcal{T}|\times K} \in \{0,1\}^{|\mathcal{T}|\times K}$, such that:

$$u_{j,i} = \begin{cases} 1 & \text{if } \mathfrak{u}_i = \mu_j \in \mathcal{U} \\ 0 & \text{otherwise} \end{cases}, \tag{4.10}$$

$$\text{and } t_{k,i} = \begin{cases} 1 & \text{if } \mathfrak{t}_i = \tau_k \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases}. \tag{4.11}$$

Furthermore, Section 4.3.2 also mentioned the collection of the public contacts and groups of all the users μ_j from the set $\mathcal{U}$. To keep the problem simple, only direct contact pairs were considered to model the back-end social structure of the users, effectively filtering out the other contacts a user μ_j had that were not in the set of interest $\mathcal{U}$, resulting in an adjacency matrix among the users $\boldsymbol{A}^{\mathcal{U}} := [a^{\mathcal{U}}_{j,j'}]_{|\mathcal{U}|\times|\mathcal{U}|} \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{U}|}, j, j' \in [1, |\mathcal{U}|]$ marking their direct friendship:

$$a^{\mathcal{U}}_{j,j'} = \begin{cases} 1 & \text{if } \mu_j \text{ and } \mu_{j'} \text{ are contacts or } j = j' \\ 0 & \text{otherwise} \end{cases}. \tag{4.12}$$

Let $\mathcal{I}(\mu_j)$ denote the set of public groups a user μ_j follows (can be an empty set if μ_j follows no group), and let $\text{IoU}(\mathcal{A},\mathcal{B})$ denote the Jaccard Index (size of Intersection over size of Union) of two generic sets $\mathcal{A},\mathcal{B}$:

$$\text{IoU}(\mathcal{A},\mathcal{B}) = \frac{|\mathcal{A}\cap\mathcal{B}|}{|\mathcal{A}\cup\mathcal{B}|+\varepsilon}, \tag{4.13}$$

where ε is a small number to avoid zero-division. Then another weighted adjacency matrix among the users could be constructed: $\boldsymbol{A}^{\mathcal{U}'} := [a^{\mathcal{U}'}_{j,j'}]_{|\mathcal{U}|\times|\mathcal{U}|} \in [0,1]^{|\mathcal{U}|\times|\mathcal{U}|}$, $j, j' \in [1, |\mathcal{U}|]$, marking the mutual interests among the users as group subscription on Flickr:

$$a^{\mathcal{U}'}_{j,j'} = \text{IoU}(\mathcal{I}(\mu_j), \mathcal{I}(\mu_{j'})). \tag{4.14}$$

To further simplify the problem, although the geo-location $\mathfrak{l}_i = (\mathfrak{x}_i, \mathfrak{y}_i)$ of each post was typically distributed in a continuous 2D geographical space, it would be beneficial to further aggregate and discretize the distribution in a topological abstraction of spatial network (Batty, 2013; Nourian, 2016; Nourian et al., 2016), which has also been proven to be effective in urban spatial analysis, including but not limited to Space Syntax (Hillier and Hanson, 1989; Penn, 2003; Ratti, 2004; Blanchard and Volchenkov, 2008). The OSMnx python library[10] was used to inquire

[10] https://osmnx.readthedocs.io/en/stable/

the simplified spatial network data on OpenStreetMap including all means of transportation (Boeing, 2017) in each city with the same centroid location and radius described in Section 4.3.2. This operation effectively saved a spatial network as an undirected weighted graph $G_0 = (V_0, E_0, \boldsymbol{w}_0)$, where $V_0 = \{v_1, v_2, ..., v_{|V_0|}\}$ is the set of all street intersection nodes, $E_0 \subseteq V_0 \times V_0$ is the set of all links possibly connecting two spatial nodes (by different sorts of transportation such as walking, biking, and driving), and $\boldsymbol{w}_0 \in \mathbb{R}_+^{|E_0|}$ is a vector with the same dimension as the cardinality of the edge set, marking the average travel time needed between node pairs (dissimilarity weights). The `distance.nearest_nodes` method of OSMnx library was used to retrieve the nearest spatial nodes to any post location $\mathfrak{l}_i = (\mathfrak{x}_i, \mathfrak{y}_i)$. By only retaining the spatial nodes that bear at least one data sample posted nearby, and restricting the link weights between nodes so that the travel time on any link is no more than 20 min, which ensures a comfortable temporal distance forming neighbourhoods and communities (Howley et al., 2009), a subgraph $G = (V, E, \boldsymbol{w})$ of G_0 could be constructed, so that $V \subseteq V_0, E \subseteq E_0$, and $\boldsymbol{w} \in [0, 20.0]^{|E|}$. As a result, another one-hot embedding matrix $\boldsymbol{S} := [s_{l,i}]_{|V| \times K} \in \{0, 1\}^{|V| \times K}$ could be obtained:

$$s_{l,i} = \begin{cases} 1 & \text{if the closest node to point } \mathfrak{l}_i \text{ is } v_l \in V \\ 0 & \text{otherwise} \end{cases}. \tag{4.15}$$

The contextual features constructed as matrices/graphs would be further used in Section 4.6 to link the posts together.

4.5 Pseudo-Label Generation

4.5.1 Heritage Values as OUV Selection Criteria

Various categories on Heritage Values (HV) have been provided by scholars (Pereira Roders, 2007; Jokilehto, 2007, 2008; Tarrafa Silva and Pereira Roders, 2010). To keep the initial step simple, this study arbitrarily applied the value definition in UNESCO WHL with regard to ten OUV selection criteria, as listed in Appendix Table A.1 with an additional class `Others` representing scenarios where no OUV selection criteria suit the scope of a sentence (resulting in an 11-class category). It must be noted that the OUV selection criteria and the corresponding Statements of OUV include elements that could be identified and categorized as either heritage values or heritage attributes. Therefore, they are not necessarily heritage value per se, a detailed discussion on which falls out of the scope of this paper. However, for pragmatic purposes of demonstrating a framework, this study omits this distinction and considers the OUV selection criteria as a proxy of HV during label generation. A group of ML models were trained and fine-tuned to make such predictions by Bai et al. (2021a) as introduced in Section 4.4.2. Except for BERT already used to generate textual features as mentioned above, a Universal Language Model Fine-tuning (UMLFiT) (Howard and Ruder, 2018) has also been trained and fine-tuned, reaching a similar performance in accuracy. Furthermore, it has been found that the average confidence by both BERT and ULMFiT models on the prediction task showed significant correlation with expert evaluation, even on social media data (Bai et al., 2021a). This suggests that it may be possible to use both trained models to generate labels about heritage values in a semi-supervised active learning setting (Prince, 2004; Zhu and Goldberg, 2009), since this task is overly knowledge-demanding for crowd-workers, yet too time-consuming for experts (Pustejovsky and Stubbs, 2012).

The pseudo-label generation step could be formulated as:

$$\boldsymbol{y}_i^{\text{BERT}} = \begin{cases} \boldsymbol{g}_{\text{BERT}}(\mathcal{S}_i|\boldsymbol{\Theta}_{\text{BERT}}) & \text{if } \mathcal{S}_i \neq \varnothing \\ \boldsymbol{0}_{11\times 1} & \text{otherwise} \end{cases}, \tag{4.16}$$

$$\boldsymbol{y}_i^{\text{ULMFiT}} = \begin{cases} \boldsymbol{g}_{\text{ULMFiT}}(\mathcal{S}_i|\boldsymbol{\Theta}_{\text{ULMFiT}}) & \text{if } \mathcal{S}_i \neq \varnothing \\ \boldsymbol{0}_{11\times 1} & \text{otherwise} \end{cases}, \tag{4.17}$$

$$\boldsymbol{Y}^{\text{HV}} := [\boldsymbol{y}_i^{\text{HV}}]_{11\times K}, \boldsymbol{y}_i^{\text{HV}} = \frac{\boldsymbol{y}_i^{\text{BERT}} + \boldsymbol{y}_i^{\text{ULMFiT}}}{2}. \tag{4.18}$$

where $\boldsymbol{g}_{(*)}$ is an end-to-end function including both pre-trained models and MLP classifiers; and $\boldsymbol{y}_i^{(*)}$ is an 11-dimensional logit vector as soft-label predictions. Let argmx$(\boldsymbol{l}, n)$ denote the function returning the index set of the largest n elements of a vector $\boldsymbol{l}$, together with the previously defined max$(\boldsymbol{l}, n)$, the confidence and [dis-]agreement of models for top-n predictions could be computed as:

$$\boldsymbol{K}^{\text{HV}} := [\boldsymbol{\kappa}_i^{\text{HV}}]_{2\times K}, \boldsymbol{\kappa}_i^{\text{HV}} := [\kappa_i^{\text{HV}(0)}, \kappa_i^{\text{HV}(1)}]^{\text{T}}, \tag{4.19}$$

$$\kappa_i^{\text{HV}(0)} = \sum_{n_0=1}^{n} \frac{\max(\boldsymbol{y}_i^{\text{BERT}}, n_0) + \max(\boldsymbol{y}_i^{\text{ULMFiT}}, n_0)}{2}, \tag{4.20}$$

$$\kappa_i^{\text{HV}(1)} = \text{IoU}(\text{argmx}(\boldsymbol{y}_i^{\text{BERT}}, n), \text{argmx}(\boldsymbol{y}_i^{\text{ULMFiT}}, n)). \tag{4.21}$$

This confidence indicator matrix $\boldsymbol{K}^{\text{HV}}$ could be presumably regarded as a filter for the labels on heritage values $\boldsymbol{Y}^{\text{HV}}$, to only keep the samples with high inter-annotator (model) agreement (Nowak and Rüger, 2010) as the "ground-truth" [pseudo-] labels, while treating the others as unlabeled (Lee et al., 2013; Sohn et al., 2020).

4.5.2 Heritage Attributes as Depicted Scenery

Heritage Attributes (HA) also have multiple categorization systems (Veldpaus and Roders, 2014; Veldpaus, 2015; Gustcoven, 2016; Ginzarly et al., 2019; UNESCO, 2020), and are arguably more vaguely defined than HV. For simplicity, this study arbitrarily combined the attribute definitions of Veldpaus (2015) and Ginzarly et al. (2019), and kept a 9-class category of tangible and/or intangible attributes visible from an image. More precisely, this category should be framed as "depicted scenery" of an image (Ginzarly et al., 2019) that heritage attributes could possibly be induced from. The depicted scenes themselves are not yet valid heritage attributes. This semantic/philosophical discussion, however, is out of the scope of this paper. The definitions of the nine categories are listed in Appendix Table A.3.

An image dataset collected in Tripoli, Lebanon and classified with expert-based annotations presented by Ginzarly et al. (2019) was used to train a few ML models to replicate the experts' behaviour on classifying depicted scenery with Scikit-learn python library (Pedregosa et al., 2011). For each image, a unique class label was provided, effectively forming a multi-class classification task. The same 512-dimensional visual representation $\boldsymbol{H}^V$ introduced in Section 4.4.1 was generated from the images as the inputs. Classifiers including Multi-layer Perceptron (MLP) (shallow neural network) (Hinton, 1990), K-Nearest Neighbour (KNN) (Altman, 1992), Gaussian Naive Bayes (GNB) (Rish et al., 2001), Support Vector Machine (SVM) (Platt et al., 1999), Random Forest (RF) (Breiman, 2001), and Bagging Classifier (Breiman, 1996a) with SVM core (BC-SVM) were first trained and tuned for optimal hyperparameters using 10-fold cross validation (CrVd) with grid search (Arlot and Celisse, 2010). Then, the individually-trained models were put into ensemble-learning settings as both a voting (Zhou, 2012) and a stacking classifier (Breiman, 1996b). All trained models were tested on validation and test datasets to evaluate their performance. Details of the machine learning models are given in Appendix B. Both ensemble models were further applied to images collected in this study. Similar to the HV labels described in Section 4.5.1, the label generation step of HA could be formulated as:

$$\boldsymbol{y}_i^{\text{VOTE}} = \boldsymbol{h}_{\text{VOTE}}(\boldsymbol{h}_i^V | \Theta_{\text{VOTE}}, \mathcal{M}, \boldsymbol{\Theta}_{\mathcal{M}}), \tag{4.22}$$

$$\boldsymbol{y}_i^{\text{STACK}} = \boldsymbol{h}_{\text{STACK}}(\boldsymbol{h}_i^V | \Theta_{\text{STACK}}, \mathcal{M}, \boldsymbol{\Theta}_{\mathcal{M}}), \tag{4.23}$$

$$\boldsymbol{Y}^{\text{HA}} := [\boldsymbol{y}_i^{\text{HA}}]_{9\times K}, \boldsymbol{y}_i^{\text{HA}} = \frac{\boldsymbol{y}_i^{\text{VOTE}} + \boldsymbol{y}_i^{\text{STACK}}}{2}. \tag{4.24}$$

where $\boldsymbol{h}_{(*)}$ is an ensemble model taking all parameters $\boldsymbol{\Theta}_{\mathcal{M}}$ from each ML model in set $\mathcal{M}$; and $\boldsymbol{y}_i^{(*)}$ is a 9-dimensional logit vector as soft-label predictions. Similarly, the confidence of models for top-n prediction is:

$$\boldsymbol{K}^{\text{HA}} = [\boldsymbol{\kappa}_i^{\text{HA}}]_{2\times K}, \boldsymbol{\kappa}_i^{\text{HA}} = [\kappa_i^{\text{HA}(0)}, \kappa_i^{\text{HA}(1)}]^T, \tag{4.25}$$

$$\kappa_i^{\text{HA}(0)} = \sum_{n_0=1}^{n} \frac{\max(\boldsymbol{y}_i^{\text{VOTE}}, n_0) + \max(\boldsymbol{y}_i^{\text{STACK}}, n_0)}{2}, \tag{4.26}$$

$$\kappa_i^{\text{HA}(1)} = \text{IoU}(\text{argmx}(\boldsymbol{y}_i^{\text{VOTE}}, n), \text{argmx}(\boldsymbol{y}_i^{\text{STACK}}, n)). \tag{4.27}$$

This matrix $\boldsymbol{K}^{\text{HA}}$ could also be regarded the filter for heritage attributes labels $\boldsymbol{Y}^{\text{HA}}$.

4.6 Multi-Graph Construction

Three types of similarities/ relations among posts were considered to compose the links connecting the post nodes: temporal similarity (posts with images taken during the same time period), social similarity (posts owned by the same people, by friends, and by people who share mutual interests), and spatial similarity (posts with images taken at the same or nearby locations). All three could be deduced from the contextual information in Section 4.4.3.

As a result, an undirected weighted multi-graph (also known as Multi-dimensional Graph) with the same node set and three different link sets could be constructed as $\mathcal{G} = (\mathcal{V}, \{\mathcal{E}^{\text{TEM}}, \mathcal{E}^{\text{SOC}}, \mathcal{E}^{\text{SPA}}\}, \{\boldsymbol{w}^{\text{TEM}}, \boldsymbol{w}^{\text{SOC}}, \boldsymbol{w}^{\text{SPA}}\})$, where $\mathcal{V} = \{v_1, v_2, ..., v_K\}$ is the node set of all the posts, $\mathcal{E}^{(*)} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of all links connecting two posts of one similarity type, and the weight vector $\boldsymbol{w}^{(*)} := [w_e^{(*)}]_{|\mathcal{E}^{(*)}|\times 1} \in \mathbb{R}_+^{|\mathcal{E}^{(*)}|}$ marks the strength of connections. The multi-graph $\mathcal{G}$ could also be easily split into three simple undirected weighted graphs $\mathcal{G}^{\text{TEM}} = (\mathcal{V}, \mathcal{E}^{\text{TEM}}, \boldsymbol{w}^{\text{TEM}})$, $\mathcal{G}^{\text{SOC}} = (\mathcal{V}, \mathcal{E}^{\text{SOC}}, \boldsymbol{w}^{\text{SOC}})$, and $\mathcal{G}^{\text{SPA}} = (\mathcal{V}, \mathcal{E}^{\text{SPA}}, \boldsymbol{w}^{\text{SPA}})$ concerning each type of similarities. Each $\mathcal{G}^{(*)}$ corresponds to a weighted adjacency matrix $\boldsymbol{A}^{(*)} := [a_{i,i'}^{(*)}]_{K\times K} \in \mathbb{R}_+^{K\times K}, i, i' \in [1, K]$, such that:

$$a_{i,i'}^{(*)} = \begin{cases} w_e^{(*)} & \text{if the } e_{\text{th}} \text{ element of } \mathcal{E} \text{ is } (v_i, v_{i'}), \\ 0 & \text{otherwise.} \end{cases} \tag{4.28}$$

The three weighted adjacency matrices could be, respectively, obtained as described in the following sections.

All graphs were constructed with NetworkX python library (Hagberg et al., 2008). The rationale under constructing various graphs was briefly described in Section 4.1: the posts close to each other (in temporal, social, or spatial contexts) could be arguably similar in their contents, and therefore, also similar in the heritage values and attributes they might convey. Instead of regarding these similarities as redundant and, e.g., removing duplicated posts by the same user to avoid biasing the analysis, such as in Ginzarly et al. (2019), this study intends to take advantage of as much available data as possible, since similar posts may enhance and strengthen the information, compensating the redundancies and/or nuances using back-end graph structures. At later stages of the analysis, the graph of posts could be even coarsened with clustering and graph partitioning methods (Karypis and Kumar, 1995; Lafon and Lee, 2006; Gao and Ji, 2019; Ma and Tang, 2021), to give an effective summary of possibly similar posts.

4.6.1 Temporal Links

Let $\mathfrak{T}_{|\mathcal{T}|\times|\mathcal{T}|}$ denote a symmetric tridiagonal matrix where diagonal entries are all 1 and off-diagonal non-zero entries are all $\alpha_{\mathcal{T}}$, where $\alpha_{\mathcal{T}} \in [0, 1)$ is a parametric scalar:

$$\mathfrak{T}_{|\mathcal{T}|\times|\mathcal{T}|} := \begin{pmatrix} 1 & \alpha_{\mathcal{T}} & 0 & \cdots & 0 & 0 \\ \alpha_{\mathcal{T}} & 1 & \alpha_{\mathcal{T}} & \cdots & 0 & 0 \\ 0 & \alpha_{\mathcal{T}} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \alpha_{\mathcal{T}} \\ 0 & 0 & 0 & \cdots & \alpha_{\mathcal{T}} & 1 \end{pmatrix}, \tag{4.29}$$

then the weighted adjacency matrix $\boldsymbol{A}^{\mathsf{TEM}}_{K\times K}$ for temporal links could be formulated as:

$$\boldsymbol{A}^{\mathsf{TEM}} = \boldsymbol{T}^{\mathsf{T}}\mathfrak{T}\boldsymbol{T}, \boldsymbol{A}^{\mathsf{TEM}} \in \{0, \alpha_{\mathcal{T}}, 1\}^{K\times K}, \tag{4.30}$$

where $\boldsymbol{T}_{|\mathcal{T}|\times K}$ is the one-hot embedding of timestamp for posts mentioned in Equation (4.11). For simplicity, $\alpha_{\mathcal{T}}$ is set to 0.5. With such a construction, all posts from which the images were originally taken in the same week would have a weight of $w_e^{\mathsf{TEM}} = 1$ connecting them in $\mathcal{G}^{\mathsf{TEM}}$, and posts with images taken in nearby weeks in a chronological order would have a weight of $w_{e'}^{\mathsf{TEM}} = 0.5$. Note, however, that the notion of “nearby” may not necessarily correspond to temporally adjacent weeks, as the interval of timestamps marking the date when a photo was taken could be months and even years in earlier time periods. In use cases sensitive to the time intervals, the value of $\alpha_{\mathcal{T}}$ could also be weighted: i.e., the longer the time interval actually is, the smaller $\alpha_{\mathcal{T}}$ becomes.

4.6.2 Social Links

Let $\mathfrak{U}_{|\mathcal{U}|\times|\mathcal{U}|}$ denote a symmetric matrix as a linear combination of three matrices marking the social relations among the users:

$$\mathfrak{U}_{|\mathcal{U}|\times|\mathcal{U}|} = \frac{\alpha_{\mathcal{U}}^{(1)}\boldsymbol{I} + \alpha_{\mathcal{U}}^{(2)}\boldsymbol{A}^{\mathcal{U}} + \alpha_{\mathcal{U}}^{(3)}(\boldsymbol{A}^{\mathcal{U}'} > \beta_{\mathcal{U}})}{\alpha_{\mathcal{U}}^{(1)} + \alpha_{\mathcal{U}}^{(2)} + \alpha_{\mathcal{U}}^{(3)}}, \tag{4.31}$$

where $\boldsymbol{I} \in \{0, 1\}^{|\mathcal{U}|\times|\mathcal{U}|}$ is a diagonal matrix of 1s for the self relation, $\boldsymbol{A}^{\mathcal{U}} \in \{0, 1\}^{|\mathcal{U}|\times|\mathcal{U}|}$ is the matrix mentioned in Equation (4.12) for the friendship relation, $(\boldsymbol{A}^{\mathcal{U}'} > \beta_{\mathcal{U}}) \in \{0, 1\}^{|\mathcal{U}|\times|\mathcal{U}|}$ is a mask on the matrix $\boldsymbol{A}^{\mathcal{U}'}$ introduced in Equation (4.14) for the common-interest relation above a certain threshold $\beta_{\mathcal{U}} \in (0, 1)$, and $\alpha_{\mathcal{U}}^{(1)}, \alpha_{\mathcal{U}}^{(2)}, \alpha_{\mathcal{U}}^{(3)} \in \mathbb{R}_+$ are parametric scalars to balance the weights of different social relations. The weighted adjacency matrix $\boldsymbol{A}^{\mathsf{SOC}}_{K\times K}$ for social links could be formulated as:

$$\boldsymbol{A}^{\mathsf{SOC}} = \boldsymbol{U}^{\mathsf{T}}\boldsymbol{\mathfrak{U}}\boldsymbol{U}, \boldsymbol{A}^{\mathsf{SOC}} \in [0,1]^{K\times K}, \tag{4.32}$$

where $\boldsymbol{U}_{|\mathcal{U}|\times K}$ is the one-hot embedding of owner/user for posts mentioned in Equation (4.10). For simplicity, the threshold $\beta_{\mathcal{U}}$ is set to 0.05 and the scalars $\alpha^{(1)}_{\mathcal{U}}, \alpha^{(2)}_{\mathcal{U}}, \alpha^{(3)}_{\mathcal{U}}$ are all set to 1. With such a construction, all posts uploaded by the same user would have a weight of $w^{\mathsf{SOC}}_{e} = 1$ connecting them in $\mathcal{G}^{\mathsf{SOC}}$, posts by friends with common interests (of more than 5% common groups subscriptions) would have a weight of $w^{\mathsf{SOC}}_{e'} = \frac{2}{3}$, and posts by either friends with little common interests or strangers with common interests would have a weight of $w^{\mathsf{SOC}}_{e''} = \frac{1}{3}$.

4.6.3 Spatial Links

Let $\boldsymbol{\mathfrak{S}} := [\mathfrak{s}_{l,l'}] \in [0,1]^{|V|\times|V|}, l, l' \in [1, |V|]$ denote a symmetric matrix computed with simple rules showing the spatial closeness (conductance) of nodes from the spatial graph $G = (V, E, \boldsymbol{w})$ mentioned in Section 4.4.3, whose weights $\boldsymbol{w} := [w_e]_{|E|\times 1} \in [0, 20.0]^{|E|}$ originally showed the distance of nodes (resistance):

$$\mathfrak{s}_{l,l'} = \begin{cases} \frac{20-w_e}{20} & \text{if the } e_{\text{th}} \text{ element of } E \text{ is } (v_l, v_{l'}), \\ 0 & \text{otherwise.} \end{cases} \tag{4.33}$$

The weighted adjacency matrix $\boldsymbol{A}^{\mathsf{SPA}}_{K\times K}$ for spatial links could be formulated as:

$$\boldsymbol{A}^{\mathsf{SPA}} = \boldsymbol{S}^{\mathsf{T}}\boldsymbol{\mathfrak{S}}\boldsymbol{S}, \boldsymbol{A}^{\mathsf{SPA}} \in [0,1]^{K\times K}, \tag{4.34}$$

where $\boldsymbol{S}_{|V|\times K}$ is the one-hot embedding of spatial location for posts mentioned in Equation (4.15). With such a construction, posts located at the same spatial node would have a weight of $w^{\mathsf{SPA}}_{e} = 1$ in $\mathcal{G}^{\mathsf{SPA}}$, and posts from nearby spatial nodes would have a weight linearly decayed based on distance within a maximum transport time of 20 min.

Additionally, the multi-graph $\mathcal{G}$ could be simplified as a simple composed graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ with a binary adjacency matrix $\boldsymbol{A} \in \{0,1\}^{K\times K}$, such that:

$$\boldsymbol{A} := (\boldsymbol{A}^{\mathsf{TEM}} > 0) \vee (\boldsymbol{A}^{\mathsf{SOC}} > 0) \vee (\boldsymbol{A}^{\mathsf{SPA}} > 0), \tag{4.35}$$

which connects two nodes of posts if they are connected and similar in at least one contextual relationship.

4.7 Analyses as Qualitative Inspection

4.7.1 Generated Visual and Textual Features

Table 4.3 shows the consistency of generated visual and textual features. The visual features compared the scene and attribute predictions on images of different sizes (150×150 and 320×240 px); and the textual features compared the OUV selection criteria with aggregated (averaged) sentence-level predictions on each sentence from set $\{s_i^{(1)}, s_i^{(2)}, ..., s_i^{(|\mathcal{S}_i|)}\}$ and paragraph-/post-level predictions on set $\mathcal{S}_i$.

TABLE 4.3 The consistency (the mean and standard deviation of top-n IoU Jaccard Index on predicted sets) of generated features. For visual features, predictions with different input image sizes (150×150 px and 320×240 px) are compared; for textual features, average sentence-level predictions and paragraph-/post-level predictions are compared. The best scores for each feature are in bold, and the selected ones for future tasks are underlined. "#" means "the number of" in the table.

Sets to Calculate IoU Jaccard Index	AMS	SUZ	VEN
# Compared Posts w. Visual Features	3727	3137	2951
Top-1 scene predictions	**0.656**	**0.676**	**0.704**
—argmx $(\boldsymbol{l}^{\mathsf{s}}, 1)$	(0.475)	(0.468)	(0.456)
Top-5 scene predictions	<u>0.615</u>	<u>0.636</u>	<u>0.635</u>
—argmx $(\boldsymbol{l}^{\mathsf{s}}, 5)$	(<u>**0.179**</u>)	(<u>**0.238**</u>)	(<u>**0.229**</u>)
Top-1 attribute predictions	**0.867**	**0.853**	**0.838**
—argmx $(\boldsymbol{l}^{\mathsf{a}}, 1)$	(0.339)	(0.354)	(0.368)
Top-10 attribute predictions	<u>0.820</u>	<u>0.802</u>	<u>0.819</u>
—argmx $(\boldsymbol{l}^{\mathsf{a}}, 10)$	(<u>**0.140**</u>)	(<u>**0.144**</u>)	(<u>**0.139**</u>)
# Compared Posts w. Textual Features	2904	754	1761
Top-1 OUV predictions	0.775	0.923	0.714
— argmx $(\boldsymbol{y}^{\mathsf{BERT}}, 1)$	(0.418)	(0.267)	(0.452)
Top-3 OUV predictions	<u>**0.840**</u>	<u>**0.938**</u>	<u>**0.791**</u>
—argmx $(\boldsymbol{y}^{\mathsf{BERT}}, 3)$	(<u>**0.246**</u>)	(<u>**0.182**</u>)	(<u>**0.266**</u>)

For both scene and attribute predictions, the means of top-1 Jaccard Index were always higher than that of top-n, however, the smaller variance proved the necessity of using top-n prediction as features. Note the attribute prediction was more stable than the scene prediction when the image shape changed, this is probably because the attributes usually describe low-level features which could appear in multiple parts in the image, while some critical information to judge the image scene may be lost during cropping and resizing in the original ResNet-18 model. Considering the relatively high consistency of model performance and the storage cost of images when the dataset would ultimately scale up (e.g., VEN-XL), the following analyses would only be performed on smaller square images of 150×150 px.

The high Jaccard Index of OUV predictions showed that averaging the textual features derived from sub-sentences of a paragraph would yield a similar performance of directly feeding the whole paragraph into models, especially when the top-3 predictions are of main interest. Note that the higher consistency in Suzhou was mainly a result of the higher proportion of posts only consisting of one sentence.

Table 4.4 gives descriptive statistics of results that were not compared against different scenarios as in Table 4.3. Only a small portion of posts had detected faces in them. While Amsterdam has the highest proportion of face pictures (17.9%), Venice has larger average area of faces on the picture (i.e., more self-taken photos and tourist pictures). These numbers are also assumed to help associate a post to human-activity-related heritage values and attributes. Considering the languages of the posts, Amsterdam showed a balance between Dutch-speaking locals and English-speaking tourists, Venice showed a balance between Italian-speaking people and non-Italian-speaking tourists, while Suzhou showed a lack of Chinese posts. This is consistent with the popularity of Flickr as social media in different countries, which also implies that data from other social media could compensate this unbalance if the provisional research questions would be sensitive to the nuance between local and tourist narratives.

TABLE 4.4 Descriptive statistics (mean and standard deviation or counts, respectively) of the facial recognition results $\boldsymbol{F}$ as visual features and original language $\boldsymbol{O}$ as textual features. "#" means "the number of" in the table.

Features	AMS	SUZ	VEN	VEN-XL
# Posts w. Faces	667	303	166	9287
# Faces detected	1.547	1.403	1.349	1.298
—$\mathfrak{f}_1$	(0.830)	(0.707)	(0.785)	(0.651)
Model Confidence	0.955	0.956	0.930	0.948
—$\mathfrak{f}_2$	(0.079)	(0.081)	(0.099)	(0.081)
Area proportion of faces	0.049	0.057	0.077	0.076
—$\mathfrak{f}_3$	(0.112)	(0.073)	(0.185)	(0.112)
# Posts w. Texts *	2904	754	1761	49,823
# Posts in English $\mathfrak{o}_1$	1488	368	640	20,271
# Posts in Native Lang $\mathfrak{o}_2$	1773	27	1215	28,633
# Posts in Other Lang $\mathfrak{o}_3$	536	413	657	21,916

* Note this is smaller than the sum of the three below, since each post can be written in multiple languages.

4.7.2 Pseudo-Labels for Heritage Values and Attributes

As argued in Section 4.5.1, the label generation process of this paper did not involve human annotators. Instead, it used thoroughly trained ML models as machine replicas of annotators and considered their confidences and agreements as a filter to maintain the "high-quality" labels as pseudo-labels. Similar operations can be found in semi-supervised learning (Zhou and Li, 2010; Lee et al., 2013; Sohn et al., 2020).

For heritage values, an average top-3 confidence of $\kappa^{\text{HV}(0)} > 0.75$ and top-3 agreement (Jaccard Index) of $\kappa^{\text{HV}(1)} > 0.50$ was used as the filter for $\boldsymbol{Y}^{\text{HV}}$. This resulted in around 40–50% of the samples with textual data in each city as "labelled", and the rest as "unlabelled". Figure 4.3 demonstrates the distribution of "labelled" data about heritage values in each city. For all cities, cultural values are far more frequent than natural values, consistent with their status of cultural WH. However, elements related to natural values could still be found and were mostly relevant. The actual OUV inscribed in WHL mentioned in Table 4.1 could all be observed as significantly present (e.g., criteria (i),(ii),(iv) for Amsterdam) except for criterion (v) in Venice and Suzhou, which might be caused by the relatively fewer examples and poorer class-level performance of criterion (v) in the original paper (Bai et al., 2021a). Remarkably, criterion (iii) in Amsterdam and criterion (vi) in Amsterdam and Suzhou were not officially inscribed, but appeared to be relevant inducing from social media, inviting further heritage-specific investigations. The distributions of Venice and Venice-large were more similar in sentence-level predictions (Kullback–Leibler Divergence $D_{\text{KL}} = 0.002$, Chi-square $\chi^2 = 39.515$) than post-level ($D_{\text{KL}} = 0.051$, $\chi^2 = 518.895$), which might be caused by the specific set of posts sub-sampled in the smaller dataset.

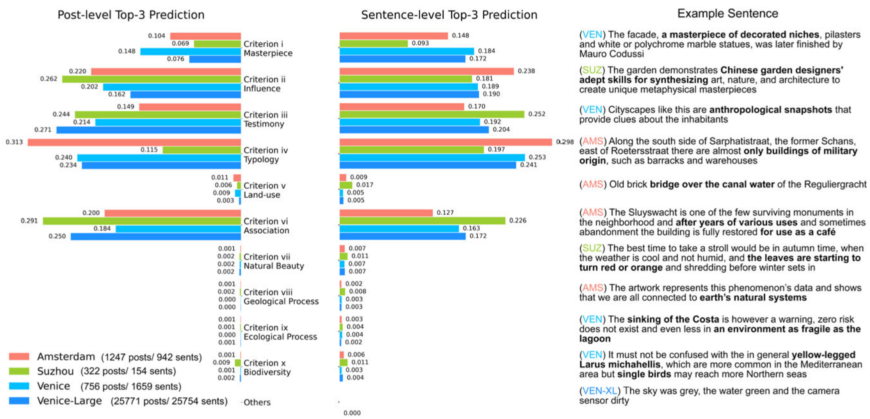


FIG. 4.3 The proportion of posts and sentences that are predicted and labeled as each heritage value (OUV selection criterion) as top-3 predictions by both BERT and ULMFiT. One typical sentence from each category is also given in the right part of the figure.

For heritage attributes, Table 4.5 shows the performance of ML models mentioned in Section 4.5.2. The two ensemble models with voting and stacking settings performed equally well and significantly better than other models (except for CrVd accuracy of SVM), proving the rationale of using both classifiers for heritage attribute label prediction. An average top-1 confidence of $\kappa^{\text{HA}(0)} > 0.7$ and top-1 agreement of $\kappa^{\text{HA}(1)} = 1$ was used as the filter for $\boldsymbol{Y}^{\text{HA}}$. This filter resulted in around 35–50% of the images in each city as "labelled", and the rest as "unlabelled". Figure 4.4 demonstrates the distribution of "labelled" data about heritage attributes in each city. It is remarkable that although the models were only trained on data from Tripoli, they performed reasonably well in unseen cases of Amsterdam, Suzhou, and Venice, capturing typical scenes of monumental buildings, architectural elements, and gastronomy, etc., respectively. Although half of the collected images were treated as "unlabelled" due to low confidence, the negative examples are not necessarily incorrect (e.g., with Monuments and Buildings). For all cities, Urban Form Elements and People's Activity and Association are the most dominant classes, consistent with the fact that most Flickr images are taken on the streets. Seen from the bar plots in Figure 4.4, the classes were relatively unbalanced, suggesting that more images from small classes might be needed or at least augmented in future applications. Furthermore, the distributions of Venice and Venice-large are similar to each other ($D_{\mathrm{KL}} = 0.076, \chi^2 = 188.241$), suggesting a good representativeness of the sampled small dataset.

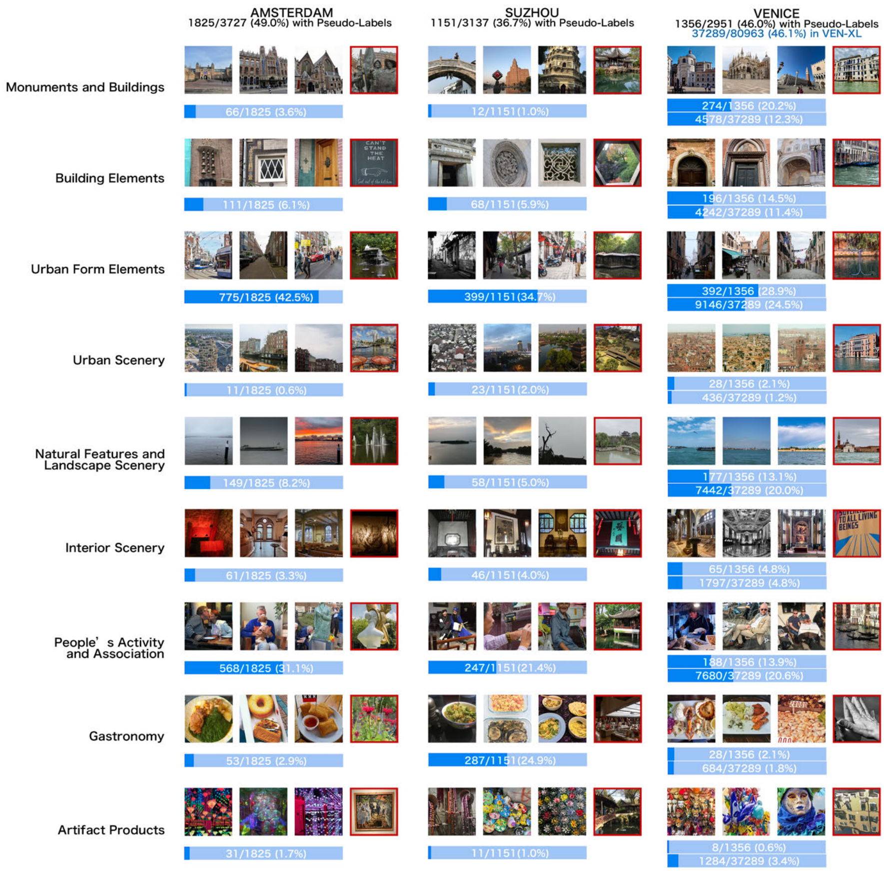


FIG. 4.4 Typical image examples in each city labelled as each heritage attribute category (depicted scene) and bar plots of their proportions in the datasets (length of bright blue background bars represent 50%). Three examples with high confidence and one negative example with low confidence (in red frame) are given. All images are 150 × 150 px "thumbnails" flagged as "downloadable".

TABLE 4.5 The performance of models during the cross validation (CrVd) parameter selection, on the validation set, and on the test set of data from Tripoli. The best two models for each performance are in bold typeface, and the best underlined.

ML Model	CrVd Acc	Val Acc	Val F1	Test Acc	Test F1
MLP	0.767	0.749	0.70	0.789	0.72
KNN	0.756	0.724	0.67	0.767	0.71
GNB	0.738	0.749	0.71	0.800	0.77
SVM	<u>0.797</u>	0.754	0.71	0.822	0.78
RF	0.766	0.734	0.68	0.789	0.72
BC-SVM	0.780	0.759	0.71	0.811	0.74
VOTE	0.788	**0.764**	**<u>0.72</u>**	**<u>0.855</u>**	**<u>0.82</u>**
STACK	**0.794**	**<u>0.768</u>**	**<u>0.72</u>**	**0.844**	**0.81**

4.7.3 Back-End Geographical Network

The back-end spatial structures of post locations as graphs $G = (V, E, \boldsymbol{w})$ were visualized in Figure 4.5. Further graph statistics in all cities were given in Table 4.6. The urban fabric is more visible in Venice than the other two cities, as there is always a dominant large component connecting most nodes in the graph, leaving fewer unconnected isolated nodes alone. While in Amsterdam, more smaller connected components exist together with a large one; and in Suzhou, the graph is even more fragmented with smaller components. This is possibly related to the distribution of tourism destinations, collectively forming bottom-up tourism districts or “tourist city” as proposed in Encalada-Abarca et al. (2022), which is also consistent with the zoning typology of WH property concerning urban morphology (Pereira Roders, 2010; Valese et al., 2020): for Venice, the Venetian islands are included together with a larger surrounding lagoon in the WH property (formerly referred to as core zone), and are generally regarded as a tourism destination as a whole; for Amsterdam, the WH property is only a part of the old city being mapped where tourists can freely wander and take photos in areas not listed yet as interesting tourism destinations; while for Suzhou, the WH properties are themselves fragmented gardens distributed in the old city, also representing the main destinations visited by (foreign) tourists.

TABLE 4.6 The statistics for the back-end Geographical Network $G = (V, E, \boldsymbol{w})$. “#” means “the number of” in the table.

Graph Features	AMS	SUZ	VEN	VEN-XL
# Nodes in V	788	230	915	3549
# Edges in E	3331	680	10,385	120,033
# Connected Components	72	38	6	13
# Nodes Largest CC *	355	50	897	3498
Graph Density	0.011	0.026	0.025	0.019
# Isolated Nodes in $V_0 \backslash V$	157	88	20	22

* Connected Components.

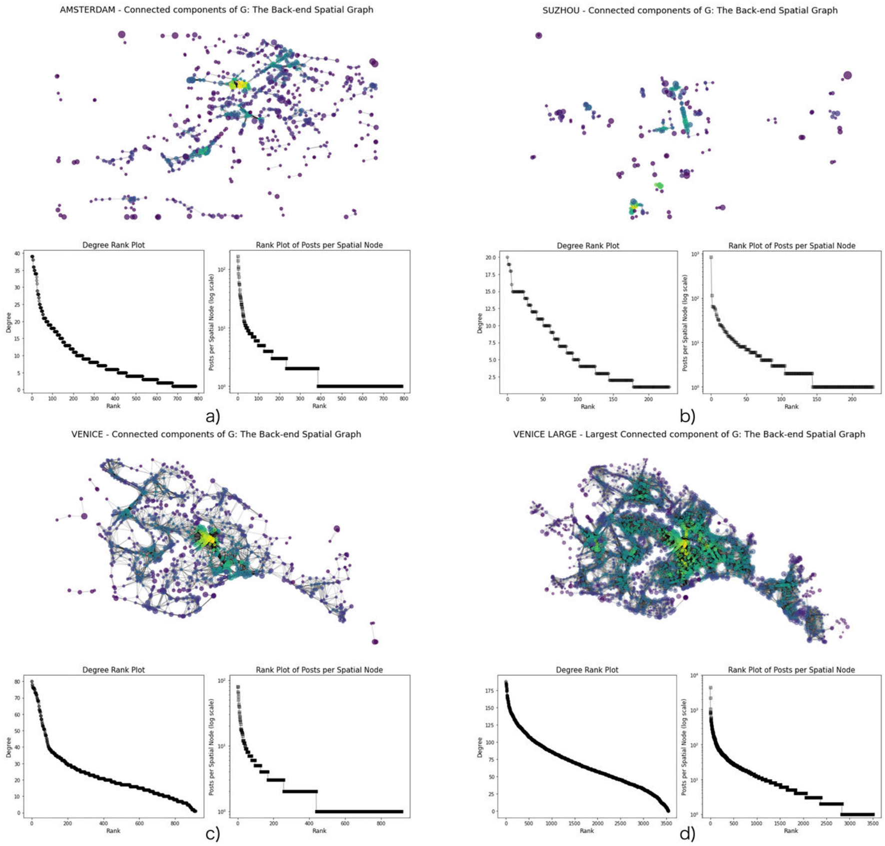


FIG. 4.5 The back-end geographical networks for three case studies, respectively, showing the graph structure, degree ranking distribution, and the ranking distribution of posts per geo-spatial node (on a logarithm scale) in Amsterdam, Suzhou, Venice, and Venice-XL. The sizes of nodes denote the number of nearby posts allocated to the nodes, and the colors of nodes illustrate the degree of the node on the graph. Each link connects two nodes reachable to each other within 20 min.

Furthermore, the two types of rank-size plots showing, respectively, the degree distribution and the posts-per-node distribution revealed similar patterns, the latter being more heavy-tailed, a typical characteristic of large-scale complex networks (Barabási, 2013; Eom and Jo, 2015), while the back-end spatial networks are relatively more regular.

4.7.4 Multi-Graphs and Sub-Graphs of Contextual Information

Table 4.7 shows graph statistics of three constructed sub-graphs $\mathcal{G}^{\text{TEM}}, \mathcal{G}^{\text{SOC}}, \mathcal{G}^{\text{SPA}}$ with different link types within the multi-graph $\mathcal{G}$, and the simple composed graph $\mathcal{G}'$ for each city, while Figure 4.6 plots their [weighted] degree distributions, respectively. The multi-graphs are further visualized in Figure 4.7.

TABLE 4.7 The statistics for the multi-graphs. "#" means "the number of" in the table.

Graph Features	AMS	SUZ	VEN
Temporal Graph $\mathcal{G}^{\text{TEM}} = (\mathcal{V}, \mathcal{E}^{\text{TEM}}, \boldsymbol{w}^{\text{TEM}})$			
# Nodes *	3727	3137	2951
# Edges	692,839	293,328	249,120
Diameter	145	116	270
Graph Density	0.100	0.060	0.057
Social Graph $\mathcal{G}^{\text{SOC}} = (\mathcal{V}, \mathcal{E}^{\text{SOC}}, \boldsymbol{w}^{\text{SOC}})$			
# Nodes **	3696	3120	2916
# Edges	877,584	602,821	242,576
# Connected Components	47	56	60
# Nodes Largest CC	2694	942	2309
Diameter Largest CC	7	6	10
Graph Density	0.129	0.124	0.057
Spatial Graph $\mathcal{G}^{\text{SPA}} = (\mathcal{V}, \mathcal{E}^{\text{SPA}}, \boldsymbol{w}^{\text{SPA}})$			
# Nodes **	3632	3102	2938
# Edges	135,079	415,049	221,414
# Connected Components	134	91	13
# Nodes Largest CC	1485	829	2309
Diameter Largest CC	22	1	22
Graph Density	0.020	0.086	0.051
Simple Composed Graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$			
# Nodes *	3727	3137	2951
# Edges	1,271,171	916,496	534,513
Diameter	4	5	4
Graph Density	0.183	0.186	0.123

* By definition a connected graph (only one connected component).
** The isolated nodes with no links are not counted here, therefore the numbers of nodes are smaller than the actual size of the node set $\mathcal{V}$.

The three link types provided heterogeneous characteristics:

- the temporal graph is by definition connected, where the highest density in Amsterdam suggested the largest number of photos taken in consecutive time periods, while the largest diameter in Venice suggested the broadest span of time;
- the social graph is structured by the relationship of users, where the largest connected components showed clusters of posts shared either by the same user, or by users who are friends or with mutual interests, the size of which in Suzhou is small because of the fewest users shown in Table 4.1;

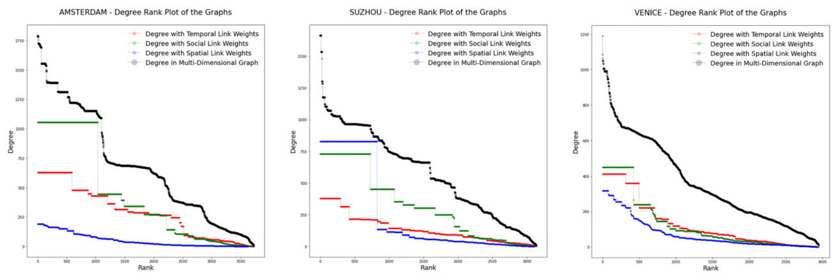


FIG. 4.6 The rank-size plots of the degree distributions in the three cases of Amsterdam, Suzhou, and Venice, with regard to the temporal links, social links, spatial links, as well as the entire multi-graph.

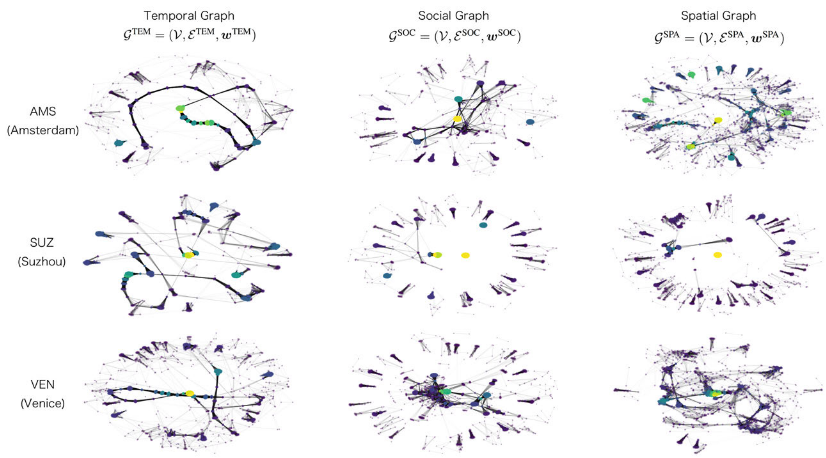


FIG. 4.7 The subgraphs of the multi-graphs in each case study city visualized using spring layout in NetworkX. The node size and colour reflect the degrees, and link thickness the edge weights.

- the spatial graph shows a similar connectivity pattern with the back-end spatial networks/graphs, where the extremely small diameter and the largest density in Suzhou reassured the fragmented positions of posts;
- although the degree distribution of three sub-graphs fluctuated due to the different socio-economic and spatiotemporal characteristics of different cities, that of the simple composed graph showed similar elbow-shaped patterns, with similar density and diameter.

Moreover, the heterogeneous graph structures suggest that different parameters and/or backbone models need to be fit and fine-tuned with each link type, a common practice for deep learning on multi-graphs.

The connected components of each type of temporal, social, and spatial links in each case study city are visualized in Figure 4.7, respectively. The `spring_layout` algorithm of NetworkX python library with the optimal distance between nodes `k` of 0.1 and random seed of `10396953` are used to output the graphs.

As the heterogeneous characteristics of constructed multi-graphs in the three cities are shown to be logically correspondent to reality, substantiating the generality of the methodological framework, they could be used as contextual information to aid future semi-supervised classification tasks concerning heritage values and attributes.

4.8 Discussion

4.8.1 Provisional Tasks for Urban Data Science

The datasets introduced could be used to answer questions from the perspectives of machine learning and social network analysis as well as heritage and urban studies. Table 4.8 gives a few provisional tasks that could be realised using the collected datasets of this paper, and further datasets to be collected using the same introduced framework.

These problems would use some or all of extracted features (visual, textual, contextual), generated labels (heritage values and attributes), constructed graph structures, and even raw data as input and output components to find the relationship function among them. Some problems are more interesting as ML/SNA problems (such as 4, 7 and 8), some are more fundamental for heritage studies and urban data science (such as 0, 1 and 6). While the former tends towards the technical and theoretical end of the whole potential range of the datasets, the latter tends towards the application end. However, to reach a reasonable performance during applications and discoveries, as is the main concern and interest for urban data science, further technical investigations and validations would be indispensable.

Even before performing such provisional tasks with the datasets created using the proposed framework in this study, the dataset creation and qualitative inspection process can already reveal interesting facts related to heritage studies, though they are performed primarily to check the quality of the created datasets in terms of their coherence and consistency. The analyses shown in Section 4.7.2 about the pseudo labels generated for the topics of heritage values and attributes provide the most trivial contribution to cultural heritage studies. On the one hand, it demonstrates that the proposed framework could transfer knowledge from pre-trained models and provide meaningful predictions as a replica of authoritative views to justify heritage values and attributes. On the other hand, the distribution of generated labels both give an expected (as examples visualized in Figures 4.3 and 4.4) and unexpected (for example the significant appearance of OUV selection criteria originally not inscribed in WHL) outcomes that could invite further heritage investigations.

TABLE 4.8 A few provisional tasks with formal problem definitions that could be performed. Potential scientific and social relevance for the Machine Learning community, and urban and/or heritage researchers, respectively, are given. The gray texts in the third column give a high-level categorization for each specific type of task in the context of machine learning.

ID	Problem Definition	Type of Task	As a *Machine Learning /Social Network Analysis* Problem	As an *Urban/Heritage Study* Question
0	$\boldsymbol{X}^{\text{vis}} \mapsto \boldsymbol{Y}^{\text{HV}} \vert \boldsymbol{K}^{\text{HV}}$	Image Classification (semi-supervised)	Using visual features to infer categories originally induced from (possibly missing) texts with co-training (Blum and Mitchell, 1998) in few-shot learning settings (Wang et al., 2020).	As the latest advances in heritage value assessment have been discovering the added value of inspecting texts (Tarrafa Silva and Pereira Roders, 2010), can values also be seen and retrieved from scenes of images?
1	$\boldsymbol{X}^{\text{tex}} \mapsto \boldsymbol{Y}^{\text{HA}} \vert \boldsymbol{K}^{\text{HA}}$	Text Classification (semi-supervised)	Using textual features to infer categories originally induced from images possibly with attention mechanisms (Vaswani et al., 2017).	How to relate the textual descriptions to certain heritage attributes (Gomez et al., 2019)? Are there crucial hints other than appeared nouns?
2	$\boldsymbol{X} := \left\{\boldsymbol{X}^{\text{vis}}, \boldsymbol{X}^{\text{tex}}\right\} \mapsto \boldsymbol{Y} := \left\{\boldsymbol{Y}^{\text{HV}} \vert \boldsymbol{K}^{\text{HV}}, \boldsymbol{Y}^{\text{HA}} \vert \boldsymbol{K}^{\text{HA}}\right\}$	Multi-modal Classification (semi-supervised)	Using multi-modal (multi-view) features to make inference, either with training joint representations or by making early and/or late fusions (Blum and Mitchell, 1998; Baltrusaitis et al., 2019).	How can heritage values and attributes be jointly inferred from the combined information of both visual scenes and textual expressions (Ginzarly et al., 2019)? How can they complement each other?
3	$\boldsymbol{X}, \boldsymbol{A} \mapsto \boldsymbol{Y}$	Node Classification (semi-supervised)	Test-beds for different graph filters such as Graph Convolution Networks (Kipf and Welling, 2016) and Graph Attention Networks (Veličković et al., 2017).	How can the contextual information of a post contribute to the inference of its heritage values and attributes? What is the contribution of time, space, and social relations (Miah et al., 2017)?
4	$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A} \mapsto \boldsymbol{A} + \boldsymbol{A}_{\text{new}}$	Link Prediction and Recommendation System (semi-supervised)	Test-beds for link prediction algorithms (Adamic and Adar, 2003) considering current graph structure and node features. What is the probability that other links also should exist?	Considering the similarity of posts, would there be heritage values and attributes that also suit the interest of another user, fit another location, and/or reflect another period of time (Majid et al., 2013)?
5	$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A} \mapsto \hat{\boldsymbol{X}}, \hat{\boldsymbol{Y}}, \hat{\boldsymbol{A}}$	Graph Coarsening (unsupervised)	Test-beds for graph pooling (Ma and Tang, 2021) and graph partitioning (Karypis and Kumar, 1995) algorithms to generate coarsened graphs (Pang et al., 2021) in different resolutions.	How can we summarize, aggregate, and eventually visualize the large-scale information from the social media platforms based on their contents and contextual similarities (Cho et al., 2022)?
6	$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A} \mapsto \boldsymbol{y}_{\mathcal{G}}^{\text{HV}} \vert \boldsymbol{Y}^{\text{HV}}, \boldsymbol{y}_{\mathcal{G}}^{\text{HA}} \vert \boldsymbol{Y}^{\text{HA}}$	Graph Classification (supervised)	Test-beds for graph classification algorithms (Zhang et al., 2018) when more similar datasets have been collected and constructed in more case study cities.	Can we summarize the social media information of any city with World Heritage property so that the critical heritage values and attributes could be directly inferred (Monteiro et al., 2014)?

TABLE 4.8 Cont.

ID	Problem Definition	Type of Task	As an *ML/SNA* Problem	As an *Urban/Heritage Study* Question
7	$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A} \mapsto \mathfrak{I}, \boldsymbol{S}$	Image/Text Generation (supervised)	Using multi-modal features to generate the missing and/or unfit images and/or textual descriptions, probably with Generative Adversarial Network (Goodfellow et al., 2014).	How can a typical image and/or textual description of certain heritage values and attributes at a certain location in a certain time by a certain type of user in a specific case study city be queried or even generated (Gomez et al., 2019)?
8	$\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{A}^{\text{TEM}}, \boldsymbol{A}^{\text{SOC}}, \boldsymbol{A}^{\text{SPA}} \mapsto \boldsymbol{R} + \boldsymbol{R}^{\text{TEM}} + \boldsymbol{R}^{\text{SOC}} + \boldsymbol{R}^{\text{SPA}}$	Attributed Multi-Graph Embedding (self-supervised)	Respectively generating a universal embedding and a context-specific embedding for each type of links in the multi-dimensional network (Ma et al., 2018), probably with random walks on graphs.	How are heritage values and attributes distributed and diffused in different contexts? Is the First Law of Geography (Tobler, 1970) still valid in the specific social, temporal and spatial graphs?
9	$\boldsymbol{X}^{(k)}, \boldsymbol{Y}^{(k)}, \boldsymbol{A}^{(k)}, \boldsymbol{T} \mapsto \boldsymbol{X}^{(k+1)}, \boldsymbol{Y}^{(k+1)}, \boldsymbol{A}^{(k+1)}$	Dynamic Prediction (self-supervised)	Given the current graph structure and its features stamped with time steps, how shall it further evolve in the next time steps (Nguyen et al., 2018; Ren et al., 2019)?	How are the current expressions of heritage values and attributes in a city influencing the emerging post contents, the tourist behaviours, and the planning decision making (Zhang and Cheng, 2020; Bai et al., 2021c)?

The analyses of generated features shown in Section 4.7.1, however, could also provide strong clues informative to heritage studies. As argued in Section 4.4, both machine-readable abstract features such as hidden-layer vectors and human-interpretable prediction categories are stored as multi-modal features. While conducting future machine learning training, sensitivity checks on such interpretable features could give insights on how and what the models learn. For example, one would expect a model predicting the heritage value of “criterion (vi)—association” and heritage attributes of “People’s Activity and Association” to pay much attention to the number and proportion of human faces in the image, and vice versa, hence the extraordinary appearance of both categories in the city of Amsterdam. As for the graph analyses in Section 4.7.4, while providing a basis for further graph-based semi-supervised learning of similar posts in nearby places, from the same time period, and by alike social groups, the spatiotemporal and socio-economic distribution of posts (as a proxy to social behaviour) already tells a story. For instance, as has been extensively argued by researchers such as Bill Hillier et al., one can often find a clear correspondence between the “buzz” or vitality of human activities in cities with the inherent centrality distributions on the network representation of the underlying space (Hillier and Hanson, 1989; Penn, 2003; Ratti, 2004). The co-appearance of large circles (large number of posts, thus high vitality) and warm colours (high centrality), and the visible clustering of warm colours (around places with good connectivity, such as the Rialto bridge and San Marco in Venice, confirming the conclusions drawn by Psarra (Psarra, 2018)) shown in Figure 4.5 could further demonstrate such findings. Culturally significant locations are often important not only due to their individual attributes but also due to the embedding in their contexts, which inevitably renders cultural heritage studies contextual.

Further advanced analyses for directly answering domain-specific questions in cultural heritage studies (such as Questions 3 and 8 about the mechanism of contextual influence of posts to the mapping, extraction, and inference of heritage values and attributes) have been categorized in Table 4.8. Note that a further distinction needs to be made within the extracted heritage values and attributes, as they may essentially be clustered into three categories:

- core heritage values and attributes officially listed and recognized that thoroughly define the heritage status;
- values and attributes relevant to conservation and preservation practice;
- other values and attributes not specifically heritage-related yet are conveyed to the same heritage property by ordinary people.

This distinction should be made clear for practitioners intending to make planning decisions based on the conclusions drawn from studying such datasets.

One advantage of the proposed framework is that it allows for the creation of multi-graphs from multiple senses of proximity or similarity in geographical, temporal, and/or the social space. In cases where one cannot easily find a ground truth, i.e., in exploratory analyses, having the possibility to treat the dataset as a set of connected data points instead of a powder-like set will be advantageous. The sense of similarity between data points by virtue of geographical/spatial proximity is arguably the oldest type of connection between them. However, when there is no exact physical sense of proximity in a geographical space, or when other forms of connection, e.g., through social media, are of influence, data scientists can benefit from other clues such as temporal connections related to the events or the social connections related to community structures. These can all inform potential questions to be answered in future studies.

Moreover, after retrieving knowledge of heritage values and attributes in case study cities from multi-modal UGC, for the sake of visualization, assessment and comparison during decision-making processes, further bundling and aggregation of individual data points would be desirable, as was briefly mentioned in Section 4.1 and also formulated in Table 4.8 as Question 5. This could be performed with all three proposed contextual information types denoting the proximity of data points. Data bundling and aggregation in the spatial domain would be the first action for creating a map. Depending on different use cases, this could be performed either on scale-dependent representations of geographical/administrative units, such as the natural islands divided by canals, or the so-called parish islands/communities in Venice (Psarra, 2018), or on identified clusters based on regular grids at different scales, such as the "tourism districts" Encalada-Abarca et al. (2022). While the use of the former (i.e., top-down boundaries) is trivial for administrative purposes, the latter (i.e., bottom-up clusters) could be arguably more generalizable in other cases, reflecting a universal collective sense of place (Encalada-Abarca et al., 2022). Data bundling and aggregation in the temporal domain would map the generated features and labels on a discrete timeline at different scales (e.g., months, years, decades, etc.), presumably of sufficiently high resolution to capture the temporal dynamics and variations of data. For example, one may find that some topics are extensively mentioned in only a short period of time, while others pertain for longer spans, suggesting different patterns of public perception and communal attention, which may also help with heritage-related event detection and contribute to further planning and management strategies (Cheng and Wicks, 2014; Bai et al., 2021c). Data bundling and aggregation in the social domain, on the other hand, could help to profile the interests of user communities or user groups (e.g., local residents and tourists), which is beneficial for instance in devising recommendation systems. As argued in Section 4.6, multiple posts by the same user were not necessarily considered redundant in this study. Instead, the consistency and/or variations revealed in posted content by the same user [community/group] profile could further categorize their preference and opinions related to the cultural significance of heritage (Majid et al., 2013).

4.8.2 Limitations and Future Steps

No thorough human evaluations and annotations were performed during the construction of the datasets presented in this paper. This manuscript provides a way to circumvent this step by using only the confidence and [dis-]agreement of presumably well-trained models as a proxy for the more conventional "inter-annotator" agreement to show the quality of datasets and generate [pseudo-]labels (Nowak and Rüger, 2010). This resembles the idea of using consistency, confidence, and disagreement to improve the model performance in semi-supervised learning (Zhou and Li, 2010; Lee et al., 2013; Sohn et al., 2020). For the purpose of introducing a general framework that could generate more graph datasets, it is preferable to exclude humans from the loop as this would function as a bottleneck limiting the process, both in time and monetary resources, and in demanded domain knowledge. However, for applications where more accurate conclusions are needed, human evaluations on the validity, reliability, and coherence of the models are still needed. In order to gain a clear sense of the performance before implementation, the inspection of some predicted results is a prudent suggestion. As the step of [pseudo-]label generation was relatively independent from the other steps introduced in this paper, higher-quality labels annotated and evaluated by experts and/or crowd-workers could still be added at a later stage as augmentation or even replacement, as an active learning process (Prince, 2004; Zhu and Goldberg, 2009; Settles, 2011). For example, future studies are invited to integrate the more recognized classification frameworks for heritage values and heritage attributes (Pereira Roders, 2007; Tarrafa Silva and Pereira Roders, 2010; Veldpaus, 2015), in response to the possible imprecision of concepts as pointed out in Section 4.5. Moreover, generating labels of heritage values and attributes was only a choice motivated by the use-case at hand which suffices to show the utility of the framework for exploratory analyses on attributed graphs in cases where the sources of data are inherently unstructured and the connections between data points are inherently multi-faceted. Yet, it is also possible to apply the same framework as well as parts of the implemented workflow while only replacing the classifiers mentioned in Section 4.5 with domain-specific modules appropriated to the use-cases, to answer other exploratory questions in urban data science and computational social sciences, as suggested in Section 4.2.

While scaling up the dataset construction process, such as from VEN to VEN-XL, a few changes need to be adopted. For data collection, an updated strategy is already described in Section 4.3.2. For feature and label generation, mini-batches and GPU computing significantly accelerated the process. However, the small graphs from case study cities containing around 3000 nodes already contained edges at the scale of millions, making it challenging to scale up in cases such as VEN-XL, the adjacency list of which would be at the scale of billions, easily exceeding the limits of computer memory. As a result, VEN-XL has not yet been constructed as a multi-graph. Further strategies such as using sparse matrices (Yuster and Zwick, 2005) and parallel computing should be considered. Moreover, the issue of scalability should also be considered for later graph neural network training, since the multi-graphs constructed in this study can become quite dense locally. Sub-graph sampling methods should be applied to avoid “neighbourhood explosion” (Ma and Tang, 2021).

Although the motivation of constructing datasets regarding heritage values and attributes from social media was to promote inclusive planning processes, the selection of social media platforms already automatically excluded those not using, or not even aware of, the platform, let alone those not using internet. The scarce usage of Flickr in China, as an example of its limitation, also suggested that conclusions drawn from such datasets may reflect perspectives from the “tourist gaze” (Urry and Larsen, 2011) rather than local communities, and therefore losing some representativeness and generality. However, the main purpose of this paper is to provide a reproducible methodological framework with mathematical definitions, not limited to Flickr as a specific instance. Images, texts, and even audio files and videos from other platforms such as Weibo, Dianping, RED, and TikTok that are more popular in China could also add complementary local perspectives. With careful adaptions, archives, official documents, news articles, academic publications, and interview transcripts could also be constructed in similar formats for fairer comparisons, which again would fit in the general framework proposed in Section 4.2 as specified instances.

4.8.3 An Additional Application in Rome Testaccio

Furthermore, the Testaccio area in Rome, Italy is chosen as an additional case study to test the methodological framework in a finer-grained smaller-scale urban area with higher resolutions, instead of the three main case studies in this chapter that are all at urban scales (Bai et al., 2023). Being at the border within the UNESCO World Heritage property “the Historic Centre of Rome”, its historic and cultural values are officially justified with Outstanding Universal Value (OUV), as shown in Appendix A.

The archaeological excavations and built heritage of great chronological and typological diversity make up the specific urban character of the area, including and not limited to the Pyramid of Cestius, Monte Testaccio, Aurelian Wall, Non-Catholic Cemetery, and the Mattatoio Slaughterhouse (De Kleijn et al., 2013). An overview of the major attractions of the area including the boundary of UNESCO World Heritage property is shown in Figure 4.8 left.

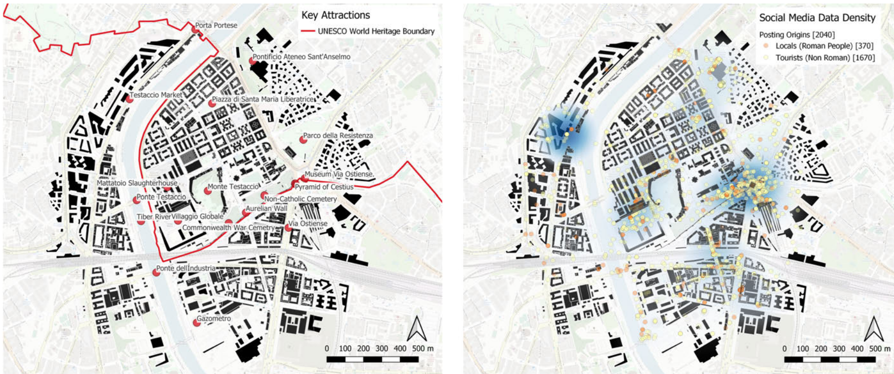


FIG. 4.8 The major tourist attractions and the distribution of social media images collected in the area of Testaccio. Left: The main tourist destinations, where the UNESCO World Heritage boundary is marked as a red line, the northeastern side of which is within the World Heritage property. Right: the posts from locals and tourists overlayed with a heatmap of all posts.

A fraction of the data collection process from this chapter is followed to extract 2000 posts in the area from Flickr (Bai et al., 2022), containing both locals and visitors, as shown in Figure 4.8 right. The images are processed with VGG-16 network pre-trained on the ImageNet dataset using Keras python library to obtain the last 4096-dimensional hidden layer vector output as their structured data representations (Simonyan and Zisserman, 2015), which is further reduced to 300D vectors with Principle Component Analysis (PCA). The PCA features are then fed into a t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm using Scikit-Learn to compute the first two components (van der Maaten and Hinton, 2008). The images are then visualized by transforming the t-SNE coordinates of the data points into a regular 2D grid using RasterFairy library, which is eventually clustered manually with their main depicted topics as proxies to the perceived urban heritage attributes. Moreover, the textual comments of the posts are processed with pre-trained natural language processing (NLP) models with the classification framework of OUV selection criteria (Bai et al., 2021a). Word Clouds are generated with the most significant and relevant OUV categories in the Testaccio area.

Analyses show a consistent representation of urban heritage images revealed on social media with official heritage values and attributes. Figure 4.9 demonstrates the visualized topic clusters of heritage attributes perceived as significant. It shows that most places mentioned above are present in the online gallery, and the most dominant contents of the area are Pyramid, Cemetery, and nightlife culture around

Monte Testaccio. An issue of interest revealed with the visualization is that the visual representation of Monte Testaccio, the Aurelian Wall, and the Tiber River are not significant on Flickr, suggesting that these formal heritage sites are not given enough attention, probably because of accessibility and visibility problems.

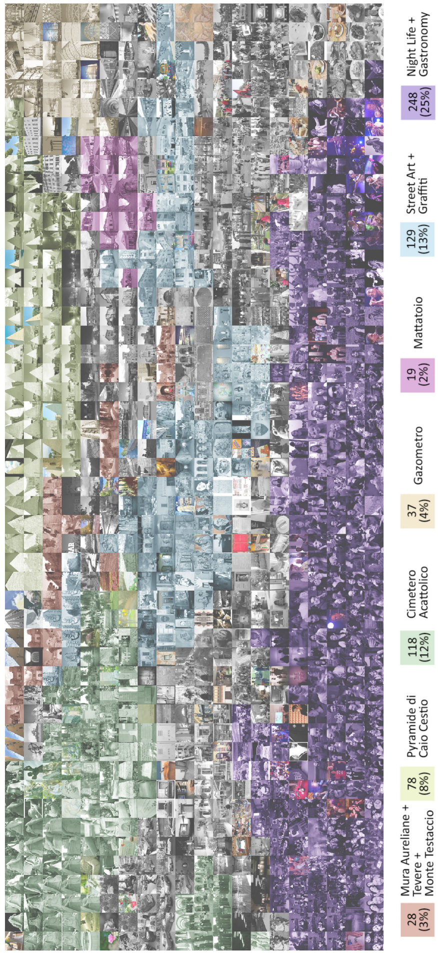


FIG. 4.9 Clustered social media images based on the image content using the t-SNE algorithm with their respective proportions.

Figure 4.10 shows the distribution of posts owned by different groups of people concerning various OUV selection criteria. The OUV-related posts are mostly concentrated in the area of the Pyramid, the Non-Catholic Cemetery, and along the Via Ostience, the Tiber River, and the Aurelian Wall. It proves that from a bottom-up perspective, both local Roman people and tourists from all over the world are actively present and eager to share their observations and experiences they have in this area. All five OUV criteria of the Roman UNESCO World Heritage property could be observed as perceived by the online community in the area, while criterion (vi) about association, criterion (iii) about testimony, and criterion (i) about masterpiece are the most representative characteristics based on classification analysis of the NLP models, as shown in the word-cloud of Figure 4.11.

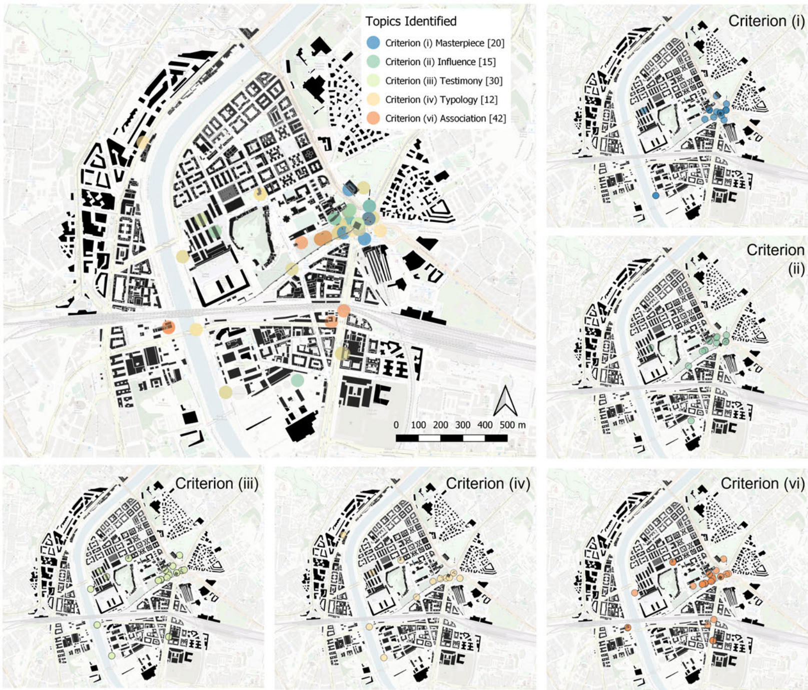


FIG. 4.10 The distribution of the sentences classified to be relevant to the OUV selection criteria.

This additional case study showcases the generalizability and application potentials of the methodological framework proposed in this chapter. Such a methodology provides an alternative perspective of viewing the urban heritage as a collection of depicted contents, to be augmented with the conventional Authorised Heritage Discourse. It can contribute as a documentation tool of collective knowledge for inclusive heritage management and local development planning.

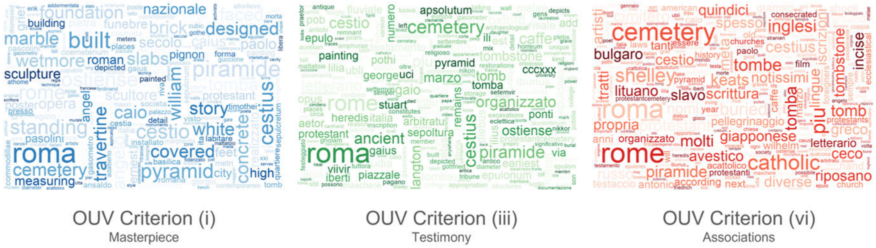


FIG. 4.11 Word clouds generated with posts classified as relevant to three significant OUV selection criteria.

4.9 Conclusions

This chapter introduced a novel methodological framework to construct graph-based multi-modal datasets Heri-Graphs concerning heritage values and attributes using data from the social media platform Flickr. Pre-trained machine learning models were applied to generate multi-modal features and domain-specific pseudo-labels. A full mathematical formulation is provided for the feature extraction, label generation, and graph construction processes. Temporal, spatial, and social relationships among the posts are used to construct multi-graphs, ready to be utilised as contextual information for further semi-supervised machine learning tasks. Three case study cities with urban areas inscribed in the UNESCO WHL, namely Amsterdam, Suzhou, and Venice, are tested with the framework to construct sample datasets, being evaluated and filtered with the consistency of models and qualitative inspections. The datasets in the three sample cities are shown to provide meaningful information concerning the spatiotemporal and socio-economic distributions of heritage values and attributes conveyed by social media users, useful for knowledge documentation and mapping for heritage and urban studies. Such understanding is strongly aligned with the Sustainable Development Goal (SDG) 11, with its ultimate objective of making the urban heritage management processes more inclusive. The datasets created through the proposed framework provide a basis for revisiting or generalizing the First Law of Geography as formulated by Tobler to include the new senses of proximity or similarity caused by crowd behaviour and other social connections through electronic media that are arguably not directly related to geographical matters. This is especially important since heritage studies in particular, urban studies, and computational social sciences, in general, are almost always concerned with contextual information, which is arguably not limited to the geographical context but also to the social and temporal contexts. Moreover, the additional case study in the Testaccio area in Rome confirms the generalizability of the dataset creation workflow proposed in this study, showing that it can also be applied in fractions of urban areas to collect smaller-scale datasets, thus not only at the scale of cities. Such datasets have the potential to be applied by both the machine learning community and urban data scientists to help answer interesting questions of scientific/technical and social relevance, which could also be applied globally with a broad range of use cases.

References

Adamic, L. A. and Adar, E. (2003). Friends and neighbors on the web. Social networks, 25(3):211–230.

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In Aggarwal, C. C., editor, Social Network Data Analytics, chapter 1, pages 1–15. SPRINGER.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3):175–185.

Amato, F., Cozzolino, G., Di Martino, S., Mazzeo, A., Moscato, V., Picariello, A., Romano, S., and Sperlí, G. (2016). Opinions analysis in social networks for cultural heritage applications. Smart Innovation, Systems and Technologies, 55:577–586.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics surveys, 4:40–79.

Bai, N., Ducci, M., Mirzikashvili, R., Nourian, P., and Pereira Roders, A. (2023). Mapping urban heritage images with social media data and artificial intelligence, a case study in testaccio, rome. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023:139–146.

Bai, N., Luo, R., Nourian, P., and Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage statements of ”Outstanding Universal Value” with soft labels. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 366–384, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2021b). “What is OUV” revisited: A computational interpretation on the statements of Outstanding Universal Value. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-M-1-2021:25–32.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2022). Heri-graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. ISPRS International Journal of Geo-Information, 11(9).

Bai, N., Nourian, P., and Pereira Roders, A. (2021c). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.

Baltrusaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2):423–443.

Barabási, A.-L. (2013). Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1987):20120375.

Batty, M. (2013). The new science of cities. MIT press.

Bekker, R. (2020). Creating insights in tourism with flickr photography, visualizing and analysing spatial and temporal patterns in venice. Master’s thesis, Rijksuniversiteit Groningen.

Blanchard, P. and Volchenkov, D. (2008). Mathematical analysis of urban spatial networks. Springer Science & Business Media.

Blum, A. and Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In Proceedings of the eleventh annual conference on Computational learning theory, pages 92–100.

Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65:126–139.

Bonci, A., Clini, P., Martin, R., Pirani, M., Quattrini, R., and Raikov, A. (2018). Collaborative intelligence cyber-physical system for the valorization and re-use of cultural heritage. Journal of Information Technology in Construction, 23(1):305–323.

Breiman, L. (1996a). Bagging predictors. Machine learning, 24(2):123–140.

Breiman, L. (1996b). Stacked regressions. Machine learning, 24(1):49–64.

Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32.

Campillo-Alhama, C. and Martinez-Sala, A.-M. (2019). Events 2.0 in the transmedia branding strategy of World Cultural Heritage Sites. PROFESIONAL DE LA INFORMACION, 28(5).

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), pages 67–74. IEEE.

Cheng, T. and Wicks, T. (2014). Event detection using twitter: A spatio-temporal approach. PloS one, 9(6):e97807.
Cho, N., Kang, Y., Yoon, J., Park, S., and Kim, J. (2022). Classifying tourists' photos and exploring tourism destination image using a deep learning model. Journal of Quality Assurance in Hospitality & Tourism, pages 1–29.
Chua, T.-S., Tang, J., Hong, R., Li, H., Luo, Z., and Zheng, Y. (2009). Nus-wide: a real-world web image database from national university of singapore. In Proceedings of the ACM international conference on image and video retrieval, pages 1–9.
Clark, K., Khandelwal, U., Levy, O., and Manning, C. D. (2019). What does BERT look at? an analysis of BERT's attention. In Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pages 276–286, Florence, Italy. Association for Computational Linguistics.
Crandall, D., Backstrom, L., Huttenlocher, D., and Kleinberg, J. (2009). Mapping the world's photos. WWW'09 - Proceedings of the 18th International World Wide Web Conference, pages 761–770.
De Kleijn, M., van Aart, C. J., Van Manen, N., Burgers, G.-J., and Scholten, H. J. (2013). Testaccio, a digital cultural biography app. In UMAP Workshops.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee.
Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
Encalada-Abarca, L., Ferreira, C. C., and Rocha, J. (2022). Measuring tourism intensification in urban destinations: An approach based on fractal analysis. Journal of Travel Research, 61(2):394–413.
Eom, Y.-H. and Jo, H.-H. (2015). Tail-scope: Using friends to estimate heavy tails of degree distributions in large-scale complex networks. Scientific reports, 5(1):1–9.
Gabrielli, L., Rinzivillo, S., Ronzano, F., and Villatoro, D. (2014). From Tweets to Semantic Trajectories: Mining Anomalous Urban Mobility Patterns. In Nin, J and Villatoro, D., editor, CITIZEN IN SENSOR NETWORKS, volume 8313 of Lecture Notes in Artificial Intelligence, pages 26–35, HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY. SPRINGER-VERLAG BERLIN.
Gao, H. and Ji, S. (2019). Graph u-nets. In international conference on machine learning, pages 2083–2092. PMLR.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2019a). Machine learning and point of interests: typical tourist Italian cities. Current Issues in Tourism, 0(0):1–13.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2019b). Using social media to identify tourism attractiveness in six Italian cities. Tourism Management, 72:306–312.
Ginzarly, M., Pereira Roders, A., and Teller, J. (2019). Mapping historic urban landscape values through social media. Journal of Cultural Heritage, 36:1–11.
Gomez, R., Gomez, L., Gibert, J., and Karatzas, D. (2019). Learning from #barcelona instagram data what locals and tourists post about its neighbourhoods. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11134 LNCS:530–544.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.
Gustcoven, E. (2016). Attributes of world heritage cities, sustainability by management–a comparative study between the world heritage cities of amsterdam, edinburgh and querétaro. Master's thesis, KU Leuven.
Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).
He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.
Hillier, B. and Hanson, J. (1989). The Social Logic of Space. Cambridge University Press.
Hinton, G. E. (1990). Connectionist learning procedures. In Machine learning, pages 555–610. Elsevier.
Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. In Gurevych, I. and Miyao, Y., editors, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers, pages 328–339. Association for Computational Linguistics.
Howley, P., Scott, M., and Redmond, D. (2009). Sustainability versus liveability: an investigation of neighbourhood satisfaction. Journal of environmental planning and management, 52(6):847–864.
Huiskes, M. J. and Lew, M. S. (2008). The MIR Flickr retrieval evaluation. Proceedings of the 1st International ACM Conference on Multimedia Information Retrieval, MIR2008, Co-located with the 2008 ACM International Conference on Multimedia, MM'08, pages 39–43.
Jokilehto, J. (2007). Aesthetics in the world heritage context. In Values and Criteria in Heritage Conservation, pages 183–194. Polistampa.
Jokilehto, J. (2008). What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties. Technical report, ICOMOS, ICOMOS Berlin.

Kang, Y., Cho, N., Yoon, J., Park, S., and Kim, J. (2021). Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos. ISPRS International Journal of Geo-Information, 10(3):137.

Karypis, G. and Kumar, V. (1995). Analysis of multilevel graph partitioning. In Supercomputing'95: Proceedings of the 1995 ACM/IEEE conference on Supercomputing, pages 29–29. IEEE.

Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

Lafon, S. and Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE transactions on pattern analysis and machine intelligence, 28(9):1393–1403.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Social science. computational social science. Science (New York, NY), 323(5915):721–723.

Lee, D.-H. et al. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML, volume 3, page 896.

Lee, H. and Kang, Y. (2021). Mining tourists' destinations and preferences through lstm-based text classification and spatial clustering using flickr data. Spatial Information Research, 29(6):825–839.

Li, D., Zhou, X., and Wang, M. (2018). Analyzing and visualizing the spatial interactions between tourists and locals: A flickr study in ten us cities. Cities, 74:249–258.

Liew, C. L. (2014). Participatory cultural heritage: A tale of two institutions' use of social media. D-Lib Magazine, 20(3-4):1–17.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision, pages 740–755. Springer.

Lu, W. and Stepchenkova, S. (2015). User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. Journal of Hospitality Marketing and Management, 24(2):119–154.

Ma, Y., Ren, Z., Jiang, Z., Tang, J., and Yin, D. (2018). Multi-dimensional network embedding with hierarchical structure. In Proceedings of the eleventh ACM international conference on web search and data mining, pages 387–395.

Ma, Y. and Tang, J. (2021). Deep learning on graphs. Cambridge University Press.

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., and Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science, 27(4):662–684.

Marine-Roig, E. and Anton Clavé, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. Journal of Destination Marketing and Management, 4(3):162–172.

Miah, S. J., Vu, H. Q., Gammack, J., and McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. Information & Management, 54(6):771–785.

Monteiro, V., Henriques, R., Painho, M., and Vaz, E. (2014). Sensing World Heritage An Exploratory Study of Twitter as a Tool for Assessing Reputation. In Murgante, B and Misra, S and Rocha, AMAC and Torre, C and Rocha, JG and Falcao, MI and Taniar, D and Apduhan, BO and Gervasi, O., editor, COMPUTATIONAL SCIENCE AND ITS APPLICATIONS - ICCSA 2014, PT II, volume 8580 of Lecture Notes in Computer Science, pages 404–419. Univ Minho; Univ Perugia; Univ Basilicata; Monash Univ; Kyushu Sangyo Univ; Assoc Portuguesa Investigacao Operac.

Nguyen, G. H., Lee, J. B., Rossi, R. A., Ahmed, N. K., Koh, E., and Kim, S. (2018). Continuous-time dynamic network embeddings. In Companion Proceedings of the The Web Conference 2018, pages 969–976.

Nourian, P. (2016). Configraphics: Graph Theoretical Methods for Design and Analysis of Spatial Configurations. TU Delft.

Nourian, P., Rezvani, S., Sariyildiz, I., and van der Hoeven, F. (2016). Spectral modelling for spatial network analysis. In Proceedings of the Symposium on Simulation for Architecture and Urban Design (simAUD 2016), pages 103–110. SimAUD.

Nowak, S. and Rüger, S. (2010). How reliable are annotations via crowdsourcing. In Proceedings of the international conference on Multimedia information retrieval, pages 557–566.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359.

Pang, Y., Zhao, Y., and Li, D. (2021). Graph pooling via coarsened graph infomax. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 2177–2181.

Patterson, G. and Hays, J. (2012). Sun attribute database: Discovering, annotating, and recognizing scene attributes. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2751–2758. IEEE.

Patterson, G., Xu, C., Su, H., and Hays, J. (2014). The sun attribute database: Beyond categories for deeper scene understanding. International Journal of Computer Vision, 108(1):59–81.

Pearson, K. (1905). The problem of the random walk. Nature, 72(1865):294–294.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.
Penn, A. (2003). Space syntax and spatial cognition: or why the axial line? Environment and behavior, 35(1):30–65.
Pentland, A. (2015). Social Physics: How social networks can make us smarter. Penguin.
Pereira Roders, A. (2007). Re-architecture: lifespan rehabilitation of built heritage. PhD thesis, Technische Universiteit Eindhoven.
Pereira Roders, A. (2010). Revealing the World Heritage cities and their varied natures. In Heritage 2010: Heritage and Sustainable Development, Vols 1 and 2, chapter Heritage a, pages 245–253. Green Lines Institute.
Pereira Roders, A. (2019). The Historic Urban Landscape Approach in Action: Eight Years Later. In Reshaping Urban Conservation, pages 21–54. Springer.
Pickering, C., Rossi, S. D., Hernando, A., and Barros, A. (2018). Current knowledge and future research directions for the monitoring and management of visitors in recreational and protected areas. Journal of Outdoor Recreation and Tourism, 21(November 2017):10–18.
Platt, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers, 10(3):61–74.
Plummer, B. A., Wang, L., Cervantes, C. M., Caicedo, J. C., Hockenmaier, J., and Lazebnik, S. (2015). Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. In Proceedings of the IEEE international conference on computer vision, pages 2641–2649.
Prince, M. (2004). Does active learning work? a review of the research. Journal of engineering education, 93(3):223–231.
Psarra, S. (2018). The Venice Variations: Tracing the Architectural Imagination. UCL press.
Pustejovsky, J. and Stubbs, A. (2012). Natural Language Annotation for Machine Learning: A guide to corpus-building for applications. " O'Reilly Media, Inc.".
Rakic, T. and Chambers, D. (2008). World Heritage: Exploring the Tension Between the National and the 'Universal'. Journal of Heritage Tourism, 2(3):145–155.
Ratti, C. (2004). Space syntax: Some inconsistencies. Environment and Planning B: Planning and Design, 31(4):487–499.
Reiter, R. (1989). Towards a logical reconstruction of relational database theory. In Readings in Artificial Intelligence and Databases, pages 301–327. Elsevier.
Ren, Y., Cheng, T., and Zhang, Y. (2019). Deep spatio-temporal residual neural networks for road-network-based data modeling. International Journal of Geographical Information Science, 33(9):1894–1912.
Rish, I. et al. (2001). An empirical study of the naive bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence, volume 3, pages 41–46.
Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 815–823.
Settles, B. (2011). From theories to queries: Active learning in practice. In Active learning and experimental design workshop in conjunction with AISTATS 2010, pages 1–18. JMLR Workshop and Conference Proceedings.
Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In Bengio, Y. and LeCun, Y., editors, 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015.
Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., and Li, C.-L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems, 33:596–608.
Sun, C., Qiu, X., Xu, Y., and Huang, X. (2019). How to Fine-Tune BERT for Text Classification? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11856 LNAI(2):194–206.
Tang, L. and Liu, H. (2009). Relational learning via latent social dimensions. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 817–826.
Tarrafa Silva, A. and Pereira Roders, A. (2010). The cultural significance of World Heritage cities : Portugal as case study. In Heritage and Sustainable Development, pages 255–263, Évora, Portugal.
Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., and Toivonen, T. (2017). Instagram, flickr, or twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Scientific reports, 7(1):1–11.
Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. Economic geography, 46(sup1):234–240.
UNESCO (1972). Convention Concerning the Protection of the World Cultural and Natural Heritage. Technical Report november, UNESCO, Paris.
UNESCO (2008). Operational guidelines for the implementation of the world heritage convention. Technical Report July, UNESCO World Heritage Centre.
UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.

UNESCO (2020). Heritage in Urban Contexts: Impact of Development Projects on World Heritage properties in Cities. Technical Report January, UNESCO World Heritage Centre.
Urry, J. and Larsen, J. (2011). The tourist gaze 3.0. Sage.
Valese, M., Noardo, F., and Pereira Roders, A. (2020). World heritage mapping in a standard-based structured geographical information system. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4-2020:81–88.
van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. Journal of machine learning research, 9(11).
van Dijck, J. (2011). Flickr and the culture of connectivity: Sharing views, experiences, memories. Memory Studies, 4(4):401–415.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 6000–6010.
Veldpaus, L. (2015). Historic urban landscapes: framing the integration of urban and heritage planning in multilevel governance. PhD thesis, Technische Universiteit Eindhoven.
Veldpaus, L. and Roders, A. P. (2014). Learning from a legacy: Venice to valletta. Change over time, 4(2):244–263.
Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.
Wang, Y., Yao, Q., Kwok, J. T., and Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys (csur), 53(3):1–34.
Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.
Williams, N. L., Inversini, A., Ferdinand, N., and Buhalis, D. (2017). Destination eWOM: A macro and meso network approach? Annals of Tourism Research, 64:87–101.
Yuster, R. and Zwick, U. (2005). Fast sparse matrix multiplication. ACM Transactions On Algorithms (TALG), 1(1):2–13.
Zeng, H., Zhou, H., Srivastava, A., Kannan, R., and Prasanna, V. (2019). Graphsaint: Graph sampling based inductive learning method. In International Conference on Learning Representations.
Zhang, F., Zhou, B., Ratti, C., and Liu, Y. (2019). Discovering place-informative scenes and objects using social media photos. Royal Society open science, 6(3):181375.
Zhang, M., Cui, Z., Neumann, M., and Chen, Y. (2018). An end-to-end deep learning architecture for graph classification. In Thirty-second AAAI conference on artificial intelligence, pages 4438–4445.
Zhang, Y. and Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Computers, Environment and Urban Systems, 79:101403.
Zhang, Z., Cui, P., and Zhu, W. (2020). Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering.
Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6):1452–1464.
Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using places database. Advances in neural information processing systems, 27.
Zhou, Y. and Long, Y. (2016). Sinogrids: a practice for open urban data in china. Cartography and Geographic Information Science, 43(5):379–392.
Zhou, Z.-H. (2012). Ensemble methods: foundations and algorithms. CRC press.
Zhou, Z.-H. and Li, M. (2010). Semi-supervised learning by disagreement. Knowledge and Information Systems, 24(3):415–439.
Zhu, X. and Goldberg, A. B. (2009). Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning, 3(1):1–130.

5 Mapping

Semi-supervised Classification of Perceived Cultural Significance on Graphs

Parts of this chapter have been published in Bai et al. (2023)
Bai, N, Nourian P, Luo R, Cheng T, Pereira Roders, A. (2023). Screening the Stones of Venice: Mapping Social Perceptions of Cultural Significance through Graph-based Semi-supervised Classification. ISPRS Journal of Photogrammetry and Remote Sensing. 203, 135-164.

ABSTRACT Mapping cultural significance of heritage properties in urban environment from the perspective of the public has become an increasingly relevant process, as highlighted by the 2011 UNESCO Recommendation on the Historic Urban Landscape (HUL). With the ubiquitous use of social media and the prosperous developments in machine and deep learning, it has become feasible to collect and process massive amounts of information produced by online communities about their perceptions of heritage as social constructs. Moreover, such information is usually inter-connected and embedded within specific socioeconomic and spatiotemporal contexts. This paper presents a methodological workflow for using semi-supervised learning with graph neural networks (GNN) to classify, summarize, and map cultural significance categories based on user-generated content on social media. Several GNN models were trained as an ensemble to incorporate the multi-modal (visual and textual) features and the contextual (temporal, spatial, and social) connections of social media data in an attributed multi-graph structure. The classification results with different models were aligned and evaluated with the prediction confidence and agreement. Furthermore, message diffusion methods on graphs were proposed to aggregate the post labels onto their adjacent spatial nodes, which helps to map the cultural significance categories in their geographical contexts. The workflow is tested on data gathered from Venice as a case study, demonstrating the generation of social perception maps for this UNESCO World Heritage property. This research framework could also be applied in other cities worldwide, contributing to more socially inclusive heritage management processes. Furthermore, the proposed methodology holds the potential of diffusing any human-generated location-based information onto spatial networks and temporal timelines, which could be beneficial for measuring the safety,

vitality, and/or popularity of urban spaces.

KEYWORDS Social Media Data, Multi-modal Machine Learning, Graph Neural Networks, Spectral Centrality, Heritage Values and Attributes, Label Diffusion

5.1 Introduction

Documenting and mapping the values (cultural significance) of cities have always been an important task in the practice of urban conservation (Zancheti and Jokilehto, 1997; ICOMOS, 2013). As an art critic, historian, writer, polymath, and a pioneer in heritage conservation, John Ruskin openly expressed and actively promoted the cultural significance of the grandiose architecture on the Venetian island in his three-volume masterpiece The Stones of Venice (Ruskin, 1879; Ruskin and Quill, 2015). Through several visits to Venice, Ruskin was attracted by the buildings, monuments, sculptures, and building elements, especially those dating from the era of Byzantine and Gothic. In fear of losing its cultural significance by industrial modernization and destructive restorations, Ruskin tirelessly documented every stone of Venice with his detailed drawings and enthusiastic guide for the readers on what to appreciate and value in future visits. However, the expressions Ruskin used can be subjective and reflect his personal tastes, which is evident in his objection against the "colourless" Renaissance buildings. Like all other visitors, the words of Ruskin describing Venice were regarded as a myth, a fiction, and a symbolic landscape, reflecting his own imagination of this idealized city (Cosgrove, 1982; Psarra, 2018). Turning the argument around, like Ruskin, all the other visitors and residents in Venice are also qualified to express the values the city conveys to them. Psarra (2018) argues that

> "[a]ny effort to describe Venice runs the risk of confusing the city with the words and the images that describe it",

bringing up another question about what these "words and images" really are about.

The modern era of Social Media has given more opportunities and challenges to the process of collecting and mapping cultural significance from the perspective of general public. This is because social media has made possible the open publication of ideas, opinions, and emotions by everyone among the online communities with their own "words and images" (Cartwright, 2010). Like the pieces of stones observed by Ruskin, those posts on social media could be understood as "digital notes of stones" to be screened and inspected to dig valuable messages. Analysing such massive data can help collect information on the cultural significance (i.e., the values of cultural heritage embodied in the places for all generations) conveyed to the general public, map knowledge from alternative perspectives other than the expert-based authorized heritage discourse, and construct an inclusive heritage management plan respecting the collective opinions (Aggarwal, 2011; ICOMOS, 2013; Amato et al., 2016; Bai et al., 2021b; Bigne et al., 2021). This aligns well with the goals and objectives set by the 2011 UNESCO Recommendation on the Historic Urban Landscape (HUL) (UNESCO, 2011; Bandarin and Van Oers, 2012; Pereira Roders, 2019). Among all the information and knowledge to be extracted and mapped, heritage values (why to conserve) and heritage attributes (what to conserve) are arguably the most informative ones to fully understand the cultural significance of a heritage property, being listed or not, e.g., see Pereira Roders (2007); Tarrafa Silva and Pereira Roders (2010); Veldpaus (2015). Ginzarly et al. (2019) demonstrates an example in this line to map the HUL values revealed on Flickr by manually checking the post contents. In the past decades, the advances in Machine Learning (ML) and Deep Learning (DL), especially Multi-modal Machine Learning focusing on fusing information from different modalities (such as texts and images), have enabled similar analyses at larger scales (LeCun et al., 2015; Baltrusaitis et al., 2019; Cao et al., 2020). In order to extract and map the most representative categories of descriptions and/or images of a place, earlier studies constructed textual and visual information from social media posts with hand-crafted or learned features (Crandall et al., 2009; Monteiro et al., 2014; Huang and Li, 2016; Lai et al., 2017; Boy and Uitermark, 2017), while recent studies have been updating the process with neural network models pre-trained on generic tasks for generalizable results (Gomez et al., 2019; Zhang et al., 2019b; Kang et al., 2021; Bai et al., 2022; Cho et al., 2022; He et al., 2022; Wang et al., 2022a; Zhang et al., 2022b).

However, two challenges remain for the approach of mapping cultural significance to be broadly applied in heritage and urban studies:

- the raw user-generated data collected from social media are usually hard to annotate especially when the labels need complex expert knowledge;
- the time-stamped and geo-tagged posts are usually scattered in space, which need to be further aggregated and summarized into higher-level spatial units, resulting in maps that are comprehensible by planners and decision-makers.

Since social media posts are embedded in socioeconomic and spatiotemporal contexts (i.e., in explicit or intrinsic graph structures denoting the connections of posts such as located in nearby places, posted in consecutive time periods, and owned by similar social groups), both challenges can be handled with the emerging fields of Semi-supervised Machine Learning on Graphs with Graph Neural Network (GNN) (Zhang and Cheng, 2020; Ma and Tang, 2021; Wu et al., 2022; Xu et al., 2022). Different from conventional supervised learning, semi-supervised learning models also have access to features from unlabelled data during training process without knowing their “true” labels (Zhou and Li, 2010). This is proved to be effective especially on graphs since neighbours on graphs are assumed to be similar both in the feature space and the label space (Zhu and Ghahramani, 2002; Kipf and Welling, 2016; Xu et al., 2022). With spatial data in physical space, such similarity is expressed as the rule of the First Law of Geography (Tobler, 1970), that nearby things are generally similar to, and therefore, more likely to influence each other.

This paper aims to explore the use of graph-based semi-supervised classification to spatially map the cultural significance categories of cities with multi-modal social media data embedded in a graph structure. To reach the aim, three research questions are explored, becoming the three main components of the workflow proposed in this paper:

1. How can graph-based semi-supervised classification help to classify a partially labelled multi-modal social media dataset concerning location-based categorical information in a city?
2. How can an ensemble of trained models help to further improve classification performance and reliability?
3. How can the labels assigned for the posts be aggregated onto the spatial network of a city in order to map the categorical information (the perceived cultural significance)?

The scope and the approach of this study are highly related to Liu and De Sabbata (2021), where the authors presented a framework for using GNN to classify multi-modal features into user-defined label sets. Whereas Liu and De Sabbata (2021) focused on exploring the effects of different graph construction methods for only one specific type of GNN model (i.e., Graph Convolutional Network) and the mapping procedure was only a showcase of randomly sampled scatter points without further spatial aggregation and application analyses, this study has the following further contributions:

- A few Deep Learning models are trained on a semi-supervised classification task about cultural significance with partially labelled multi-modal graph-based datasets, and the soft-label predictions of individual models were aggregated into ensemble results, keeping track of the confidence and agreement of the models, as a measure of reliability;
- The obtained post labels are further aggregated into spatial nodes and diffused on a spatial network based on the geographical/topological proximity, effectively summarizing the information into a set of spatial maps for cultural significance categories;
- Detailed analyses on the spatial and aspatial distributions of the cultural significance categories, as well as the association of input features and output categories are provided, informative for future inclusive heritage management processes.

The workflow demonstrated in this paper with the special case of heritage cultural significance can be easily generalized in other use cases for spatially diffusing and mapping any human-generated features and labels, which can be extended to the evaluation of spatial safety, vitality, and/or architectural style in urban spaces (Cheng and Wicks, 2014; Sun et al., 2022; Zhang et al., 2022a).

5.2 Data and Materials

5.2.1 Case Study: Venice

To relate to the metaphor of the title and its relationship with Ruskin's controversial masterpiece The Stones of Venice (Ruskin, 1879; Ruskin and Quill, 2015), this study selects Venice as a case study to test the methodological framework. Venice and its Lagoon was inscribed in the UNESCO World Heritage List in 1987 fulfilling all first six selection criteria of Outstanding Universal Value (OUV) related to cultural heritage (UNESCO, 1972, 2008; Jokilehto, 2007). Despite its status as a cultural heritage property, its special urban typology and intimate relationship with the water give the city strong clues of natural values (Bai et al., 2022), making it a popular tourism destination of diverse interests, which also means that it may suffer from the mass-tourism (Urry and Larsen, 2011; Bertocchi and Visentin, 2019). Meanwhile, Venice can be found in various academic publications and non-academic fictions, as well as voluntary comments on social media platforms, providing abundant information from all sorts of perspectives (Calvino, 1978; Cosgrove, 1982; Bigne et al., 2021). The city itself is also a product of top-down conscious city planning (state-craft) and bottom-up collective community building (city-craft) (Psarra, 2018), both firmly embedded in a spatiotemporal and socioeconomic context. All these characteristics make Venice a representative case study to demonstrate the utility of the proposed framework. Yet, it is also important to notice that the selection of Venice as the case study is only a pragmatic choice, and hypothetically the framework should be generalizable in other cities with urban areas inscribed in the UNESCO WHL, similar to Psarra's argument, that Venice could be considered as a prototype of other global cities (Psarra, 2018).

5.2.2 Data Usage

This study uses the open datasets Heri-Graphs-Venice (VEN) and Venice-Large (VEN-XL) introduced by Bai et al. (2022), where multi-modal information from the social media platform Flickr is collected, containing visual and textual features, temporal, social, and spatial contexts (as a multi-graph), as well as partially-labelled pseudo-labels for cultural significance categories based on model confidence. In their definition, cultural significance was specified with two concepts as soft labels, effectively providing two probability distribution vectors: an 11-class OUV selection criteria (referred to from here on as OUV for simplicity) category (UNESCO, 1972, 2008; Jokilehto, 2008; Bai et al., 2021a), and a 9-class heritage attributes (HA) category (Veldpaus, 2015; Gustcoven, 2016; Ginzarly et al., 2019), both listed in Table 5.2. Since Flickr is an image-sharing platform and textual information is not mandatory during posting, both datasets collected therefrom were better equipped with visual features as 982-dimensional stacked vectors of a few pre-trained model outputs, and only about half of data samples contained valid BERT-based textual features as 771-dimensional vectors.

Within the two datasets, the lite version VEN was already formatted as a multi-graph with three types of undirected weighted links (temporal, social, and spatial) showing the contextual connections among the nodes representing posts on Flickr. However, the larger version VEN-XL was only provided with the nodal features because of the large memory requirement to construct adjacency matrices with a huge number of nodes. Following the guidelines given by Bai et al. (2022), this paper also constructed multi-graph mini-batches for VEN-XL in Pytorch-Geometric library (Fey and Lenssen, 2019) using sparse matrices as graph structure (Yuster and Zwick, 2005). An overview of both datasets is given in Table 5.1. The label rates (.122/.143) of the datasets are comparable with common semi-supervised learning datasets in graph neural networks such as Citeer (.036) and Cora (.052) (Kipf and Welling, 2016; Yang et al., 2016). Note VEN-XL has a larger average degree for nodes with all types of links, yet the multi-graphs are less dense than the lite VEN dataset.

TABLE 5.1 Descriptive overview of the data used for this study previously collected by Bai et al. (2022)

Dataset	VEN			VEN-XL		
	Count	Rate/Proportion		Count	Rate/Proportion	
Nodes	2951	-		80,963	-	
Nodes with Visual Features	2951	100%		80,963	100%	
Nodes with Textual Features	1761	59.7%		49,823	61.5%	
Nodes with OUV Selection Criteria Labels	756	25.6%		25,771	31.8%	
Nodes with Heritage Attribute Labels	1356	45.9%		37,289	46.1%	
Nodes with Both Types of Labels	361	12.2%		11,569	14.3%	
	Count	Average Degree	Density	Count	Average Degree	Density
Temporal Links	249,120	84.4	.057	35,527,354	438.8	.011
Social Links	242,576	82.2	.056	38,170,651	471.5	.012
Spatial Links	221,414	75.0	.051	101,046,098	1248.1	.031
Simple Composed Links*	534,513	181.1	.123	145,005,270	1791.0	.044

*Multiple links among two nodes leads to only one link in the simple composed graph.

As a summary, the datasets in this study have three challenges for the semi-supervised classification task: 1) only partial labels are available for the categories of interest, requiring the unlabelled nodes to be tagged; 2) only partial features are available for some nodes, requiring the models to learn as much as possible from their neighbours on graphs; 3) the VEN-XL dataset is too large to conduct training and inference directly, requiring sampling of subgraphs. All these characteristics of the datasets entail that both transductive (training and inference on the same graph) and inductive (inference on unseen [sub-] graphs) semi-supervised learning on graphs (Yang et al., 2016; Liu and De Sabbata, 2021) are indispensable, reflecting the scope and necessity of this study. For both datasets, the nodes with both types of labels (OUV and HA) are treated as the training sets (361 for VEN; 11,569 for VEN-XL), and the nodes with only one type of labels are randomly and evenly separated as validation sets (695; 19,961) and test sets (695; 19,961), while the remainder of the nodes is considered as unlabelled data (1200; 29,472). In the training sets, all essential categories are present, though the distribution is unbalanced, as presented in Table 5.2.

TABLE 5.2 The distribution of cultural significance categories as OUV selection criteria and heritage attributes in the training sets.

Dataset	VEN	VEN-XL
Categories of OUV Selection Criteria	**(361)**	**(11,569)**
(within top-3 entries)		
Criterion (i) - Masterpiece	172 (15.9%)	2463 (7.1%)
Criterion (ii) - Influence	188 (17.4%)	4704 (13.6%)
Criterion (iii) - Testimony	247 (22.8%)	9864 (28.4%)
Criterion (iv) - Typology	261 (24.1%)	8578 (24.7%)
Criterion (v) - Land-use	7 (0.6%)	54 (0.2%)
Criterion (vi) - Association	205 (18.9%)	8921 (25.7%)
Criterion (vii) - Natural Beauty	1 (0.1%)	58 (0.2%)
Criterion (viii) - Geological Process	0 (0.0%)	18 (0.1%)
Criterion (ix) - Ecological Process	1 (0.1%)	19 (0.1%)
Criterion (x) - Bio-diversity	1 (0.1%)	28 (0.1%)
Others - Not related	0 (0.0%)	0 (0.0%)
Categories of Heritage Attributes	**(361)**	**(11,569)**
(within top-1 entries)		
Monument and Buildings	69 (19.1%)	1507 (13.0%)
Building Elements	71 (19.7%)	1501 (13.0%)
Urban Form Elements	101 (28.0%)	2636 (22.8%)
Urban Scenery	6 (1.7%)	113 (1.0%)
Natural Features and Landscape Scenery	30 (8.3%)	2051 (17.7%)
Interior Scenery	25 (6.9%)	480 (4.1%)
People's Activity and Association	49 (13.6%)	2457 (21.2%)
Gastronomy	9 (2.5%)	139 (1.2%)
Artifact Products	1 (0.3%)	685 (5.9%)

5.2.3 General Notations

Most of the notations in this chapter are consistent with that in Chapter 4, with a certain level of simplification. Since the data structure is exactly the same for VEN and VEN-XL except for the sample size, this section will describe the general notation system eligible for both datasets. For each dataset, an undirected multi-graph $\mathcal{G} = (\mathcal{V}, \{\mathcal{E}^{\text{TEM}}, \mathcal{E}^{\text{SPA}}, \mathcal{E}^{\text{SOC}}\})$ with three types of links (temporal, spatial, and social, as mentioned in Section 5.2.2) represents its contextual structure, where $\mathcal{V} = \{v_i\}$, $i \in [0, K)$ is the node set of all the posts collected and K is the total number of posts, and $(v_i, v_{i'}) \in \mathcal{E}^{(*)} \subseteq \mathcal{V} \times \mathcal{V}, \mathcal{E}^{(*)} \in \{\mathcal{E}^{\text{TEM}}, \mathcal{E}^{\text{SPA}}, \mathcal{E}^{\text{SOC}}\}$ is a link marking one type of contextual relations among the posts. For simplicity, the link weights in Bai et al. (2022) are omitted, resulting in binary adjacency matrices $\boldsymbol{A}^{(*)} := [A^{(*)}_{i,i'}] \in \{0,1\}^{K \times K}$, $\boldsymbol{A}^{(*)} \in \{\boldsymbol{A}^{\text{TEM}}, \boldsymbol{A}^{\text{SPA}}, \boldsymbol{A}^{\text{SOC}}\}$, where all the links $(v_i, v_{i'})$ with an original weight larger than 0 will lead to $A^{(*)}_{i,i'} = 1$, otherwise $A^{(*)}_{i,i'} = 0$. Moreover, a simple composed graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E})$ could be obtained by merging the adjacency matrices into $\boldsymbol{A}$, so that $\boldsymbol{A} = \left(\boldsymbol{A}^{\text{TEM}} > 0\right) \bigvee \left(\boldsymbol{A}^{\text{SPA}} > 0\right) \bigvee \left(\boldsymbol{A}^{\text{SOC}} > 0\right) \in \{0,1\}^{K \times K}$. In this simple composed graph $\mathcal{G}'$, a link would exist if at least one contextual type of links exists between two nodes in the multi-graph $\mathcal{G}$.

For all the nodes in the graph $\mathcal{G}$, a 2D feature array $\boldsymbol{X} := [\boldsymbol{x}_i]_{i \in [0,K)} = \begin{bmatrix} \boldsymbol{X}^{\text{vis}} \\ \boldsymbol{X}^{\text{tex}} \end{bmatrix} \in \mathbb{R}^{1753 \times K}$ would exist, where $\boldsymbol{x}_i \in \mathbb{R}^{1753 \times 1}$ is a vector representing the features of node v_i, $\boldsymbol{X}^{\text{vis}} \in \mathbb{R}^{982 \times K}$, $\boldsymbol{X}^{\text{tex}} \in \mathbb{R}^{771 \times K}$ are respectively the visual and textual features, and $\begin{bmatrix} \cdot \\ \cdot \end{bmatrix}$ is the vertical concatenation operation of arrays. In cases where no textual data was available for a post node, the corresponding entries in vector $\boldsymbol{x}_i$ would be all zeros, dividing the nodes $\mathcal{V}$ into two sub-clusters $\mathcal{V}_{\text{tex+}}, \mathcal{V}_{\text{tex-}} \subset \mathcal{V}$, with or without textual data.

Since pseudo-labels for posts were respectively provided for a different subset of $\mathcal{V}$ concerning OUV and HA, four sub-clusters $\mathcal{V}_{\text{V+,A+}}, \mathcal{V}_{\text{V+,A-}}, \mathcal{V}_{\text{V-,A+}}, \mathcal{V}_{\text{V-,A-}} \subset \mathcal{V}$ could be categorized, as they have different label arrays:

- For nodes with both labels in $\mathcal{V}_{\text{V+,A+}}$, the label array would be $\boldsymbol{Y}_{\text{V+,A+}} = \begin{bmatrix} \boldsymbol{y}_i^{\text{OUV}} \\ \\ \boldsymbol{y}_i^{\text{HA}} \end{bmatrix}_{v_i \in \mathcal{V}_{\text{V+,A+}}}$, where $\boldsymbol{y}_i^{\text{OUV}} \in [0,1]^{11 \times 1}, \boldsymbol{y}_i^{\text{HA}} \in [0,1]^{9 \times 1}$ are respectively a column-stochastic vector denoting the soft labels of node v_i for OUV and HA categories;
- For nodes with only OUV labels in $\mathcal{V}_{\text{V+,A-}}$, the label array would be $\boldsymbol{Y}_{\text{V+,A-}} = \left[\boldsymbol{y}_i^{\text{OUV}}\right]_{v_i \in \mathcal{V}_{\text{V+,A-}}}$;
- For nodes with only HA labels in $\mathcal{V}_{\text{V-,A+}}$, the label array would be $\boldsymbol{Y}_{\text{V-,A+}} = \left[\boldsymbol{y}_i^{\text{HA}}\right]_{v_i \in \mathcal{V}_{\text{V-,A+}}}$;
- For nodes with in $\mathcal{V}_{\text{V-,A-}}$, there is no label array.

Note the following relationship holds for the sub-clusters:

$$\begin{aligned}(\mathcal{V}_{\text{V+,A+}} \cup \mathcal{V}_{\text{V+,A-}}) &\subset \mathcal{V}_{\text{tex+}},\\ (\mathcal{V}_{\text{V-,A+}} \cup \mathcal{V}_{\text{V-,A-}}) \cap \mathcal{V}_{\text{tex+}} &\neq \emptyset,\\ (\mathcal{V}_{\text{V-,A+}} \cup \mathcal{V}_{\text{V-,A-}}) \cap \mathcal{V}_{\text{tex-}} &\neq \emptyset, \end{aligned} \tag{5.1}$$

meaning that having textual features as input is a necessary but not sufficient condition of having the OUV label.

5.3 Problem Definition

The workflow proposed in this paper is visualized in Figure 5.1. The input data from two databases VEN and VEN-XL are:

- a partially-labelled attributed multi-graph about the inter-related social media posts;
- an assignment bipartite graph with relations mapping the posts to their closest street intersections (spatial nodes);
- a topological representation of the spatial network as a weighted undirected graph marking the proximity of the street intersections.

After three main components, i.e.,

1. semi-supervised learning of multiple models co-trained in a classification task (Section 5.3.1),
2. aggregating the prediction outputs as soft labels of those models (Section 5.3.2),
3. aggregating and diffusing the post-level labels on the spatial graph (Section 5.3.3),

two outputs are obtained

- a graph fully-labelled on all post-level nodes together with confidence and agreement scores based on model performance;
- a graph fully-labelled on spatial-level nodes summarizing the information of nearby posts and proximate spatial neighbours.

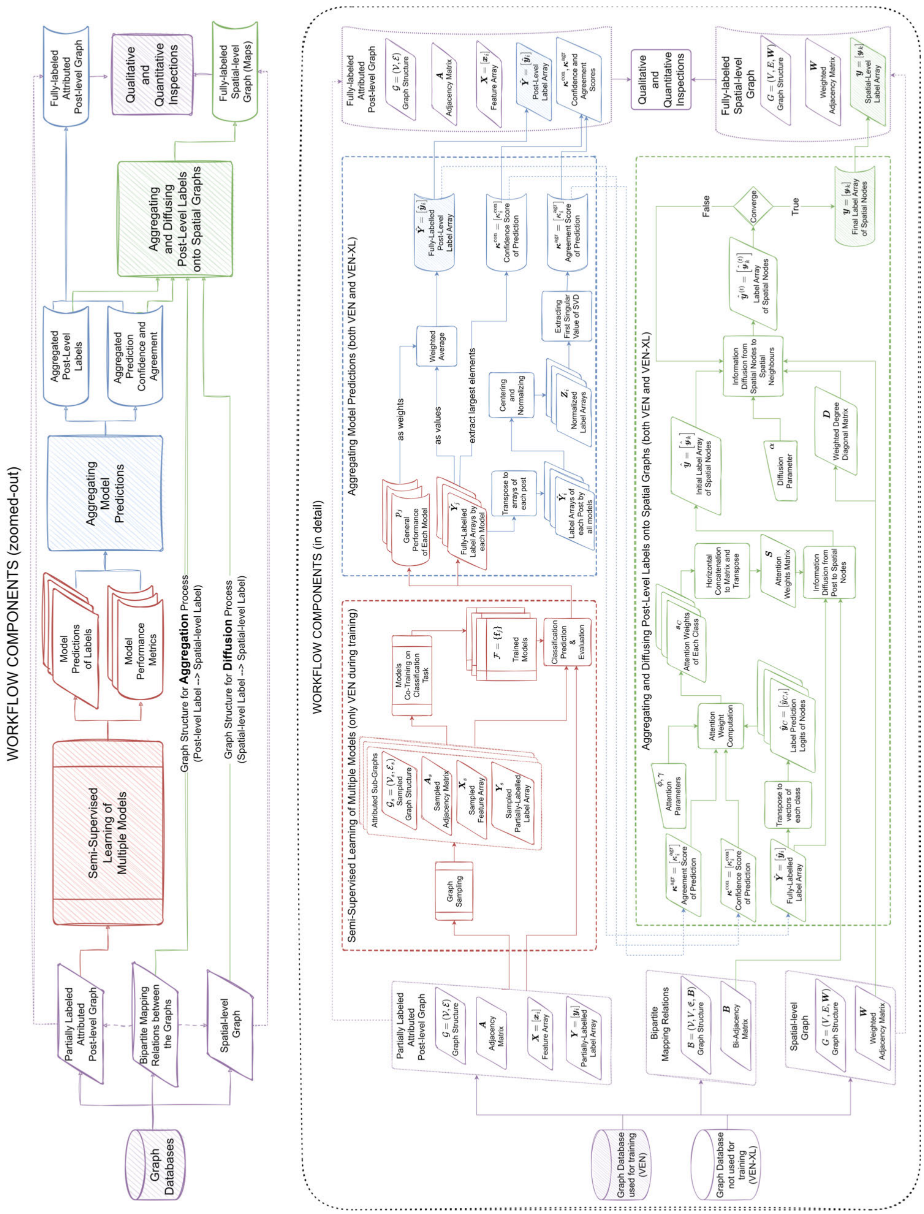


FIG. 5.1 The general methodological workflow proposed in this paper, both as zoomed-out high-level modulated framework in the upper part, and as a detailed workflow with mathematical notations in the lower part to be instantiated in the texts. Only the lite dataset VEN is used to train the models in the first step of semi-supervised learning, while the large dataset VEN-XL is directly used for inference and later steps. The indices i, j, k are respectively a generic example of the posts $v_i \in \mathcal{V}$, the models $\mathbf{f}_j \in \mathcal{F}$, and the spatial intersection nodes $\nu_k \in V$.

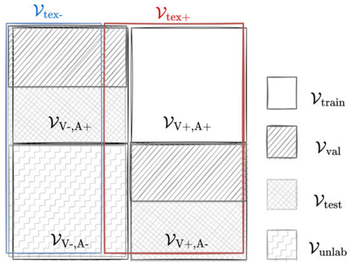


FIG. 5.2 The Venn Diagram showing the logic relations of the three types of sub-clustering of nodes in $\mathcal{V}$. The relationship described in Equation (5.1) and (5.2) are visualized.

Both outcomes are tested with qualitative and quantitative inspections (Section 5.6). The graph structures are conceptually visualized in Figure 5.3. The process will be formally described in the following Sections. The relevant works concerning the proposed workflow will be discussed in Section 5.7.4.

5.3.1 Semi-Supervised Training on Sampled Graphs

As described in Section 5.2.2, the nodes in $\mathcal{V}$ are further split into training set $\mathcal{V}_{\text{train}}$, validation set $\mathcal{V}_{\text{val}}$, test set $\mathcal{V}_{\text{test}}$, and unlabelled set $\mathcal{V}_{\text{unlab}}$, where:

$$\begin{aligned}
&\mathcal{V}_{\text{train}} = \mathcal{V}_{\text{V+,A+}}, \\
&\mathcal{V}_{\text{unlab}} = \mathcal{V}_{\text{V-,A-}}, \\
&\mathcal{V}_{\text{val}} \cup \mathcal{V}_{\text{test}} = \mathcal{V}_{\text{V+,A-}} \cup \mathcal{V}_{\text{V-,A+}}, \\
&|\mathcal{V}_{\text{val}}| = |\mathcal{V}_{\text{test}}|.
\end{aligned} \tag{5.2}$$

The semi-supervised learning task in this paper is to use the training nodes $\mathcal{V}_{\text{train}}$ and teach a group of models to learn the mapping functions within a candidate model set $\mathcal{F} = \{\mathbf{f}_j\}, j \in [0, |\mathcal{F}|)$ from input features $\boldsymbol{X}$ to output labels $\boldsymbol{Y}$, tune the hyper-parameters and select the optimal models based on their performance on the validation nodes $\mathcal{V}_{\text{val}}$, evaluate the generalizability of the models on unseen test data on $\mathcal{V}_{\text{test}}$, and apply the trained models to generate predicted soft labels $\hat{\boldsymbol{Y}} = [\hat{\boldsymbol{y}}_i]_{v_i \in \mathcal{V}}$ for all nodal data including the ones in $\mathcal{V}_{\text{unlab}}$. The logic relations among the three types of clustering of the node set $\mathcal{V}$ mentioned in Equations (5.1) and (5.2) are illustrated in the Venn Diagram of Figure 5.2.

For both efficiency and generalizability, sub-graphs are strategically sampled from the original graphs to train the models: $\mathcal{G}_s = (\mathcal{V}_s, \{\mathcal{E}_s^{\text{TEM}}, \mathcal{E}_s^{\text{SPA}}, \mathcal{E}_s^{\text{SOC}}\})$ or $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ with respectively sampled adjacency matrices $\boldsymbol{A}_s^{(*)}, \boldsymbol{A}_s$ and feature array $\boldsymbol{X}_s$, where $\mathcal{V}_s \subseteq \mathcal{V}, \mathcal{E}_s \subseteq \mathcal{E}, \mathcal{E}_s^{(*)} \subseteq \mathcal{E}^{(*)}$, depending on if the models would use the multi-graph structure or the simple composed one. For each training epoch, non-repetitive mini-batches of nodes $\mathcal{V}_{\text{batch}} \subset \mathcal{V}_s$ are used as base nodes to sample several different sub-graphs $\mathcal{G}_s$. Then the training loss $\mathcal{L}_{\text{train}}$ of any model $\mathbf{f}_j$ with model parameter $\boldsymbol{\Theta}_j$

for each mini-batch $\mathcal{V}_{\text{batch}}$ could be described as:

$$\mathcal{L}_{\text{train}}(\boldsymbol{\Theta}_j, \mathcal{V}_{\text{batch}}) = \sum_{v_i \in \mathcal{V}_{\text{batch}} \cap \mathcal{V}_{\text{train}}} \left(\ell(\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}}, \boldsymbol{y}_i^{\text{OUV}}) + \omega_{\text{V/A}} \ell(\hat{\boldsymbol{y}}_{j,i}^{\text{HA}}, \boldsymbol{y}_i^{\text{HA}}) \right), \tag{5.3}$$

$$\hat{\boldsymbol{y}}_{j,i} := \begin{bmatrix} \hat{\boldsymbol{y}}_{j,i}^{\text{OUV}} \\ \\ \hat{\boldsymbol{y}}_{j,i}^{\text{HA}} \end{bmatrix} = \begin{bmatrix} \text{softmax}(\boldsymbol{z}_{j,i}^{\text{OUV}}) \\ \\ \text{softmax}(\boldsymbol{z}_{j,i}^{\text{HA}}) \end{bmatrix}, \tag{5.4}$$

$$\mathbf{1}_{11\times 1}^{\mathsf{T}} \hat{\boldsymbol{y}}_{j,i}^{\text{OUV}} = \mathbf{1}_{9\times 1}^{\mathsf{T}} \hat{\boldsymbol{y}}_{j,i}^{\text{HA}} = 1, \tag{5.5}$$

$$\text{and } \boldsymbol{z}_{j,i} := \begin{bmatrix} \boldsymbol{z}_{j,i}^{\text{HV}} \\ \\ \boldsymbol{z}_{j,i}^{\text{HA}} \end{bmatrix} = \mathbf{f}_j(\boldsymbol{A}_s, \boldsymbol{X}_s; \boldsymbol{\Theta}_j)_i, \tag{5.6}$$

where ℓ is a loss function comparing the similarity of two vectors, such as cross-entropy (Rubinstein and Kroese, 2013), $\omega_{\text{V/A}}$ is a scalar parameter balancing the importance of OUV and HA categories during training, $\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}} \in [0,1]^{11\times 1}$, $\hat{\boldsymbol{y}}_{j,i}^{\text{HA}} \in [0,1]^{9\times 1}$ are respectively predicted stochastic label vectors for OUV and HA by the j_{th} model on the i_{th} example, and $\boldsymbol{z}_{j,i}^{\text{OUV}} \in \mathbb{R}^{11\times 1}$, $\boldsymbol{z}_{j,i}^{\text{HA}} \in \mathbb{R}^{9\times 1}$ are respectively components of the model output vector $\boldsymbol{z}_{j,i} \in \mathbb{R}^{20\times 1}$. Notice that the two objectives of classifying OUV and HA are trained together with a shared model architecture and are only distinguished before final loss computation, instead of having two separate models. This is assumed to be more generalizable and could capture more information on the associations between the two closely-related topics.

While evaluating the model performance on validation set $\mathcal{V}_{\text{val}}$ (and eventually on test set $\mathcal{V}_{\text{test}}$), the computation of the scores $\mathcal{L}_{\text{val}}^{\text{OUV}}$ and $\mathcal{L}_{\text{val}}^{\text{HA}}$ respectively on OUV and HA categories would be further distinguished as:

$$\mathcal{L}_{\text{val}}^{\text{OUV}}(\boldsymbol{\Theta}_j) = \frac{\sum_{\mathcal{V}_{\text{batch}} \subset \mathcal{V}_{\text{val}}} \sum_{v_i \in \mathcal{V}_{\text{batch}} \cap \mathcal{V}_{\text{V+,A-}}} \ell_{\text{V}}(\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}}, \boldsymbol{y}_i^{\text{OUV}})}{|\mathcal{V}_{\text{val}} \cap \mathcal{V}_{\text{V+,A-}}|} \tag{5.7}$$

$$\mathcal{L}_{\text{val}}^{\text{HA}}(\boldsymbol{\Theta}_j) = \frac{\sum_{\mathcal{V}_{\text{batch}} \subset \mathcal{V}_{\text{val}}} \sum_{v_i \in \mathcal{V}_{\text{batch}} \cap \mathcal{V}_{\text{V-,A+}}} \ell_{\text{A}}(\hat{\boldsymbol{y}}_{j,i}^{\text{HA}}, \boldsymbol{y}_i^{\text{HA}})}{|\mathcal{V}_{\text{val}} \cap \mathcal{V}_{\text{V-,A+}}|}, \tag{5.8}$$

where ℓ_{V} and ℓ_{A} are topic-specific evaluation metrics for both classification tasks which will be introduced in Section 5.4.3. For each batch $\mathcal{V}_{\text{batch}} \subset \mathcal{V}_{\text{val}}$, a new sample sub-graph $\mathcal{G}_s$ is used to compute the soft labels $\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}}, \hat{\boldsymbol{y}}_{j,i}^{\text{HA}}$.

5.3.2 Aggregating Prediction Outputs

Assume the semi-supervised learning process mentioned in Section 5.3.1 trains all models in $\mathcal{F} = \{\mathbf{f}_j\}$ properly and they generate a set of well-fit label arrays

$\left\{\hat{\boldsymbol{Y}}_j := \left[\hat{\boldsymbol{y}}_{j,i}\right]_{v_i \in \mathcal{V}}\right\}_{\mathsf{f}_j \in \mathcal{F}}$, where $\hat{\boldsymbol{Y}}_j \in [0,1]^{20 \times K}$ is the predicted label array on the entire dataset $\mathcal{V}$ by the model $\mathbf{f}_j$. Practice in ensemble learning has shown that a group of trained models would usually perform better than an individual model and could yield more reliable predictions (Zhou, 2012). Therefore, this study considers a soft voting mechanism to conclude the final node labels $\hat{\boldsymbol{Y}} := [\hat{\boldsymbol{y}}_i]_{v_i \in \mathcal{V}}$, $\hat{\boldsymbol{Y}} \in [0,1]^{20 \times K}$, such that: $\hat{\boldsymbol{y}}_i = (\sum_{\mathsf{f}_j \in \mathcal{F}} p_j \hat{\boldsymbol{y}}_{j,i})/(\sum_{\mathsf{f}_j \in \mathcal{F}} p_j)$, or in the matrix form, $\hat{\boldsymbol{Y}} = (\sum_{\mathsf{f}_j \in \mathcal{F}} p_j \hat{\boldsymbol{Y}}_j)/(\sum_{\mathsf{f}_j \in \mathcal{F}} p_j)$, where $\hat{\boldsymbol{Y}}$ is a weighted average of the label arrays by all models whose column-sum pertains 2 for each post, and the weight p_j is the general performance score (e.g., accuracy, which will be discussed in Section 5.4.3) of model $\mathbf{f}_j$ on validation set.

Furthermore, the confidence of model prediction and the agreement/coherence among the different models also provide information for the reliability of the predictions (Zhou and Li, 2010). The former is trivial as the model confidence on all data points $\boldsymbol{\kappa}^{\mathsf{con}} := [\kappa_i^{\mathsf{con}}] \in [0,1]^{K \times 1}$ could be defined as the sum of top-n entries of the label vectors divided by two (since the sum of each label vector $\hat{\boldsymbol{y}}_i$ is two, as defined in Equation (5.4)). The latter is also trivial when only two models are concerned since the agreement of two vectors could be easily computed with any distance measure (e.g., cosine similarity, Euclidean distance, Jaccard Index, and/or cross-entropy). When $|\mathcal{F}| > 2$, this becomes a problem of measuring the general linear dependence of a group of vectors composing the array $\hat{\boldsymbol{Y}}_i := \left[\hat{\boldsymbol{y}}_{j,i}\right]_{\mathsf{f}_j \in \mathcal{F}}$, $\hat{\boldsymbol{Y}}_i \in [0,1]^{20 \times |\mathcal{F}|}$ for each node v_i. Inspired by GeoMatt22 (2020), this study computes the model agreement $\boldsymbol{\kappa}^{\mathsf{agr}} =: \left[\kappa_i^{\mathsf{agr}}\right] \in [0,1]^{K \times 1}$ from the first singular value $\sigma_{\boldsymbol{Z}_i,1}$ of the centred (subtracted by row-means) and normalized (divided by vector lengths) label matrix $\boldsymbol{Z}_i := [\boldsymbol{z}_{j,i}/\|\boldsymbol{z}_{j,i}\|]_{\mathsf{f}_j \in \mathcal{F}}$, $\boldsymbol{z}_{j,i} = \hat{\boldsymbol{y}}_{j,i} - \sum_j \hat{\boldsymbol{y}}_{j,i}/|\mathcal{F}|$ based on its Singular Value Decomposition (SVD) results, so that:

$$\kappa_i^{\mathsf{agr}} = \frac{\sigma^2_{\boldsymbol{Z}_i,1} - 1}{|\mathcal{F}| - 1}. \tag{5.9}$$

This is effective since the first several singular values measure how much variance of the matrix could be explained by its low-rank approximation, which is equivalent to eigenvalues in Principal Component Analysis (PCA) in statistics. The value of $\boldsymbol{\kappa}^{\mathsf{agr}}$ ranges theoretically from the largest possible value (i.e., 1) when there are $|\mathcal{F}|$ completely parallel vectors in $\boldsymbol{Z}_i$, to the smallest possible value (i.e., 0) when all vectors are orthogonal (under the condition that $|\mathcal{F}| < 20$).

5.3.3 Spatial Diffusion of Node Labels

In order to map the predicted node labels on the topological/ geographical space, the label array $\hat{\boldsymbol{Y}}$ computed in Section 5.3.2 is further aggregated spatially, going one step further than the research conducted in Liu and De Sabbata (2021), where the labels of individual post nodes were directly drawn on maps. In Bai et al. (2022), the mapping relations of the posts to spatial nodes are also provided. For a city, an undirected weighted graph $G = (V, E, \boldsymbol{W})$ denotes its geographical representation obtained from Open Street Map (Boeing, 2017), where $V = \{\nu_k\}, k \in [0, |V|)$ is the node set of spatial intersections in a walkable network, $(\nu_k, \nu_{k'}) \in E \subseteq V \times V$ is a link marking if two spatial nodes are reachable to each other within 20 minutes by all means of transportation, and $\boldsymbol{W} := [W_{k,k'}] \in [0,1]^{|V|\times|V|}$ is a non-negative weighted adjacency matrix whose diagonal entries $W_{k,k}$ are all 1, recording the temporal closeness (i.e., the shorter time it takes to travel, the closer this weight gets to 1) between any pair of nodes ν_k and $\nu_{k'}$, where $W_{k,k'} = 0$ when the nodes are not connected (not reachable within 20 minutes). Moreover, $\boldsymbol{B} := [B_{i,k}] \in \{0,1\}^{K\times|V|}$ records the one-hot mapping relation from posts nodes $\mathcal{V}$ to spatial nodes V, effectively a binary bi-adjacency matrix of a bipartite graph $\mathcal{B} = (\mathcal{V}, V, \mathscr{E}, \boldsymbol{B})$ connecting both node sets, where $(v_i, \nu_k) \in \mathscr{E} \subset \mathcal{V} \times V$ marks the link if a post is located nearby a spatial node. Note that the following relationship holds according to Bai et al. (2022): $\boldsymbol{A}^{\mathsf{SPA}} = \left(\boldsymbol{B}\boldsymbol{W}\boldsymbol{B}^{\mathsf{T}} > 0\right) = \boldsymbol{B}\left(\boldsymbol{W} > 0\right)\boldsymbol{B}^{\mathsf{T}} \in \{0,1\}^{K\times K}$.

Without loss of generality, the processes of spatially aggregating and diffusing the node labels are visualized in Figure 5.3, taking the neighbours of a generic spatial node ν_k in both the spatial graph G as $\mathcal{N}_G(\nu_k) := \{\nu_{k'} | (\nu_k, \nu_{k'}) \in E$ or $W_{k,k'} > 0\} \subset V$ and in the bipartite graph $\mathcal{B}$ as $\mathcal{N}_{\mathcal{B}}(\nu_k) := \{v_i | (v_i, \nu_k) \in \mathscr{E}$ or $B_{i,k} = 1\} \subset \mathcal{V}$. The procedure takes place in two consecutive steps:

- Aggregating the predicted soft labels of all the posts nearby a spatial node $\hat{\boldsymbol{Y}}_{\mathcal{N}_{\mathcal{B}}(\nu_k)} := [\hat{\boldsymbol{y}}_i]_{v_i \in \mathcal{N}_{\mathcal{B}}(\nu_k)}$ to get the spatial node label $\hat{\mathbf{y}}_k \in [0,1]^{20\times 1}$, forming a 2D array $\hat{\boldsymbol{\mathcal{Y}}} := [\hat{\mathbf{y}}_k], \hat{\boldsymbol{\mathcal{Y}}} \in [0,1]^{20\times|V|}$;
- Diffusing the labels of all the spatial nodes to their spatial neighbours $\hat{\boldsymbol{\mathcal{Y}}}_{\mathcal{N}_G(\nu_k)} := [\hat{\mathbf{y}}_{k'}]_{\nu_{k'} \in \mathcal{N}_G(\nu_k)}$ based on their proximity iteratively, and vice versa, to get the final label $\mathbf{y}_k \in [0,1]^{20\times 1}$, with the label array $\boldsymbol{\mathcal{Y}} := [\mathbf{y}_k], \boldsymbol{\mathcal{Y}} \in [0,1]^{20\times|V|}$.

For the first step, the aggregation process should consider not only the respective values of the neighbouring labels, but also their importance (how dominant is the value compared to all the other nodes), prediction confidence (how confident are models predicting the label vectors containing this value) and prediction agreement (how reliable is this value). As it highly resembles the graph pooling operations in GNN, inspirations have been taken from literature (Li et al., 2015; Knyazev et al., 2019; Lee et al., 2019; Ma and Tang, 2021) to use an attention-based computation on each label category channel (as one instance among the 11 OUV or 9 HA categories) $\hat{\boldsymbol{y}}_C := \hat{\boldsymbol{Y}}^{\mathsf{T}}\boldsymbol{e}_C, \hat{\boldsymbol{y}}_C \in [0,1]^{K\times 1}$ to summarize the labels, where $\boldsymbol{e}_C \in \{0,1\}^{20\times 1}$ is a one-hot unit vector only marking its C_{th} entry as 1. The attention value $\boldsymbol{s}_C \in [0,1]^{K\times 1}$ of all nodes v_i for any label category channel C could be computed as:

$$\boldsymbol{s}_C = \frac{\exp\left(\hat{\boldsymbol{y}}_C \odot (\boldsymbol{\kappa}^{\mathsf{con}})^{1/\phi} \odot (\boldsymbol{\kappa}^{\mathsf{agr}})^{1/\gamma}\right)}{\mathbf{1}_{K\times 1}^{\mathsf{T}}\exp\left(\hat{\boldsymbol{y}}_C \odot (\boldsymbol{\kappa}^{\mathsf{con}})^{1/\phi} \odot (\boldsymbol{\kappa}^{\mathsf{agr}})^{1/\gamma}\right)}, \tag{5.10}$$

where $\boldsymbol{\kappa}^{\mathsf{con}}$ and $\boldsymbol{\kappa}^{\mathsf{agr}}$ are model-level confidence and agreement scores on each node computed in Section 5.3.2, $\phi, \gamma \in \mathbb{R}$ are respectively parameters to adjust the contribution of confidence and agreement in the attention computation, such that when they get larger, high values of $\boldsymbol{\kappa}$ will be pushed closer to 1, $\odot$ is an element-wise Hadamard multiplication of vectors and arrays, and $\mathbf{1}_{K\times 1}$ is a K-dimensional vector of all 1s. Note that $\boldsymbol{s}_C$ is a stochastic vector over all the nodes.

Concatenating vectors $\boldsymbol{s}_C^\mathsf{T}$ for all category channels vertically together, an attention-based weight matrix $\boldsymbol{S} \in [0,1]^{20\times K}$ is obtained. This is then used as the weight of label array $\hat{\boldsymbol{\mathcal{Y}}}$ during the aggregation operation:

$$\hat{\boldsymbol{\mathcal{Y}}}' := \begin{bmatrix} \hat{\boldsymbol{\mathcal{Y}}}^{\mathsf{OUV}}_{11\times|V|} \\ \\ \hat{\boldsymbol{\mathcal{Y}}}^{\mathsf{HA}}_{9\times|V|} \end{bmatrix} = \left(\left(\boldsymbol{S}\odot\hat{\boldsymbol{Y}}\right)\boldsymbol{B}\right) \oslash (\boldsymbol{S}\boldsymbol{B}),$$

$$\hat{\boldsymbol{\mathcal{Y}}} = \begin{bmatrix} \hat{\boldsymbol{\mathcal{Y}}}^{\mathsf{OUV}} \oslash \left(\mathbf{1}_{11\times 1}\mathbf{1}_{11\times 1}^{\mathsf{T}}\hat{\boldsymbol{\mathcal{Y}}}^{\mathsf{OUV}}\right) \\ \\ \hat{\boldsymbol{\mathcal{Y}}}^{\mathsf{HA}} \oslash \left(\mathbf{1}_{9\times 1}\mathbf{1}_{9\times 1}^{\mathsf{T}}\hat{\boldsymbol{\mathcal{Y}}}^{\mathsf{HA}}\right) \end{bmatrix} \tag{5.11}$$

where $\oslash$ is the element-wise Hadamard division of two arrays, and the outcome of any spatial node $\hat{\mathbf{y}}_k$ is effectively a special form of weighted-average of the label vectors of all its neighbours $\hat{\boldsymbol{Y}}_{\mathcal{N}_{\mathcal{B}}(\nu_k)}$, scaled differently by the attention matrix $\boldsymbol{S}$ on each label category channel C. Similar to $\hat{\boldsymbol{Y}}$, the array $\hat{\boldsymbol{\mathcal{Y}}}$ is also a stack of two column-stochastic arrays for the OUV and HA labels, respectively.

Once the initial spatial node labels $\hat{\boldsymbol{\mathcal{Y}}}$ are computed, they could be used as the input state of an iterative diffusion process at the second step, during which each spatial node obtains information from its spatial neighbours and updates its own label while being reminded of its original state, until the labels converge at a steady state. This process resembles the graph filtering operation in GNN (Hamilton et al., 2017; Ma and Tang, 2021; Wu et al., 2022). For each spatial node ν_k, its initial label is $\hat{\mathbf{y}}_k^{(0)} = \hat{\mathbf{y}}_k$. Assume the label is $\hat{\mathbf{y}}_k^{(t)}$ at the t_{th} iteration, then its next state after a diffusion step could be described as:

$$\hat{\mathbf{y}}_k^{(t+1)} = (1-\alpha)\hat{\mathbf{y}}_k + \alpha\frac{\sum_{\nu_{k'}\in\mathcal{N}_G(\nu_k)} W_{k,k'}\hat{\mathbf{y}}_{k'}^{(t)}}{\sum_{\nu_{k'}\in\mathcal{N}_G(\nu_k)} W_{k,k'}}, \tag{5.12}$$

or in its matrix form:

$$\hat{\boldsymbol{\mathcal{Y}}}^{(t+1)} = (1-\alpha)\hat{\boldsymbol{\mathcal{Y}}} + \alpha\hat{\boldsymbol{\mathcal{Y}}}^{(t)}\left(\boldsymbol{W}\boldsymbol{D}^{-1}\right), \tag{5.13}$$

where $\boldsymbol{D}$ is a diagonal matrix each entry of which records the degree (row-sum or column-sum) of the weighted symmetrical matrix $\boldsymbol{W}$, $\boldsymbol{W}\boldsymbol{D}^{-1}$ is the column-normalized stochastic matrix of $\boldsymbol{W}$, $\hat{\boldsymbol{\mathcal{Y}}}^{(t)} := \left[\hat{\boldsymbol{y}}_k^{(t)}\right] \in [0,1]^{20\times|V|}$ is the label array at the t_{th} iteration, and $\alpha \in [0,1)$ is a parameter controlling the importance of neighbouring nodes in the diffusion process. Even though label array $\hat{\boldsymbol{\mathcal{Y}}}$ only needs to be computed once needless of iterating, the rules described in Equations (5.12) and (5.13) enforce the spatial nodes to remember its original state at each iteration step, which could be effectively understood as that the spatial node ν_k is pulling information both from its spatial neighbours $\mathcal{N}_G(\nu_k)$ (the second term in the Equations) and from its bipartite post neighbours $\mathcal{N}_\mathcal{B}(\nu_k)$ (the first term in the Equations) simultaneously on two respective graphs G and $\mathcal{B}$.

For the steady state, the following equations hold:

$$\boldsymbol{\mathcal{Y}} = (1-\alpha)\hat{\boldsymbol{\mathcal{Y}}} + \alpha\boldsymbol{\mathcal{Y}}\left(\boldsymbol{W}\boldsymbol{D}^{-1}\right), \tag{5.14}$$

$$\boldsymbol{\mathcal{Y}}\left(\boldsymbol{I} - \alpha\boldsymbol{W}\boldsymbol{D}^{-1}\right) = (1-\alpha)\hat{\boldsymbol{\mathcal{Y}}}, \tag{5.15}$$

$$\text{therefore, } \boldsymbol{\mathcal{Y}} = (1-\alpha)\hat{\boldsymbol{\mathcal{Y}}}\left(\boldsymbol{I} - \alpha\boldsymbol{W}\boldsymbol{D}^{-1}\right)^{-1}. \tag{5.16}$$

For each row $\mathbf{y}_C^\mathsf{T} \in [0,1]^{1\times|V|}$ of $\boldsymbol{\mathcal{Y}}$ marking the distribution of one label category channel, the following also holds:

$$\mathbf{y}_C^\mathsf{T} = (1-\alpha)\hat{\mathbf{y}}_C^\mathsf{T}\left(\boldsymbol{I} - \alpha\boldsymbol{W}\boldsymbol{D}^{-1}\right)^{-1}, \tag{5.17}$$

where $\hat{\mathbf{y}}_C^\mathsf{T} := \boldsymbol{e}_C^\mathsf{T}\hat{\boldsymbol{\mathcal{Y}}}, \hat{\mathbf{y}}_C^\mathsf{T} \in [0,1]^{1\times|V|}$ is the C_{th} row of initial label array $\hat{\boldsymbol{\mathcal{Y}}}$. Note that the final array $\boldsymbol{\mathcal{Y}}$ is no longer a stack of two column-stochastic arrays respectively for OUV and HA labels since the sum of the “labels” of each spatial node can fluctuate around two, depending on the significance of the spatial nodes for each category channel. Also note that in the following equation:

$$\mathbf{y}_C = \left(\hat{\mathbf{y}}_C^\mathsf{T}(1-\alpha)\left(\boldsymbol{I} - \alpha\boldsymbol{W}\boldsymbol{D}^{-1}\right)^{-1}\right)^\mathsf{T} = \left((1-\alpha)\left(\boldsymbol{I} - \alpha\boldsymbol{W}\boldsymbol{D}^{-1}\right)^{-1}\right)^\mathsf{T}\hat{\mathbf{y}}_C, \tag{5.18}$$

the first component is clearly related to the generalized Katz Centrality (Benzi and Klymko, 2014; Zhan et al., 2017):

$$C_{\text{Katz}} = \beta\left(\boldsymbol{I} - \alpha\boldsymbol{A}^\mathsf{T}\right)^{-1}\mathbf{1}, \tag{5.19}$$

where the bias constant β is replaced with a constrained $1-\alpha$. Equation (5.19) performs one more step of summation of Equation (5.18) to obtain a centrality value. In other words, the calculation here uses an intermediate component of Katz centrality computation to weight the spatial labels (Nourian, 2016; Nourian et al., 2016; Zhan et al., 2017).

When $\alpha = 0$, no diffusion happens and the label vectors remain the same in all the steps. For Equations (5.16) and (5.17) to be solvable, the parameter α has to be chosen so that it is smaller than the reciprocal of the absolute value of the largest eigenvalue of $\boldsymbol{W}\boldsymbol{D}^{-1}$, i.e. $1/|\lambda|$, similar to the attenuation value for Katz Centrality computation. If this largest value is chosen, Equation (5.19) becomes a standard eigenvector centrality (Gould, 1967; Bonacich, 1972). Moreover, by adjusting the local diffusion rule in Equations (5.12) and (5.13), the computation could be easily adjusted to other variants of spectral-based centrality such as PageRank (Page et al., 1999) and standard Katz Centrality (Katz, 1953). Note that the term $\hat{\mathbf{y}}_k^{(t)}$ denoting the last state of the nodes are not included in Equations (5.12) and (5.13). Equations (B.8) to (B.13) in Appendix B will prove that adding such a term would end up calculating the same result in Equations (5.16) and (5.17) under certain constraints.

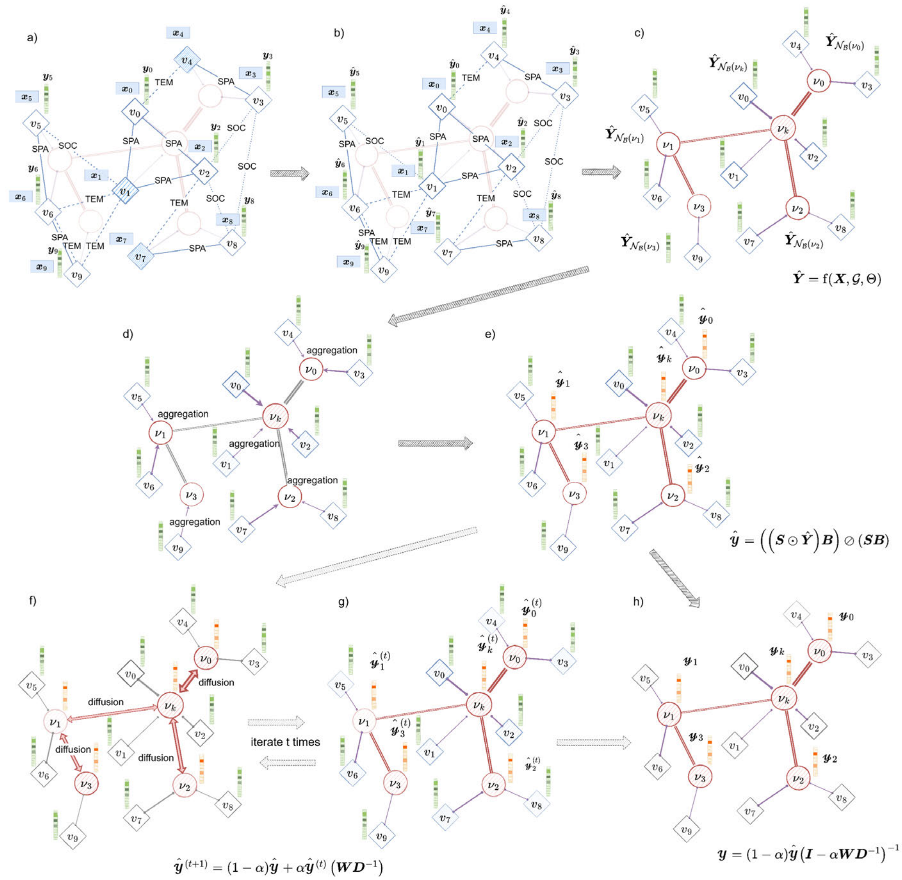


FIG. 5.3 The conceptually visualized semi-supervised learning, aggregation, and diffusion processes of node labels on a Post-level Attributed Multi-Graph (blue), a Post-Spatial Bipartite Graph (purple), and a Spatial Graph (red). Post nodes are represented with cylinders and spatial nodes with circles. a) All posts are connected with temporal, spatial, or social links in a partially labelled attributed multi-graph, where each node has a complete feature array $\boldsymbol{x}_i$ and only some nodes have initial labels $\boldsymbol{y}_i$; b) An estimated label vector $\hat{\boldsymbol{y}}_i$ is obtained for each post node with semi-supervised learning; c) All posts neighbouring the spatial nodes ν_k are labelled with $\hat{\boldsymbol{Y}}_{\mathcal{N}_{\mathcal{B}}(\nu_k)}$; d) Each spatial node aggregates (a single-sided process) the labels of neighbouring post nodes in the bipartite graph; e) The initial label for each spatial node $\hat{\mathbf{y}}_k^{(0)} = \hat{\mathbf{y}}_k$ is obtained; f) Each spatial node diffuses (a double-sided process) the labels of neighbouring spatial nodes in the spatial graph; g) An intermediate state at step t of label diffusion on the spatial graph to obtain the label vector $\hat{\mathbf{y}}_k^{(t)}$; h) The steady state when the spatial node label vector $\mathbf{y}_k$ converges. Note the iterative processes of f) and g) can be skipped by direct algebraic calculation in h).

5.4 Experiments

5.4.1 Selected Models and Baselines

As described in Section 5.3.1, a group of models in a candidate set $\mathcal{F}$ will be trained on the datasets, and the best-performing model f_j of each type will be selected to output the model-specific predictions $\hat{\boldsymbol{Y}}_j$ to be further aggregated. To make the model ensemble various enough for its best effect (Zhou, 2012), the following diverse model types that are shown to be effective in literature are illustratively used:

Random Classifier Using Prior Distributions

- RDC - Random Classifier (RDC), a Random Dummy Classifier baseline disregarding input features that generates random outputs based on the category distribution (prior) in the training set as shown in Table 5.2 (Baumer et al., 2015).

Graph-free Classifiers Using Multi-modal Features

- MLP - Multi-Layer Perceptron Classifiers with visual and textual features (Gardner and Dorling, 1998).

Homogeneous-graph GNN Classifiers

- GCN - The Graph Convolution Network (GCN) with initial residual connections and identity mapping (GCNII) proposed by Chen et al. (2020) as an extension for the vanilla GCN proposed by Kipf and Welling (2016).
- GAT - The Graph Attention Network (GAT) proposed by Veličković et al. (2017) with attention mechanism.
- GSA - Graph Sample and Aggregate (GraphSAGE) Models (GSA) proposed by Hamilton et al. (2017), which is especially effective for inductive learning, where knowledge learnt on one [sub-]graph is generalized across other unseen [sub-]graphs.

Heterogeneous-graph GNN Classifiers

- HGSA - Heterogeneous GraphSAGE Network (HGSA), the heterogeneous GNN that handles each type of links separately with a different GraphSAGE sub-model, where results are aggregated when multiple types of links point to the same destination node (Zhang et al., 2019a).
- HGT - The Heterogeneous Graph Transformers (HGT) proposed by Hu et al. (2020) that incorporates each type of links with an attention-based Transformer module (Vaswani et al., 2017).

During initial trials on the model structures, adding a linear layer in most graph-based models (except for GCN and GSA) and concatenating its output with that of the graph filters was found to boost the classification performance on VEN dataset. This is probably because the three types of links in VEN, i.e., the temporal, social, and spatial connections of the posts are all weak relations so that concatenating the neighbour features with the learnt feature of the node itself could overcome possible "over-smoothing" problem on these GNN, where individual features of all the nodes are forgotten and replaced by a universal aggregated one (Li et al., 2018). Also note that the Relational Graph Convolution Networks (Schlichtkrull et al., 2018) are not used as candidate models, as they assume that there only exists at most one type of relations between any two nodes, which is not the case in VEN, as two posts can be taken by the same person (socially similar) at the same place (spatially similar) in the same week (temporally similar).

5.4.2 Sub-sampling of Graphs

The `NeighborLoader` in PyTorch Geometric (PyG) library (Fey and Lenssen, 2019), which is based on the Neighbour Sampler introduced by Hamilton et al. (2017), is used to generate sub-graphs $\mathcal{G}_s$ for all graph-based classifiers. A mini-batch of 32 post nodes are used as the input nodes $\mathcal{V}_{\mathsf{batch}}$ for all sorts of subsets in $\mathcal{V}_{\mathsf{train}}, \mathcal{V}_{\mathsf{val}}$, $\mathcal{V}_{\mathsf{test}}$, and $\mathcal{V}_{\mathsf{unlab}}$. To make the GNN models compatible, for Homogeneous-graph GNN Classifiers (GCN, GAT, GSA), 75 neighbours are sampled for each node for two iterations, and for heterogeneous-graph GNN Classifiers (HGSA, HGT), 25 neighbours are sampled for each node and link type for two iterations. This effectively reduces the size of sub-graphs: the total number of links from the order of 1×10^6 in VEN and 1×10^8 in VEN-XL all to the order of 1×10^5 in the sub-graphs. This is especially desirable for datasets at scales such as VEN-XL for it to fit in computer memory during training and inference.

5.4.3 Evaluation Metrics

Cross-Entropy of the soft labels are used as the loss functions $\ell_{\mathsf{V}}, \ell_{\mathsf{A}}$ for both OUV and HA classifications, while the parameter $\omega_{\mathsf{V/A}}$ mentioned in Equation (5.4) is set to 1 for simplicity during training.

For OUV classification, Top-1 Accuracy ($p^{\mathsf{OUV(1)}}$), Top-n Accuracy ($p^{\mathsf{OUV(n)}}$), and Order-n Jaccard Index ($p^{\mathsf{OUV(nJ)}}$) are used as general evaluation metrics, while for HA classification, only Top-1 Accuracy ($p^{\mathsf{HA(1)}}$) is used, since HA categories were assumed to be more precise in Bai et al. (2022). Let topk$(\boldsymbol{v}, n)$ denote a function returning an ordered set containing the indices of the top-n entries of a generic vector $\boldsymbol{v}$, then the evaluation metrics on any subset $\mathcal{V}_* \in \{\mathcal{V}_{\mathsf{val}}, \mathcal{V}_{\mathsf{test}}\}$ by model f_j can be respectively described as:

$$p_{*,j}^{\text{OUV(1)}} = \frac{\sum_{v_i \in \mathcal{V}_* \cap \mathcal{V}_{\text{V+,A-}}} \left(\text{topk}(\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}}, 1) = \text{topk}(\boldsymbol{y}_i^{\text{OUV}}, 1)\right)}{|\mathcal{V}_* \cap \mathcal{V}_{\text{V+,A-}}|} \tag{5.20}$$

$$p_{*,j}^{\text{OUV(n)}} = \frac{\sum_{v_i \in \mathcal{V}_* \cap \mathcal{V}_{\text{V+,A-}}} \left(\text{topk}(\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}}, 1) \in \text{topk}(\boldsymbol{y}_i^{\text{OUV}}, n)\right)}{|\mathcal{V}_* \cap \mathcal{V}_{\text{V+,A-}}|} \tag{5.21}$$

$$p_{*,j}^{\text{OUV(nJ)}} = \frac{\sum_{v_i \in \mathcal{V}_* \cap \mathcal{V}_{\text{V+,A-}}} \frac{\left|\left(\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}} > \frac{1}{n+1}\right) \wedge \left(\boldsymbol{y}_i^{\text{OUV}} > \frac{1}{n+1}\right)\right|}{\left|\left(\hat{\boldsymbol{y}}_{j,i}^{\text{OUV}} > \frac{1}{n+1}\right) \vee \left(\boldsymbol{y}_i^{\text{OUV}} > \frac{1}{n+1}\right)\right|}}{|\mathcal{V}_* \cap \mathcal{V}_{\text{V+,A-}}|} \tag{5.22}$$

$$p_{*,j}^{\text{HA(1)}} = \frac{\sum_{v_i \in \mathcal{V}_* \cap \mathcal{V}_{\text{V-,A+}}} \left(\text{topk}(\hat{\boldsymbol{y}}_{j,i}^{\text{HA}}, 1) = \text{topk}(\boldsymbol{y}_i^{\text{HA}}, 1)\right)}{|\mathcal{V}_* \cap \mathcal{V}_{\text{V-,A+}}|}, \tag{5.23}$$

where Equation (5.22) computes the Intersection over Union (Jaccard Index) of two sets of indices pointing to vector entries with values larger than a threshold (e.g., when $n = 3$, the computation is about logits larger than $.25$), being an effective way of evaluating soft label classification.

Furthermore, the per-class metrics of precision, recall, F1 score (harmonic average of precision and recall), and confusion matrix are used to inspect the model performance on each OUV and HA category channel. Moreover, since VEN and VEN-XL are unbalanced datasets as mentioned in Section 5.2.2 where some small classes only exist in top-n rather than top-1 labels, they are never counted in per-class metrics calculation as “true-positive” instances. As an explorative treatment, top-n per-class metrics are computed with the Algorithm 1, where the predicted and “ground-truth” top-n classes are permuted to obtain n^2 confusion matrices, which are further summed and normalized. Note the diagonal entries of normalized confusion matrix $\tilde{\boldsymbol{M}}$ are effectively top-n F1 scores of top-n precision and recall. A similar explanation applies to the off-diagonal entries.

5.4.4 Implementations of Experiments

As briefly described in Section 5.3.1, the training procedure consists of the following steps:

1. for each model type, hyper-parameter searching was performed on sampled sub-graphs of VEN for 300-1000 epochs of training on $\mathcal{V}_{\text{train}}$ with grid search in small ranges, where early-stopping was implemented based on the overall performance on validation set $\mathcal{V}_{\text{val}}$;
2. the hyper-parameter configuration of the selected best models are used to re-train model checkpoints to be stored and used for inference;
3. the stored models are evaluated with metrics mentioned in Section 5.4.3 on both validation set $\mathcal{V}_{\text{val}}$ and test set $\mathcal{V}_{\text{test}}$ with 10 runs of different random seeds since some GPU-based models do not generate exactly same outcomes given a fix random seed;

Algorithm 1: Computing Top-n Per-Class Metrics

Data: Number of Classes N, $1 \le n \le N$, a $N \times K$ Label Array $\boldsymbol{Y}$, a $N \times K$ Predicted Label Array $\hat{\boldsymbol{Y}}$, Standard Confusion Matrix Function of Index Arrays ConfMat$(\boldsymbol{d}, \hat{\boldsymbol{d}})$

Result: Normalized Top-n Confusion Matrix $\tilde{\boldsymbol{M}}$, Top-n Precision $\boldsymbol{p}$, Top-n Recall $\boldsymbol{r}$, Top-n F1 Score $\boldsymbol{f}$

$\epsilon \leftarrow 0.0000001$;
$i, j, l, m \leftarrow 0$;
$\boldsymbol{M}, \tilde{\boldsymbol{M}} \leftarrow N \times N$ arrays of 0s;
$\boldsymbol{D}, \hat{\boldsymbol{D}} \leftarrow K \times n$ arrays of 0s;
v, $\boldsymbol{p}, \boldsymbol{r}, \boldsymbol{f} \leftarrow N \times 1$ arrays of 0s;
d, $\hat{\text{d}} \leftarrow K \times 1$ arrays of 0s;
$\boldsymbol{D} \leftarrow$ topk$(\boldsymbol{Y}, n)$;
$\hat{\boldsymbol{D}} \leftarrow$ topk$(\hat{\boldsymbol{Y}}, n)$; `/Indices of top-`$n$` entries`
for $i \in [0, n)$ **do**
    $\boldsymbol{d} \leftarrow \boldsymbol{D}[:, i]$; `/Indices of `$i_{\text{th}}$` largest entries`
    for $j \in [0, n)$ **do**
        $\hat{\boldsymbol{d}} \leftarrow \hat{\boldsymbol{D}}[:, j]$;
        $\boldsymbol{M} \leftarrow \boldsymbol{M} +$ ConfMat$(\boldsymbol{d}, \hat{\boldsymbol{d}})$;
    end
end
$\boldsymbol{v} = \boldsymbol{M}$.diagonal(); `/The diagonal entries`
for $l \in [0, N)$ **do**
    $\boldsymbol{p}[l] \leftarrow \boldsymbol{v}[l] / (\boldsymbol{M}[l, :].\text{sum}() - (n-1) \times \boldsymbol{v}[l] + \epsilon)$;
    $\boldsymbol{r}[l] \leftarrow \boldsymbol{v}[l] / (\boldsymbol{M}[:, l].\text{sum}() - (n-1) \times \boldsymbol{v}[l] + \epsilon)$;
    $\boldsymbol{f}[l] \leftarrow 2 \times \boldsymbol{p}[l] \times \boldsymbol{r}[l] / (\boldsymbol{p}[l] + \boldsymbol{r}[l] + \epsilon)$;
    for $m \in [0, N)$ **do**
        $\tilde{\boldsymbol{M}}[l, m] = 2 \times \boldsymbol{M}[l, m] / (\boldsymbol{M}[l, :].\text{sum}() + \boldsymbol{M}[:, m].\text{sum}() - 2 \times (n-1) \times \boldsymbol{M}[l, m] + \epsilon)$;
    end
end

4 once the overall performance of a model type is acceptable, it is used to predict the final label arrays $\hat{\boldsymbol{Y}}_j$ on the entire dataset $\mathcal{V}$ to be further aggregated;

5 Instead of repeating the same training process for VEN-XL, the model checkpoints obtained in step 2 are directly evaluated with $\mathcal{V}_{\text{train}}$, $\mathcal{V}_{\text{val}}$ and $\mathcal{V}_{\text{test}}$ of VEN-XL (all practically test sets) and used to predict label arrays since it is assumed that the model checkpoints are generalizable in inductive learning.

All models are implemented using building blocks provided by PyTorch Geometric (PyG) library. The datasets are structured and stored respectively as `Data` and `HeteroData` classes in PyG for different model types. More details of the training settings can be found in Appendix B.

To aggregate the predicted label arrays and perform SVD for the agreement score $\boldsymbol{\kappa}^{\text{agr}}$, PyTorch (Paszke et al., 2019) is used. The sum of Top-1 HA Accuracy and Order-3 OUV Jaccard Index on both validation and test sets are used as the weight p_j for aggregation. To compute the confidence score $\boldsymbol{\kappa}^{\text{con}}$, the top-$4$ entries of the aggregated label array $\hat{\boldsymbol{Y}}$ are used. For simplicity, parameters ϕ, γ in Equation (5.10) are both set to 2 to compute the attention array $\boldsymbol{S}$. As for the spatial diffusion process, the parameter $\alpha \in [0, \min(1/|\lambda|, 1))$ is tested with 10 different values evenly dividing its theoretical lower and upper bounds (smaller than 1 for Equation (5.13) to be meaningful) to test its effect on the distribution of the final label array $\boldsymbol{\mathcal{Y}}$ on the spatial network.

5.5 Ablation Studies

5.5.1 Sensitivity on Alternative Conditions

To reflect on the assumption that graph-based models can better deal with semi-supervised learning tasks with a large proportion of missing features and/or labels, the trained model checkpoints are directly evaluated on an altered validation set $\mathcal{V}_{\text{val}}$ where the visual or textual features of the mini-batches are masked and clipped to 0, while all the other nodes in the sampled graphs $\mathcal{G}_s$ are intact.

The usefulness of three link types $\{\boldsymbol{A}^{\text{TEM}}, \boldsymbol{A}^{\text{SPA}}, \boldsymbol{A}^{\text{SOC}}\}$ are also experimented. For homogeneous graph models, the simple composed links $\boldsymbol{A}$ are replaced by each sub-link type to sample the sub-graphs for evaluation on $\mathcal{V}_{\text{val}}$ in mini-batches. For heterogeneous graph models, only one link type is kept or masked during sub-graph sampling, yielding six different alternative performance scores on $\mathcal{V}_{\text{val}}$.

As an alternative to the original graph links provided by Bai et al. (2022), a k-Nearest Neighbour (KNN) graph structure based on features is also tested for homogeneous graph models. Since textual features have missing values, only visual features $\boldsymbol{X}^{\text{vis}}$ are used to compute an adjacency matrix $\boldsymbol{A}^{\text{KNN}} \in \{0, 1\}^{K \times K}$, where each entry $A^{\text{KNN}}_{i,i'} = 1$ only if the post node $v_{i'}$ is within the 3 nearest neighbours of v_i based on cosine similarity. The KNN graph structure is computed with the `knn_graph` function of PyG library.

5.5.2 Interpreting the Association of Input Features

For the final post-level label array $\hat{\boldsymbol{Y}}$ and the initial spatial-level label array $\hat{\boldsymbol{\mathcal{Y}}}$ before diffusion, rectangular co-occurrence matrices $\boldsymbol{O} \in \mathbb{N}^{11 \times 9}$ of top-3 OUV and top-1 HA categories are computed, where each matrix entry is normalized by dividing the total number of examples used for computation. When computing $\boldsymbol{O}$ for post-level label $\hat{\boldsymbol{Y}}$, only the posts whose sum of confidence score $\boldsymbol{\kappa}^{\text{con}}$ and agreement score $\boldsymbol{\kappa}^{\text{agr}}$ were above the 25% quantile are considered. These matrices can be used to explain the association of OUV and HA categories as well as their general distributions. When two categories from OUV and HA have high co-occurrence, they could be well-associated, informative for further heritage study investigations.

Furthermore, GNNExplainer (Ying et al., 2019) is illustratively used for GAT and GSA on $\mathcal{V}_{\text{train}}, \mathcal{V}_{\text{val}}, \mathcal{V}_{\text{test}}$ to compute the relative importance of all visual and textual features for each OUV and HA category, among which 473 features out of 1753 are more explainable with physical meanings, e.g., scene categories (Zhou et al., 2017), SUN attribute categories (Patterson et al., 2014), number of faces (Schroff et al., 2015), and origin of languages. For all nodes considered, GNNExplainer predicted the relative importance of all features for classifying each node in sampled sub-graph mini-batches for 200 epochs. The explainable features mentioned above that entered the top-250 rankings by each node are counted for each OUV and HA category. A bipartite graph connecting the features with the categories is visualized in Gephi with Force Atlas algorithm (Bastian et al., 2009; Jacomy et al., 2014), which could be considered as an interpretable lexicon of the cultural significance categories.

5.5.3 Statistical Tests and Spatial Mapping

T-Tests and Analysis of Variance (ANOVA) are conducted on the difference of model performance, confidence scores, and agreement scores between datasets VEN and VEN-XL and among subsets $\mathcal{V}_{\text{train}}, \mathcal{V}_{\text{val}}, \mathcal{V}_{\text{test}}, \mathcal{V}_{\text{unlab}}$ to check the coherence and consistency of trained models. All statistical tests are performed with Pingouin library (Vallat, 2018).

For each category channel of the final spatial label array $\mathbf{y}_C$ with each value of α as in Equation (5.17), the global Moran's I is computed as the spatial auto-correlation measure (Moran, 1950; Rogerson and Sun, 2001; Rogerson, 2021) of each OUV and HA category, showing the effect of spatial diffusion on the final label distribution, such that:

$$I_C = \frac{|V|(\mathbf{y}_C - \bar{\mathrm{y}}_C\mathbf{1})^\mathsf{T}\boldsymbol{W}(\mathbf{y}_C - \bar{\mathrm{y}}_C\mathbf{1})}{\mathbf{1}^\mathsf{T}\boldsymbol{W}\mathbf{1} \times (\mathbf{y}_C - \bar{\mathrm{y}}_C\mathbf{1})^\mathsf{T}(\mathbf{y}_C - \bar{\mathrm{y}}_C\mathbf{1})}, \tag{5.24}$$

where $\mathbf{1}$ is a $|V|$-dimensional vector of all 1s, $\bar{\mathrm{y}}_C$ is the mean of vector $\mathbf{y}_C$, and $\boldsymbol{W}$ is the spatial closeness matrix mentioned in Section 5.3.3, thus not a conventional weight matrix with zero diagonal entries (Chen, 2021).

The spatial clustering effect of hot spots (clusters of high values) of each category channel is found with the computation of local Moran's I and the simulated p values based on random re-assignment of values on the spatial nodes (Anselin, 1995; Rogerson and Sun, 2001), such that:

$$\boldsymbol{I}_C = (\mathbf{y}_C - \bar{\mathrm{y}}_C\mathbf{1}) \odot \boldsymbol{W}(\mathbf{y}_C - \bar{\mathrm{y}}_C\mathbf{1}). \tag{5.25}$$

The spatial statistics global and local Moran's I are computed using the ESDA:

Exploratory Spatial Data Analysis tool of PySAL library (Rey and Anselin, 2007) with doubly-standardized weight transformation together with 9999 permutations to generate simulated distributions for estimating two-tailed p values with Bonferroni correction (VanderWeele and Mathur, 2019), where all the other parameters are kept as default. This computation would return the same results as implementing Equations (5.24) and (5.25). Afterward, the values of OUV and HA categories on spatial nodes are mapped using QGIS (QGIS Development Team, 2023).

5.6 Results

5.6.1 Classification Performance

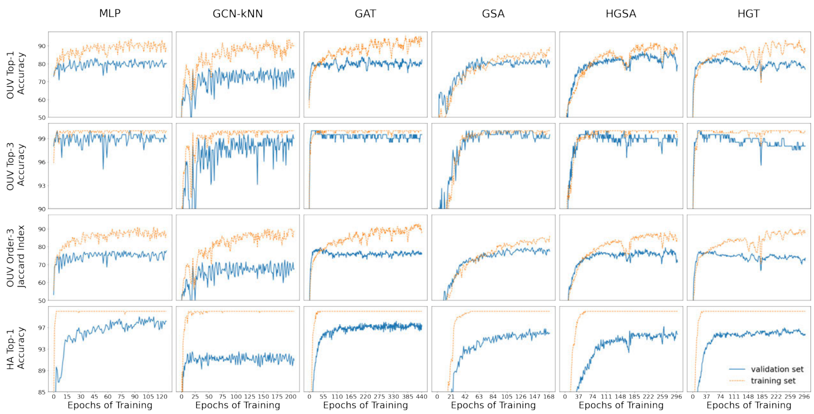


FIG. 5.4 The training curves of the stored model checkpoints on the four main evaluation metrics for OUV and HA classification tasks. The dashed curves in orange show the performance of models on training set for each epoch, and the continuous curves in blue show the performance on validation set.

The classification performance of all the models is shown in Table 5.3 for VEN and in Table 5.4 for VEN-XL, while detailed performance curves of each model checkpoint during training can be found in Figure 5.4. The selected candidate models all performed reasonably well, as they all appeared in the best two instances at least once among the evaluation metrics on VEN, far exceeding the random classifier RDC. Note only GCN selected is based on KNN graph structure mentioned in Section 5.5.1, since it performed better as will be shown in Figure 5.8. Since different random seeds would change the configuration of sampled sub-graphs and the group of neighbours a node can learn from, the classification performance can be affected. Still, except for the top-1 OUV accuracy for HGSA, other variances are generally small. Furthermore, as the goal of this study is not to select the best model architecture, but to have stable and reliable performance, no single model was selected as the "final" one to predict labels. Rather, the aggregated prediction of all models was used in further steps. In VEN, aggregated prediction performed well in all evaluation metrics, either being among the best two models or performing considerably to the best ones. Yet in VEN-XL where models were directly evaluated without further training or fine-tuning, the aggregated prediction performed best for all metrics in all subsets. It is remarkable that GAT performed arguably the best among the individual models both in VEN and VEN-XL, suggesting that it has decent generalizability. Note the general performance of selected models including the aggregated prediction on all evaluation metrics dropped significantly from VEN to VEN-XL on their respective validation and test sets according to one-sided paired T-Test, $t(55) = 4.517, p < .0001$, yet the effect size of this drop is minimum (Cohen's $d = 0.096$), suggesting that the knowledge learned on the small VEN dataset during training has been successfully transferred and generalized to the large unseen VEN-XL dataset.

TABLE 5.3 The performance (%) of each model type in VEN dataset on validation and test sets as mean ± standard deviation, computed using the stored model checkpoints with ten runs of evaluation with different random seeds. The best two models on each metric are marked in bold.

Model	$p_{\text{val}}^{\text{OUV(1)}}$	$p_{\text{test}}^{\text{OUV(1)}}$	$p_{\text{val}}^{\text{OUV(3)}}$	$p_{\text{test}}^{\text{OUV(3)}}$	$p_{\text{val}}^{\text{OUV(3J)}}$	$p_{\text{test}}^{\text{OUV(3J)}}$	$p_{\text{val}}^{\text{HA(1)}}$	$p_{\text{test}}^{\text{HA(1)}}$
RDC	18.79±3.12	18.75±3.08	57.14±2.19	56.46±3.69	21.92±1.16	22.67±1.85	17.56±1.67	18.09±1.15
MLP*	80.79±0.00	80.21±0.00	**99.51±0.00**	99.48±0.00	75.79±0.00	74.13±0.00	**98.98±0.00**	**98.21±0.00**
GCN-KNN*	74.38±0.00	72.92±0.00	**99.51±0.00**	**98.44±0.00**	69.21±0.00	68.40±0.00	91.87±0.00	97.38±0.00
GAT	80.39±0.43	**82.55±0.42**	**99.51±0.00**	**99.48±0.00**	76.32±0.21	**76.11±0.29**	**98.07±0.10**	97.38±0.08
GSA	80.69±0.72	79.06±0.65	**99.51±0.15**	**99.48±0.00**	**77.17±0.38**	75.48±0.49	95.71±0.21	97.08±0.22
HGSA	**84.73±1.14**	77.86±0.35	99.11±0.20	99.11±0.33	**77.33±0.60**	71.74±0.42	96.63±0.24	95.65±0.30
HGT*	79.31±0.00	78.65±0.00	98.03±0.00	**99.48±0.00**	73.81±0.00	74.05±0.00	96.95±0.00	96.42±0.00
Aggregated	**84.23**	**81.77**	99.01	**100.00**	76.77	**76.30**	97.56	**98.21**

*Deterministic outputs on GPU by the stored model checkpoint with different random seeds.

TABLE 5.4 The performance (%) of each model type in VEN-XL dataset on train, validation, and test sets, computed directly using the stored model checkpoints trained on VEN as inductive learning setting. The best two models on each metric are marked in bold.

Model	$p_{\text{train}}^{\text{OUV(1)}}$	$p_{\text{val}}^{\text{OUV(1)}}$	$p_{\text{test}}^{\text{OUV(1)}}$	$p_{\text{train}}^{\text{HA(1)}}$	$p_{\text{val}}^{\text{HA(1)}}$	$p_{\text{test}}^{\text{HA(1)}}$
MLP	79.16	80.53	80.52	91.58	**96.86**	**96.79**
GCN-KNN	76.01	75.54	76.43	85.93	91.41	91.24
GAT	**80.04**	**80.88**	**80.90**	**93.32**	96.28	96.01
GSA	75.92	78.19	78.21	90.09	94.69	94.10
HGSA	77.12	78.81	78.48	90.66	95.10	94.62
HGT	77.58	78.34	78.92	91.36	95.40	95.25
Aggregated	**80.54**	**81.49**	**81.81**	**91.62**	**96.54**	**96.11**
Model	$p_{\text{train}}^{\text{OUV(3)}}$	$p_{\text{val}}^{\text{OUV(3)}}$	$p_{\text{test}}^{\text{OUV(3)}}$	$p_{\text{train}}^{\text{OUV(3J)}}$	$p_{\text{val}}^{\text{OUV(3J)}}$	$p_{\text{test}}^{\text{OUV(3J)}}$
MLP	**98.67**	98.70	**98.86**	74.42	75.25	75.22
GCN-KNN	96.80	96.67	96.53	70.65	71.65	71.67
GAT	98.47	**98.72**	98.61	**74.09**	73.50	73.44
GSA	98.44	98.69	98.37	72.73	**75.55**	**75.28**
HGSA	98.49	98.41	98.41	70.63	70.53	69.85
HGT	97.95	98.04	98.20	72.66	72.48	72.39
Aggregated	**98.67**	**98.77**	**98.83**	**75.93**	**76.57**	**76.45**

The per-class metrics of OUV and HA categories by the aggregated prediction array $\hat{\boldsymbol{Y}}$ on both datasets can be seen in Table 5.5 and Table 5.6. For most cultural OUV selection criteria except for Criterion (v) about Land-use and almost all HA categories except for Artificial Products, the aggregated prediction performed reasonably well in both VEN used for training, and VEN-XL completely new to the models. The poor performance of OUV Criteria (v), (viii), (ix) and HA category Artificial Products is clearly related to their scarce presence in the training set of VEN shown in Table 5.2, where the models had to learn the key features of a category using less than 10 examples. Specifically, even though there are a few training data labelled as Criteria (v)(ix)(x), no data from validation and test sets are labelled with them, thus resulting blanks ('-') in Table 5.5. Future data augmentation is expected to teach the models specifically on these scarce classes. Under the same condition of scarcity, the prediction on Criterion (x) - Bio-diversity, Urban Scenery, and Gastronomy performed remarkably well, suggesting that these classes are probably more clearly separated from the others in the feature space, easy for models to learn even with few-shot learning.

TABLE 5.5 The per-class performance metrics of OUV Selection Criteria classes in VEN and VEN-XL datasets. When no correct predictions were made for a class, the score would be 0.00; yet when no examples of a class were available, the score is marked as "-". The class "Others" is omitted since no examples were assigned to it.

Metrics	Precision	Recall	F1	Top3 Precision	Top3 Recall	Top3 F1
i. Masterpiece	0.94 | 0.87	0.89 | 0.79	0.92 | 0.82	0.86 | 0.64	0.81 | 0.64	0.84 | 0.64
ii. Influence	0.76 | 0.59	0.65 | 0.87	0.70 | 0.70	0.63 | 0.27	0.53 | 0.74	0.58 | 0.39
iii. Testimony	0.68 | 0.69	0.80 | 0.79	0.74 | 0.74	0.73 | 0.93	0.61 | 0.74	0.66 | 0.83
iv. Typology	0.88 | 0.70	0.79 | 0.76	0.83 | 0.73	0.65 | 0.75	0.76 | 0.64	0.70 | 0.69
v. Land-use	- | 0.00	- | 0.00	- | 0.00	0.11 | 0.03	1.00 | 0.38	0.20 | 0.06
vi. Association	0.78 | 0.94	0.88 | 0.82	0.82 | 0.87	0.63 | 0.89	0.75 | 0.76	0.68 | 0.82
vii. Natural Beauty	1.00 | 0.16	1.00 | 0.55	1.00 | 0.24	0.25 | 0.17	1.00 | 0.94	0.40 | 0.28
viii. Geological Process	- | 0.00	- | 0.00	- | 0.00	0.00 | 0.00	0.00 | 0.00	0.00 | 0.00
ix. Ecological Process	- | 0.00	- | 0.00	- | 0.00	- | 0.00	- | 0.00	- | 0.00
x. Bio-diversity	- | 0.66	- | 0.73	- | 0.69	0.00 | 0.39	0.00 | 1.00	0.00 | 0.56

TABLE 5.6 The per-class performance metrics of Heritage Attributes classes in VEN and VEN-XL datasets.

Metrics	Precision	Recall	F1
Monument and Buildings	0.99 | 0.98	0.99 | 0.98	0.99 | 0.98
Building Elements	1.00 | 0.98	0.98 | 0.96	0.99 | 0.97
Urban Form Elements	0.99 | 0.99	0.98 | 0.97	0.98 | 0.98
Urban Scenery	0.91 | 0.74	1.00 | 1.00	0.95 | 0.85
Natural Features and Landscape Scenery	0.99 | 0.99	0.97 | 0.99	0.98 | 0.99
Interior Scenery	0.95 | 0.90	1.00 | 0.96	0.97 | 0.93
People's Activity and Association	0.96 | 0.99	1.00 | 0.88	0.98 | 0.93
Gastronomy	0.95 | 0.92	0.82 | 0.83	0.88 | 0.87
Artifact Products	0.29 | 0.08	0.67 | 0.93	0.40 | 0.15

The top-n per-class metrics proposed in Algorithm 1 is especially useful to evaluate scarce classes, as they may be absent as top-1 yet appear as top-n classes in validation and test sets, which can be seen in the cases of Criteria (v), (viii), (x) for VEN in Table 5.5. Such metrics are arguably stricter than standard per-class metrics

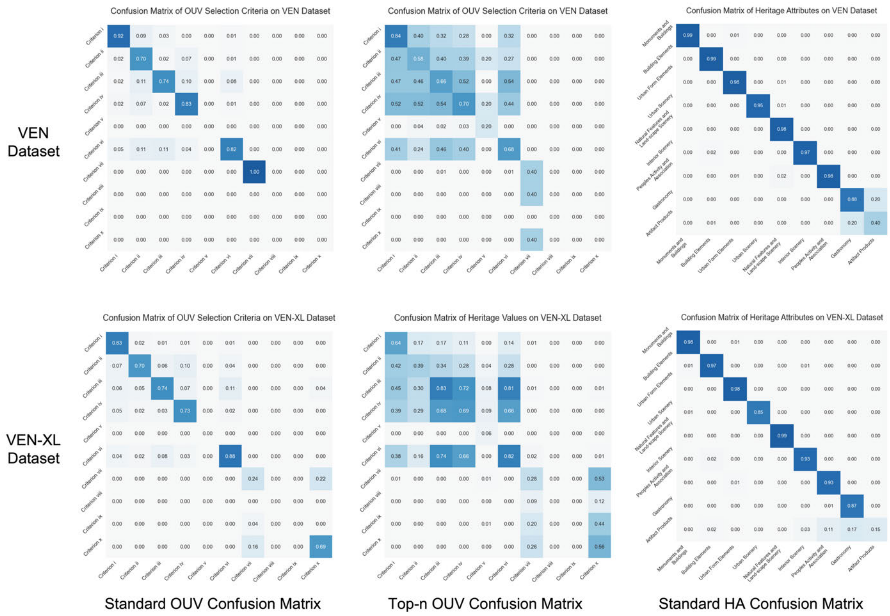


FIG. 5.5 The normalized top-1 and top-n confusion-matrix heatmaps of OUV selection criteria and Heritage Attributes classification of the aggregated prediction on both VEN and VEN-XL datasets. Note that these confusion matrices are not stochastic, and the entries represent the extent of confusion, where the diagonal entries are F1 scores in Tables 5.5 and 5.6.

in the sense that it evaluates the overlap of all top-n predictions with top-n labels (only when they are all the same, the metrics get to their theoretical maximum of 1), which could be seen as an extension of top-n accuracy with soft labels.

Moreover, the top-n per-class metrics allow a deeper observation of the confusion among the classes, as shown in Figure 5.5. While Criterion (v) - Land Use is absent in standard OUV confusion matrices (for the same reason mentioned above that no data in validation and test sets of VEN has a top-1 label of it), the values in top-n confusion matrices give a hint on how other classes are confused and thus related with it: posts about land-use in Venice also concern with the influence of Venice to the world and its special architectural style near the canals. Posts concerning Criteria (iii), (iv), and (vi) are easily confused with each other, meaning that when people post about Venice on Flickr, themes about testimony of the past, architectural typology and the association of architectural and urban elements with human activity usually come together. The same goes for Criteria (vii) and (x) about natural beauty of the city and the living animals and plants indicating bio-diversity. For HA, Artifact Products can be confused with Gastronomy and People's Activity, which also makes sense as all three topics usually depict human and human-related objects. Such associations will be further elaborated in Section 5.6.4.

5.6.2 Consistency of Predictions

The confidence score $\boldsymbol{\kappa}^{\text{con}}$ and the agreement score $\boldsymbol{\kappa}^{\text{agr}}$ mentioned in Sections 5.3.2 have similar distributions for VEN and VEN-XL datasets as in Figure 5.6. Two-way ANOVA F-Tests on the level of datasets and on the level of zoomed-in subsets $\mathcal{V}_{\text{train}}, \mathcal{V}_{\text{val}}, \mathcal{V}_{\text{test}}$, and $\mathcal{V}_{\text{unlab}}$ is showed in Table 5.7. All effects are statistically significant, yet only the main effect of subset has large effect sizes η^2, and the main effect of the dataset and the interaction effect are all minimum, which can also be seen with Cohen's d computed with independent T-Tests with Welch's correction. The very small effect sizes on the level of dataset indicate that the significant drops of both scores from VEN to VEN-XL are mainly caused by the large sample size in VEN-XL, suggesting that the models function consistently and coherently in both datasets.

TABLE 5.7 Means, Standard Deviations, and Two-Way ANOVA Statistics on the Confidence and Agreement scores. An Independent T-Test with Welch's correction is also performed on the level of two datasets.

Score	**VEN**	**VEN-XL**			**ANOVA**		
	M (SD)	M (SD)	**Effect**	df	F	p	η^2
Confidence Score $\boldsymbol{\kappa}^{\text{con}}$							
$-\mathcal{V}_{\text{train}}$	0.795 (0.042)	0.744 (0.076)	Dataset	1	59.938	<.0001	.0004
$-\mathcal{V}_{\text{val}}$	0.666 (0.076)	0.663 (0.080)	Subset	3	17,336.251	<.0001	.3827
$-\mathcal{V}_{\text{test}}$	0.667 (0.077)	0.664 (0.080)	Dataset × Subset	3	32.388	<.0001	.0012
$-\mathcal{V}_{\text{unlab}}$	0.573 (0.084)	0.563 (0.083)	Residual	83,906			
(Overall)	0.644 (0.105)	0.638 (0.102)	t(3158.402) = 2.910, p=.004, Cohen's d=0.056				
Agreement Score $\boldsymbol{\kappa}^{\text{agr}}$							
$-\mathcal{V}_{\text{train}}$	0.741 (0.033)	0.664 (0.099)	Dataset	1	110.854	<.0001	.0008
$-\mathcal{V}_{\text{val}}$	0.604 (0.110)	0.589 (0.115)	Subset	3	16,195.095	<.0001	.3662
$-\mathcal{V}_{\text{test}}$	0.604 (0.111)	0.589 (0.116)	Dataset × Subset	3	27.723	<.0001	.0001
$-\mathcal{V}_{\text{unlab}}$	0.444 (0.129)	0.427 (0.129)	Residual	83,906			
(Overall)	0.556 (0.152)	0.541 (0.149)	t(3160.154) = 5.235, p<.0001, Cohen's d=0.100				

TABLE 5.8 The post hoc comparison of the main effect of four different subsets for the confidence score $\boldsymbol{\kappa}^{\text{con}}$ and the agreement score $\boldsymbol{\kappa}^{\text{agr}}$ using the Tukey HSD Test.

Score	**Group A**	**Group B**	M**(Group A)**	M**(Group B)**	$\Delta(M)$	T	**Tukey** p	**Cohen's** d
$\boldsymbol{\kappa}^{\text{con}}$	$\mathcal{V}_{\text{train}}$	$\mathcal{V}_{\text{val}}$	0.746	0.663	0.083	89.315	<.0001	1.027
	$\mathcal{V}_{\text{train}}$	$\mathcal{V}_{\text{test}}$	0.746	0.664	0.082	88.638	<.0001	1.019
	$\mathcal{V}_{\text{train}}$	$\mathcal{V}_{\text{unlab}}$	0.746	0.564	0.182	210.015	<.0001	2.266
	$\mathcal{V}_{\text{val}}$	$\mathcal{V}_{\text{test}}$	0.663	0.664	-0.001	-0.792	.858	-0.008
	$\mathcal{V}_{\text{val}}$	$\mathcal{V}_{\text{unlab}}$	0.663	0.564	0.099	137.655	<.0001	1.239
	$\mathcal{V}_{\text{val}}$	$\mathcal{V}_{\text{unlab}}$	0.664	0.564	0.100	138.521	<.0001	1.247
$\boldsymbol{\kappa}^{\text{agr}}$	$\mathcal{V}_{\text{train}}$	$\mathcal{V}_{\text{val}}$	0.666	0.590	0.076	55.832	<.0001	0.642
	$\mathcal{V}_{\text{train}}$	$\mathcal{V}_{\text{test}}$	0.666	0.590	0.076	55.805	<.0001	0.642
	$\mathcal{V}_{\text{train}}$	$\mathcal{V}_{\text{unlab}}$	0.666	0.427	0.238	186.453	<.0001	2.012
	$\mathcal{V}_{\text{val}}$	$\mathcal{V}_{\text{test}}$	0.590	0.590	0.000	-0.032	.999	-0.000
	$\mathcal{V}_{\text{val}}$	$\mathcal{V}_{\text{unlab}}$	0.590	0.427	0.162	152.187	<.0001	1.370
	$\mathcal{V}_{\text{val}}$	$\mathcal{V}_{\text{unlab}}$	0.590	0.427	0.162	152.187	<.0001	1.370

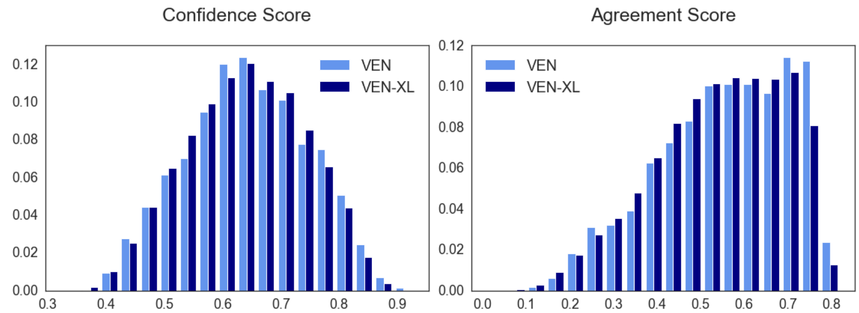


FIG. 5.6 The distribution of the confidence score $\boldsymbol{\kappa}^{\text{con}}$ and the agreement score $\boldsymbol{\kappa}^{\text{agr}}$ on both VEN (light blue) and VEN-XL (dark blue) datasets, both as density-based histograms.

Post hoc comparisons using the Tukey HSD test indicated that the scores in $\mathcal{V}_{\text{train}}$ are always significantly higher than $\mathcal{V}_{\text{val}}$, and $\mathcal{V}_{\text{test}}$, and the scores in $\mathcal{V}_{\text{unlab}}$ are always significantly lower than all the others with large effect size, while there is no significant difference between $\mathcal{V}_{\text{val}}$ and $\mathcal{V}_{\text{test}}$, as shown in Table 5.8. This again shows the consistency and coherence of the model performance. When further aggregating the labels into spatial nodes, those posts with high prediction confidence and agreement (thus are more reliable) contribute more to attention score computation. Note the scores on the training set gets closer to the validation and test sets in VEN-XL than in VEN with lower means and larger standard deviations. This is probably because the models are not trained on VEN-XL, and the training set, therefore, becomes another validation/test set, as pointed out in Section 5.4.4.

5.6.3 Robustness of Models

Figure 5.7 shows the performance of selected models while masking the visual or textual features of the sub-sampled validation mini-batches. Masking visual features significantly lowers the HA scores, and masking textual features significantly lowers the OUV scores. This is a natural and consistent behaviour considering how those labels were originally derived: in Bai et al. (2022), HA labels were generated using images only and OUV labels were generated using texts only. In this study, however, the models have access to both textual and visual features when making classifications on both HA and OUV categories. GCN-KNN was the most robust model against the masking of visual features since the KNN graph structure $\boldsymbol{A}^{\text{KNN}}$ was computed before masking, unconsciously leaking the association information of visually similar images (and possibly their HA labels) to the models being trained. All graph-based models performed better than the graph-free MLP at HA classification while masking visual features, whereas the homogeneous models remained better than random classifier RDC. For OUV classification, Order-3 Jaccard Index of all models became extremely vulnerable and got far worse than RDC after masking textual features, since the requirement of being larger than $1/(n+1)$ in

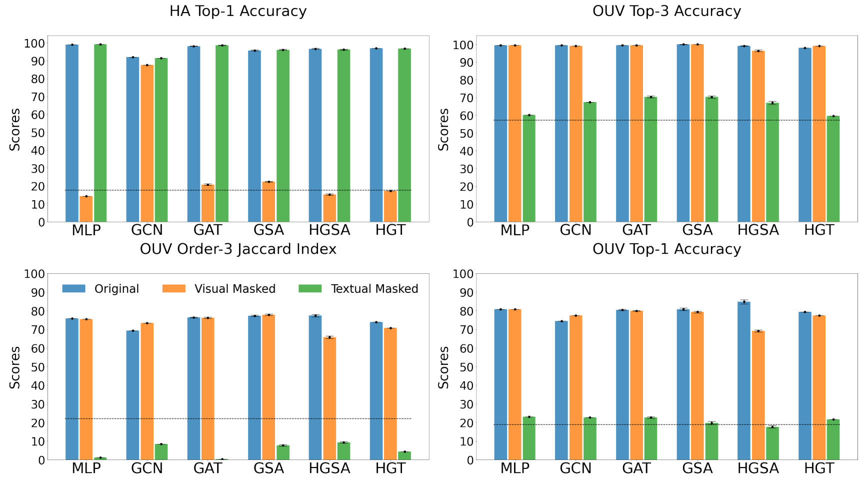


FIG. 5.7 The performance of all selected model checkpoints on the evaluation metrics when masking visual or textual features of mini-batches. The performance of the prior-based random classifier RDC in Table 5.3 is marked with dashed lines.

Equation (5.22) cannot be easily fulfilled when models get uncertain of their predictions. Top-3 OUV Accuracy shows that almost all graph-based models (except for HGT) performed better than MLP (which was also better than RDC) while masking textual features, implying that those models managed to learn the missing textual information of a post from its neighbours, which is only possible on graphs. However, such an effect is not obvious for Top-1 OUV Accuracy, where most models performed only slightly better than RDC.

Figure 5.8 shows the relative performance change of all graph-based models using different graph structures, compared to the original links. GCN trained on KNN graph $\boldsymbol{A}^{\mathrm{KNN}}$ performed significantly better than the original links in all metrics, while GAT and GSA performed slightly worse on KNN graph, suggesting the necessity of using GCN-KNN as the selected candidate model in Table 5.3 and 5.4. Changing graph structure only slightly lowers the performance on GAT and GSA, while not affecting HGT at all. This seems to suggest that these models work as long as there is some graph structure marking the relationship of data points, indifferent of the type of links. Meanwhile, GCN and HGSA are more dependent on the links used for inference.

The various behaviours imply that the selected models are divergent enough, suggesting that aggregating the prediction results to form an ensemble is both necessary and beneficial. The discussion on the complex effects of the model performance, however, falls out of the scope of this paper and invites further investigations in future studies.

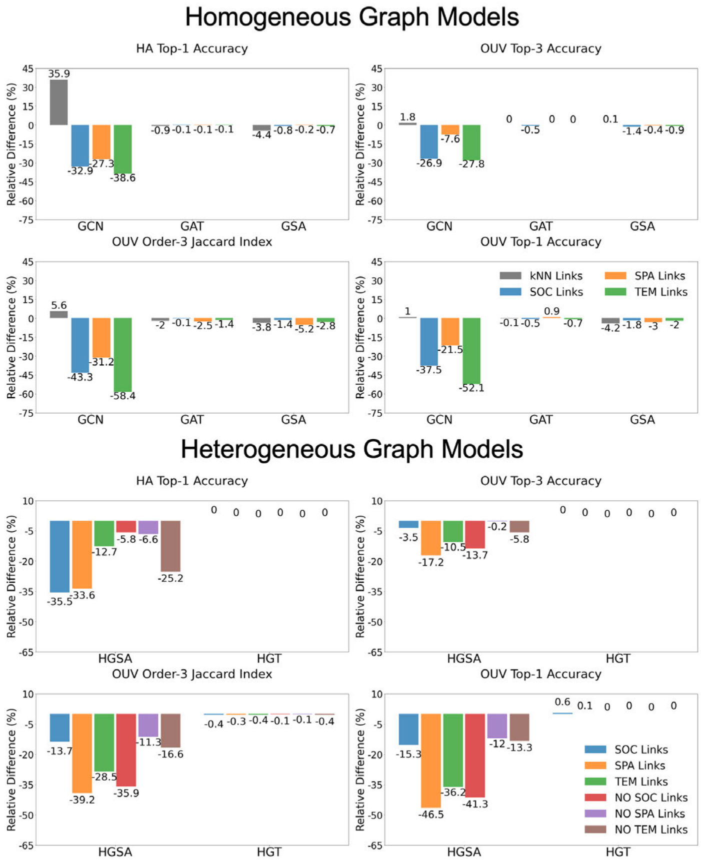


FIG. 5.8 The relative performance change of homogeneous and heterogeneous graph models directly evaluated on sub-graphs with one or two of the link types in $\{\boldsymbol{A}^{\text{TEM}}, \boldsymbol{A}^{\text{SPA}}, \boldsymbol{A}^{\text{SOC}}\}$, compared to the original composed links $\boldsymbol{A}$. The models with KNN links $\boldsymbol{A}^{\text{KNN}}$ were trained separately.

5.6.4 Association of Features and Labels

Figure 5.9 shows the co-occurrence matrices of OUV and HA categories as heatmaps, where frequent OUV-HA pairs imply the association of abstract OUV selection criteria and substantial Heritage Attributes. The four matrices on both post-level labels $\hat{\boldsymbol{Y}}$ and spatial-level labels $\hat{\boldsymbol{\mathcal{Y}}}$ in both VEN and VEN-XL datasets are similar to each other. The spatial-level distribution on VEN-XL is the most sparse (and concentrated) among the four matrices where most OUV-HA pairs focused on the large classes, i.e., Criteria (iii) and (vi) for OUV and Urban Form Element for HA. A similar yet more extreme pattern can be observed in Figure B.1 in Appendix B when the parameter α gets larger, pushing the diffused spatial nodes label array $\boldsymbol{\mathcal{Y}}$ to a uniform-like distribution, suggesting possible “over-smoothing”. A few OUV-HA pairs always stand out as associated categories in those co-occurrence matrices: 1) As the most

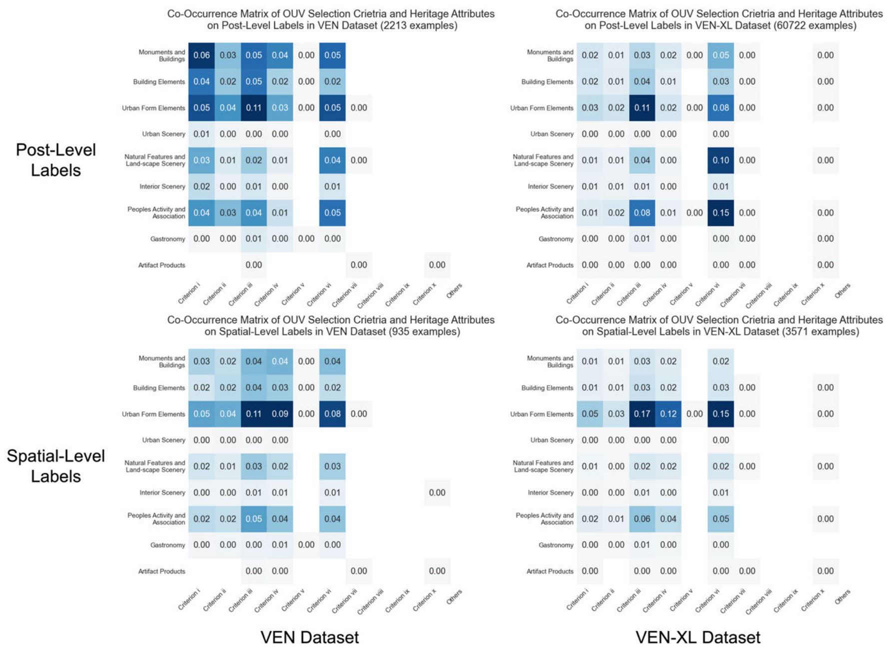


FIG. 5.9 The normalized co-occurrence matrix heatmaps $\boldsymbol{O}$ of the OUV and HA categories in post-level label array $\hat{\boldsymbol{Y}}$ and spatial-level label array $\hat{\boldsymbol{\mathcal{Y}}}$ in both VEN and VEN-XL datasets.

common HA category, the Urban Form Elements always associate strongly with Criteria (iii), (iv) and (vi), suggesting that when people post about testimony of past, architecture type, and human-life-related traditions, they are usually immersed in the urban context of streets and squares; 2) The second largest HA category about People's Activity also associate strongly with Criteria (iii) and (vi), since they have obvious connections with human; 3) As expected, the most associated OUV category with Monuments and Building is Criterion (iv) about architecture typology, and that with Building Element is Criterion (iii) about testimony for a [possibly lost] tradition; 4) The most unexpected associations are the ones for Natural Features and Landscape Scenery, where the most relevant Criterion (vii) about natural beauty is always present but not in a dominant position, which has also been taken by Criterion (iii) and (vi). The pattern of OUV and HA category distribution will be further dis-aggregated and mapped spatially in Section 5.6.5 for detailed inspection.

Figure 5.10 visualizes the explainable features that are shown to be important for classifying the nodes into each OUV and HA category, effectively forming a lexicon of features for the categories as a bipartite graph. The contribution of features is interrelated to OUV/HA categories. For example, the recognized scene of "Canals in Urban Environment" and the SUN attribute of "Open Area" from an image both contribute generally to almost all OUV/HA categories, especially on Criteria (iii)(vi) and "Urban Form Element", while "Open Area" has less to do with "Interior Scenery", "Building Elements", and "People's Activity and Association". While HA category "Interior Scenery" could be inferred with a limited range of features such as

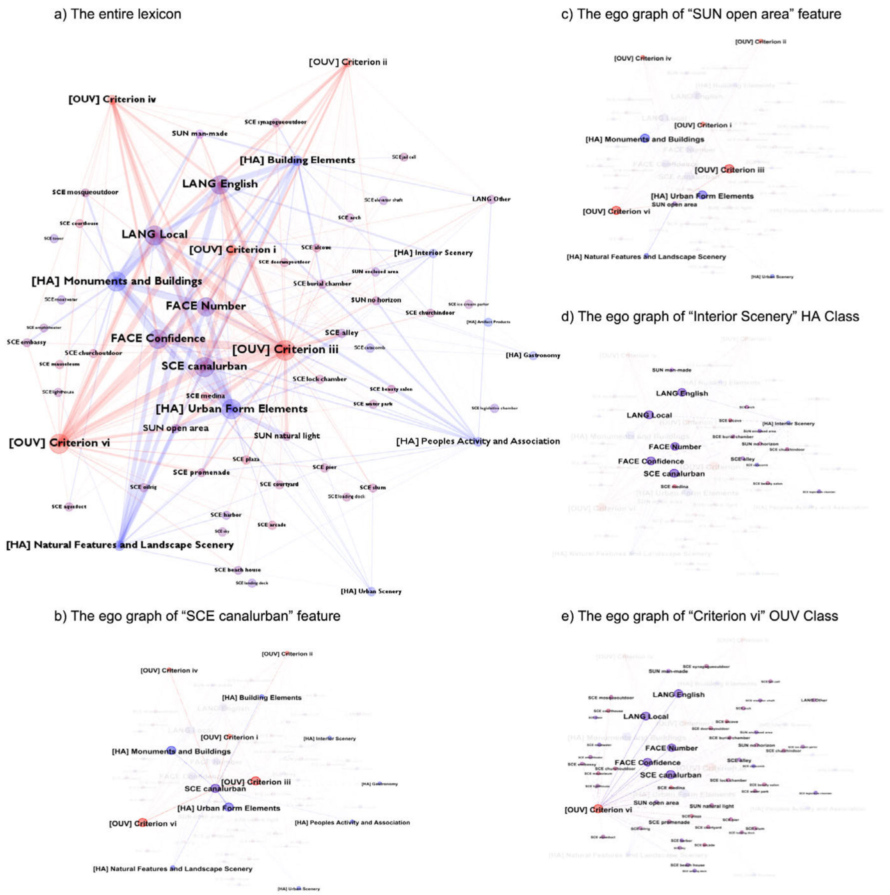


FIG. 5.10 The bipartite graph of feature nodes and OUV/HA category nodes showing the relative importance for explainable features while classifying the nodes belonging to each OUV and HA category. The larger a feature node is, the more this feature appeared in the top-250 important features while classifying a node based on GNNExplainer. The edge weights show the number of times the features contributed to the categories. Only nodes with a larger weighted degree of 8 are shown. Red lines are associations for OUV classes and blue lines for HA. Sub-figures b-e show ego graphs (a sub-graph of the entire lexicon in sub-figure a) around a specific feature or category node. "SCE" denotes scene category within Zhou et al. (2017); "SUN" denotes SUN attribute category in Patterson and Hays (2012); "LANG" denotes the detected language and "FACE" denotes face recognition results from Bai et al. (2022).

"Enclosed Area" and "Arch", OUV Criterion (vi) could be inferred from a large variety of visual and textual features, depending on the type of human activity taking place. The face recognition and language detection results appear to contribute universally to the classification of most categories, which could be possibly explained that the presence of human faces and the original languages of posts provide additional information that could not be inferred from features extracted with scene recognition models originally trained with images with few people and language models trained with English texts. However, among all visual and textual features, explainable ones are usually less informative than the higher-level hidden features, as can be seen in Figure B.2. More concrete investigations are invited to explain this complex pattern

5.6.5 Mapping of Heritage Cultural Significance

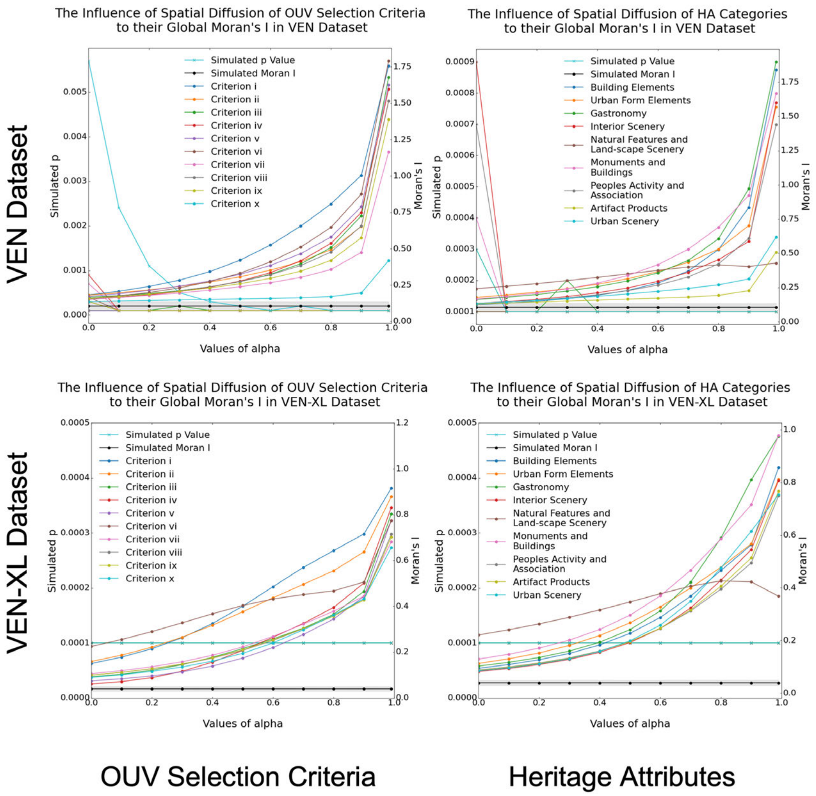


FIG. 5.11 The change of global Moran's I of each OUV and HA category when the diffusion parameter α changes in VEN and VEN-XL. A simulated distribution of expected values of I based on 9999 permutations is used to estimate the p values.

Figure 5.11 demonstrates that the global Moran's I for OUV and HA categories gradually increase as the diffusion parameter α ascends. For most categories in VEN and all in VEN-XL, a spatial auto-correlation is significant after Bonferroni correction ($p < .025/20$) even before diffusion compared to the permutated distributions, confirming the First Law of Geography. For smaller α values, the increases in Moran's I are not drastic, yet effectively further decrease the simulated p values. The largest value of $\alpha = 0.99$ yields extreme I values larger than 1 in VEN. This suggests that choosing a relatively small value for α could enhance the spatial pattern of the categories without disturbing their distributions too much. Note the expected value (mean) of I according to simulation is not the conventional $-1/(N-1)$, since the weight matrix $\boldsymbol{W}$ used here has non-zero diagonal entries and is not row-standardized. However, Figure B.3 shows a similar pattern with the conventional weight matrix for computing Moran's I as defined in Equation (B.14).

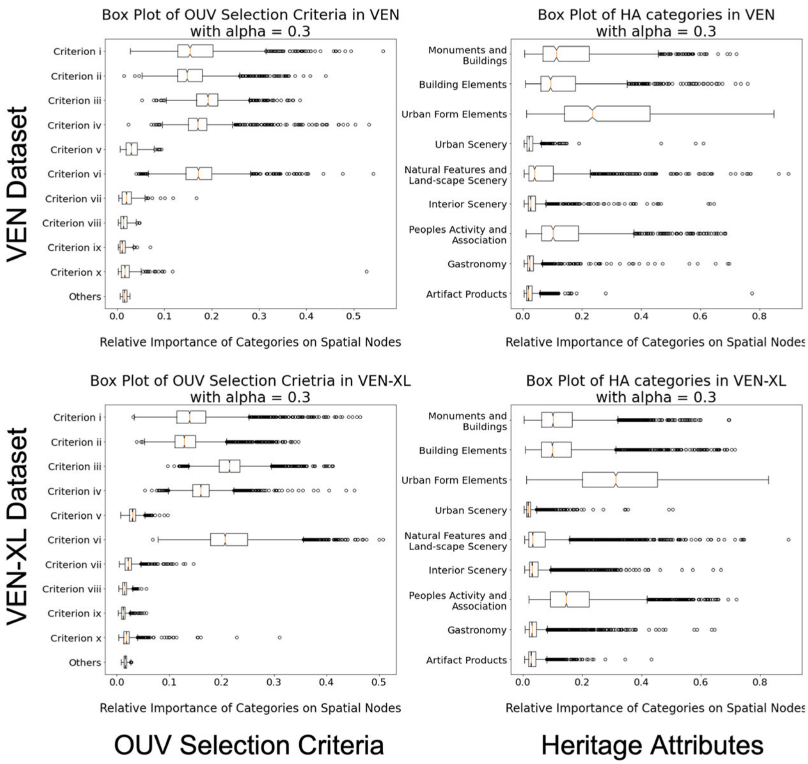


FIG. 5.12 The box plots of each OUV and HA category demonstrating the distributions of spatial node labels $\mathcal{Y}$ in both VEN and VEN-XL datasets.

FIG. 5.13 The geographical distribution of OUV categories in VEN-XL based on the spatial diffusion of labels. The nodes with high ranges of value for each category under equal-interval division are visualized as circles, the size of which demonstrates the number of posts distributed near the spatial node, while those nodes with a significant local Moran's I are shown with dashed borders. Three demonstrative photos and one comment from "hotspot" areas of categories are given below each map.

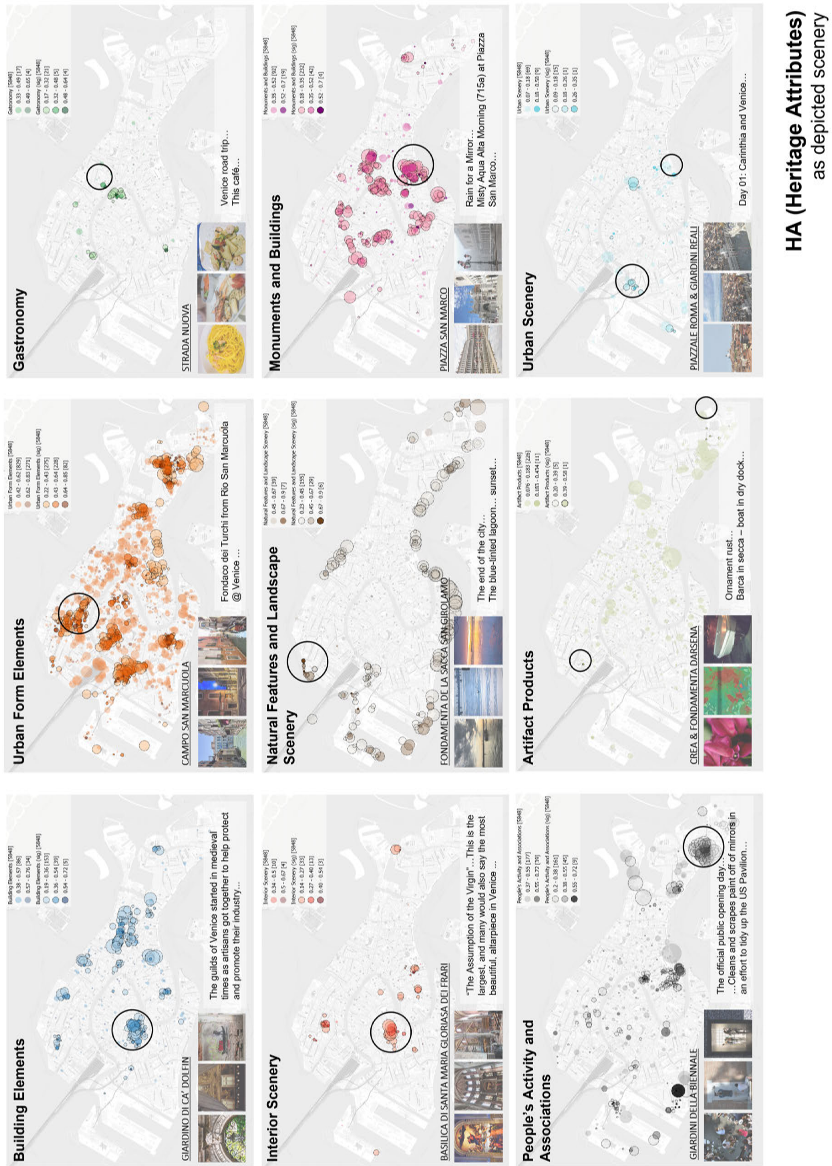


FIG. 5.14 The geographical distribution of HA categories in VEN-XL based on the spatial diffusion of labels. The nodes with high ranges of value for each category under equal-interval division are visualized as circles, the size of which demonstrates the number of posts distributed near the spatial node, while those nodes with a significant local Moran's I are shown with dashed borders. Three demonstrative photos and one comment from "hotspot" areas of categories are given below each map.

The following sections will use $\alpha = 0.3$ for demonstrative purposes of exploratory spatial data analysis. The distribution of spatial node labels $\mathcal{Y}$ in Figure 5.12 also demonstrates a consistent pattern in VEN and VEN-XL: 1) five OUV and HA categories are relatively more dominant than the others; 2) the confidence of OUV labels for spatial nodes are generally lower than HA labels since OUV categories have to be sometimes inferred without textual information; 3) whereas the less dominant categories have lower means and quantile values, the “outliers” point to the exceptional spatial nodes representing specific OUV and HA categories. It further shows that although none of Criteria (vii) - (x) are inscribed with Venice in WHL, scarce cases related to Criteria (vii) and (x) can still be found.

Figures 5.13 and 5.14 demonstrate the final maps of OUV and HA categories identified from Flickr showing their spatial distributions and auto-correlation patterns, together with illustrative examples. The magnitude of HA categories is generally higher than OUV, as also pointed out in Figure 5.12. Almost all categories display spatial patterns of “hotspots” of high values appearing at nearby places, justified with significant local Moran’s $\boldsymbol{I}$. Some categories are spread all over Venice, e.g., OUV Criterion (iii) about Testimony and HA Urban Form Elements, due to their universal nature, while others are much more concentrated at dedicated spots, e.g., OUV Criterion (iv) about Architecture Typology and HA People’s Activity and Associations. Even though some categories are less present with far more limited range, e.g., OUV Criterion (v) about Land-Use and HA Artifact Product, the methodology does manage to find relevant spatial spots with posts of images and/or comments related to the topic. The OUV-HA pairs generally believed to associate with each other, such as Criterion (iv) about Architecture Typology and HA Monuments and Buildings, Criterion (vi) about Human Association and HA People’s Activity and Associations, and Criterion (vii) about Natural Beauty and HA Natural Features and Landscape Scenery, partly overlap with each other, yet not totally identical, showing the nuances of the concepts reflected in social media posts. Interestingly, the hotspot visualization and illustrated examples prove that Venice is more than conventionally popular destinations such as the Piazza San Marco and Ponte di Rialto. Other places including churches, piazza, campo, gardens, exhibition venues, and even normal streets are also attracting people and making them realize the beauty of the city with different focal points.

Additionally, Figure 5.15 visualizes some typical posts of each OUV and HA category irrespective of their geographical locations, which can also be beneficial information for heritage scholars. Further visualizations, comparisons, and discussions of the spatial mapping of OUV and HA categories identified with the proposed methodology can be found in Appendix B with Figures B.4 till B.6.

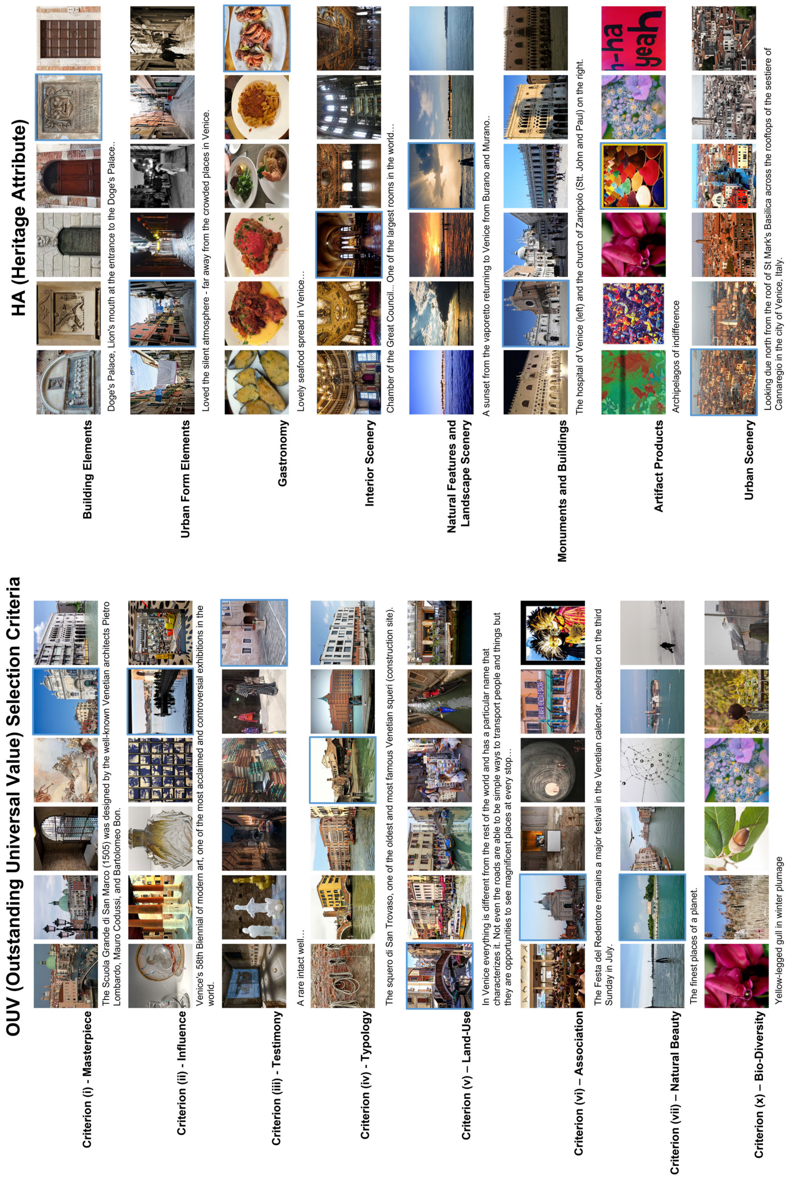


FIG. 5.15 Post-level demonstrations of images and/or comments that have the largest logits for OUV and HA categories. For each category, six typical images and one comment are visualized, both are mostly among top-10 entries. The corresponding image to the comment is highlighted with a blue frame. No images from HA category People's Activities and Association are shown since the typical images always have a large portion of human faces on them.

5.7 Discussion

5.7.1 Documenting Knowledge for Heritage Studies

The initial motivation for conducting this research is to propose a "knowledge documentation and mapping tool of cultural and natural heritage characteristics", especially the heritage values and attributes, for the "recognition of cultural significance and diversity", in support of the HUL approach (UNESCO, 2011). Instead of actively engaging the civil society to contribute to the narratives with their knowledge and values a city they live in or visit conveys to them, this study makes use of the existing information on social media with a real-world dataset to make exploratory analyses. The term "exploratory" is crucial for interpreting the findings and applying the methodology in practice. It functions as a complementary tool to help heritage managers and authorities explore the voices of the public on social media, either to confirm or to challenge/ adjust their hypotheses over the spatial distribution of the cultural significance in a city. For example, one could be affirmative ahead of time that tourists are over-crowded in only a few popular spots in Venice such as San Marco and Rialto, and that the beauties hidden in the other places are easily over-looked. However, the mapping practice in Figures 5.13 and 5.14 suggests that Flickr users are indeed exploring a broad range of places all over the island, attracted by different types of cultural significance reflecting various heritage values and attributes. Heritage experts and practitioners could inspect the social media posts located nearby unexpected places revealed with cultural significance to get inspiration for further planning actions in pursuit of social inclusion (Waterton et al., 2006; Bai et al., 2021b).

In order to fully reflect the need for inclusive heritage management processes, further studies are needed to: 1) quantitatively and qualitatively collect ideas from broader communities, especially from those who do not use social media, for a fair comparison to justify the representativeness of similar studies; 2) apply the same methodology and test the models in a wider selection of case studies in different geographical and topological contexts, as to evaluate the generalizability of the proposed workflow; 3) update the OUV selection criteria and Heritage Attributes label categories with other frameworks, tailor-made for the research interests and objectives in their own usage scenarios. Furthermore, UNESCO Statements of OUV are assumed to include elements from both heritage values and attributes. This study completes one side of the puzzle of analysing the association between OUV Selection Criteria and Heritage Attributes and further mapping them spatially. Future studies could complete the other end by employing analyses and mapping practices under the classification framework of Heritage Values (Pereira Roders, 2007; Tarrafa Silva and Pereira Roders, 2010; Foroughi et al., 2022).

5.7.2 A Mapping Tool for Urban Explorations

Nevertheless, as a mapping tool in full mathematical details, the application scenarios of this study could go beyond heritage studies. In principle, given a back-end spatial network, the mathematical constructs of attention-based information aggregation and graph diffusion processes described in Section 5.3.3 could also be fed with any sort of input feature array obtained from posts instead of only the output labels to be aggregated and mapped on spatial nodes. For example, one could map the SUN attribute feature of "biking" or "socializing" to explore the activities distributed in a city or map the number and proportion of faces in the posted images to observe the crowdedness, or even map some low-level visual features to mine the patterns of architectural style (Sun et al., 2022). In this sense, the proposed methodology could be generalized in applications of measuring safety (by diffusing crime rate), vitality (by mapping diversity of human activity), and popularity of urban spaces (by plotting the crowdedness), where it diffuses any sort of human-generated information onto a spatial network with inherent connectivity patterns. It is clearly related to the location-led place profiling approach in Lai (2019), whereas the categories in this study go beyond the text-only clustering of urban activities.

When making spatial statistical inferences, like other similar spatial analyses, the result is dependent on how the spatial connectivity and weights are measured. An interesting alternative could be aggregating the posts on regular spatial grids of different resolutions and using queen/rook-based contiguity as weight matrix to perform the diffusion (Anselin, 2003; Rogerson, 2021). As such, the label information will be rasterized and can be easily overlayed and collated in GIS platforms with other global and local datasets (Esch et al., 2017; Bekker, 2020). Moreover, the diffusion-mapping process proposed by this paper can be seen as an alternative and supplement to the conventional kernel-density heatmaps, which is further elaborated upon in Appendix B.

Even though there are originally three types of graph links in Heri-Graphs (Bai et al., 2022), this study only discovers the mapping, aggregation, and diffusion on the spatial-level nodes for pragmatic reasons, since spatial mapping is the most desired option. However, other than diffusing spatial-level node labels, mapping the foci and interests to temporal nodes (time periods in history) and social nodes (groups of social media users) that are derivable from $\boldsymbol{A}^{\mathsf{TEM}}$, $\boldsymbol{A}^{\mathsf{SOC}}$ can also answer interesting research questions. For instance, other than the spatial bipartite relation $\boldsymbol{B}$ mentioned in Section 5.3.3, the temporal bipartite relation $\boldsymbol{B}^{\mathsf{TEM}}$ (mapping the posts to the unique sorted weekly timestamps) and the tri-diagonal temporal adjacency matrix $\boldsymbol{W}^{\mathsf{TEM}}$ (recording the consecutive patterns of the weekly timestamps) can be used to substitute the aggregation computation in Equation (5.11) and the diffusion computation in Equation (5.16). Here a similar relationship also holds according to Bai et al. (2022): $\boldsymbol{A}^{\mathsf{TEM}} = \left(\boldsymbol{B}^{\mathsf{TEM}}\boldsymbol{W}^{\mathsf{TEM}}\boldsymbol{B}^{\mathsf{TEM}\mathsf{T}} > 0\right) = \boldsymbol{B}^{\mathsf{TEM}}\left(\boldsymbol{W}^{\mathsf{TEM}} > 0\right)\boldsymbol{B}^{\mathsf{TEM}\mathsf{T}} \in \{0,1\}^{K\times K}$. Every other module of the methodological framework visualized in Figure 5.1 is still valid, except that the aggregation and diffusion would be conducted on the temporal-level graph. Analogue to the 2-dimensional mapping of spatial labels presented in this study, 1-dimensional mapping of temporal labels could result in attributed timelines showing the development of different label and/or feature categories. Similar mapping computations can be conducted for the social graph (social network of users on social media). These effects will be discovered in follow-up studies in various use cases.

5.7.3 A Machine Learning Application

It is worth noting that the labels generated in VEN and VEN-XL datasets were originally not annotated by humans, but rather by a few ML models, or more specifically, MLP models as connectors between hidden features and output soft-label vectors (Bai et al., 2022). Therefore, using more complex graph-based GNN models in this study to replicate labels generated by simple MLP seems a reversed knowledge distillation process (i.e., confident students teaching a group of teachers) (Gou et al., 2021). It has also been shown in the most recent literature that simple MLPs using a Bag of Words could outperform most graph-based models in text classification tasks (Galke and Scherp, 2022). This trend is again visible here for some of the metrics in Table 5.3 and 5.4. However, this paper also shows that GNN models have other benefits in terms of inductive learning and missing input data, as demonstrated in Figure 5.7. Considering that the pseudo-labels of training and validation sets came from data-points of high prediction confidence (with high top-n prediction logits) and consistency (with similar prediction results by different trained models), the philosophy behind the training process in this paper also resembles the self-training strategy, where the originally unlabelled samples that end up with top prediction confidence in one round of training are added to the next round as labelled ones (Li et al., 2018; Sun et al., 2020; Wang et al., 2022b). The indications of such similarities mentioned above to the methodology and results are, however, out of the scope of this paper.

The classification performance can be further improved by adding humans in the loop with active learning (Prince, 2004). An important challenge given by the Heri-Graphs dataset that is not yet solved in this study is the imbalance of categories and the extreme sparsity in some small classes. This is a pragmatic difficulty since Heri-Graphs were originally created with real-world social media data for an application in heritage studies and did not enforce the categories to be balanced (Bai et al., 2022). However, future studies could implement data augmentation on the small classes in the unbalanced training data to further improve the classification performance. Few-shot learning and Zero-shot learning techniques can also be implemented (Sung et al., 2018). Further specific investigations are also invited to discover the effect of different graph structures, e.g., the original weighted adjacency matrices instead of binary ones, for the training and diffusion processes.

While applying the obtained model from this study to other case study cities in the world, such as Amsterdam and Suzhou also collected by Bai et al. (2022), two options could be considered, following the conventional GNN terminology of transductive and inductive learning (Kipf and Welling, 2016; Yang et al., 2016; Hamilton et al., 2017; Veličković et al., 2017). By stacking the graphs of different datasets together before sampling sub-graphs, the pre-trained models could be used to fine-tune the new models while the test data could be seen together with training data, entailing a transductive learning setting. On the contrary, directly applying the trained model here to other cases would mean that the new test data are totally unobserved during training, entailing an inductive learning setting. Researchers are welcome to explore the advantages and drawbacks of either option according to their own application scenarios.

5.7.4 Related Works about the Workflow

The proposed workflow in Figure 5.1 takes inspiration from many different fields.

The first main component, i.e., semi-supervised learning of multiple models (Section 5.3.1), was the initial motivation of Graph Neural Networks (Kipf and Welling, 2016) and has been a topic extensively studied in computer science, with or without a graph structure (Blum and Mitchell, 1998; Zhou and Li, 2010; Yang et al., 2016; Hamilton et al., 2017; Veličković et al., 2017; Li et al., 2018; Ma and Tang, 2021). The extra complexity of this study from a real-world dataset is that the semi-supervised learning process needs to react to two modalities (visual and textual, among which the textual features might be missing) and perform well in two classification tasks (OUV and HA) with a multi-graph structure (composed of spatial, temporal, and social links). The most closely relevant study in the literature is Liu and De Sabbata (2021), which did not include the other two components, as already mentioned in Section 5.1.

The second main component, i.e., aggregating model predictions (Section 5.3.2), leverages the concept from Ensemble Learning (Schapire and Singer, 1998; Zhou, 2012; Sagi and Rokach, 2018). The approach of computing an aggregated prediction vector as a weighted average of multiple models is similar to the “soft voting” mechanism (Zhou, 2012). Outside the field of computer science, aggregating the opinions of multiple actors based on their agreement and confidence is also an active topic in decision science (Stone, 1961; Budescu and Rantilla, 2000; Budescu and Yu, 2007). However, it is a technical innovation in this study to assign a class-level agreement vector to each aggregated prediction by computing SVD on the matrices composed of the original predictions of models in the ensemble, which is informative for evaluating the effect of aggregation.

The third main component, i.e., aggregating and diffusing post-level labels onto spatial graphs (Section 5.3.3), contains the most methodological innovations of the proposed workflow. As already pointed out in Section 5.3.3, the processes of aggregating and diffusing information on graphs resemble the operations of graph pooling and graph filtering, respectively (Ma and Tang, 2021), thus the

Equations (5.10) and (5.12) can be formally similar to the ones in Graph Neural Network literature (Veličković et al., 2017; Lee et al., 2019; Knyazev et al., 2019). However, they are for different purposes: instead of computing intermediate representations for the training loop, in this paper, these Equations are used to summarize the post-level information and assign it to spatial nodes, which were initially unlabelled in nature. The exchange of label information on bipartite graphs as shown in Equation (5.11) also makes it different from the Label Propagation Algorithm (Zhu and Ghahramani, 2002; Huang et al., 2020; Wang and Leskovec, 2021), albeit the latter approach has the same spirit of diffusing soft labels based on the connectivity of nodes. Even though plenty of studies attempted to draw the label categories of social media posts on spatial maps, the majority of them either directly plotted the posts as unconnected data points (Huang et al., 2019; Liu and De Sabbata, 2021), or provided only the predominant categories or word-clouds for each detected/predefined cluster (Hu et al., 2015; Lai et al., 2017; Ginzarly et al., 2019), or created a kernel-density heatmap to show the distribution without a mathematical expression for the spatial nodes (Lansley and Longley, 2016; Bekker, 2020; Kang et al., 2021). The proposed method has the benefit of keeping a soft label structure (as probability distribution) for each discrete spatial unit (street intersections), which is also algebraically derivable. Further advantages of the proposed mapping process with label diffusion will be elaborated with Figures B.4, B.5, and associative discussions in Appendix B.

Interestingly, even though the process of aggregating and diffusing labels is rare in spatial mapping, an essentially similar approach can be found on social networks for developing recommendation systems, where information is diffused on a tripartite graph of user-image-tag (Mao et al., 2016; Zhang et al., 2017; Wang et al., 2018), which could be regarded an analogue of the space-post-label triplet in this study. Furthermore, an interesting connection can also be found in a few recent studies with label diffusion processes during semantic segmentation on point clouds (Mascaro et al., 2021; Deng et al., 2022; Liao et al., 2022) and in a study predicting the effect of drug-disease association using diffusion on a bipartite graph (Xie et al., 2021).

Despite all the resemblances mentioned above, an additional innovation in this study is to bring all the components from different fields together in a holistic workflow and adapt them accordingly to solve a real-world research problem: mapping cultural significance categories obtained from social media platforms. To the best of the authors' knowledge, this study is the first to combine all these aspects with interdisciplinary knowledge, especially as the label category of interest is a unique example from the field of heritage studies, dominated by expert-based qualitative approach.

5.8 Conclusions

This paper proposes a workflow to obtain social perception maps concerning the cultural significance of places located in an urban spatial network using social media information. Several graph neural network models are trained with semi-supervised learning on attributed graph datasets with visual and textual nodal features of user-generated posts, effective on various evaluation metrics. The predicted post-level soft labels are aggregated considering the confidence and agreement of models, which are further aggregated and diffused on a back-end spatial network to obtain spatial-level labels. The distributions of spatial labels on heritage-related cultural significance categories are tested with spatial statistics and mapped with examples. The entire workflow is mathematically explained in detail and tested with the case study of Venice, shown to provide reasonable maps of cultural significance. The workflow can also be applied to other cities worldwide as a knowledge documentation tool collecting the voices of communities posting on the internet, with the ultimate goal of promoting socially inclusive heritage management processes, as suggested by the UNESCO Historic Urban Landscape approach. Moreover, the proposed methodology of diffusing human-generated location-based information onto the spatial network also has the potential for broader use scenarios in different domains of urban studies.

References

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In Aggarwal, C. C., editor, Social Network Data Analytics, chapter 1, pages 1–15. SPRINGER.

Amato, F., Cozzolino, G., Di Martino, S., Mazzeo, A., Moscato, V., Picariello, A., Romano, S., and Sperlí, G. (2016). Opinions analysis in social networks for cultural heritage applications. Smart Innovation, Systems and Technologies, 55:577–586.

Anselin, L. (1995). Local indicators of spatial association—lisa. Geographical analysis, 27(2):93–115.

Anselin, L. (2003). An introduction to spatial autocorrelation analysis with geoda. Spatial Analysis Laboratory, University of Illinois, Champagne-Urbana, Illinois.

Bai, N., Luo, R., Nourian, P., and Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage statements of "Outstanding Universal Value" with soft labels. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 366–384, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bai, N., Nourian, P., Luo, R., Cheng, T., and Roders, A. P. (2023). Screening the stones of venice: Mapping social perceptions of cultural significance through graph-based semi-supervised classification. ISPRS Journal of Photogrammetry and Remote Sensing, 203:135–164.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2022). Heri-graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. ISPRS International Journal of Geo-Information, 11(9).

Bai, N., Nourian, P., and Pereira Roders, A. (2021b). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.

Baltrusaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2):423–443.

Bandarin, F. and Van Oers, R. (2012). The historic urban landscape: managing heritage in an urban century. John Wiley & Sons.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In Proceedings of the International AAAI Conference on Web and Social Media, volume 3, pages 361–362.

Baumer, E., Elovic, E., Qin, Y., Polletta, F., and Gay, G. (2015). Testing and comparing computational approaches for identifying the language of framing in political news. In Proceedings of the 2015 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, pages 1472–1482.

Bekker, R. (2020). Creating insights in tourism with flickr photography, visualizing and analysing spatial and temporal patterns in venice. Master's thesis, Rijksuniversiteit Groningen.

Benzi, M. and Klymko, C. (2014). A matrix analysis of different centrality measures. arXiv preprint arXiv:1312.6722.

Bertocchi, D. and Visentin, F. (2019). "the overwhelmed city": Physical and social over-capacities of global tourism in venice. Sustainability, 11(24):6937.

Bigne, E., Ruiz, C., Cuenca, A., Perez, C., and Garcia, A. (2021). What drives the helpfulness of online reviews? a deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. Journal of Destination Marketing & Management, 20:100570.

Blum, A. and Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In Proceedings of the eleventh annual conference on Computational learning theory, pages 92–100.

Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65:126–139.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. Journal of mathematical sociology, 2(1):113–120.

Boy, J. D. and Uitermark, J. (2017). Reassembling the city through instagram. Transactions of the Institute of British Geographers, 42(4):612–624.

Budescu, D. V. and Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. Acta psychologica, 104(3):371–398.

Budescu, D. V. and Yu, H.-T. (2007). Aggregation of opinions based on correlated cues and advisors. Journal of Behavioral Decision Making, 20(2):153–177.

Calvino, I. (1978). Invisible cities. Houghton Mifflin Harcourt.

Cao, R., Tu, W., Yang, C., Li, Q., Liu, J., Zhu, J., Zhang, Q., Li, Q., and Qiu, G. (2020). Deep learning-based remote and social sensing data fusion for urban region function recognition. ISPRS Journal of Photogrammetry and Remote Sensing, 163:82–97.

Cartwright, W. E. (2010). Addressing the value of art in cartographic communication. ISPRS Journal of Photogrammetry and Remote Sensing, 65(3):294–299.

Chen, M., Wei, Z., Huang, Z., Ding, B., and Li, Y. (2020). Simple and deep graph convolutional networks. In International Conference on Machine Learning, pages 1725–1735. PMLR.

Chen, Y. (2021). An analytical process of spatial autocorrelation functions based on moran's index. PloS one, 16(4):e0249589.

Cheng, T. and Wicks, T. (2014). Event detection using twitter: A spatio-temporal approach. PloS one, 9(6):e97807.

Cho, N., Kang, Y., Yoon, J., Park, S., and Kim, J. (2022). Classifying tourists' photos and exploring tourism destination image using a deep learning model. Journal of Quality Assurance in Hospitality & Tourism, pages 1–29.

Cosgrove, D. (1982). The myth and the stones of venice: an historical geography of a symbolic landscape. Journal of Historical Geography, 8(2):145–169.

Crandall, D., Backstrom, L., Huttenlocher, D., and Kleinberg, J. (2009). Mapping the world's photos. WWW'09 - Proceedings of the 18th International World Wide Web Conference, pages 761–770.

Deng, Y., Wang, M., Yang, Y., and Yue, Y. (2022). Hd-ccsom: Hierarchical and dense collaborative continuous semantic occupancy mapping through label diffusion. In 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 2417–2422. IEEE.

Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., and Strano, E. (2017). Breaking new ground in mapping human settlements from space–the global urban footprint. ISPRS Journal of Photogrammetry and Remote Sensing, 134:30–42.

Fey, M. and Lenssen, J. E. (2019). Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428.

Foroughi, M., de Andrade, B., and Pereira Roders, A. (2022). Peoples' values and feelings matter: Participatory heritage management using social media. In Muntañola, J., editor, Artificial Intelligence and Architectural Design, volume 33, pages 107–120. Oficina de Publicacions Acadèmiques Digitals de la UPC.

Galke, L. and Scherp, A. (2022). Bag-of-words vs. graph vs. sequence in text classification: Questioning the necessity of text-graphs and the surprising strength of a wide MLP. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 4038–4051, Dublin, Ireland. Association for Computational Linguistics.

Gardner, M. W. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric environment, 32(14-15):2627–2636.
GeoMatt22 (2020). Similarity metrics for more than two vectors? Stack Exchange. (version: 2020-12-10) (access date: 2022-08-31).
Ginzarly, M., Pereira Roders, A., and Teller, J. (2019). Mapping historic urban landscape values through social media. Journal of Cultural Heritage, 36:1–11.
Gomez, R., Gomez, L., Gibert, J., and Karatzas, D. (2019). Learning from #barcelona instagram data what locals and tourists post about its neighbourhoods. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11134 LNCS:530–544.
Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2021). Knowledge distillation: A survey. International Journal of Computer Vision, 129(6):1789–1819.
Gould, P. R. (1967). On the geographical interpretation of eigenvalues. Transactions of the Institute of British Geographers, pages 53–86.
Gustcoven, E. (2016). Attributes of world heritage cities, sustainability by management–a comparative study between the world heritage cities of amsterdam, edinburgh and querétaro. Master's thesis, KU Leuven.
Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in neural information processing systems, 30.
He, Z., Deng, N., Li, X., and Gu, H. (2022). How to "read" a destination from images? machine learning and network methods for dmos' image projection and photo evaluation. Journal of Travel Research, 61(3):597–619.
Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., and Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. Computers, Environment and Urban Systems, 54:240–254.
Hu, Z., Dong, Y., Wang, K., and Sun, Y. (2020). Heterogeneous graph transformer. In Proceedings of The Web Conference 2020, pages 2704–2710.
Huang, Q., He, H., Singh, A., Lim, S.-N., and Benson, A. R. (2020). Combining label propagation and simple models out-performs graph neural networks. arXiv preprint arXiv:2010.13993.
Huang, W. and Li, S. (2016). Understanding human activity patterns based on space-time-semantics. ISPRS journal of photogrammetry and remote sensing, 121:1–10.
Huang, X., Wang, C., Li, Z., and Ning, H. (2019). A visual–textual fused approach to automated tagging of flood-related tweets during a flood event. International Journal of Digital Earth, 12(11):1248–1264.
ICOMOS, A. (2013). The Burra Charter: The Australia ICOMOS charter for places of cultural significance 2013. Australia ICOMOS Incorporated.
Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. PloS one, 9(6):e98679.
Jokilehto, J. (2007). Aesthetics in the world heritage context. In Values and Criteria in Heritage Conservation, pages 183–194. Polistampa.
Jokilehto, J. (2008). What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties. Technical report, ICOMOS, ICOMOS Berlin.
Kang, Y., Cho, N., Yoon, J., Park, S., and Kim, J. (2021). Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos. ISPRS International Journal of Geo-Information, 10(3):137.
Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika, 18(1):39–43.
Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
Knyazev, B., Taylor, G. W., and Amer, M. (2019). Understanding attention and generalization in graph neural networks. Advances in neural information processing systems, 32.
Lai, J. (2019). Urban Place Profiling Using Geo-Referenced Social Media Data. PhD thesis, UCL (University College London).
Lai, J., Cheng, T., and Lansley, G. (2017). Improved targeted outdoor advertising based on geotagged social media data. Annals of GIS, 23(4):237–250.
Lansley, G. and Longley, P. A. (2016). The geography of twitter topics in london. Computers, Environment and Urban Systems, 58:85–96.
LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature, 521(7553):436–444.
Lee, J., Lee, I., and Kang, J. (2019). Self-attention graph pooling. In International conference on machine learning, pages 3734–3743. PMLR.
Li, Q., Han, Z., and Wu, X.-M. (2018). Deeper insights into graph convolutional networks for semi-supervised learning. In Thirty-Second AAAI conference on artificial intelligence, pages 1–8.
Li, Y., Tarlow, D., Brockschmidt, M., and Zemel, R. (2015). Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493.
Liao, L., Chen, W., Xiao, J., Wang, Z., Lin, C.-W., and Satoh, S. (2022). Unsupervised foggy scene understanding via self spatial-temporal label diffusion. IEEE Transactions on Image Processing, 31:3525–3540.
Liu, P. and De Sabbata, S. (2021). A graph-based semi-supervised approach to classification learning in digital geographies. Computers, Environment and Urban Systems, 86:101583.
Ma, Y. and Tang, J. (2021). Deep learning on graphs. Cambridge University Press.

Mao, J., Lu, K., Li, G., and Yi, M. (2016). Profiling users with tag networks in diffusion-based personalized recommendation. Journal of Information Science, 42(5):711–722.
Mascaro, R., Teixeira, L., and Chli, M. (2021). Diffuser: Multi-view 2d-to-3d label diffusion for semantic scene segmentation. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pages 13589–13595. IEEE.
Monteiro, V., Henriques, R., Painho, M., and Vaz, E. (2014). Sensing World Heritage An Exploratory Study of Twitter as a Tool for Assessing Reputation. In Murgante, B and Misra, S and Rocha, AMAC and Torre, C and Rocha, JG and Falcao, MI and Taniar, D and Apduhan, BO and Gervasi, O., editor, COMPUTATIONAL SCIENCE AND ITS APPLICATIONS - ICCSA 2014, PT II, volume 8580 of Lecture Notes in Computer Science, pages 404–419. Univ Minho; Univ Perugia; Univ Basilicata; Monash Univ; Kyushu Sangyo Univ; Assoc Portuguesa Investigacao Operac.
Moran, P. A. (1950). Notes on continuous stochastic phenomena. Biometrika, 37(1/2):17–23.
Nourian, P. (2016). Configraphics: Graph Theoretical Methods for Design and Analysis of Spatial Configurations. TU Delft.
Nourian, P., Rezvani, S., Sariyildiz, I., and van der Hoeven, F. (2016). Spectral modelling for spatial network analysis. In Proceedings of the Symposium on Simulation for Architecture and Urban Design (simAUD 2016), pages 103–110. SimAUD.
Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.
Patterson, G. and Hays, J. (2012). Sun attribute database: Discovering, annotating, and recognizing scene attributes. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2751–2758. IEEE.
Patterson, G., Xu, C., Su, H., and Hays, J. (2014). The sun attribute database: Beyond categories for deeper scene understanding. International Journal of Computer Vision, 108(1):59–81.
Pereira Roders, A. (2007). Re-architecture: lifespan rehabilitation of built heritage. PhD thesis, Technische Universiteit Eindhoven.
Pereira Roders, A. (2019). The Historic Urban Landscape Approach in Action: Eight Years Later. In Reshaping Urban Conservation, pages 21–54. Springer.
Prince, M. (2004). Does active learning work? a review of the research. Journal of engineering education, 93(3):223–231.
Psarra, S. (2018). The Venice Variations: Tracing the Architectural Imagination. UCL press.
QGIS Development Team (2023). QGIS Geographic Information System. Open Source Geospatial Foundation.
Rey, S. J. and Anselin, L. (2007). PySAL: A Python Library of Spatial Analytical Methods. The Review of Regional Studies, 37(1):5–27.
Rogerson, P. and Sun, Y. (2001). Spatial monitoring of geographic patterns: an application to crime analysis. Computers, Environment and Urban Systems, 25(6):539–556.
Rogerson, P. A. (2021). Spatial Statistical Methods for Geography. SAGE Publications Ltd.
Rubinstein, R. Y. and Kroese, D. P. (2013). The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning. Springer Science & Business Media.
Ruskin, J. (1879). The stones of Venice. Crowell.
Ruskin, J. and Quill, S. (2015). Ruskin's Venice: The Stones Revisited. Lund Humphries.
Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4):e1249.
Schapire, R. E. and Singer, Y. (1998). Improved boosting algorithms using confidence-rated predictions. In Proceedings of the eleventh annual conference on Computational learning theory, pages 80–91.
Schlichtkrull, M., Kipf, T. N., Bloem, P., Berg, R. v. d., Titov, I., and Welling, M. (2018). Modeling relational data with graph convolutional networks. In European semantic web conference, pages 593–607. Springer.
Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 815–823.
Stone, M. (1961). The opinion pool. The Annals of Mathematical Statistics, pages 1339–1342.
Sun, K., Lin, Z., and Zhu, Z. (2020). Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 5892–5899.
Sun, M., Zhang, F., Duarte, F., and Ratti, C. (2022). Understanding architecture age and style through deep learning. Cities, 128:103787.
Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H., and Hospedales, T. M. (2018). Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1199–1208.
Tarrafa Silva, A. and Pereira Roders, A. (2010). The cultural significance of World Heritage cities : Portugal as case study. In Heritage and Sustainable Development, pages 255–263, Évora, Portugal.
Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. Economic geography, 46(sup1):234–240.

UNESCO (1972). Convention Concerning the Protection of the World Cultural and Natural Heritage. Technical Report november, UNESCO, Paris.

UNESCO (2008). Operational guidelines for the implementation of the world heritage convention. Technical Report July, UNESCO World Heritage Centre.

UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.

Urry, J. and Larsen, J. (2011). The tourist gaze 3.0. Sage.

Vallat, R. (2018). Pingouin: statistics in python. J. Open Source Softw., 3(31):1026.

VanderWeele, T. J. and Mathur, M. B. (2019). Some desirable properties of the bonferroni correction: is the bonferroni correction really so bad? American journal of epidemiology, 188(3):617–618.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 6000–6010.

Veldpaus, L. (2015). Historic urban landscapes: framing the integration of urban and heritage planning in multilevel governance. PhD thesis, Technische Universiteit Eindhoven.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.

Wang, H. and Leskovec, J. (2021). Combining graph convolutional neural networks and label propagation. ACM Transactions on Information Systems (TOIS), 40(4):1–27.

Wang, L., Han, X., He, J., and Jung, T. (2022a). Measuring residents' perceptions of city streets to inform better street planning through deep learning and space syntax. ISPRS Journal of Photogrammetry and Remote Sensing, 190:215–230.

Wang, P., Luo, H., Obaidat, M. S., and Wu, T.-Y. (2018). The internet of things service recommendation based on tripartite graph with mass diffusion. In 2018 IEEE International Conference on Communications Workshops (ICC Workshops), pages 1–6. IEEE.

Wang, Y., Jin, W., and Derr, T. (2022b). Graph neural networks: Self-supervised learning. In Wu, L., Cui, P., Pei, J., and Zhao, L., editors, Graph Neural Networks: Foundations, Frontiers, and Applications, pages 391–420. Springer Singapore, Singapore.

Waterton, E., Smith, L., and Campbell, G. (2006). The utility of discourse analysis to heritage studies: The burra charter and social inclusion. International journal of heritage studies, 12(4):339–355.

Wu, L., Cui, P., Pei, J., and Zhao, L. (2022). Graph Neural Networks: Foundations, Frontiers, and Applications. Springer Singapore, Singapore.

Xie, G., Li, J., Gu, G., Sun, Y., Lin, Z., Zhu, Y., and Wang, W. (2021). Bgmsdda: a bipartite graph diffusion algorithm with multiple similarity integration for drug–disease association prediction. Molecular Omics, 17(6):997–1011.

Xu, Y., Zhou, B., Jin, S., Xie, X., Chen, Z., Hu, S., and He, N. (2022). A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. Computers, Environment and Urban Systems, 95:101807.

Yang, Z., Cohen, W., and Salakhudinov, R. (2016). Revisiting semi-supervised learning with graph embeddings. In International conference on machine learning, pages 40–48. PMLR.

Ying, Z., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. (2019). Gnnexplainer: Generating explanations for graph neural networks. Advances in neural information processing systems, 32.

Yuster, R. and Zwick, U. (2005). Fast sparse matrix multiplication. ACM Transactions On Algorithms (TALG), 1(1):2–13.

Zancheti, S. M. and Jokilehto, J. (1997). Values and urban conservation planning: some reflections on principles and definitions. Journal of architectural conservation, 3(1):37–51.

Zhan, J., Gurung, S., and Parsa, S. P. K. (2017). Identification of top-k nodes in large networks using katz centrality. Journal of Big Data, 4(1):1–19.

Zhang, C., Song, D., Huang, C., Swami, A., and Chawla, N. V. (2019a). Heterogeneous graph neural network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 793–803.

Zhang, F., Zhou, B., Ratti, C., and Liu, Y. (2019b). Discovering place-informative scenes and objects using social media photos. Royal Society open science, 6(3):181375.

Zhang, J., Yang, Y., Tian, Q., Zhuo, L., and Liu, X. (2017). Personalized social image recommendation method based on user-image-tag model. IEEE Transactions on Multimedia, 19(11):2439–2449.

Zhang, Y. and Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Computers, Environment and Urban Systems, 79:101403.

Zhang, Y., Li, Y., Zhang, E., and Long, Y. (2022a). Revealing virtual visiting preference: Differentiating virtual and physical space with massive tiktok records in beijing. Cities, 130:103983.

Zhang, Y., Zhang, F., and Chen, N. (2022b). Migratable urban street scene sensing method based on vision language pre-trained model. International Journal of Applied Earth Observation and Geoinformation, 113:102989.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6):1452–1464.

Zhou, Z.-H. (2012). Ensemble methods: foundations and algorithms. CRC press.

Zhou, Z.-H. and Li, M. (2010). Semi-supervised learning by disagreement. Knowledge and Information Systems, 24(3):415–439.

Zhu, X. and Ghahramani, Z. (2002). Learning from labeled and unlabeled data with label propagation. Tech. Rep., Technical Report CMU-CALD-02–107, Carnegie Mellon University.

PART D

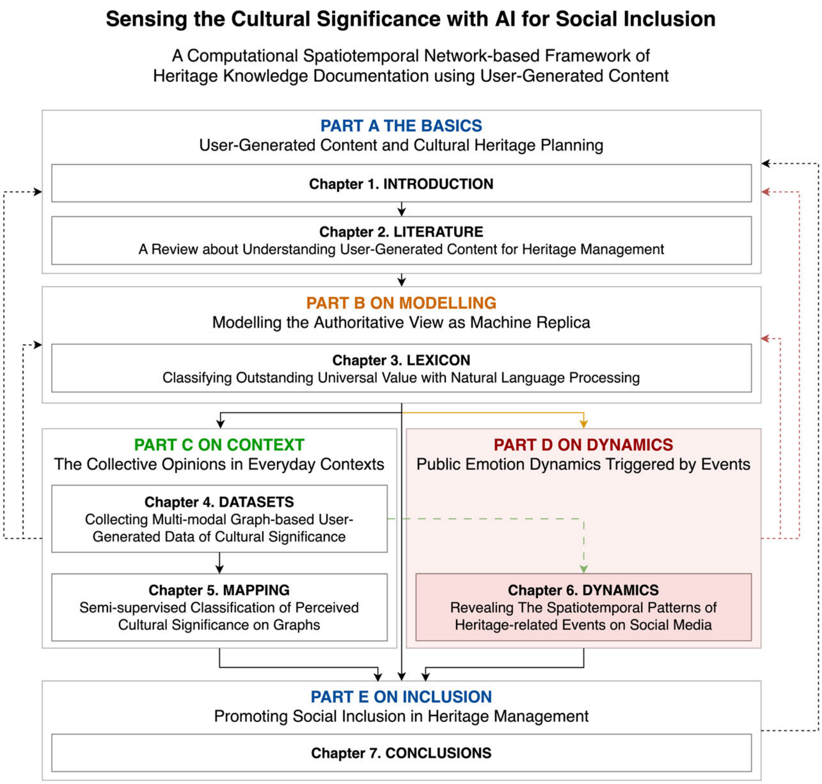
On Dynamics

Public Emotion Dynamics Triggered by Events

On a parallel line to the work in PART C, this part of dissertation focuses on the activated event-triggered scenario when online communities extensively join the discussion of a well-known cultural heritage property under events, possibly at risk. A methodological framework is proposed to explore and describe the spatiotemporal dynamics of both the intensity of posting behaviour and the semantic information of the discussions before, during, and after radical Heritage-related Events (HREs). The expressed emotions and proposed actions of a temporally-formed heritage community are obtained with the aid of pre-trained deep learning models including the machine replica developed in PART B and pre-defined topic modelling algorithms. This part of dissertation combines the content, structure, and context of social media posts based on the knowledge system in PART A. The timelines showing the development of online discussions during HREs both confirms known knowledge and discovers new knowledge, informative to heritage management especially from a global perspective.

One chapter is included in this part:

Chapter 6 Mechanisms - Revealing the Spatiotemporal Patterns of Heritage-Related Events on Social Media.


Sensing the Cultural Significance with AI for Social Inclusion
A Computational Spatiotemporal Network-based Framework of
Heritage Knowledge Documentation using User-Generated Content
PART A THE BASICS
User-Generated Content and Cultural Heritage Planning
Chapter 1. INTRODUCTION
Chapter 2. LITERATURE
A Review about Understanding User-Generated Content for Heritage Management
PART B ON MODELLING
Modelling the Authoritative View as Machine Replica
Chapter 3. LEXICON
Classifying Outstanding Universal Value with Natural Language Processing
PART C ON CONTEXT
The Collective Opinions in Everyday Contexts
Chapter 4. DATASETS
Collecting Multi-modal Graph-based User-Generated Data of Cultural Significance
Chapter 5. MAPPING
Semi-supervised Classification of Perceived Cultural Significance on Graphs
PART D ON DYNAMICS
Public Emotion Dynamics Triggered by Events
Chapter 6. DYNAMICS
Revealing The Spatiotemporal Patterns of Heritage-related Events on Social Media
PART E ON INCLUSION
Promoting Social Inclusion in Heritage Management
Chapter 7. CONCLUSIONS

6 Mechanisms

Revealing the Spatiotemporal Patterns of Heritage-Related Events on Social Media

Parts of this chapter have been published in Bai et al. (2023a) and will be submitted as Bai, et al. (2024b).
Bai, N, Cheng T, Nourian P, Pereira Roders, A. (2023a). An Exploratory Data Analysis of the Spatiotemporal Patterns of Heritage-Related Events on Twitter. In The 30th International Conference on Geoinformatics (CPGIS 2023), July 19-21, University College London, London, UK.
Bai, N, Nourian P, Cheng T, Pereira Roders, A. (2024). Semantic-Augmented Network-based Spatiotemporal Mapping of Heritage-Related Events Detected on Social Media. (Under Preparation).

ABSTRACT Triggered by radical Heritage-Related Events, communities around the world are being actively involved on social media to share the cultural significance they convey to heritage properties including their opinions and emotional attachments. This chapter presents the results of exploratory data analysis on a new graph-based spatiotemporal dataset collected from Twitter concerning events happening in UNESCO World Heritage properties that triggered global concerns with cases of the Notre Dame Paris fire and the Venice flood, both in 2019. The spatiotemporal patterns of tweeting behaviours of online communities before, during, and after the event demonstrate a clear distinction of activation levels caused by the events. The dominant emotions and topics of people during the online debate have been detected and visualized with pre-trained deep-learning models and unsupervised clustering algorithms. Clear spatiotemporal dynamics can be observed from the data collected in both case studies, while each case also demonstrated its specific characteristics due to the severity of the event. The methodological framework proposed and the analytical outcomes obtained in this chapter could be used both in urban studies to mine the public opinions about heritage-related events, and by the Geo-AI community to test spatiotemporal clustering algorithms.

KEYWORDS World Heritage, Spatiotemporal Analysis, Social Network Analysis, Urban Analytics, Event Detection, Topic Modelling

6.1 Introduction

Triggered by radical (not necessarily negative or disastrous) Heritage-Related Events (abbreviated hereinafter as HRE), such as the fire in Notre Dame de Paris burning down the tower designed by Eugène Viollet-le-Duc in April 2019[1], the Parade moving the ancient Egyptian Pharaoh mummies into a new museum in April 2021[2], the opening of a grandiose exhibition in Rijksmuseum, Amsterdam, assembling paintings of the Delft-based artist Johannes Vermeer from all over the world in February 2023[3], the terrible earthquake in Turkey and Syria destroying ancient UNESCO World Heritage sites in February 2023[4], or the more regular occasions of floods, festivals, and/or Biennial exhibitions in Venice, communities around the world are being actively involved on social media platforms, such as Twitter, Weibo, and TikTok, to share their opinions and emotional attachments (Monteiro et al., 2014; Chianese et al., 2016; Bai et al., 2021b). In the digital age, the Internet and social media have eased, accelerated, magnified, and even sometimes polarized the expressing and sharing mechanism (Tucker et al., 2018; Stevens et al., 2020). Shortly after an event, related information is spread contagiously and collective emotions (anger and sorrow in negative events, or happiness in positive events) are triggered (Zhai et al., 2020). While sharing experiences, giving opinions, and expressing emotions concerning an HRE, the participating public may not be deliberately talking about the cultural significance per se, some of which may not even be aware of the concept of cultural significance or the status of cultural heritage, they are still unconsciously sending messages revealing the cultural significance they convey to heritage properties (Bai et al., 2021b). The concept of “heritage community" also gets further expanded in an online environment, transcending the geographical boundary (Council of Europe, 2005; Zagato et al., 2015), still pertaining to its original definition in Faro Convention:

> “a heritage community consists of people who value specific aspects of cultural heritage which they wish, within the framework of public action, to sustain and transmit to future generations”.

Albeit bearing the risk of enhancing “mediatisation of heritage” and biasing the cultural significance (Garduño Freeman and Gonzalez Zarandona, 2021), such opinions and emotions containing information about the perceived cultural significance, as well as the dynamics of messages spreading on an intrinsic social network composed of temporally-founded heritage community, could help heritage managers and urban planners make more informed and inclusive decisions (Lipizzi et al., 2015; Zhai et al., 2020). Furthermore, all the geo-tagged and time-stamped posts on social media, as well as the corresponding Heritage-Related Events (HREs)

[1] https://www.bbc.com/news/world-europe-47971044, accessed 05 May 2023
[2] https://bbc.com/news/world-middle-east-56508475, accessed 05 May 2023
[3] https://www.dutchnews.nl/news/2023/02/now-or-never-vermeer-exhibition-opens-at-rijksmuseum/, accessed 05 May 2023
[4] https://www.archdaily.com/996027/a-major-earthquake-hits-turkey-and-syria-destroying-a-2000-year-old-unesco-world-heritage-site, accessed 05 May 2023

themselves, are unavoidably embedded in their spatiotemporal and social contexts (Zhang and Cheng, 2020; Bai et al., 2022). Aggregating information on social media and mapping the spikes on both a discrete timeline and a spatial representation could yield visualizations that could be easily understandable for decision-makers to make assessments of the impact caused by an HRE and draw conclusions on what to do next to better support urban conservation, following the Recommendation on the Historic Urban Landscape (UNESCO, 2011; Pereira Roders, 2019).

In the computer science literature, event detection is an important task in the field of computer vision and natural language processing, while an **event** is formally defined (Li and Fei-Fei, 2007; Liu et al., 2016) as:

> “a semantically meaningful human activity, taking place within a selected environment and containing a number of necessary objects”.

In the social media era, events could also be regarded as a collection/archive of User-Generated Content concerning certain issues within “a specific structure and limit", “completely initiated and organized by users through social media" (Marine-Roig et al., 2017). Since an event on social media is essentially a group of semantically related posts bounded by space and time, studies have been using geo-tagged tweets to identify meaningful clusters that correspond to well-known “ground truths" and/or previously unknown real-world events (Cheng and Wicks, 2014; Huang et al., 2018; Arjona, 2020; Farnaghi et al., 2020; Kersten and Klan, 2020; George et al., 2021; Afyouni et al., 2022; Rani and Kaushal, 2022). Cheng and Wicks (2014) demonstrated in their case study in London that only by using the spatiotemporal information without adding any verbal/semantic hints, meaningful events can already emerge from the data, since “people will tweet more than expected in order to describe the event and spread information". Following the same logic, many studies in spatiotemporal event detection (Huang et al., 2018; Shi and Pun-Cheng, 2019; Kersten and Klan, 2020; George et al., 2021) first apply a clustering algorithm considering both spatial and temporal proximity, e.g., ST-DBSCAN (Spatiotemporal Density-Based Spatial Clustering of Applications with Noise) (Birant and Kut, 2007; Huang et al., 2018; Kersten and Klan, 2020), STSS (Space-Time Scan Statistic) (Kulldorff et al., 2005; Cheng and Wicks, 2014), STKDE (Space-Time Kernel Density Estimation) (Hu et al., 2018; Kersten and Klan, 2020), OPTICS (Ordering Points To Identify the Clustering Structure) (Ankerst et al., 1999), and Poisson Model (George et al., 2021), followed by summarizing the keywords, drawing word clouds, or conducting a Topic Modelling algorithm such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003) to study the semantic features of the identified spatiotemporal clusters. In other words, only spatiotemporal proximity but not the semantic proximity of posts was considered during the clustering and event detection procedure in these studies, whereas some recent studies also integrated the similarity of textual vector representations (Farnaghi et al., 2020; Rani and Kaushal, 2022), the social distance of people (Yanenko and der Weberei, 2019), or other high-dimensional feature distance (Choi and Hong, 2021) into the clustering metrics. Moreover, in some other application scenarios other than event detection, the conventional spatiotemporal clustering algorithms on Euclidean distance can also be extended to distance on spatial or social networks (Martínez-López et al., 2009;

Costa and Kulldorff, 2014; Wang and Phoa, 2016; Adepeju, 2017; Shen, 2018).

Zooming into the detection and analyses of Heritage-related Events, conceptually, two types of HREs can exist - the events in heritage, and the heritage in events. The former are the events that particularly happen to/in a built cultural heritage, such as the Notre Dame fire, the Pharaoh Parade, as well as the Biennials and floods in Venice mentioned at the beginning of this chapter. The latter are the events that happen outside of a heritage property but have a broader influence regionally or globally, eventually affecting the heritage, such as the Turkey-Syria earthquake mentioned earlier (Meghraoui and Sbeinati, 2023), as well as the global event of Covid-19 pandemic (Sofaer et al., 2021; Ginzarly and Srour, 2022; Naramski et al., 2022; Tenzer, 2022). Even though both can be understood as HREs, this chapter will only focus on the first type (i.e., events in heritage) for demonstrative purposes, since the reactions on social media are assumed to be more focused on the heritage itself, thus more informative for deriving the expressed cultural significance. For the second type, however, the consecutive work in Kumar (2019) has also demonstrated the application of crowd-sourcing and social media sensing to facilitate heritage management in disaster. Local archives including letters, telegrams, newspaper articles, magazines, as well as social media texts and images have been used to study the actions, reactions, and purposes of public engagement after disastrous events such as the 1966 Florence Flood (Kumar, 2020a) and 2015 Nepal Earthquake (Kumar et al., 2020; Kumar, 2020b). Specifically, manual coding schemes and/or machine learning models have been used to distinguish if a piece of information reflected heritage and/or demonstrated damage or not (Kumar et al., 2020), and to classify the response of users into showing the "Situation" and the state of heritage after the event, conveying a "Message" with heritage as background information, recalling a past "Memory" before the event, demonstrating the "Practices" of how people used the heritage after the event (Kumar, 2020b), calling for contribution as "Action", and/or expressing the "Sentiment" for the loss of heritage (Kumar, 2020a).

Furthermore, for a single event that happened in a specific heritage such as the Notre Dame fire, online discussions could span far beyond the core "heritage community" and trigger a variety of sub-topics in different places at different times. For example, people may extensively post their witness accounts of the event and share their sorrow when they first heard about the news (Garduño Freeman and Gonzalez Zarandona, 2021; Padilha et al., 2021a,b). At some specific moment, a group of people may suddenly start talking about their random guesses on who to blame for such a tragedy, which could get spread with anger as fake news (Passaro et al., 2022). In parallel, some other groups of people may start suggesting future development scenarios and proposing redesign projects, which could also get resisted and trigger another round of discussion "wave" (Lupo, 2021; Molina and Molina, 2021). All these imaginary and/or realistic scenarios could be regarded as sub-events taking a slightly different perspective of the same event with different spatiotemporal bounds (Card et al., 2015; Roy and Goldwasser, 2020; van Eck et al., 2020), suggesting the necessity of treating an HRE as a collection of spatiotemporal sub-event clusters, probably in a hierarchical structure. While presumably only a few sub-event clusters may be directly referring to the cultural significance of heritage with events happening therein, many weakly related clusters can still contribute to inclusive heritage management processes, as a valuable and inspiring source of systematic inputs from the public for knowledge documentation, as suggested by the UNESCO 2011 Recommendation on the Historic Urban Landscape (UNESCO, 2011).

This chapter aims to explore the spatiotemporal patterns of online public discussions on social media in Heritage-Related Events (HREs) and propose a methodological framework to extract critical information that is useful for heritage managers from unstructured social media data. Two case studies, i.e., the fire at Notre Dame and the flood in Venice, are tested with the workflow (see Section 6.2.1). Exploratory data analysis on both the spatiotemporal distributions and the semantic focuses of online discussions concerning HREs has been conducted. Three questions are going to be reflected by the collected empirical data:

1. What are the spatiotemporal and social patterns of the posting behaviour at a global scale before, during, and after a major HRE?
2. How are the emotions being expressed in social media posts and evolving over time?
3. What are the main semantic topics being discussed and spread and how can heritage managers learn and benefit from the discussions?

6.2 Data and Materials

6.2.1 Case Studies: Notre-Dame Paris Fire and Venice Flood

This chapter takes HREs for two UNESCO World Heritage properties as case studies: the fire in Notre Dame de Paris, France on 15 April 2019, and the unpreceded flood in Venice, Italy on 12 November 2019. The original statements of Outstanding Universal Value demonstrating the official cultural significance of both properties can be found in Appendix A. Whereas the severe fire in Notre Dame was described as “a catastrophe for the humankind” (Praticò et al., 2020; Garduño Freeman and Gonzalez Zarandona, 2021; Molina and Molina, 2021), the exceptionally severe flood in Venice was also reported as unpreceded in 50 years (Ferrarin et al., 2021; Lorini et al., 2022). According to the categorization of HRE in Section 6.1, both case studies are assumed to be events in heritage, which is valid for Venice since the entire Venetian island together with its broader surrounding Lagoon area are inscribed as UNESCO World Heritage.

For both case studies, datasets have been collected, processed, and analyzed in the past few years from social media and state media containing texts and images

(Padilha et al., 2021a,b; Lorini et al., 2022; Passaro et al., 2022). However, none of these studies focused on the cultural significance of heritage. Rather, they demonstrated the application of classifying images into their spatial direction (Padilha et al., 2021b), sorting them with a chronological order (Padilha et al., 2021a), distinguishing fake news from the real ones (Passaro et al., 2022), and determining the severeness of flooding revealed in images while accurately geo-coding them in corresponding locations on site (Lorini et al., 2022). For simplicity, this study collected a group of new text-based datasets for the purpose of exploring the spatiotemporal patterns of public reactions during HREs. However, theoretically, these existing datasets can also be integrated into the framework at a later stage as complementary information.

6.2.2 Data Collection Strategy

The "`full-archive search`" endpoint of the Twitter API v2[5] with an Academic Research Access was used to collect tweets about both case studies for a period of two weeks (one week before the event, and one week after), i.e., 08 - 22 April 2019 for the Notre Dame fire, and 05 - 19 November 2019 for the Venice flood. A two-step procedure was followed to collect the raw data:

1. A local search first queries for geo-tagged tweets (with the query "`has:geo`") in a fixed radius ("`point_radius`") from the hypothetical core of HREs, i.e., 1.5 km from the Cathedral of Notre-Dame de Paris (48.852737N 2.350699E) and 8 km from the centre of Venetian Island (45.438759N 12.327145E);
2. A global search then queries for geo-tagged tweets that also mentioned the name of the place (not the event, thus not with words "flood" and "fire"), i.e., "Notre-Dame OR notredame OR (notre dame) OR 巴黎圣母院 OR 巴黎聖母院" and "venice OR venezia OR venedig OR venise OR venicia OR veneza OR 威尼斯", respectively.

In both steps, collected tweet data typically include the following elements: the timestamp at UTC (Coordinated Universal Time) time-zone, the pseudo-anonymized user ID, the name-based geo-location with its name and ID, the IDs of original tweets it interacted with (replying to, quoting, or referring to the original tweet), the textual contents, as well as the language code[6].

Furthermore, after the pre-processing of the collected raw data, the original tweets (not necessarily geo-tagged) that are associated with (being replied to, being quoted, or being the original tweet of a series of conversations) each tweet mentioned above are also collected using their IDs, as a round of supplemental search, obtaining the same types of information as local and global searches.

[5] https://developer.twitter.com/en/docs/twitter-api/tweets/search/introduction, accessed 08 May 2023
[6] https://developer.twitter.com/en/docs/twitter-for-websites/supported-languages, accessed 08 May 2023

The data collection took place respectively on 6 December 2022 (local and global searches) and 31 January - 1 February 2023 (supplemental search). Note that the Twitter API rules have been updated on 29 March 2023, and the previous access plans were deprecated. The readers are suggested to refer to the updated version of Twitter documentation for reproducibility.

6.2.3 Geo-coding and Pre-processing of Collected Data

For each case study, the tweets collected with local, global, and supplemental searches mentioned in Section 6.2.2 are merged to make a universal collection of data potentially related to the HRE of interest. Different from earlier versions of Twitter APIs where detailed numerical geographical locations (latitude and longitude) are provided for each geo-tagged tweet, such as in Cheng and Wicks (2014), the current API only provides a geo-ID indicating a name-based categorical geo-location that was selected by the user when posting. These geo-tags of tweets vary in different scales ranging from micro-level POI and neighbourhood to meso-level town and city, to macro-level province and country. Since this chapter aims to study all tweets collected globally, a combined geo-coding (providing the latitudes and longitudes with the given name of places) and reverse geo-coding (providing the name/address of places at different administrative levels with the given latitudes and longitudes) (Kounadi et al., 2013) procedure to unify the resolution of geo-locations is necessary for comparison and aggregation. For pragmatic purposes, the level of "cities" is selected as a balance point for such unification. The tweets with more detailed geo-locations (POIs, neighbourhoods, and towns) are simply relocated in the cities where they are posted, while the tweets with more coarsened geo-locations (provinces, regions, and countries) are arbitrarily mapped to the capital cities (if any). Using [reverse] geo-coding Python libraries GeoPy[7] with OpenStreetMap Nominatim[8] as geocoder (Clemens, 2015), CountryInfo[9], and Wikipedia-API[10], all places are merged and mapped to the city level. As a consequence, a list of cities participating in the discussion of HRE was obtained for each case study. The same set of geo-coding Python libraries is again consulted to obtain the latitude, longitude, and country names of all posting cities. Cities whose names were originally written in another language were translated into English using Google Translator API from the Deep Translator python library[11]. This procedure effectively provides the refined numerical geo-location of the geo-tagged tweets.

The timestamps are grouped into three clusters: before, during, and after the HRE, divided by the interval of the first four days of the event, i.e., 15-18 April 2019 for the Notre-Dame fire, and 12-15 November 2019 for the Venice flood.

[7] https://geopy.readthedocs.io/en/stable/, accessed 08 May 2023
[8] https://github.com/osm-search/Nominatim, accessed 08 May 2023
[9] https://github.com/porimol/countryinfo, accessed 08 May 2023
[10] https://github.com/martin-majlis/Wikipedia-API, accessed 08 May 2023
[11] https://deep-translator.readthedocs.io/en/latest/, Accessed 08 May 2023

Moreover, the collected tweets as raw textual data are in different languages and highly unstructured, which were fed into a pre-processing pipeline:

1. The tweet sentences are tokenized with the TweetTokenizer[12] of NLTK Python library (Bird et al., 2009);
2. The tokens are normalized by turning the letters to lowercase, transforming any '@' sign into '`@USER`' denoting the special token for a user ID, changing any internet link into '`HTTPURL`' denoting the special token for a hyperlink, and de-emojizing the emojis into their corresponding verbal description in English using the demojize tool of Emoji Python library[13];
3. The normalized tokens are joined back as sentences and translated into English using Google Translator API from the Deep Translator library.

Note that no "stopwords" were removed at the stage of pre-processing, since Transformer-based Natural Language Processing models such as BERT prefer texts to appear in their original contexts (Devlin et al., 2019).

After the geo-coding and pre-processing, each collected tweet can be organized as a structured tuple. Let i be the index of a generic sample of the dataset for one HRE case study, then the tweets could be represented as a tuple $\mathfrak{d}_i = (\mathcal{S}_i, \mathcal{O}_i, \mathfrak{u}_i, \mathfrak{t}_i, \mathfrak{l}_i)$, $\mathfrak{d}_i \in \mathfrak{D} = \{\mathfrak{d}_0, \mathfrak{d}_1, ..., \mathfrak{d}_{K-1}\}$, where K is the sample size of the dataset in a case study, $\mathcal{S}_i = \{s_i^{(0)}, s_i^{(1)}, ..., s_i^{(|\mathcal{S}_i|-1)}\}$ is a set of normalized and translated English tweet sentences, $\mathcal{O}_i = \{\mathfrak{d}_{i'} | \mathfrak{d}_{i'} \in \mathcal{D}\}$ or $\mathcal{O}_i = \varnothing$ is the set of all collected related tweets to $\mathfrak{d}_i$, where $\mathfrak{d}_i$ referred to $\mathfrak{d}_{i'}$ in either way of interaction mentioned in Section 6.2.2, which can also be empty when the tweet stands alone, $\mathfrak{u}_i \in \mathcal{U}$ is a user ID that is one instance from the user set $\mathcal{U} = \{\mu_0, \mu_1, ..., \mu_{|\mathcal{U}|-1}\}$, $\mathfrak{t}_i \in \mathcal{T}$ is a timestamp that is one instance from the ordered set of all the unique timestamps $\mathcal{T} = \{\tau_0, \tau_1, ..., \tau_{|\mathcal{T}|-1}\}$ from the dataset at the level of hours, and $\mathfrak{l}_i = (\mathfrak{x}_i, \mathfrak{y}_i, \mathfrak{c}_i)$ or $\mathfrak{l}_i = \varnothing$ is the geo-location of the city where the tweet was posted, which could be empty if the tweet was not geo-tagged, including its geographical coordinate of latitude ($\mathfrak{y}_i$) and longitude ($\mathfrak{x}_i$), and name of the city $\mathfrak{c}_i \in \mathcal{C}$ that is one instance from the set of unique cities $\mathcal{C} = \{\zeta_0, \zeta_1, ..., \zeta_{|\mathcal{C}|-1}\}$. For any city $\zeta_j \in \mathcal{C}$, a corresponding geo-location (x_j, y_j) is stored as the metadata of the city. If $\mathfrak{c}_i = \zeta_j$ for a post $\mathfrak{d}_i$, it automatically entails that $\mathfrak{x}_i = x_j, \mathfrak{y}_i = y_j$.

In the case of the Notre Dame fire, the total number of tweets $K = |\mathfrak{D}| = 198,061$, the number of users $|\mathcal{U}| = 42,036$, and the number of cities $|\mathcal{C}| = 4968$, while in the case of Venice flood, $K = |\mathfrak{D}| = 15,641$, $|\mathcal{U}| = 3541$, and $|\mathcal{C}| = 835$.

[12] https://www.nltk.org/api/nltk.tokenize.casual.html, Accessed 08 May 2023
[13] https://carpedm20.github.io/emoji/docs/, Accessed 08 May 2023

6.3 Methodology

6.3.1 Overview of the Workflow

Figure 6.1 visualizes the general framework proposed in this chapter to collect, structure, and analyze the spatiotemporal dynamics of public discussion on Twitter for a previously-known HRE (Heritage-related Event). In the context of this chapter, the "Prior Knowledge of a Heritage-related Event" refers to the case studies of the Notre Dame fire and Venice flood mentioned in Section 6.2.1, yet it can also be extended in future studies about any HRE or general social events of interest.

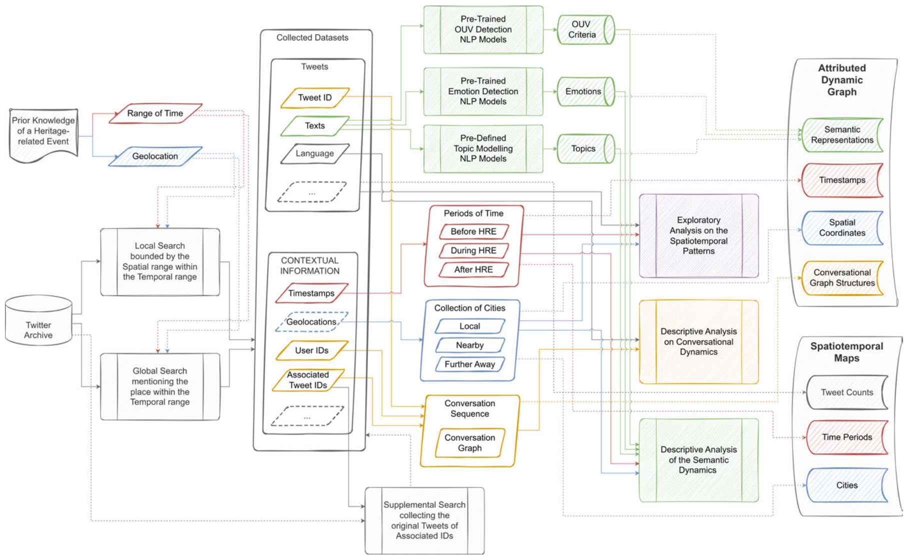


FIG. 6.1 The general methodological workflow proposed in this chapter.

Through the data collection step, the textual and contextual information of all tweets makes up a complete dataset. By inputting the textual data of tweets into several pre-trained models and pre-defined algorithms, semantic information about cultural significance, emotions, and topics is obtained as [pseudo-]labels. The contextual information of tweets is respectively used to distinguish the period of time relative to the HRE (before, during, and after), determine the position of cities relative to the city where the HRE happened (same city as local, nearby cities within a given radius, and global cities further away), and construct a directed network marking the conversation sequence on Twitter. Thereafter, the number of tweets in all cities along the timeline is analyzed through exploratory data analysis to describe the general spatiotemporal pattern on Twitter during HRE. Furthermore, descriptive analyses are conducted both on the conversational graph to distinguish the behavioural changes with respect to HREs, and on the semantic labels to uncover the dynamic associations among the expressed emotions, discussed topics, and entailed cultural significance in relation to the spatial and temporal bounds.

6.3.2 Spatiotemporal Dynamics

The number of tweets is first aggregated temporally and spatially to grasp the general spatiotemporal patterns of posting in time of HREs. Temporally, the tweets are counted every hour as in $\mathcal{T}$ from one week before the day when the event happened till one week after resulting a vector $\boldsymbol{t} := [t_k]_{|\mathcal{T}|\times 1} \in \mathbb{N}^{|\mathcal{T}|\times 1}, t_k = |\{\mathfrak{d}_i | \mathfrak{t}_i = \tau_k\}|$ to draw a timeline on the volume and intensity of the discussion globally. Since some of the original posts (collected through the supplemental search) spanned far before the day when the HRE happened, those posted before the starting date (08 April 2023 and 05 November 2023, respectively) were filtered out from further analysis. Spatially, the tweets are counted in the level of cities as in $\mathcal{C}$ (mentioned in Section 6.2.3) resulting a vector $\boldsymbol{c} := [c_j]_{|\mathcal{C}|\times 1} \in \mathbb{N}^{|\mathcal{C}|\times 1}, c_j = |\{\mathfrak{d}_i | \mathfrak{c}_i = \zeta_j\}|$, and further grouped one level higher to the level of countries. Spatial and temporal intervals are then considered together to further separate the posts. Specifically, the set of timestamps $\mathcal{T}$ is divided into $\mathcal{T}_{\mathsf{B}} \subset \mathcal{T}$ before the main HREs, $\mathcal{T}_{\mathsf{D}} \subset \mathcal{T}$ during the HREs up to four days after the event happened, and $\mathcal{T}_{\mathsf{A}} \subset \mathcal{T}$ after the main HREs upon one week after, as mentioned briefly in Section 6.2.3. And the set of cities $\mathcal{C}$ is divided into $\mathcal{C}_0 = \{\zeta_0\} \subset \mathcal{C}$ which is the host city of the HRE (Paris or Venice), $\mathcal{C}_1 \subset \mathcal{C}$ which contains cities from the same country (France or Italy), and $\mathcal{C}_2 \subset \mathcal{C}$ containing cities from other part of the world. This categorization of time is referred to as "Periods" and that of cities as "Locality" in this chapter.

Considering the different Periods, the vector $\boldsymbol{c}$ can be disaggregated into three vectors $\boldsymbol{c}_{\mathsf{B}}, \boldsymbol{c}_{\mathsf{D}}, \boldsymbol{c}_{\mathsf{A}} \in \mathbb{N}^{|\mathcal{C}|\times 1}$, where $\boldsymbol{c}_{\mathsf{B}} + \boldsymbol{c}_{\mathsf{D}} + \boldsymbol{c}_{\mathsf{A}} = \boldsymbol{c}$, respectively counting the number of tweets posted in each city before, during, and after HREs, the entries of which can be 0 when tweets are only posted in a city for specific periods, very common in cities with $\mathcal{C}_2$ before the HRE. The entries of the vectors $\boldsymbol{c}_{\mathsf{B}}, \boldsymbol{c}_{\mathsf{D}}, \boldsymbol{c}_{\mathsf{A}}$ are sorted in descending order, generating ranks of the cities in each period, where the ranks of cities with a tie are arbitrarily assigned. Then the numbers of tweets in all cities in each period are plotted against their rankings $\boldsymbol{n} = [1, 2, 3..., |\mathcal{C}|]^{\mathsf{T}}$ in a log-log scale, resulting in Rank-size plots. This is to check if the fat-tailed distribution in the seminal Zipf's Law or the more general power law still holds in terms of tweeting behaviour in HRE and if there is a pattern shift among the different periods (Cristelli et al., 2012; Moreno-Sánchez et al., 2016). Moreover, for each tweet $\mathfrak{d}_i$, the geodesic distance (the arc length on Earth surface) $\boldsymbol{d} := [d_i]_{K\times 1} \in \mathbb{R}^{K\times 1}$ of the city $\mathfrak{c}_i$ where it is posted and the city ζ_0 where the HRE actually happened can be computed, using their respective latitudes and longitudes $(\mathfrak{x}_i, \mathfrak{y}_i)$ and (x_0, y_0). The distributions of vector $\boldsymbol{d}$ are also plotted while distinguishing the tweets in different periods (thus 3 distributions for each HRE case study). The geodesic distance is computed using GeoPy Python library.

6.3.3 Social Connections as Graphs

Conversational sequence of the discussion on Twitter is modeled as a directed multi-graph $\mathcal{G}^{\text{MULT}} = (\mathcal{V}, \{\mathcal{E}^{\text{CONV}}, \mathcal{E}^{\text{USER}}\})$. Two types of links are considered:

1. For any tweet $\mathfrak{d}_i$, as long as the corresponding set $\mathcal{O}_i$ of associated tweets is not empty, the tweet $\mathfrak{d}_i$ and all $\mathfrak{d}_{i'} \in \mathcal{O}_i$ are added to the node set $\mathcal{V}$, and the links pointing from the tweet to its associated ones $(\mathfrak{d}_i, \mathfrak{d}_{i'})$ are added to the link set $\mathcal{E}^{\text{CONV}}$.
2. For any user $\mu_u \in \mathcal{U}$, all the tweets posted by this same user are assembled as $\{\mathfrak{d}_i | \mathfrak{u}_i = \mu_u\}$ and added to the node set $\mathcal{V}$ if any of them is still not there, which are then ranked in chronological order. The links pointing from the later tweets to their immediate temporal neighbour are added to the link set $\mathcal{E}^{\text{USER}}$.

The nodes inherit all the data from the tweets as their node attributes, including text, language, user information, timestamp, period, and [possibly] latitude, longitude, and locality of its posting city. For any link $(\mathfrak{d}_i, \mathfrak{d}_{i'})$, the temporal lag $\mathfrak{t}_i - \mathfrak{t}_{i'}$, the spatial geodesic distance between $\mathfrak{c}_i$ and $\mathfrak{c}_{i'}$, as well as the period of the interaction, i.e., the period of $\mathfrak{t}_i$ are recorded as its link attribute.

Graph properties such as degree distributions, density, number and size of [weakly] connected components, betweenness centrality, and PageRank are computed to describe the general features of the constructed conversational networks on Twitter during HREs (Wasserman and Faust, 1994; Page et al., 1999; Aggarwal, 2011; Barabási, 2013; Nourian et al., 2016; Bai et al., 2023b). For simplicity, the two types of links are merged in most analyses if not mentioned explicitly, making a simple directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{E} = \mathcal{E}^{\text{CONV}} \bigcup \mathcal{E}^{\text{USER}}$. In the end, the node sets $\mathcal{V} \subseteq \mathcal{D}$ are smaller than the raw tweet sets $\mathcal{D}$. In the case of Notre-Dame fire, the number of tweets appearing as nodes in the conversational network $|\mathcal{V}| = 179,758$ (79.3% of $|\mathcal{D}|$), the number of conversational links including interactions of different users and consecutive posting of the same users $|\mathcal{E}| = 221,285$, while in the case of Venice flood, $|\mathcal{V}| = 11,961$ (76.5% of $|\mathcal{D}|$), and $|\mathcal{E}| = 12,106$.

6.3.4 Semantics on Cultural Significance, Emotions, and Topics

For each tweet $\mathfrak{d}_i \in \mathcal{V} \subseteq \mathcal{D}$, the translated English sentences $\mathcal{S}_i$ are fed into pre-trained models and pre-defined algorithms to obtain their semantic meanings.

Concerning the cultural significance, the BERT (Devlin et al., 2019) and ULMFiT (Howard and Ruder, 2018) models pre-trained and finetuned with the UNESCO Statements of OUV in Bai et al. (2021a) and Chapter 3 are used to predict the most probable OUV selection criteria (check Table A.1) mentioned in the tweets, such that:

$$\boldsymbol{y}_i^{\text{BERT}} = \boldsymbol{g}_{\text{BERT}}(\mathcal{S}_i|\boldsymbol{\Theta}_{\text{BERT}}), \tag{6.1}$$

$$\boldsymbol{y}_i^{\text{ULMFiT}} = \boldsymbol{g}_{\text{ULMFiT}}(\mathcal{S}_i|\boldsymbol{\Theta}_{\text{ULMFiT}}), \tag{6.2}$$

$$\boldsymbol{Y}^{\text{OUV}} := [\boldsymbol{y}_i^{\text{OUV}}]_{11\times|\mathcal{V}|}, \boldsymbol{y}_i^{\text{OUV}} = \frac{\boldsymbol{y}_i^{\text{BERT}} + \boldsymbol{y}_i^{\text{ULMFiT}}}{2}. \tag{6.3}$$

where $\boldsymbol{g}_{(*)}$ is the end-to-end function of pre-trained models, $\boldsymbol{\Theta}_{(*)}$ is the parameter of the models, and $\boldsymbol{y}_i^{(*)}$ is an 11-dimensional logit vector as soft-label predictions. Let top-n$(\boldsymbol{l}, n)$ denote the function returning the index set of the largest n elements of a vector $\boldsymbol{l}$, let max$(\boldsymbol{l}, n)$ denotes the function returning the value of the n_{th} largest element of vector $\boldsymbol{l}$, and let IoU$(\mathcal{A}, \mathcal{B})$ denotes the Intersection over Union of any two generic sets $\mathcal{A}, \mathcal{B}$ (see also Equation 4.13), then the confidence and [dis-]agreement of both models for top-n predictions could be computed as:

$$\boldsymbol{K}^{\text{OUV}} := [\boldsymbol{\kappa}_i^{\text{OUV}}]_{2\times|\mathcal{V}|}, \boldsymbol{\kappa}_i^{\text{OUV}} := [\kappa_i^{\text{OUV}(0)}, \kappa_i^{\text{OUV}(1)}]^{\text{T}}, \tag{6.4}$$

$$\kappa_i^{\text{OUV}(0)} = \sum_{n_0=1}^{n} \frac{\max(\boldsymbol{y}_i^{\text{BERT}}, n_0) + \max(\boldsymbol{y}_i^{\text{ULMFiT}}, n_0)}{2}, \tag{6.5}$$

$$\kappa_i^{\text{OUV}(1)} = \text{IoU}(\text{top-n}(\boldsymbol{y}_i^{\text{BERT}}, n), \text{top-n}(\boldsymbol{y}_i^{\text{ULMFiT}}, n)), \tag{6.6}$$

where $\kappa_i^{\text{OUV}(0)}$ denotes the confidence of both models predicting a certain probability distribution of OUV selection criteria, and $\kappa_i^{\text{OUV}(1)}$ denotes the agreement of the models in their categorical predictions. Since it was noted that the models work better with top-3 predictions (Bai et al., 2021a), only the tweets that are predicted by both models with higher top-3 confidence of .75 and top-3 agreement of .50 are considered as truly expressing information related to cultural significance, making up a subset of tweet nodes $\mathcal{V}^{\text{OUV}} \subset \mathcal{V} \subset \mathcal{D}$. Note this approach is very similar to Equations (4.16) and (4.19) in Bai et al. (2022) and Chapter 4. After filtering with confidence and agreement, the number of tweets classified as mentioning cultural significance in Notre-Dame fire is $|\mathcal{V}^{\text{OUV}}| = 61,550$ $(34.2\%$ of $|\mathcal{V}|)$, and for Venice flood $|\mathcal{V}^{\text{OUV}}| = 3628$ $(30.3\%$ of $|\mathcal{V}|)$. And the predicted categorical top-3 [pseudo-]labels of each tweet with respect to OUV selection criteria can be described as an array of sets:

$$\boldsymbol{\mathcal{Y}}^{\text{OUV}} = \left[\mathcal{Y}_i^{\text{OUV}}\right] = \left[\left\{\text{top-n}(\boldsymbol{y}_i^{\text{OUV}}, 3)|\mathfrak{v}_i \in \mathcal{V}^{\text{OUV}}\right\} \text{ or } \varnothing\right]. \tag{6.7}$$

Concerning the emotions expressed in the tweets, pre-trained models on both Sentiment Analysis and Emotion Detection tasks are considered (Acheampong et al., 2020), both of which have become important tasks in the field of Natural Language Processing (Eisenstein, 2018; Rao and McMahan, 2019; Jurafsky and Martin, 2020). Whereas the former only classifies texts into a polarity of negative, positive, and neutral, the latter considers the full spectral of 6 basic human emotions by Paul Ekman (Ekman, 1992), i.e., joy (happiness), sadness, fear, disgust, anger, and surprise. Sentiment Analysis is a relatively easier task compared to the Emotion Detection, and has already been broadly applied in the analysis of User-Generated Content for heritage and tourism studies (Mazloom et al., 2017; Afzaal et al., 2019; Taecharungroj and Mathayomchan, 2019). Emotion Detection is more complex, less applied, but can potentially provide more inspiring information for heritage management (Dickinger and Lalicic, 2016; Nenko and Petrova, 2018; Pan et al., 2019). Similar to the pre-trained models on cultural significance (Bai et al., 2021a), preliminary studies show that the Emotion Detection models are not robust enough with different input data to produce consistently high-quality predictions, even if possibly outputting a high confidence. Therefore, this study decides to integrate predictions on several Emotion Detection and Sentiment Analysis models, and only keep the ones with high consistency across tasks. On one hand, pysentimiento Python toolkit is used to predict both the sentiment[14] and the emotion[15] categories of the tweets (García-Vega et al., 2020; Pérez et al., 2021; Pérez et al., 2022). On the other hand, additional models for sentiment analysis[16] and emotion detection[17] are respectively used, both of which are finetuned with the BERTweet as base models (Nguyen et al., 2020; Pérez et al., 2021). All the predictions are conducted with the “text classification” pipeline in Huggingface Transformer Python library (Wolf et al., 2020). It is worth noting that pysentimiento emotion detection enables an additional class of “others” aside from the original 6 basic emotions, not forcing the model to predict one emotion category even if the sentence can be indeed neutral. Similar to Equation (6.1), the two emotion logic vectors and two sentiment logic vectors could be respectively computed as:

$$\boldsymbol{y}_i^{\mathsf{EM(0)}} \in [0,1]^{7\times 1}, \boldsymbol{y}_i^{\mathsf{EM(1)}} \in [0,1]^{6\times 1}, \boldsymbol{y}_i^{\mathsf{SE(0)}} \in [0,1]^{3\times 1}, \boldsymbol{y}_i^{\mathsf{SE(1)}} \in [0,1]^{3\times 1}. \tag{6.8}$$

$$\kappa_i^{\mathsf{EM}} = \begin{cases} \text{top-n}(\boldsymbol{y}_i^{\mathsf{EM(1)}},1) = \text{top-n}(\boldsymbol{y}_i^{\mathsf{EM(0)}},1) & \text{if top-n}(\boldsymbol{y}_i^{\mathsf{EM(0)}},1) \neq \text{‘others’} \\ \text{top-n}(\boldsymbol{y}_i^{\mathsf{EM(1)}},1) \in \text{top-n}(\boldsymbol{y}_i^{\mathsf{EM(0)}},2) & \text{otherwise} \end{cases}, \tag{6.9}$$

$$\kappa_i^{\mathsf{SE}} = \begin{cases} 1 & \text{if top-n}(\boldsymbol{y}_i^{\mathsf{SE(0)}},1) = \text{‘NEU’} \\ 1 & \text{if top-n}(\boldsymbol{y}_i^{\mathsf{SE(1)}},1) = \text{‘NEU’} \\ \text{top-n}(\boldsymbol{y}_i^{\mathsf{SE(1)}},1) = \text{top-n}(\boldsymbol{y}_i^{\mathsf{SE(0)}},1) & \text{otherwise} \end{cases}, \tag{6.10}$$

$$\boldsymbol{k}^{\mathsf{EMS}} = [\kappa_i^{\mathsf{EMS}}]_{1\times|\mathcal{V}|}, \kappa_i^{\mathsf{EMS}} = \kappa_i^{\mathsf{EM}} \wedge \kappa_i^{\mathsf{SE}}, \text{where } \kappa_i^{\mathsf{EMS}}, \kappa_i^{\mathsf{EM}}, \kappa_i^{\mathsf{SE}} \in \{0,1\}. \tag{6.11}$$

The emotion labels are only considered as consistent ($\kappa_i^{\mathsf{EM}} = 1$) when the top-1 predictions of both models are the same, or in case pysentimiento considers a tweet as containing “other” neutral emotions, the second most significant emotion is the same as the other model. And the sentiment labels are considered as similar ($\kappa_i^{\mathsf{SE}} = 1$) when the top-1 predictions of both models are the same or when either model predicts “NEU” (neutral) as the polarity of the sentence. Only when the tweet has a consistent emotion detection result and a similar sentiment detection result, the emotion prediction of it is considered as valid ($\kappa_i^{\mathsf{EMS}} = 1$), resulting in a subset of tweet nodes $\mathcal{V}^{\mathsf{EMS}} \subset \mathcal{V} \subset \mathcal{D}$ expressing emotions. After filtering with consistency, the number of tweets classified as expressing consistent emotions in Notre-Dame fire is $|\mathcal{V}^{\mathsf{EMS}}| = 93,616$ (52.1% of $|\mathcal{V}|$), among which $27,375$ displayed an explicit emotion other than ‘others’ (15.2% of $|\mathcal{V}|$), while in Venice flood, the numbers are respectively $|\mathcal{V}^{\mathsf{EMS}}| = 6235$ (also 52.1% of $|\mathcal{V}|$, coincidentally) and 1573 with explicit emotions

[14] https://huggingface.co/pysentimiento/robertuito-sentiment-analysis, accessed 12 May 2023
[15] https://huggingface.co/pysentimiento/robertuito-emotion-analysis, accessed 12 May 2023
[16] https://huggingface.co/finiteautomata/bertweet-base-sentiment-analysis, accessed 12 May 2023
[17] https://huggingface.co/Emanuel/bertweet-emotion-base, accessed 12 May 2023

(13.2% of $|\mathcal{V}|$). The predicted categorical emotion and sentiment [pseudo-]labels of each tweet can therefore be described as an array of sets:

$$\boldsymbol{\mathcal{Y}}^{\text{EMS}} = \left[\mathcal{Y}_i^{\text{EMS}}\right] = \left[\left\{\text{top-n}(\boldsymbol{y}_i^{\text{EM}}, 1), \text{top-n}(\boldsymbol{y}_i^{\text{SE}}, 1) | \kappa_i^{\text{EMS}} = 1 \wedge \mathfrak{v}_i \in \mathcal{V}^{\text{EMS}}\right\} \text{ or } \varnothing\right]. \tag{6.12}$$

Moreover, concerning the topics of discussions, BERTopic[18] Python library is used to conduct unsupervised topic modelling (Grootendorst, 2022). BERTopic is a modular pipeline with six main components, i.e., document embedding making the most use of pre-trained large language models, such as Sentence Transformers (Reimers and Gurevych, 2019); dimensionality reduction transforming the high-dimensional embedding vectors into lower dimensions to help cluster models; document clustering with HDBSCAN (McInnes et al., 2017) to group similar documents together, word tokenization within each document cluster counting the appearance of words or N-grams (N continuous words); topic representation calculating the significant words/N-grams that can differentiate one cluster from the topics using class-based Tf-Idf (Term Frequency - Inverse Document Frequency) algorithm; and eventually, fine-tuning on the topic representation to further improve the generated topics. BERTopic takes full use of state-of-the-art word embeddings of pre-trained models and is context-aware, which is different from the conventional topic modelling algorithm Latent Dirichlet Allocation (LDA) (Blei et al., 2003), only considering documents as bag-of-words (Grootendorst, 2022). For each HRE, BERTopic pipeline is called upon the translated English tweets $\mathcal{S}_i$ to generate topics, where each topic is characterized with 10 keywords containing single words and 2-grams. Each topic needs to appear more than 45 times for Notre-Dame fire and 25 times for Venice flood. The number of topics was not determined to any arbitrary value, but was rather merged automatically with HDBSCAN[19]. Afterwards, all the generated topics (denoted as $\mathcal{Z} = \{z_m | m = 0, 1, ..., |\mathcal{Z}| - 1\}$) with their associated keywords are gone through manually by an expert to select the ones as a subset $\mathcal{Z}^{\text{s}}$ presumably "interesting" and informative for heritage management, under the categories such as description of the incidence, expression of emotions, and call for actions. The outcome of BERTopic topic modelling is effectively a probability distribution for each tweet over all topics (including the 'noise' topic usually excluded from further analyses):

$$\boldsymbol{y}_i^{\text{TOP}} = \left[y_{i,m}^{\text{TOP}}\right]_{|\mathcal{Z}|\times 1} \in [0,1]^{|\mathcal{Z}|\times 1}, \mathbf{1}_{|\mathcal{Z}|\times 1}^{\mathsf{T}} \boldsymbol{y}_i^{\text{TOP}} = \sum y_i^{\text{TOP}} = 1, \tag{6.13}$$

where $y_{i,m}^{\text{TOP}}$ refers to the probability of the i_{th} tweet being categorized as the m_{th} topic within $\mathcal{Z}$, and $\mathbf{1}$ is a vector of all 1s. Keeping only the predictions where there is high confidence ($y_{i,m}^{\text{TOP}} > 0.5$) for the interesting topics $z_m \in \mathcal{Z}^{\text{s}}$, a subset of tweet nodes $\mathcal{V}^{\text{TOP}} \subset \mathcal{V} \subset \mathcal{D}$ referring to heritage-informative topics can be obtained. For Notre-Dame fire, the number of obtained topics after topic modelling is $|\mathcal{Z}| = 260$, the number of interesting topics $|\mathcal{Z}^{\text{s}}| = 57$ (52.1% of $|\mathcal{Z}|$), and the number of tweets referring to interesting topics $|\mathcal{V}^{\text{TOP}}| = 77,007$ (42.8% of $|\mathcal{V}|$), among which 8206 are not within the 'noise' topic (4.6% of $|\mathcal{V}|$). And for Venice flood, the numbers are respectively $|\mathcal{Z}| = 45$, $|\mathcal{Z}^{\text{s}}| = 22$ (48.9% of $|\mathcal{Z}|$), $|\mathcal{V}^{\text{TOP}}| = 5515$ (46.1% of $|\mathcal{V}|$), among

[18] https://maartengr.github.io/BERTopic/index.html, accessed May 13 2023

[19] https://maartengr.github.io/BERTopic/getting_started/topicreduction/topicreduction.html, accessed 15 May 2023

which 1836 are not within the 'noise' topic (15.3% of $|\mathcal{V}|$). The eventual topic [pseudo-]labels of each tweet can be described as an array of sets:

$$\boldsymbol{\mathcal{Y}}^{\text{TOP}} = \left[\mathcal{Y}_i^{\text{TOP}}\right] = \left[\{z_m | y_{i,m}^{\text{TOP}} > 0.5 \wedge \mathfrak{d}_i \in \mathcal{V}^{\text{TOP}} \wedge z_m \in \mathcal{Z}^s\} \text{ or } \varnothing\right]. \tag{6.14}$$

After the semantic [pseudo-]labels $\boldsymbol{\mathcal{Y}}^{\text{OUV}}, \boldsymbol{\mathcal{Y}}^{\text{EMS}}, \boldsymbol{\mathcal{Y}}^{\text{TOP}}$ have all been obtained, the timelines demonstrating the temporal development of each type of semantic topic are visualized, and the instances of tweets under different Periods and Localities are counted. Both steps are similar to the operations previously described in Section 6.3.2. The implementation details of the topic modelling procedure using BERTopic can be found in Appendix B.

6.4 Results

6.4.1 General Spatiotemporal Patterns

The temporal distribution of tweets (vector $\boldsymbol{t}$) is visualized in Figure 6.2. It shows a clear daily pattern that Notre-Dame is generally talked more of on Twitter than Venice, while HREs triggered the discussion and raised the scale of tweeting behaviour to a significantly higher level. However, the peaks also died out quickly after a few days, dropping to the scale before the event. This effect is more obvious in

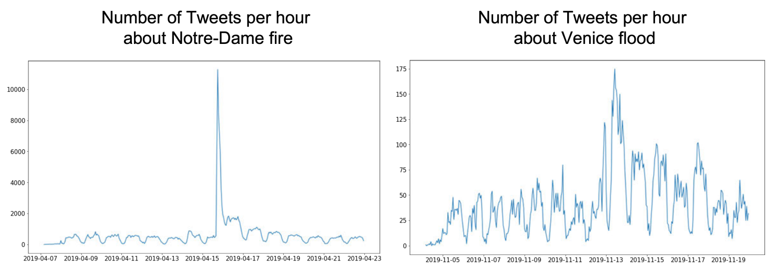


FIG. 6.2 The temporal pattern of tweets throughout the data collection period concerning heritage-related events aggregated to the hourly level.

Notre-Dame de Paris (almost 10 folds) than in Venice (about 3 folds), possibly because even though exceptionally severe in 50 years, Venice undergoes and recovers from floods almost annually, making this HRE incomparable with the fire in Notre-Dame de Paris that shocked the entire world drastically. Both the aggregated spatial patterns of both HREs regardless of periods (vector $\boldsymbol{c}$) and the ones disaggregated with different periods relative to the happening of events (vectors $\boldsymbol{c}_{\mathrm{B}}, \boldsymbol{c}_{\mathrm{D}}, \boldsymbol{c}_{\mathrm{A}}$) are visualized in Figures 6.3 and 6.4, respectively.

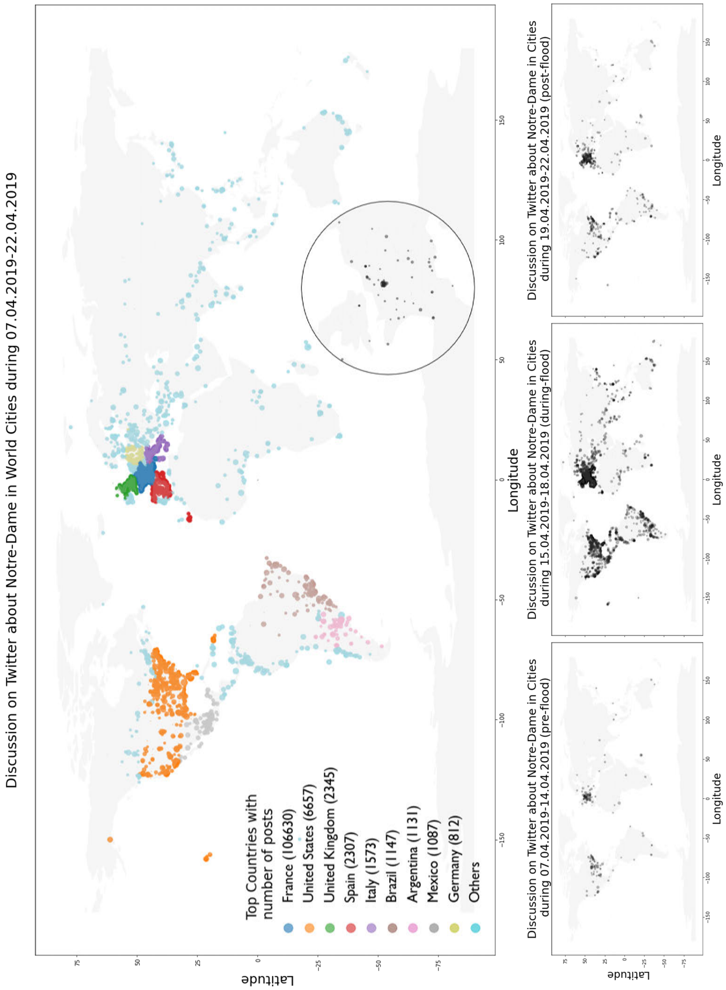


FIG. 6.3 The global spatial pattern of tweets throughout the data collection period in the case of Notre-Dame fire. The larger size a node, the more tweets located in the city it represents. Nodes are colored by the top 9 countries contributing to the tweet-scape. The spatial pattern is further disaggregated in periods before, during, and after the events.

Besides the fact that the case of Notre-Dame fire had a much larger scale than Venice flood spreading to more cities world-wide, they both demonstrated similar patterns. The figures confirmed the hypothesis previously mentioned in Chapter 1 that the online discussions on Twitter triggered by HREs would most probably go beyond the geographical boundaries, forming a global community caring about World Heritage. Naturally, the tweets posted from the same country (France and Italy for both cases) made the largest contribution to the discussion composing almost half of the tweet-scape, while other countries nearby (e.g., European countries) and far away in both senses of geographical and cultural distances also participated substantially. Interestingly, United States, United Kingdom, France, and Italy all entered the top-5

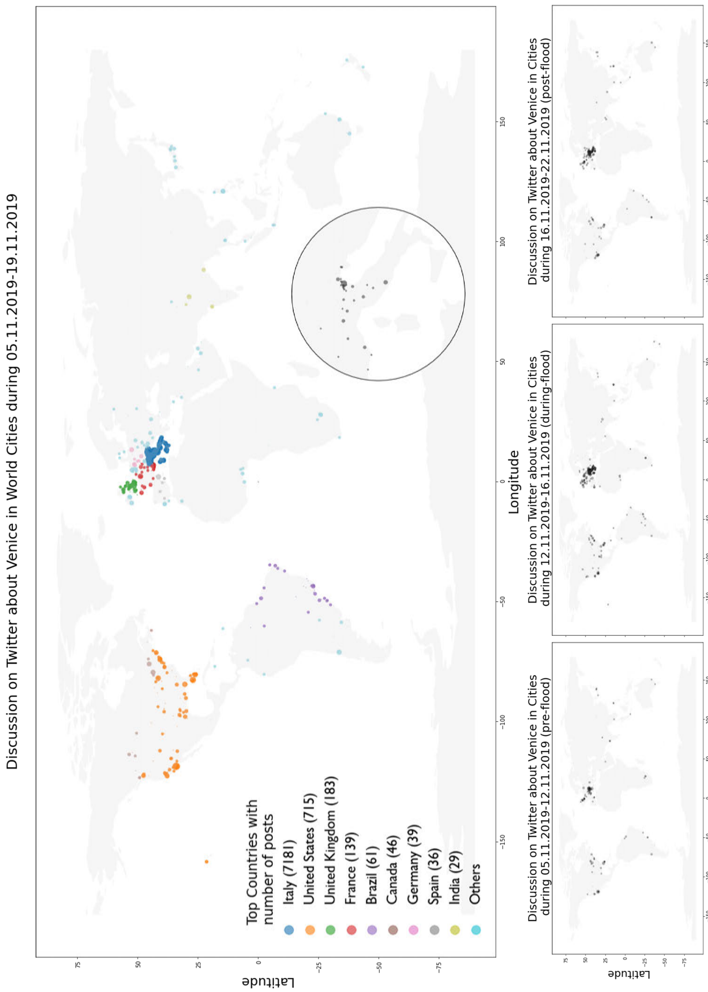


FIG. 6.4 The global spatial pattern of tweets throughout the data collection period in the case of Venice flood.

posting countries in both cases, indicating the concentration of heritage-aware people in these places. However, this spatial pattern also strongly correlates with the number of active users of Twitter in each country, the main target group, and major purpose of usage in different regions. For example, the voice of China is significantly missing from the discussion, since people there mainly used Weibo and Wechat to fulfill the similar purposes of instant reaction and personal blogging. Moreover, through data exploration, the large amount of discussion on "Venice" in United States before and after the flood might be related to the place in Florida's Gulf Coast that has the same name (also known as Venice Beach), thus an unexpected outlier, which were not initially excluded during the data collection.

The spatial patterns of before, during, and after HREs could be compared by plotting the ranked vectors $\boldsymbol{c}_{\mathrm{B}}, \boldsymbol{c}_{\mathrm{D}}, \boldsymbol{c}_{\mathrm{A}}$ against their ranking $\boldsymbol{n}$ in a log-log scale, as shown in Figure 6.5. The ubiquitous linear pattern of the rank-size plots in social sciences and urban studies indicating a quasi power law of the sizes can again be observed, except for the extremely huge number of tweets (heavy head) in the highest-ranking city (i.e., the city where the event happened), that cannot get fitted linearly even on a logarithm scale. By excluding the highest-ranking city, a more reasonable line can be fitted (the dash lines rather than the dotted lines in Figure 6.5), using the Maximum Likelihood algorithm to characterize the relationship between the ranking $\boldsymbol{n}$ and the sizes $\boldsymbol{c}_{\mathrm{B}}, \boldsymbol{c}_{\mathrm{D}}, \boldsymbol{c}_{\mathrm{A}}$. It is also evident that the online participation was spread to more cities globally during HREs with the longest tails in Figure 6.5, while the posting behaviour of the post-event period did not yet fully recover to the pre-event level, implying possible aftermath effects. It is also noted that even though the numbers of posts are almost always highest during HREs in a city, followed by after HREs, and then before HREs, this is not the case for the highest-ranking cities. For them, the posts before HREs are even higher than that during HREs in both case studies, which is probably a logical outcome due to the number of days in each period (7 days before HREs, 4 days during HREs, and 4 days after HREs), suggesting the popularity of the place as a heritage property and tourist destination under an everyday/baseline scenario.

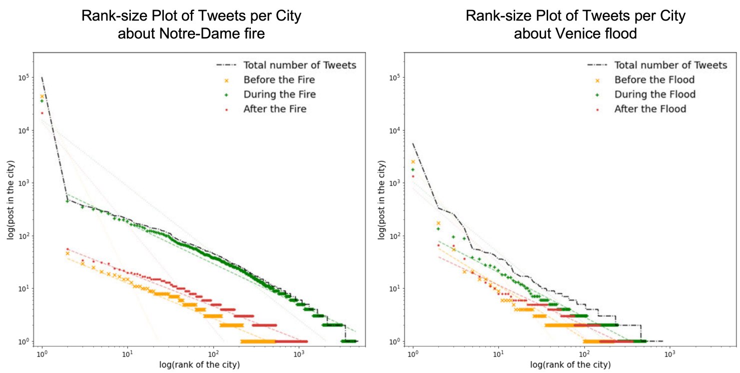


FIG. 6.5 The log-scale rank-size plot of tweets per city in periods before, during and after the events in Notre-Dame de Paris and Venice. Two lines are fitted to the points using the Maximum Likelihood algorithm, while the dotted ones included the highest-ranking city for the fitting, and the dashed lines excluded them.

6.4.2 Conversation Dynamics

Some key graph statistics of both conversational graphs $\mathcal{G}$ and $\mathcal{G}^{\mathrm{MULT}}$ in Notre-Dame and Venice are shown in Table 6.1. Due to the network structures based on tweeting behaviours, the graphs for both case studies are sparse with low density and disconnected with many weakly connected components. Yet the sizes of the largest weakly-connected components consisting of around 1/3 to 1/2 of all nodes suggest that people joining the collective discussion from different local perspectives are likely to merge as a whole. Figure 6.6 visualizes the distribution of some key node-level graph statistics such as degree, betweenness centrality and PageRank on either the original conversational graph $\mathcal{G}$ or the largest weakly-connected component of it. It can be observed that the distributions are very similar in shape to each other in the both case studies of Notre-Dame fire and Venice flood, albeit the larger graph size in Notre-Dame.

TABLE 6.1 Key Statistics of the conversational graphs in both case studies.

Case Study	***Notre-Dame de Paris fire***		***Venice flood***	
Statistics	**Number/Count**	**Rate/Proportion**	**Number/Count**	**Rate/Proportion**
Nodes $\mathcal{V}$	179,758		11,961	
Merged Links $\mathcal{E}$	198,061		12,106	
Conversational Links $\mathcal{E}^{\text{CONV}}$	83,593	42.2%*	4786	39.5%*
User Links $\mathcal{E}^{\text{USER}}$	137,998	69.7%*	8426	69.6%*
Nodes with geo-tags $\{\mathfrak{v}_i \mid \mathfrak{l}_i \neq \varnothing\}$	132,073	73.5%	8745	73.1%
Graph Density	6.13e-6		8.46e-5	
Weakly-Connected Components (WCC)	17,323		1585	
Nodes in Largest WCC	87,680	48.8%	3375	28.2%
Graph Density in Largest WCC	1.48e-5		3.68e-4	

*summation larger than 100% because of links both as multiple types of conversational links and as user links.

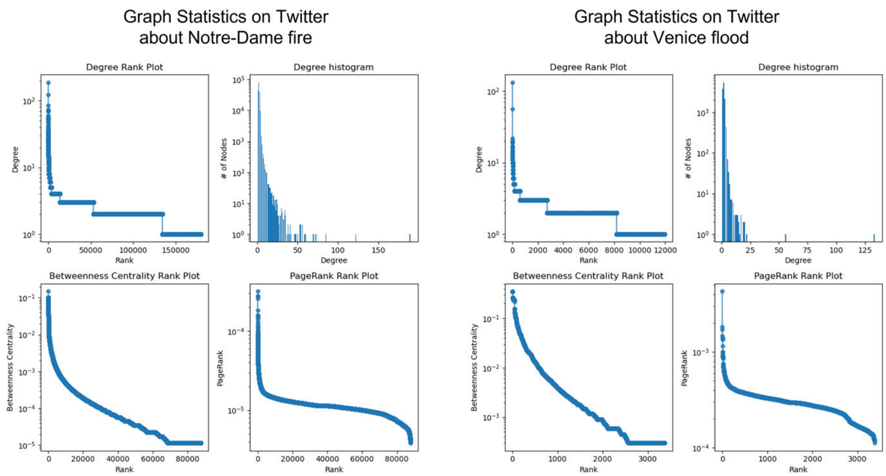


FIG. 6.6 The graph statistics on the conversation graph and/or its largest weakly-connected component in Notre-Dame and Venice, which includes: the log-scale rank-size plot of degree distribution in the entire graph, the log-scale degree histogram, the log-scale rank-size plot of the betweenness centrality on the largest component, and the log-scale rank-size plot of PageRank on the largest component.

When inspecting the tweet nodes with highest centralities, a result that was rather counter-intuitive emerged - from the perspective of heritage management, nodes with highest degrees were far less informative than those with a high betweenness centrality, while the significance of PageRank stands in between. In the graph of Notre-Dame, the two nodes with largest degrees (one of them also became the node with largest PageRank) are respectively a personal travel log and a post about football match. Only the nodes with the 3rd and 4th largest degree (one of them was also the node with second largest PageRank) discussed about the emotion attachment (‘horrible’, ‘sad’) and immediate actions (‘flying water tankers could be used… Must act quickly’) towards the HREs. All the nodes with top-3 betweenness centrality were about donations for the ‘reconstruction’, either expressing gratitude (‘you touched me so much’) or bringing up controversies within online debates (‘it’s beyond comprehensible’). Interestingly, the node with the 3rd largest PageRank discussed about the ethical and humanitarian necessity of Notre Dame as a World Heritage Site (‘what kind of humanity… for whom it was made?’). In the graph of Venice, the pattern gets more extreme. None of nodes with top-4 degrees are about the flood. All three nodes with highest betweenness centrality were blaming politicians and MOSE[20], the project that aimed at protecting Venice from flooding (‘don’t even know if it works’, ‘reverse MOSE effect’). Yet the nodes with top-3 PageRank were all spreading information of the flood as an event/incidence (‘second historical high tide… 82% of the city under water’). From the perspective of message passing, the significance of betweenness centrality and PageRank is also logical, since these two metrics imply how frequent/easy a message passing route would go through/stay on a node, thus influential for the entire information spreading mechanism. On the contrary, the nodes with high degrees can simply be a natural outcome of the graph construction process mentioned in Section 6.3.3. They can be simply the original posts that many other tweets referred to in the period of interest, whereas they themselves were posted long before the event, thus not directly relevant to the HREs.

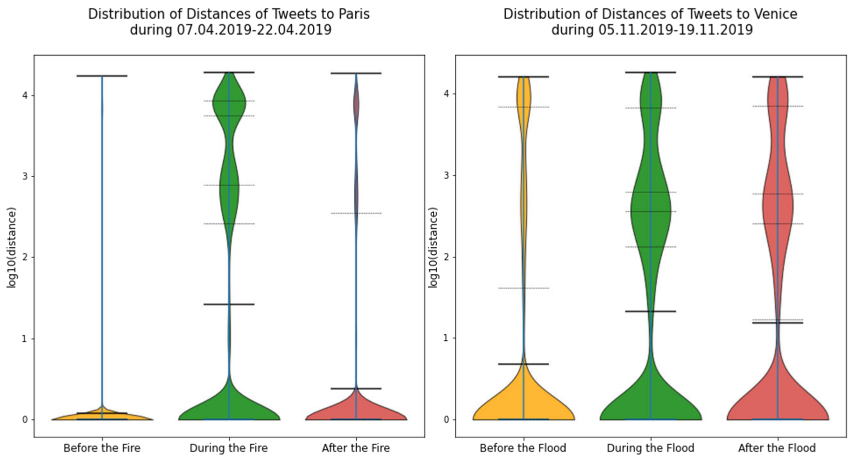


FIG. 6.7 The violin plots showing the distribution of distances of tweets to the core of a HRE before, during and after the event in Notre-Dame and Venice. The mean, 60%, 70%, 80%, and 90% deciles are visualized.

Furthermore, the distribution of the distances $\boldsymbol{d}$ from each tweet node (if $\mathfrak{c}_i \neq \varnothing$) to

[20] MOdulo Sperimentale Elettromeccanico in Italian, literally translated as ’Experimental Electromechanical Module’ https://en.wikipedia.org/wiki/MOSE, accessed 16 May 2023

the city where the HREs happened is visualized in Figure 6.7. It demonstrates the changes of global engagement. By testing on the ordinal variable of locality ('0' if a tweet was posted in the same city, '1' if they are from the same country, or '2' if they are from far beyond), Kruskal-Wallis H-tests showed a significant difference across the periods ($\boldsymbol{c}_{\mathrm{B}}$ before, $\boldsymbol{c}_{\mathrm{D}}$ during, and $\boldsymbol{c}_{\mathrm{A}}$ after HREs), $H(2) = 26,449.3, p < .001$ in Notre-Dame and $H(2) = 374.2, p < .001$ in Venice. Post-hoc two-tailed Mann–Whitney U-tests in Table 6.2 showed significant differences in the medians of locality among all pairs of HREs periods, where the period during HREs shows the broadest span of locality of tweets, possibly from far beyond. Almost all comparisons are justified with a medium Rank Biserial Correlation (i.e., the difference between the proportions of favorable and unfavorable evidence[21]) effect size with a larger absolute value of 0.1, except for the small effect size for the difference between during HREs and after HREs in Venice flood. Both statistics are calculated using Pingouin[22] Python library.

TABLE 6.2 Post-hoc Mann-Whitney U-tests comparing the median of ordinal variable Locality in different periods before, during, and after HREs.

Case Study		Notre-Dame de Paris fire			Venice flood		
Statistics		U-value	p-value	RBC*	U-value	p-value	RBC*
Before HREs	During HREs	840,735,944.0	<.001	-.409	4,090,061.5	<.001	-.230
Before HREs	After HREs	602,097,762.0	<.001	-.097	4,135,609.0	<.001	-.173
During HREs	After HREs	511,615,038.5	<.001	.318	3,532,576.5	<.001	.053

*Rank Biserial Correlation as effect size.

6.4.3 Detected Cultural Significance, Emotions, and Key Topics

As mentioned already in Section 6.3.4, the number of tweets with a non-empty pseudo-label for OUV selection criteria $\boldsymbol{\mathcal{Y}}^{\mathrm{OUV}}$, emotions $\boldsymbol{\mathcal{Y}}^{\mathrm{EMS}}$, and key topics $\boldsymbol{\mathcal{Y}}^{\mathrm{TOP}}$ are all smaller than the entire dataset and are different from each other. The relations of overlapping (in terms of number and proportion of tweets) of the three types of semantic labels for both case studies are visualized in Figure 6.8. It can be noted that the proportions demonstrate a very similar pattern with a significant Spearman correlation of $\rho = .976, p < .001$, where pure emotional expressions without mentioning cultural significance and key topics are consistently the majority, and the tweets with all three types of labels are always the minority. For all three types of semantic labels, more tweets had overlapping labels than standing alone, implying the associative nature of the cultural significance, emotions, and topics, focusing on classifying/clustering the tweets from a different perspective.

Among the detected OUV selection criteria $\boldsymbol{\mathcal{Y}}^{\mathrm{OUV}}$ possibly mentioned in the tweets, Criterion (vi) about people's association and activity, Criterion (iii) about the testimony of a [religious/cultural] tradition, and Criterion (iv) about the architectural

[21] https://pingouin-stats.org/build/html/generated/pingouin.mwu.html, accessed 10 Aug 2023
[22] https://pingouin-stats.org/build/html/index.html, accessed 16 May 2023

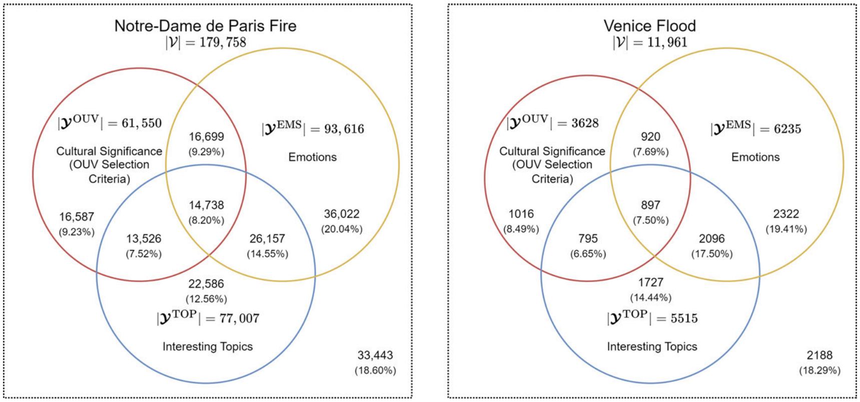


FIG. 6.8 The Venn Diagram of the number of tweets with each type of semantic label.

typology are consistently the three most significant ones, both in Notre-Dame and in Venice. Then there always followed Criterion (i) about a masterpiece, Criterion (ii) about values and influence, and Criterion (vii) about natural beauty, in a slightly different order. Even though "Paris, Banks of the Seine" as a UNESCO World Heritage property including Notre-Dame was only officially justified with OUV selection criteria (i)(ii)(iv), and "Venice and Its Lagoon" was only justified with the cultural criteria (i)(ii)(iii)(iv)(v)(vi), the appearance of tweets related to unjustified criteria, such as criteria (vii) in both cases, is not surprising. This is because the NLP models employed in this study only read the sentences literally and try to find the best-matching OUV selection criteria, with the knowledge of the wordings of the whole UNESCO World Heritage List. The labels given are therefore only an indication and not necessarily correct, especially as there lacks the step of comparison studies to justify if the described element with certain value is "universally outstanding". However, the number of tweets detected as relevant to OUV selection criteria also follows a logical order of how laypeople perceive the cultural significance of a city, especially during HREs: as tourism destinations for activities (vi), as a traditional landmark at risk of losses (iii), as a collection of grandiose buildings (iv) and masterworks (i), as a representation of cultural influences (ii), and as a scenery spot (vii) despite being cultural heritage.

Among the detected emotions and sentiments $\mathcal{Y}^{\text{EMS}}$ possibly expressed in the tweets, the emotion category 'others' and 'neutral' sentiment are consistently dominant in both case studies. In both cases, 'joy' and 'sadness' followed as equally sub-dominant explicit emotions, respectively pointing to the general sharing behaviour of people in an everyday context and the triggered sorrow after knowing the existence of a radical HRE. 'Anger' was also consistently the 3rd most expressed emotion, although less significant in Notre-Dame (roughly 1/3 of sadness) than in Venice (more than 1/2 of sadness). The existence of 'anger' as a main emotion being expressed is quite reasonable, as people could start looking for the actors to blame for an event (can be a group of people, some politicians, or a costly infrastructure) immediately after knowing it. 'Fear' and 'surprise' were both detected in Notre-Dame and Venice, but were not significant in either case and 'disgust' was never detected as the main emotion of a tweet.

The detected topics of interest $\mathcal{Z}^{s}$ can be grouped within six main themes in a hierarchical structure:

- **Emotions** that are mainly composed of words reflecting an explicit emotion, or repeatedly using certain emojis. This is the most significant topic cluster in both Notre-Dame and Venice.
- **Heritage** that explicitly or implicitly mentioned certain heritage values or heritage attributes considered as meaningful, such as “spire”, “rose window”, “architectural monument”, and “artefact” in Notre-Dame, and “Venetians” and “holiness” in Venice.
- **Incidence** that reported the development and severeness of the event, such as the description of “fire”, “burn”, “collapsed spire”', and “destroyed ashes” in Notre-Dame, and the description of “tide”, “high water”, “flood”, “climate change”, and a specific “bookstore” with “destroyed books” in Venice.
- **Actions** that either reflected on who and what to blame, such as “MOSE” in Venice, or called for further actions as monetary and emotional supports, such as “help”, “donation”, “rebuilt”, “reconstruct”, “laser scanner”, and “local management” in Notre-Dame, and “receive support” and “help Venice” in Venice.
- **Other Sites** that extensively mentioned and compared another associated and/or unrelated place or person, such as “Louvre”, “Victor Hugo”, “Eiffel Tower”, “Vatican”, and “national Museum” in Notre-Dame, and “Biennale” and “Venice Beach” in Venice.
- **Politics** that referred to a politician, a party, a movement, or a celebrity that can be possibly relevant, such as “Emmanuel Macron”, “elected officials”, “yellow vest” and “Henri Pinault” in Notre-Dame, while none of the politics-related topics seemed to relate to Venice flood.

Moreover, even though the conventional practice of topic modelling using BERTopic would disregard the remaining documents that cannot be clustered into any existing topics, it is found in this study that the keywords generated from such a ‘noise’ topic have a clear connection to heritage management. For example, the words “heritage” and “San Marco” respectively appeared in the ‘noise’ topic of Notre-Dame and Venice. Therefore, the ‘noise’ topic is kept and renamed as **Base**. A full list of keywords for each sub-topic within the six themes can be found in Appendix B.

6.4.4 The Spatiotemporal Dynamics of Semantics

The temporal development of the detected semantic information, i.e., the cultural significance $\boldsymbol{\mathcal{Y}}^{\mathrm{OUV}}$, the emotions $\boldsymbol{\mathcal{Y}}^{\mathrm{EMS}}$, and the key topics $\boldsymbol{\mathcal{Y}}^{\mathrm{TOP}}$ along with the HREs can be inspected with timelines. A selection of highly-relevant types of semantic information (Cultural Significance, Emotions, and Topics of “incidence” and “action”) is visualized in Figure 6.9 and 6.10. The full collection of timelines with all detected semantic topics can be found in Appendix B.

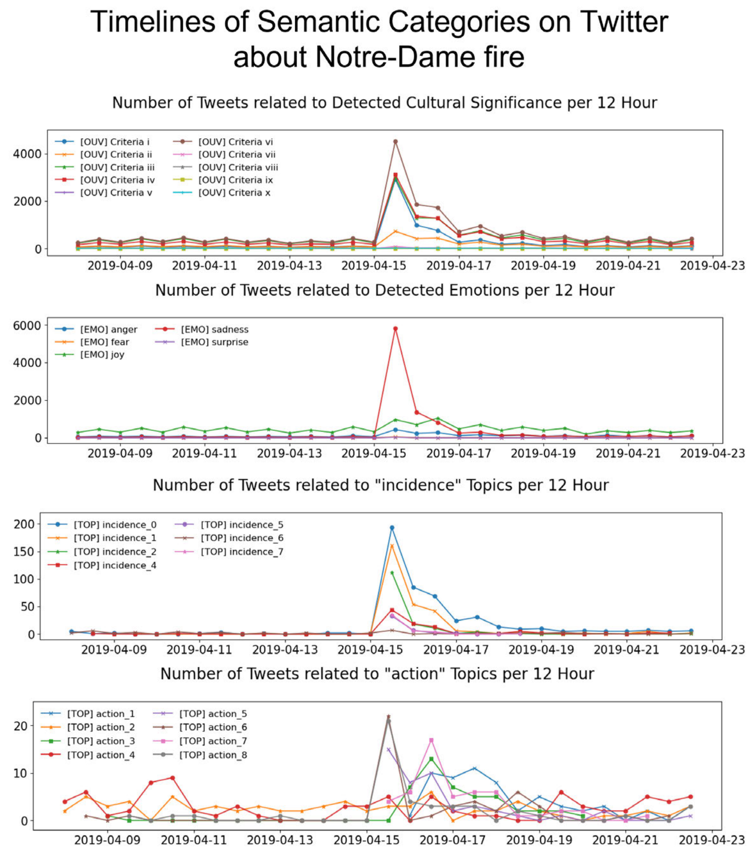


FIG. 6.9 A selection of timelines showing the temporal development of semantic information along with the HREs in the case of Notre-Dame fire.

Different from Figure 6.2 where the number of tweets is counted every hour, the tweets classified as related to each type of semantic information are counted every 12 hours in Figure 6.9 and 6.10 to allow for general temporal trends to emerge. Therefore, Figure 6.9 and 6.10 can be treated as both the smoothed (with a longer time window) and the factorized (as subsets of entire tweets) version of the general timeline. For almost all types of selected semantic information, the timelines demonstrate similar patterns: the tweets classified as related to the semantic information remained at a low level until the HREs happened when the intensity rose to a very high level for a short period; afterwards the intensity drew back to the normal level resembling the pre-HREs periods. This pattern is more obvious in the case of Notre-Dame fire as the contrast of intensity was extremely high.

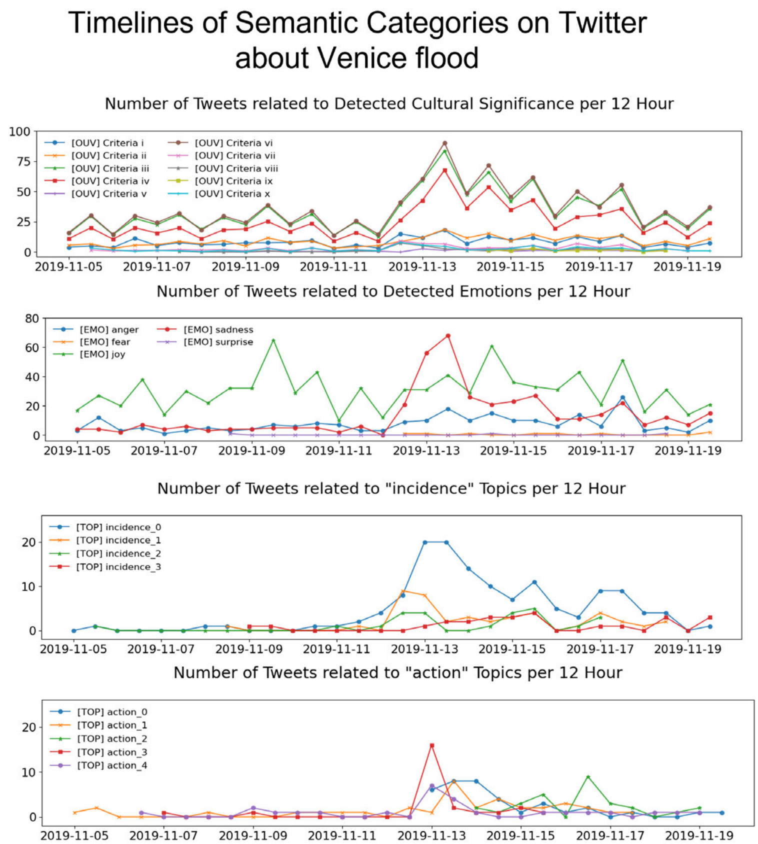


FIG. 6.10 A selection of timelines showing the temporal development of semantic information along with the HREs in the case of Venice flood.

For both cases, the OUV selection criteria mentioned in tweets that rose the most during HREs were criterion (vi) about people's association, criterion (iii) about testimony, and criterion (iv) about architectural typology. The probable reasons have been discussed in Section 6.4.3. Yet for Notre-Dame, the discussion concerning criterion (i) about masterpieces also increased significantly together with the other three, since people cared about the architectural monument (e.g., the spire and the rose window) and the important artefacts that can possibly be destroyed by the fire, which was slightly less worrisome in the case of flood in Venice.

The emotions of sadness and anger were both triggered to rise during both HREs. The sadness in Notre-Dame became 100-fold and got extremely dominant during the HREs, while in Venice the triggered sadness was only 10-fold, both of which dropped to about 2- to 3-fold of baseline periods before HREs. Even though not as significant as sadness, the anger in both cases also remained higher since HREs happened compared to the calm baseline periods before HREs. Interestingly, the dominant emotion of joy before the HREs also remained at a moderate level throughout the HREs and was back to dominance in the last days, and in the case of Notre-Dame, even rose a little bit on the days when the fire took place. This could be simply a result of a higher amount of tweets being posted.

The majority of discussions describing the incidence and actions emerged on the same day when the HREs happened. In the case of Notre-Dame, the descriptions of the fire taking place in the cathedral (incidences 1, 2, 5) and the courage of firefighters (incidence 4) mainly appeared on 15 and 16 April UTC and diminished quickly afterwards, while incidence 0 also mentioning the historic and symbolic meaning of Notre-Dame had another wave on 17 April, probably corresponding to the action 1 about the donations of French billionaires to rebuild the destroyed parts. Similarly, actions about rebuilding the Notre-Dame (action 5), the collapsed arrow in an identically modernized version (action 6) and other facilities with the help of local management (action 8) already existed immediately after the fire started on April 15, but were brought back to sight on 16 and 17 April when donations were made (action 1, 3). Remarkably, on 17 April, another wave of discussion went dominant (action 7) mentioning the late Belgian art historian Andrew Tallon and his work of using 3D laser scanning to build a digital model of Notre-Dame, as a prosperous source for the sake of restoration[23]. Other more general actions such as thinking (action 4) and appraising (action 2) did not demonstrate a clear temporal pattern related to the fire as HREs. In the case of Venice, on the other hand, the most dominant description of the incidence as the worst flooding in 50 years (incidence 0) extended to a few days after the starting point on 13 November, probably because the topic was also concerned with climate change and global warming as the hypothesized cause of the event. In the later days of the flooding, a specific topic emerged reporting the damaged books by the flood in Bertoni bookshop located in San Marco (incidence 4). From the first days of the flooding, the MOSE project was mentioned a lot (action 0,3) and criticized as a failure costing billions of euros. Interestingly, starting on 14 November and reaching its climax on 16 November, an online campaign to save Venice by donating one euro for each selfie made was initiated by the Comune di Venezia (action 2), following the discussion of support made by companies (action 1).

Aggregating the number of tweets under each type of semantic information for different periods (before, during, and after HREs) and different localities (same city, same country, and further away) in the same period (i.e., during HREs), the distributions can be visualized as the heatmap in Figure 6.11. The semantic categories that are too over-representative (the ‘base’ topic in Notre-Dame) or too scarce (the OUV selection criteria viii-x, and the emotions fear and surprise) are omitted from the visualization. Visually, it can already be observed that the distributions varied with different periods and localities. For example, in both cases, the emotion of sadness and the topic of incidences were significantly higher during HREs than before and after, and significantly more posted (proportionally, not necessarily numerically) in the same city than further away. However, the two case studies also demonstrated different spatiotemporal patterns concerning the distribution of semantic information. In the case of the Notre-Dame fire, significantly

[23]https://www.vassar.edu/stories/2019/190417-notre-dame-andrew-tallon.html, accessed 22 May 2023

more tweets concerning OUV criterion (i) about masterpieces were detected during HREs posted by people from France; people from Paris and France expressed extensively their sorrow as they reported the fire as an incidence and possible damage to this heritage property, while people from further away tried to suggest and/or take various actions to help Notre-Dame. In the case of the Venice flood, on the contrary, more anger (probably associated with MOSE), action-related, and heritage-related discussions were detected in Italy, while significantly more emotions- and/or emoji-related tweets were posted further away.

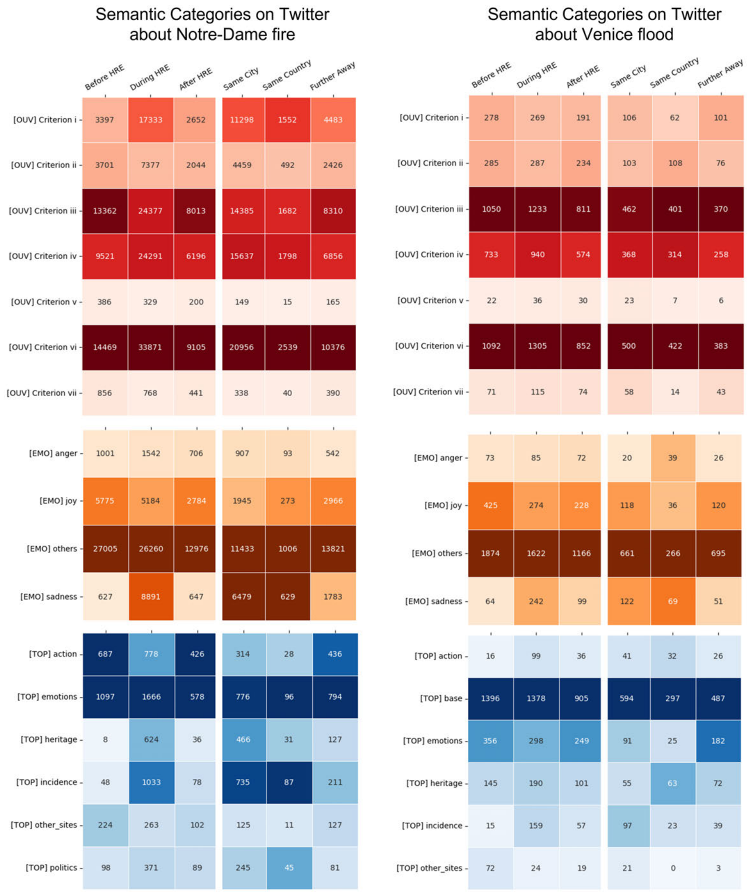


FIG. 6.11 The distribution of categorized top-3 OUV selection criteria, detected emotions, and key topics under each theme for different periods and for different localities all during HREs in Notre-Dame and Venice. The number of tweets belonging to the semantic information type (row) and the period/locality type (column) is annotated as a value matrix in the cell, while the colours of the heatmap are painted using the corresponding column-normalized matrix denoting the proportions. Note the row-sum of cells with different localities equals the cell of 'during HREs' of the same row.

Such observations have been further justified with Chi-square Contingency tests[24], a non-parametric version of two-way ANOVA tests for categorical variables, as reported in Table 6.3. Even though all the independent Chi-square tests showed a significant difference in distributions across different periods and different localities with $p < .01$, the effect size - Cramer's V in the tests of OUV was generally small, meaning that the difference may only be significant because of the large sample sizes. In the case study of Notre-Dame fire, the differences in the distributions of emotions and topics were mostly having a medium effect size, while for the Venice flood, the effect sizes were always small. This complex spatiotemporal dynamic of distributions for each type of semantic information invites further investigations.

TABLE 6.3 Independent Chi-square tests on the distributions of semantic labels across different periods and localities. The effect size Cramer's V is calculated as $V = \sqrt{\chi^2/(n \times df*)}$ following Gravetter et al. (2020), where $df*$ is the minimum of the number of rows or columns minus 1 (consistently $df* = 2$ in this case).

Statistics		Notre-Dame de Paris fire				Venice flood			
		χ^2	n	df	V	χ^2	n	df	V
OUV	Periods	3639.9***	182,689	12	.100††	24.7*	10,482	12	.034†
	Localities	646.2***	108,346	12	.055†	50.2***	4185	12	.077††
Emotions	Periods	8584.0***	93,398	6	.214†††	151.8***	6224	6	.110††
	Localities	3245.8***	41,877	6	.197††	95.2***	2223	6	.146††
Topics	Periods	1209.6***	8206	10	.271†††	221.7***	5515	10	.142††
	Localities	470.7***	4735	10	.223†††	154.6***	2148	10	.190††

*p<.05, ***p<.001, †very small effect size, ††small effect size, †††medium effect size.

6.5 Discussion

6.5.1 Indications for Heritage Management

Through the analyses in this chapter, two outcomes are reached that are meaningful for heritage management:

- Well-known knowledge and "common sense" have been confirmed using empirical data. This includes the fact that people will extensively express sadness and relentlessly share information about the damage during HREs and that the HREs will trigger discussions online and involve concerned people from far beyond, transcending geographical boundaries. Specifically, the pattern in Chapter 1

[24] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.chi2_contingency.html, accessed 20 May 2022

demonstrated with the Google trend search engine was successfully confirmed and restated with Twitter data.

- Previously less-known and/or surprising information also emerged in the detected and summarized key topics as valuable discoveries. This included the criticism of the MOSE project in Venice, the campaign of "oneeuroforoneselfie" by the Venice municipality, the rediscovery of the work by Professor Andrew Tallon for the Notre-Dame Cathedral, and the volume of tweets expressing joy and anger during HREs in both case studies.

The confirmation of well-known knowledge is meaningful for heritage management in the sense that it can support the decisions made efficiently based on past experience and heuristics. It also shows the necessity for expanding the definition of heritage community suggested in Faro Convention (Council of Europe, 2005) in the time of radical events, by including the temporally-founded communities bonded by the HREs into the scope. Moreover, the discovery of less-known pieces of information is even more valuable and informative for heritage management. On the one hand, not all scholars and practitioners studying an HRE are acquainted with the heritage property and the event in depth, especially in the case of global collaborations (with geospatial and sociopolitical distance) and retrospective historic investigations (with temporal and cultural distance). On the other hand, even for local managers and sophisticated scholars studying a heritage property for years, some specific aspects of knowledge can still be overlooked if they are, by any chance, trapped in their own information bubbles with confirmation bias (Bozdag et al., 2014; Suzuki and Yamamoto, 2020). The workflow demonstrated in this chapter provides the possibility for end users to acquire new information at a large scale, being an effective and transferable knowledge documentation tool that has the potential of being applied globally, possibly useful for inclusive heritage management and planning as suggested by HUL (UNESCO, 2011). Interestingly, from a retrospective view, MOSE was not yet fully completed and put in use in 2019, and it managed to prevent Venice from an even larger flooding in 2022[25]. In this case, the proposed framework can also help engineers and historians of technology to reveal the dynamics and mechanisms of the public reactions concerning a major project. The tools and workflow proposed in this chapter can understood as an "observatory" of specific heritage properties, which can be eventually turned into a dashboard or "thermometer" to monitor the reactions and social sentiments of public about built heritage.

When commenting on the usage of online media during a radical event in the digital era, Garduño Freeman and Gonzalez Zarandona (2021) brought up the examples of the Notre Dame fire and Palmyra destruction. Whereas the search volume on Google Ads increased 60-fold in response to the former event, it only increased seven and a half times for the latter. Garduño Freeman and Gonzalez Zarandona (2021) criticised that this seemed to suggest that:

> "one site was mourned by more people than the other, so much so that it has created the impression of a **Notre-Dame effect**."

[25] News article Marea a Venezia. Il Mose salva la città, l'acqua tocca 204 centimetri, accessed 26 May 2023.

They further argued that the so-called Notre-Dame effect and the broader concept of "mediatisation of heritage" entailed digital colonialism, challenged the equality and equity of UNESCO World Heritage properties co-existing in the same list, and created spectres reflecting on the aesthetic, economic, social, and political values of the European culture, only composing a subset of heritage values proposed by Pereira Roders (2007) and Tarrafa Silva and Pereira Roders (2010). Even though the case studies in this chapter, Notre-Dame de Paris and Venice, both come from the "Canon" of architectural and urban history in Europe, thus not able to be simply abstracted and explained with ideological and ethnic divisions, a similar effect of one HREs raising more attention than the other can be observed. The consistencies, similarities, nuances, and significant differences between the two case studies from different aspects suggest that there might be some general rules behind people's online actions and reactions in any HRE, which is applied at a different level and adapted to the specificity of the event. Therefore, extra caution and consciousness are needed by researchers and practitioners during the interpretation of the results in future applications of the methodological framework proposed in this chapter in other case studies of HREs distributed globally. Especially, the cultural significance of a World Heritage property should not be over-simplified during planning and decision-making as a set of "valuable linguistic metonyms" (keywords) targeted at only specific groups of audiences (Garduño Freeman and Gonzalez Zarandona, 2021).

6.5.2 Limitations and Future Studies

Since one of the main research interests of this study is to investigate the spatiotemporal patterns of the tweets posted during HREs, the data collected in the study was naturally restricted to the ones that are either initially given a geo-location (first two rounds of search as mentioned in Section 6.2.2) or directly connected to the tweets with geo-locations (last round of "supplemental search"). This restriction also automatically limited the scope of the study since it has been shown that only a small proportion of tweets are accompanied by geo-locations while posting (Cheng and Wicks, 2014). There exist massive online interactions between tweets that are not geo-coded. Future studies could lift up this restriction on geo-tags and collect a more inclusive initial dataset in the second step, i.e., the "global search", and query for all tweets both directly and/or indirectly (with a distance of 2-3 network steps) responding to and being responded by the seed tweets in the third step of "supplemental search". Afterwards, Named-Entity Recognition (Won et al., 2018) and other relevant techniques could be used to infer the geo-locations of the tweets if they are not explicitly given (Zhang and Gelernter, 2014). As such, the modules of constructing graphs and generating semantic labels are still valid, whereas a more comprehensive view of the dynamic conversation behaviour could be obtained, albeit probably less focused on the spatial aspects.

Different from the conventional event detection studies utilizing spatiotemporal information (Cheng and Wicks, 2014; Kersten and Klan, 2020; George et al., 2021), where spatiotemporal clustering algorithms are first used to find significant clusters before feeding the textual information to topic models such as LDA to semantically describe each cluster, this chapter skipped the step of spatiotemporal clustering. Instead, the tweets were clustered implicitly with their semantic information when conducting topic modelling, as an internal module of BERTopic (Grootendorst, 2022). This was only a pragmatic choice as it was assumed that the tweets standing alone from any significant cluster can still contribute to the online discussion arena and form the temporary heritage community with their semantic expressions. However, including an additional step of spatiotemporal clustering either before or after the topic modelling could give another layer of interpretation to the results. The answered questions would therefore become “What are the expressed emotions and main semantic topics being discussed within each cluster that is significantly distinguishable by its spatiotemporal density”, which could also be an interesting topic for future studies.

Traditional topic modelling algorithms such as LDA are known as unstable against different configurations, hard to reproduce, and work badly with short texts such as tweets (Dahal et al., 2019). Merging the tweets at the level of users into user documents can be an easy strategy to resolve the problem. The usage of BERTopic partly resolves the issue and makes it possible to obtain clear topics and fine-grained tweet-level predictions, well-fitting the purpose of this study. Depending on the specific questions of interest, the tweets collected in this study could also be merged at the level of users, communities, interest groups, cities, countries, and/or spatiotemporal clusters to be detected. Moreover, the BERTopic models are still not totally stable and reproducible when being run multiple times, indicating that it can still not yet be a fully automatic algorithm and human experts are always needed to control the quality of topics and select the relevant ones for further interpretation. End users applying the methodological framework proposed in this study should be informed of this limitation.

Furthermore, starting from the collected dataset and conducted exploratory analyses, many more interesting questions in the fields of heritage studies, urban studies, social sciences, computer science, and Geo-AI research could be answered. By repeating the same procedure in other case studies of HREs concerning World Heritage properties with different geopolitical and cultural backgrounds happening in different years (Kumar, 2020a,b), possibly also with positive events, general rules of online interaction discussed in 6.5.1 could be verified, resulting in a handbook for heritage managers on how to act and react on social media with concerned people during events. By digging into the semantic development embedded in the conversational graph structure, the mechanisms of information spreading, stance changing, and interactional framing could be further revealed (Lipizzi et al., 2015; Luo et al., 2020; van Eck et al., 2020). The time zones of the posting locations, the language being used, and the social interests of users can all be possibly used as grouping variables to describe and explain the spatiotemporal patterns of posting behaviour, the development of semantic information, as well as the interaction mechanisms behind them. Such mechanisms are supposed to be generalizable across fields beyond the scope of heritage, but can also be used to explain discussions and debates on other societal issues triggering public interactions, such as sustainability actions, climate change campaigns, and global pandemics (Dewulf and Bouwen, 2012; Roy and Goldwasser, 2020; Stevens et al., 2020). Only textual information has been analysed for the semantic meaning, yet a multi-modal representation including images, memes, audio, and videos can possibly add other contradictory or complementary information (Bai et al., 2022; Rojas-Padilla et al., 2022). The emojis used in tweets were not thoroughly investigated in this chapter, similar operations as in Tenzer (2022) could also be conducted to compare the change of emoji usage before, during, and after the HREs. The results of this study could be combined with other similar and/or relevant works collecting information using social media in the case of Notre-Dame fire (Padilha et al., 2021a; Passaro et al., 2022) and Venice flood (Andrade, 2022; Lorini et al., 2022), in order to construct a multi-layer digital archive concerning the event.

6.6 Conclusions

This chapter presents a methodological framework proposed to investigate the collective behaviour of people on social media when radical Heritage-related events (HREs) happen. It applies a few pre-trained natural language processing models to obtain pseudo-labels of tweets in the time of HREs for their semantic meanings in terms of conveyed cultural significance, expressed emotions, and discussed topics. The conversational sequences and the spatiotemporal contexts are modelled in a graph structure. Two case studies that both happened in 2019, the fire in Notre-Dame de Paris and the flood in Venice, are used as demonstrative examples to showcase the framework. Exploratory data analysis and statistical tests are conducted to describe the spatiotemporal dynamics of the actions and reactions of the online public from the same city, the same country, and far beyond within the periods before, during, and after HREs. Results show that the online discussions went far beyond the local heritage community and triggered vivid expressions of emotions and action proposals globally, even though the reactions drew back quickly after the HREs. The methodological framework can be also applied in other similar cases of events happening to heritage globally and can facilitate inclusive heritage management processes as an information gathering and eventually a knowledge documentation tool to confirm known facts and discover new ones.

References

Acheampong, F. A., Wenyu, C., and Nunoo-Mensah, H. (2020). Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports, 2(7):e12189.

Adepeju, M. (2017). Modelling of sparse spatio-temporal point process (STPP): an application in predictive policing. PhD thesis, UCL (University College London).

Afyouni, I., Khan, A., and Aghbari, Z. A. (2022). Deep-eware: spatio-temporal social event detection using a hybrid learning model. Journal of Big Data, 9.

Afzaal, M., Usman, M., Fong, A. C., and Fong, S. (2019). Multiaspect-based opinion classification model for tourist reviews. Expert Systems, 36(2):e12371.

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In Aggarwal, C. C., editor, Social Network Data Analytics, chapter 1, pages 1–15. SPRINGER.

Andrade, B. (2022). I can see through the waters eyes. covid-19 in heritage cities: Citizen participation and self-organization for greater conservation and sustainability: The case of venezia pulita (clean venice). In Living (World) Heritage Cities: Opportunities, challenges, and future perspectives of people-centered approaches in dynamic historic urban landscapes, pages 199–212. Sidestone Press.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. (1999). Optics: Ordering points to identify the clustering structure. ACM Sigmod record, 28(2):49–60.

Arjona, J. O. (2020). Analysis of the space-temporal patterns of events from twitter data: The case of madrid 2017 world pride. Estudios Geograficos, 81.

Bai, N., Cheng, T., Nourian, P., and Pereira Roders, A. (2023a). An exploratory data analysis of the spatiotemporal patterns of heritage-related events on twitter. The 30th International Conference on Geoinformatics (CPGIS 2023), London, UK.

Bai, N., Luo, R., Nourian, P., and Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage statements of "Outstanding Universal Value" with soft labels. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 366–384, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2022). Heri-graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. ISPRS International Journal of Geo-Information, 11(9).

Bai, N., Nourian, P., and Pereira Roders, A. (2021b). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.

Bai, N., Nourian, P., Pereira Roders, A., Bunschoten, R., Huang, W., and Wang, L. (2023b). Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data. Environment and Planning B: Urban Analytics and City Science, 50(1):94–116.

Barabási, A.-L. (2013). Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1987):20120375.

Birant, D. and Kut, A. (2007). St-dbscan: An algorithm for clustering spatial–temporal data. Data & knowledge engineering, 60(1):208–221.

Bird, S., Klein, E., and Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022.

Bozdag, E., Gao, Q., Houben, G.-J., and Warnier, M. (2014). Does offline political segregation affect the filter bubble? an empirical analysis of information diversity for dutch and turkish twitter users. Computers in human behavior, 41:405–415.

Card, D., Boydstun, A., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 438–444.

Cheng, T. and Wicks, T. (2014). Event detection using twitter: A spatio-temporal approach. PloS one, 9(6):e97807.

Chianese, A., Marulli, F., and Piccialli, F. (2016). Cultural Heritage and Social Pulse: A Semantic Approach for CH Sensitivity Discovery in Social Media Data. In Proceedings - 2016 IEEE 10th International Conference on Semantic Computing, ICSC 2016, pages 459–464. Institute of Electrical and Electronics Engineers Inc.

Choi, C. and Hong, S.-Y. (2021). Mdst-dbscan: A density-based clustering method for multidimensional spatiotemporal data. ISPRS International Journal of Geo-Information, 10(6):391.

Clemens, K. (2015). Geocoding with openstreetmap data. GEOProcessing 2015, page 10.

Costa, M. A. and Kulldorff, M. (2014). Maximum linkage space-time permutation scan statistics for disease outbreak detection. International journal of health geographics, 13(1):1–14.

Council of Europe (2005). Convention on the value of cultural heritage for society (faro convention). Technical report, Council of Europe, Faro.
Cristelli, M., Batty, M., and Pietronero, L. (2012). There is more than a power law in zipf. Scientific reports, 2(1):1–7.
Dahal, B., Kumar, S. A., and Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. Social network analysis and mining, 9:1–20.
Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
Dewulf, A. and Bouwen, R. (2012). Issue framing in conversations for change: Discursive interaction strategies for "doing differences". The Journal of Applied Behavioral Science, 48(2):168–193.
Dickinger, A. and Lalicic, L. (2016). An analysis of destination brand personality and emotions: a comparison study. Information Technology & Tourism, 15:317–340.
Eisenstein, J. (2018). Natural Language Processing. MIT Press.
Ekman, P. (1992). An argument for basic emotions. Cognition & emotion, 6(3-4):169–200.
Farnaghi, M., Ghaemi, Z., and Mansourian, A. (2020). Dynamic spatio-temporal tweet mining for event detection: a case study of hurricane florence. International Journal of Disaster Risk Science, 11:378–393.
Ferrarin, C., Bajo, M., Benetazzo, A., Cavaleri, L., Chiggiato, J., Davison, S., Davolio, S., Lionello, P., Orlić, M., and Umgiesser, G. (2021). Local and large-scale controls of the exceptional venice floods of november 2019. Progress in Oceanography, 197:102628.
García-Vega, M., Díaz-Galiano, M., García-Cumbreras, M., Del Arco, F., Montejo-Ráez, A., Jiménez-Zafra, S., Martínez Cámara, E., Aguilar, C., Cabezudo, M., Chiruzzo, L., et al. (2020). Overview of tass 2020: Introducing emotion detection. In Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, pages 163–170.
Garduño Freeman, C. and Gonzalez Zarandona, J. A. (2021). Digital spectres: the notre-dame effect. International Journal of Heritage Studies, 27(12):1264–1277.
George, Y., Karunasekera, S., Harwood, A., and Lim, K. H. (2021). Real-time spatio-temporal event detection on geotagged social media. Journal of Big Data, 8(1):91.
Ginzarly, M. and Srour, F. J. (2022). Cultural heritage through the lens of covid-19. Poetics, 92:101622.
Gravetter, F. J., Wallnau, L. B., Forzano, L.-A. B., and Witnauer, J. E. (2020). Essentials of statistics for the behavioral sciences. Cengage Learning.
Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv preprint arXiv:2203.05794.
Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. In Gurevych, I. and Miyao, Y., editors, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers, pages 328–339. Association for Computational Linguistics.
Hu, Y., Wang, F., Guin, C., and Zhu, H. (2018). A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. Applied geography, 99:89–97.
Huang, Y., Li, Y., and Shan, J. (2018). Spatial-temporal event detection from geo-tagged tweets. ISPRS International Journal of Geo-Information, 7(4):150.
Jurafsky, D. and Martin, J. H. (2020). Speech and language processing An Intoduction to Natural Language Processing, Computational Liguistics, and Speech Recognition. Third Edition Draft.
Kersten, J. and Klan, F. (2020). What happens where during disasters? a workflow for the multifaceted characterization of crisis events based on twitter data. Journal of Contingencies and Crisis Management, 28(3):262–280.
Kounadi, O., Lampoltshammer, T. J., Leitner, M., and Heistracher, T. (2013). Accuracy and privacy aspects in free online reverse geocoding services. Cartography and Geographic Information Science, 40(2):140–153.
Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R., and Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. PLoS medicine, 2(3):e59.
Kumar, P. (2019). Learning from the Past and Preparing for the Future: Cases and Tools for Cultural Heritage during Disasters. PhD thesis, IMT School for Advanced Studies Lucca.
Kumar, P. (2020a). Crowdsourcing to rescue cultural heritage during disasters: A case study of the 1966 florence flood. International Journal of Disaster Risk Reduction, 43:101371.
Kumar, P. (2020b). Twitter, disasters and cultural heritage: A case study of the 2015 nepal earthquake. Journal of Contingencies and Crisis Management, 28(4):453–465.
Kumar, P., Ofli, F., Imran, M., and Castillo, C. (2020). Detection of disaster-affected cultural heritage sites from social media images using deep learning techniques. Journal on Computing and Cultural Heritage (JOCCH), 13(3):1–31.
Li, L.-J. and Fei-Fei, L. (2007). What, where and who? classifying events by scene and object recognition. In 2007 IEEE 11th international conference on computer vision, pages 1–8. IEEE.
Lipizzi, C., Iandoli, L., and Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using twitter streams. International Journal of Information Management, 35(4):490–503.

Liu, M., Liu, X., Li, Y., Chen, X., Hauptmann, A. G., and Shan, S. (2016). Exploiting feature hierarchies with convolutional neural networks for cultural event recognition. Proceedings of the IEEE International Conference on Computer Vision, 2016-Febru:274–279.

Lorini, V., Rufolo, P., and Castillo, C. (2022). Venice was flooding... one tweet at a time. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2):1–16.

Luo, Y., Card, D., and Jurafsky, D. (2020). Detecting stance in media on global warming. arXiv preprint arXiv:2010.15149.

Lupo, B. M. (2021). Patrimônio cultural e cat´ ástrofe: Os concursos internacionais não-oficiais realizados para a notre dame de paris após o incêndio de 2019. Herança-Revista de História, Património e Cultura, 4(2):018–038.

Marine-Roig, E., Martin-Fuentes, E., and Daries-Ramon, N. (2017). User-generated social media events in tourism. Sustainability, 9(12):2250.

Martínez-López, B., Perez, A., and Sánchez-Vizcaíno, J. (2009). Combined application of social network and cluster detection analyses for temporal-spatial characterization of animal movements in salamanca, spain. Preventive veterinary medicine, 91(1):29–38.

Mazloom, M., Hendriks, B., and Worring, M. (2017). Multimodal context-aware recommender for post popularity prediction in social media. In Proceedings of the on Thematic Workshops of ACM Multimedia 2017, pages 236–244.

McInnes, L., Healy, J., and Astels, S. (2017). hdbscan: Hierarchical density based clustering. J. Open Source Softw., 2(11):205.

Meghraoui, M. and Sbeinati, R. (2023). Archeoseismology and the lost villages in northern syria, the impact of large earthquakes on cultural heritage. In Sustainable Conservation of UNESCO and Other Heritage Sites Through Proactive Geosciences, pages 445–461. Springer.

Molina, S. O. and Molina, M. Á. O. (2021). Notre-dame de parís. dov'era e com'era: la réplica que habita en la ruina. Loggia, Arquitectura & Restauración, 34:8–27.

Monteiro, V., Henriques, R., Painho, M., and Vaz, E. (2014). Sensing World Heritage An Exploratory Study of Twitter as a Tool for Assessing Reputation. In Murgante, B and Misra, S and Rocha, AMAC and Torre, C and Rocha, JG and Falcao, MI and Taniar, D and Apduhan, BO and Gervasi, O., editor, COMPUTATIONAL SCIENCE AND ITS APPLICATIONS - ICCSA 2014, PT II, volume 8580 of Lecture Notes in Computer Science, pages 404–419. Univ Minho; Univ Perugia; Univ Basilicata; Monash Univ; Kyushu Sangyo Univ; Assoc Portuguesa Investigacao Operac.

Moreno-Sánchez, I., Font-Clos, F., and Corral, Á. (2016). Large-scale analysis of zipf's law in english texts. PloS one, 11(1):e0147073.

Naramski, M., Szromek, A. R., Herman, K., and Polok, G. (2022). Assessment of the activities of european cultural heritage tourism sites during the covid-19 pandemic. Journal of Open Innovation: Technology, Market, and Complexity, 8(1):55.

Nenko, A. and Petrova, M. (2018). Emotional geography of st. petersburg: detecting emotional perception of the city space. In Digital Transformation and Global Society: Third International Conference, DTGS 2018, St. Petersburg, Russia, May 30–June 2, 2018, Revised Selected Papers, Part II 3, pages 95–110. Springer.

Nguyen, D. Q., Vu, T., and Nguyen, A. T. (2020). Bertweet: A pre-trained language model for english tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 9–14.

Nourian, P., Rezvani, S., Sariyildiz, I., and van der Hoeven, F. (2016). Spectral modelling for spatial network analysis. In Proceedings of the Symposium on Simulation for Architecture and Urban Design (simAUD 2016), pages 103–110. SimAUD.

Padilha, R., Andaló, F. A., Lavi, B., Pereira, L. A., and Rocha, A. (2021a). Temporally sorting images from real-world events. Pattern Recognition Letters, 147:212–219.

Padilha, R., Andaló, F. A., Pereira, L. A., and Rocha, A. (2021b). Unraveling the notre-dame cathedral fire in space and time: an x-coherence approach. In Crime Science and Digital Forensics, pages 3–19. CRC Press.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Pan, J., Mou, N., and Liu, W. (2019). Emotion analysis of tourists based on domain ontology. In Proceedings of the 2019 International Conference on Data Mining and Machine Learning, pages 146–150.

Passaro, L. C., Bondielli, A., Dell'Oglio, P., Lenci, A., and Marcelloni, F. (2022). In-context annotation of topic-oriented datasets of fake news: A case study on the notre-dame fire event. Information Sciences, 615:657–677.

Pereira Roders, A. (2007). Re-architecture: lifespan rehabilitation of built heritage. PhD thesis, Technische Universiteit Eindhoven.

Pereira Roders, A. (2019). The Historic Urban Landscape Approach in Action: Eight Years Later. In Reshaping Urban Conservation, pages 21–54. Springer.

Pérez, J. M., Furman, D. A., Alonso Alemany, L., and Luque, F. M. (2022). RoBERTuito: a pre-trained language model for social media text in Spanish. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 7235–7243, Marseille, France. European Language Resources Association.

Praticò, Y., Ochsendorf, J., Holzer, S., and Flatt, R. J. (2020). Post-fire restoration of historic buildings and implications for notre-dame de paris. Nature Materials, 19(8):817–820.

Pérez, J. M., Giudici, J. C., and Luque, F. (2021). pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks.
Rani, M. and Kaushal, S. (2022). Geoclust: Feature engineering based framework for location-sensitive disaster event detection using ahp-topsis. Expert Systems with Applications, 210.
Rao, D. and McMahan, B. (2019). Natural Language Processing with PyTorch - Build Intelligent Language Applications Using Deep Learning. O'Reilly Media, Inc.
Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.
Rojas-Padilla, E., Metze, T., and Termeer, K. (2022). Seeing the visual: A literature review on why and how policy scholars would do well to study influential visualizations. Policy Studies Yearbook, 12(1):103–136.
Roy, S. and Goldwasser, D. (2020). Weakly supervised learning of nuanced frames for analyzing polarization in news media. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 7698–7716.
Shen, J. (2018). Profiling and Grouping Space-time Activity Patterns of Urban Individuals. PhD thesis, UCL (University College London).
Shi, Z. and Pun-Cheng, L. S. (2019). Spatiotemporal data clustering: a survey of methods. ISPRS international journal of geo-information, 8(3):112.
Sofaer, J., Davenport, B., Sørensen, M. L. S., Gallou, E., and Uzzell, D. (2021). Heritage sites, value and wellbeing: learning from the covid-19 pandemic in england. International Journal of Heritage Studies, 27(11):1117–1132.
Stevens, T. M., Aarts, N., and Dewulf, A. (2020). Using emotions to frame issues and identities in conflict: farmer movements on social media. Negotiation and Conflict Management Research.
Suzuki, M. and Yamamoto, Y. (2020). Analysis of relationship between confirmation bias and web search behavior. In Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services, pages 184–191.
Taecharungroj, V. and Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tourism Management, 75(July):550–568.
Tarrafa Silva, A. and Pereira Roders, A. (2010). The cultural significance of World Heritage cities : Portugal as case study. In Heritage and Sustainable Development, pages 255–263, Évora, Portugal.
Tenzer, M. (2022). Tweets in the peak: Twitter analysis-the impact of covid-19 on cultural landscapes. Internet Archaeology, 59.
Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., and Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018).
UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.
van Eck, C. W., Mulder, B. C., and Dewulf, A. (2020). Online climate change polarization: Interactional framing analysis of climate change blog comments. Science Communication, 42(4):454–480.
Wang, T.-C. and Phoa, F. K. H. (2016). A scanning method for detecting clustering pattern of both attribute and structure in social networks. Physica A: Statistical Mechanics and its Applications, 445:295–309.
Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.
Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In Liu, Q. and Schlangen, D., editors, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020 - Demos, Online, November 16-20, 2020, pages 38–45.
Won, M., Murrieta-Flores, P., and Martins, B. (2018). ensemble named entity recognition (ner): evaluating ner tools in the identification of place names in historical corpora. Frontiers in Digital Humanities, 5:2.
Yanenko, O. and der Weberei, A. (2019). Introducing social distance to st-dbscan. In Proceedings of the 22nd AGILE Conference.
Zagato, L. et al. (2015). The notion of "heritage community" in the council of europe's faro convention. its impact on the european legal framework. Adell Nicolas, pages 141–168.
Zhai, X., Luo, Q., and Wang, L. (2020). Why tourists engage in online collective actions in times of crisis: Exploring the role of group relative deprivation. Journal of Destination Marketing and Management, 16(August 2019).
Zhang, W. and Gelernter, J. (2014). Geocoding location expressions in twitter messages: A preference learning method. Journal of Spatial Information Science, 9:37–70.
Zhang, Y. and Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Computers, Environment and Urban Systems, 79:101403.

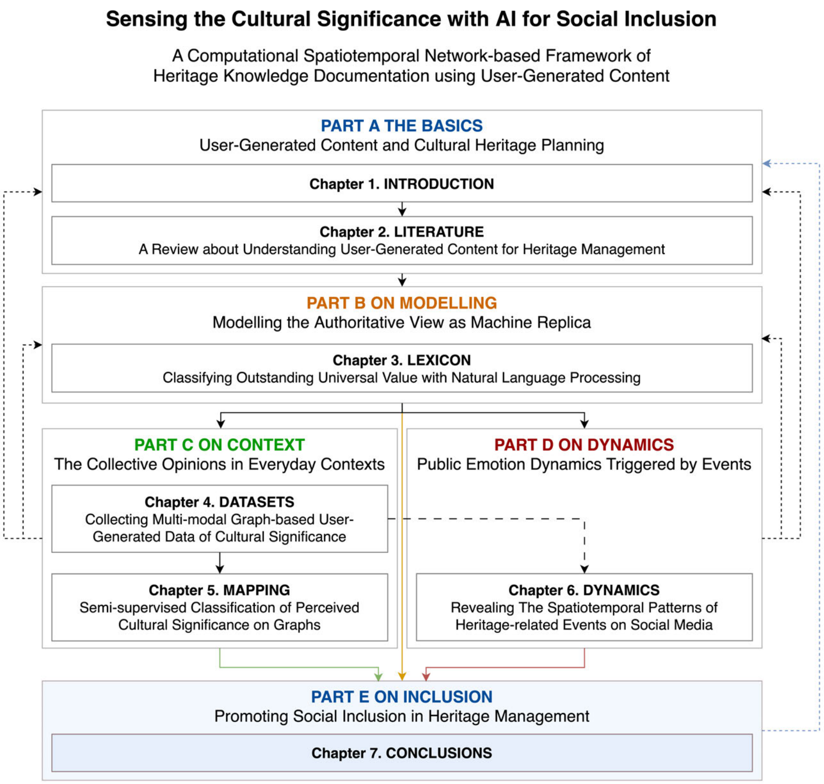
PART E

On Inclusion

Promoting Social Inclusion in Heritage Management

This part of dissertation concludes the research by summarizing the main outcomes, reflecting on the research questions, and pointing to future directions. The modelling of the machine replica in PART B, the mapping of everyday contexts of cultural significance under baseline scenarios in PART C, and the descriptions of discussion dynamics triggered by radical events under activated scenarios in PART D are respectively used to respond to the research aim. The basis of the dissertation brought up in PART A is revisited with evidence obtained throughout the dissertation. As a whole, it contributes as a knowledge documentation tool under the Recommendation on the Historic Urban Landscape, in pursuit of inclusive heritage management processes in the future.

One chapter is included in this part:


Sensing the Cultural Significance with AI for Social Inclusion
A Computational Spatiotemporal Network-based Framework of
Heritage Knowledge Documentation using User-Generated Content
PART A THE BASICS
User-Generated Content and Cultural Heritage Planning
Chapter 1. INTRODUCTION
Chapter 2. LITERATURE
A Review about Understanding User-Generated Content for Heritage Management
PART B ON MODELLING
Modelling the Authoritative View as Machine Replica
Chapter 3. LEXICON
Classifying Outstanding Universal Value with Natural Language Processing
PART C ON CONTEXT
The Collective Opinions in Everyday Contexts
Chapter 4. DATASETS
Collecting Multi-modal Graph-based User-Generated Data of Cultural Significance
Chapter 5. MAPPING
Semi-supervised Classification of Perceived Cultural Significance on Graphs
PART D ON DYNAMICS
Public Emotion Dynamics Triggered by Events
Chapter 6. DYNAMICS
Revealing The Spatiotemporal Patterns of Heritage-related Events on Social Media
PART E ON INCLUSION
Promoting Social Inclusion in Heritage Management
Chapter 7. CONCLUSIONS

7 Conclusions

Parts of this chapter have been published in Bai et al. (2023a).
Bai N, Ducci M, Mirzikashvili R, Nourian P, Pereira Roders, A. (2023a). Mapping Urban Heritage Images with Social Media Data and Artificial Intelligence, A Case Study in Testaccio, Rome. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023. p. 139–146.

ABSTRACT This chapter concludes the dissertation. It provides a reflection on the proposed interdisciplinary methodology and the main findings obtained in each chapter concerning the use of Artificial Intelligence to understand User-Generated Content and therefrom summarize the perceived and expressed cultural significance of cities conveyed to social media users. The research questions are revisited and addressed with the added value of this research. The scientific and societal contributions and the key limitations are re-stated. Finally, suggestions for future research agendas are given on how to utilize the research outcomes in inclusive heritage management practices, as well as computational urban and social studies.

KEYWORDS Inclusive Heritage Management, Artificial Intelligence, Computational Social Sciences, Cultural Significance, Historic Urban Landscape

7.1 Summary of Main Outcomes

This research started with the observation that the reactions of the online public to cultural heritage on social media vary in **baseline scenarios** when people calmly share their travelling/living experiences in the cities they visit and/or reside and in **activated scenarios** when radical events such as the fire in Notre-Dame de Paris happened. Both scenarios would hypothetically demonstrate different patterns and intensities of social inclusion within discussions concerning heritage, informative for summarizing the cultural significance conveyed by the public from alternative views. In order to build a knowledge documentation tool concerning cultural significance with User-Generated information on social media, which is hypothetically more socially inclusive compared to the traditional approach, as called for by the UNESCO Recommendation on the Historic Urban Landscape (UNESCO, 2011; Pereira Roders, 2019), both scenarios are indispensable. Contrary to the conventional expert-based case-specific heritage investigations where conclusions are mainly drawn from extrinsic expertise and intrinsic professional knowledge gradually built up through years of training (UNESCO, 1972, 2008), summarizing, analysing, and eventually mapping cultural significance from massive user-generated data in a global context requires a different set of skills from various disciplines. Considering the large amount of unstructured data (mostly texts and images) available, it is pragmatically hard, if not impossible, to process them manually and qualitatively, promising both efficiency and quality. The cutting-edge Artificial Intelligence models that have been extensively developed and pre-trained, showing the ability to be transferred and generalized in other types of tasks (Pan and Yang, 2010), offer the opportunity to augment the research field of heritage studies from the bottom up with large-scale evidence that can be reproduced efficiently in other contexts.

This dissertation is among the very first examples of bringing knowledge from both Artificial Intelligence and Social Media Analysis to the field of heritage studies, in pursuit of socially inclusive heritage management processes. Following is a summary of the results and takeaways obtained from each previous chapter of this dissertation.

+ **PART A** built up the theoretical and methodological basis of the entire dissertation.

- **Chapter 1** set up the scope of this dissertation as an interdisciplinary exploration combining heritage studies, urban studies, computer science, spatial analysis, and social sciences. It formally defined the concepts of baseline (everyday) scenarios and activated (event-triggered) scenarios for the discussion about cultural heritage properties on social media, forming the theoretical foundation of this research, which linked back to the principles and steps for implementing the Recommendation on the Historic Urban Landscape (UNESCO, 2011; Pereira Roders, 2019). Research objectives and questions were formulated, and an overview of the datasets and case studies employed in the dissertation was given.
- The systematic literature review in **Chapter 2** investigated the understanding of user-generated content on social media platforms in the broad field of heritage management. 431 research articles, conference papers, and book chapters were initially collected and eventually reduced to 73 studies for qualitative synthesis. A systematic coding scheme was developed and applied to the studies, under the themes of research context, research content, and research methodology, which were later visualized in a 2D space using the Multi-Dimensional Scaling algorithm to show the associations among the aspects. In addition to the conventional approaches of social network analysis focusing on the "structure" and "content" (Aggarwal, 2011), a third category of "context" is applied to classify the studies. It was found that the questions of interest were being studied with a complex and interdisciplinary approach. Several methods, models, and datasets that were frequently used in the literature were summarized, which inspired and were also applied throughout the dissertation. It further showed that with all the methodological challenges, the development of heritage-specific computational tools to deal with large-scale data was urgently needed, proving the necessity of this research.

+ **PART B** prepared for further analysis of social media data by first training a machine replica of the authoritative view.

– The official Statement of Outstanding Universal Value (OUV) of UNESCO World Heritage List was collected, processed, and structured as a multi-class single-label classification dataset "WHOSe Heritage" in **Chapter 3**. The co-justification pattern of OUV selection criteria was mathematically translated as a hierarchical label structure for each sentence in the dataset, as a response to the association among the OUV selection criteria. State-of-the-art natural language processing models were trained on the dataset, where label smoothing was adapted in the training process to combine the actual label and 'parental labels' of all data points. The best-performing models BERT and ULMFiT both reached a top-3 accuracy of around 94%, both of which became the main outcomes of this chapter and were further used in all following chapters. Albeit not perfect, the performance of the models was approved by experts as sufficiently reliable. As a by-product, a heritage lexicon was obtained, capturing the essential concepts of OUV. The chapter also showed that the OUV selection criteria were consistently associated with each other in different similarity metrics and that some of the association patterns discovered by Jokilehto (2008) needed a revisit and amendment with the recent inscriptions. The machine replica obtained in this Chapter could be used to verify and improve the consistency and coherence of future inscription documents for UNESCO World Heritage.

+ **PART C** zoomed in to the baseline scenarios when people calmly share their thoughts and experiences about the cities they visit or live in. It presented a methodological framework for mapping the collective opinions of cultural significance the cities conveyed to people in everyday contexts.

– The image-sharing social media platform Flickr was used to collect a multi-modal graph-based dataset in three cities, Amsterdam, Suzhou, and Venice, concerning their cultural significance in **Chapter 4**. The unstructured raw images and texts were processed through pre-trained deep-learning models to generate structured vectors as multi-modal representations. Pseudo-labels concerning the heritage attributes and OUV selection criteria categories were also generated using pre-trained models including the ones from Chapter 3, based on the confidence and agreement of predictions by different models. The spatial proximity, temporal sequence, and social similarity were modelled as contextual graphs of the multi-modal data points. The final outcomes of this chapter were four (two for Venice and one for the other two cities) partially labelled attributed-graph datasets, the Heri-Graphs. Qualitative inspections showed that the datasets were comparable and consistent and that the pseudo-labels captured the main elements of cultural significance albeit generated under a transfer-learning setting. The entire procedure was described with mathematical details. And the potentials of Heri-Graph datasets in both machine learning and heritage studies were thoroughly discussed. An additional test case in the Rome Testaccio area demonstrated that the methodological framework was also applicable to smaller urban areas. Specifically, this chapter revisited and reused the manually annotated data of Ginzarly et al. (2019) and updated the mapping procedure with the aid of artificial intelligence.

– The two Heri-Graph datasets in Venice (one small and one large) were taken as the inputs for graph-based semi-supervised classification tasks in **Chapter 5**. An ensemble of several Graph Neural Network models was co-trained on the partially labelled datasets with semi-supervised learning, the predictions of which were then aggregated as the soft labels for all post-level data points. Going one step beyond the direct mapping of post-level labels on maps such as in Liu and De Sabbata (2021), they were diffused onto the nodes of a spatial network where the posts were geographically located. The initial spatial labels were further diffused on the spatial network so that the eventual label of a spatial node combined the characteristics of both its nearby posts and its spatial neighbours. The resulting label distributions both reflected the user-generated information about a place and satisfied the assumption of the First Law of Geography. Maps showing the distribution of each sub-category of cultural significance in Venice were eventually created based on the auto-correlation patterns of various spatial labels. More places other than the most popular tourist destinations in Venice (i.e., San Marco Square and Rialto Bridge) emerged from the maps, displaying different focal places of cultural significance in the city perceived and expressed by Flickr users. Again, mathematical details were provided for the full process. The maps generated could be considered the main outcomes of the methodological framework in PART C. They indicated that the proposed framework was an effective knowledge documentation tool in the baseline scenarios. As a by-product, this chapter also provided numerical benchmarks for semi-supervised classification tasks on Heri-Graph datasets based on both conventional machine learning metrics and new metrics that are adapted and proposed in this chapter.

+ **PART D** switched to the activated scenarios when radical events triggered reactions of people concerning heritage properties at risk. It presented a methodological framework for describing the spatiotemporal patterns of the intensity and semantics of online discussions during Heritage-related Events.

– The instant social networking platform Twitter was used to collect a text dataset enriched with spatial and temporal features in **Chapter 6** about two radical events related to heritage: the fire in Notre-Dame de Paris in April 2019 and the flood in Venice in November 2019. The numbers of tweets were counted spatially for each city and temporally for every hour, and further aggregated in three localities (from the same city where the event happened, from the same country, or far beyond) and three periods (before, during, and after the events). Exploratory analysis of the tweet counts showed that the intensity of tweeting behaviour significantly increased during heritage-related events, transcending the geographical boundaries. Conversation and interaction sequences of Twitter users were constructed as graphs, where the nodes with the highest betweenness centrality values were shown to have provided informative messages for heritage management. Pre-trained deep learning models including the ones from Chapter 3 and pre-defined topic modelling algorithms were used to obtain the semantics of tweets in terms of the type of cultural significance they related to, the emotions they expressed, and the key topics they discussed. The distributions of the semantic categories also demonstrated clear spatiotemporal divergence according to statistical inferences, providing more information on the contexts where discussions happened. As the main outcome of PART D, the timelines of semantic topics revealed the dynamics of the dominant emotions being expressed and the key actions being mentioned and proposed by the online public. Albeit forming temporary online communities that are concerned with heritage properties, the peaks of discussions triggered and activated by the events only remained for 3-4 days, and almost fell back to the level of baseline afterwards. This chapter presented a complete workflow for describing and analysing the spatiotemporal patterns of heritage-related events. The proposed framework proved to be an effective knowledge documentation tool in the activated scenarios since it provided opportunities for end users (heritage managers, scholars, and decision-makers) to both confirm known knowledge and discover new knowledge. This chapter also showed that people distributed globally would form a temporary "heritage community" when radical events happened, calling for a possible extension of the definition by the Faro Convention (Council of Europe, 2005).

7.2 Revisiting Research Questions

In the Introduction of this dissertation (Chapter 1), four research questions have been raised to approach the aim of this research. They will be respectively revisited in the following section with the knowledge obtained from conducting this research, as summarized in Section 7.1.

The first sub-question is: **"How can mathematical and/or computational modelling help to construct a machine replica of the authoritative view of the cultural significance of UNESCO World Heritage properties as the basis for analyzing User-Generated Content?"**

The "WHOSe Heritage" developed in Chapter 3, both as a dataset and as a group of trained natural language processing model checkpoints, could be regarded as a machine replica of the authoritative view of cultural significance. The computational models have been "taught" with the Statements of Outstanding Universal Value (OUV) justifying the cultural significance of World Heritage properties, written and approved by thousands of heritage experts from UNESCO, ICOMOS, and IUCN. The models are capable of taking a generic sentence and outputting the probability distribution of how the 10 OUV selection criteria might relate to this sentence. The models also show a sufficient ability to differentiate between positive and negative classes, consistent with the evaluation of experts. That is, the models could interpret every sentence they "see" with the knowledge they have "learned" inductively from the authoritative documents, replicating the "justification" process of the inscription of UNESCO World Heritage, mainly based on semantic similarities of sentences. Mathematical modelling is used in the process to reflect the associative nature of the OUV selection criteria and is shown to improve the prediction accuracy of the computational models for the OUV selection criteria categories.

Similarly, the several machine learning models trained in Chapter 4 on the images previously annotated by Ginzarly et al. (2019) could be understood as another such machine replica, taught with image inputs and trying to infer the depicted heritage attribute therein. Again, the mathematical modelling in Chapter 4 and 5 focusing on the spatial, temporal, and social similarities and associations of social media posts as prior knowledge helped realize the graph neural network models.

To the best of the author's knowledge, the datasets and models from Chapters 3 to 5 are the very first open-source computational tools that specifically focused on the classification of texts and images into categories related to UNESCO World Heritage, OUV selection criteria, and heritage attributes. They provided a unique opportunity for facilitating the conventional approach of heritage justification and investigation with AI that can process and analyze massive data efficiently, which is indispensable for including social media users in heritage management at scale. Naturally, the models as "machine replicas" are not truly making solid justifications for the cultural significance of a potential heritage property through site visits, value assessments, historic investigations, and comparative studies, like what the experts usually do (Jokilehto, 2008; UNESCO, 2008; Veldpaus, 2015). Unlike humans, the computational models in this dissertation make assumptions with knowledge learned from massive empirical data. On the other hand, unlike individual experts writing the Statements of OUV alone who can be focused too much on the specific case study and trapped with partial and incomplete insights, the computational models approached more from a holistic viewpoint with collective knowledge. Moreover, the assumed labels by the models are sometimes simply a depiction of the reality mirrored on social media, weakly relevant to the cultural significance of the place and can hardly be justified as a valuable heritage to be preserved for future generations. Whether or not the cultural significance can be conceptually reduced to numbers - the probability distribution under a fixed category system - is, however, a question out of the scope of this dissertation yet worth debating on. Still, from a pragmatic point of view, the machine replicas suffice the initial need to scale up the analytical process that can otherwise only be conducted manually, if not impossible. They become an effective starting point and basis for analyzing User-Generated Content from the perspective of cultural significance.

The second sub-question is: **"As for a baseline scenario, how can a computational method help to map the spatiotemporal and social contexts of the public opinions about the cultural significance in a normal everyday setting?"**

The "Heri-Graphs" datasets and models developed in Chapter 4 and 5 showed the effects of how a systematic workflow mainly composed of computational methods can help to map the contexts of cultural significance perceived and expressed by people in baseline scenarios. The effects are three-fold.

– Firstly, transfer learning of pre-trained machine and deep learning models are effective in transforming the massive unstructured texts and images that are only understandable by humans qualitatively into high-dimensional vectors (both as semantic representations and as probabilities concerning cultural significance categories) that could be analysed mathematically. This is a desirable characteristic both for scaling up the analyses and for reproducibility. The massive User-Generated Content produced by various individuals in different forms is all brought to the same abstract mathematical space, thus the vagueness of human interpretation that often hinders reproducibility can be omitted.

- Secondly, the construction of graphs marking the spatial, temporal, and social connections of social media posts gives the term “context” an operational meaning (Zimmermann et al., 2007). These connections are no longer only a background concept, but also meaningful matrices that messages could be propagated through and projected on.
- Thirdly, the diffusion process in Chapter 5 aggregating the cultural significance category labels (or any other type of categories) of posts onto spatial networks completes the mapping practice with visualized “maps”. User-Generated Content, therefore, is no longer just a “bag” of unrelated images and sentences posted here and there but also summarized information reflecting the collective opinions over time interlinking with each other.

Since baseline scenarios cover all the other time except for when a radical event happens, as defined in Section 1.1, the data are naturally distributed sparsely within longer time spans. As a pragmatic choice, the mapping for the baseline scenarios was restricted to urban or sub-urban scales. This also allows for fine-grained explorations over the distribution of cultural significance in space. In this dissertation, cultural significance categories are eventually mapped on street intersections. Maps of perceived cultural significance at this resolution can in turn only be achievable by aggregating the posts of every individual throughout the long time span. In other words, if the time period taken into consideration was too short, there might not be sufficient data points available for summarizing a collective knowledge generated by individuals of all interests. Moreover, even though one main purpose of having a systematic workflow is to ensure that the same outcome (maps here) can be repeatedly obtained by different users, it does not mean that the mapping with computational methods produces only one “true” answer. Instead, the computational methods are almost always accompanied by adjustable parameters, allowing for flexibility. This shows another benefit of computational models in response to the sub-question: the variations of outcome based on different design choices of users can also be reproducible. Moreover, it was sometimes asserted in the field of heritage studies that the “authorized heritage discourse” would automatically and unavoidably differ from the community view. Yet another interesting finding from the dissertation is that the two are not necessarily contradictory to each other. Even though possibly distributed in different spots in the city, the official elements of cultural significance - heritage values and attributes are all present in the maps drawn with social media user-generated data.

The third sub-question is: **“As for an activated scenario, how can the dynamics and mechanism of the emotion/information spreading on social media platforms be described when some radical events happen about a heritage property?”**

Unlike the baseline scenarios that can expand to years, the peak of the activated discussion usually only lasts a few days before it goes back to baselines. The main characteristics of event-triggered activated scenarios, i.e., short-lasting in time, far-reaching in space, and strong-affecting in society, indicate that it is both pragmatic and necessary to study them at a global scale. Therefore, developing a separate variant of the methodological framework during heritage-related events is also a necessity. The exploratory analyses on the “HREs” datasets developed in Chapter 6 provided some examples of how the dynamics of public reactions can be described during the time of radical events. Consistent with the categorization of Aggarwal (2011) and Bai et al. (2021), the descriptions can be made on the context, structure, and content of the social network.

- From the view of the context, social media posts are embedded in spatiotemporal bounding boxes, even though the geo-tags for them are not always explicitly available. The counts of posts aggregated in any arbitrary spatial unit (in this dissertation, a city) and temporal unit (in this dissertation, an hour) describe the dynamics of the discussion intensity on social media around the period of a radical heritage-related event.
- From the view of the structure, the actions and reactions of social media users always form a communication network, on which emotions and other information are being spread. Network sciences and graph theory have proved with abundant examples from various fields that the configuration of nodes and links on these networks influences the mechanism of how information spreading happens (Barabási et al., 2016). An intuitive way that was also employed in this dissertation of describing the discussions is to compute the centrality indicators of social media posts as nodes and inspect the characteristics of those with the highest centralities, which was also shown effective in the two events investigated in this dissertation.
- From the view of the content, transfer learning of pre-trained deep learning models and topic modelling can be used to describe explicit and implicit semantic meanings of social media posts. Depending on the application scenario, such semantic information can include but is not limited to the relevant cultural significance categories, the expressed emotions, the descriptions and assumptions on the event itself, the proposed immediate and future actions, as well as political discussions.

Combining context and content, the temporal development and spatial divergence of different semantic topics can also be revealed. By describing triggered discussions on social media platforms with their contextual, structural, and content-related semantic features, the collective knowledge of the heritage properties under events is also documented in a systematic and reproducible manner.

The last sub-question is: **“How can the evidence-based research findings improve the power and degree of social inclusion in future heritage management in broader cases?”**

This question was not answered explicitly with a separate chapter in this dissertation but has been touched upon in the “Discussion” sections of all content chapters. Four levels of usage of the research findings and evidence obtained in this research can be applied in future heritage management research and practices. It is worth emphasizing that there is no implication that “higher-level” usages are more advanced and superior than the “lower-level” ones. The levels here only relate to how much overlapping there can be between the use case and this dissertation.

- The most direct and obvious usage is by exhibiting and interpreting the written texts, visualized maps, timelines, and/or other statistics and graphics. Various examples in the different case study cities can be presented to different stakeholders to check if the results are aligned with or against their expert knowledge and intuitions.
- The second level of usage of the findings is by exploring the collected raw data and processed datasets from this dissertation, which are all shared with open access. Other data mining techniques and directions not covered in this research can be discovered with specific objectives. Since the datasets are collected in this research, the scope would still be restricted to the case studies discussed here.
- The third level of usage would be repeating a part of or the entire methodological workflow in another city containing World Heritage or another heritage-related event as a case study. This will be a validation step on the generalizability and effectiveness of the workflow proposed in this research. As such, datasets and results under the same format allow for comparison and generalization.
- The last level of usage would be adapting and customizing the framework proposed in this study for other purposes in urban studies and heritage management.

All steps can be implemented by the local heritage managers, the global heritage organizations, the local residents, the global tourists, and anyone from the online communities who finds it relevant. The information on how online communities' perceptions differ from the official document can be informative for policymakers, thus offering a chance for participation and enhancing the degree of democratization. And the message of how policymakers' reactions reflect the public's opinions can be encouraging for society. Knowing that they are among the others who post similar topics on social media, which becomes part of the collective knowledge can hypothetically increase the sense of belonging; while realizing that the collective knowledge can actually make a change in the decision-making process in heritage management and urban planning actions can hypothetically increase the sense of authenticity. As the senses of belonging and authenticity are both strengthened, a higher level of social inclusion could be reached by definition (Jansen et al., 2014).

Moreover, social media could also enable observations of how the current ongoing planning actions could gradually alter first the physical spaces and then the "digital twin" of the study area by collecting freshly added posts by residents and tourists along the timeline and merging them into the same dataset. This new collection of data will be processed and analysed with newer generations of more powerful AI algorithms in the future. Then a new round of research could be conducted, coupled with integrated analyses of mixed methods, possibly again pointing to new planning directions. This would suggest an abductive and iterative system with a data-driven feedback mechanism for decision-making, integrating a diversity of data sources in an inclusive and participatory planning process (Dubois and Gadde, 2002).

After answering the sub-questions with the outcome of this dissertation, the aim of this research, i.e., **to explore the use of AI in a methodological framework to include the contribution of a larger and more diverse group of participants and facilitate the knowledge documentation of cultural significance in cities with user-generated social media data** is generally met.

With the aid of various AI-based algorithms and models, an interdisciplinary heritage-centred methodological framework for knowledge documentation of cultural significance using user-generated social media data has been proposed and tested for cities with urban areas inscribed in the UNESCO WHL. Two variations of the methodological framework targeted respectively the baseline and the activated scenarios. Both variations were tackled with a complete workflow fully described with mathematical details, starting from data collection, to feature engineering and representation, statistical analysis, outcome aggregation, and eventually to visualizations. Specifically, spatiotemporal and social contexts were always treated as an additional layer of semantic information in the framework. Case studies such as Venice (both in baseline and activated scenarios), Amsterdam, Rome, Suzhou (only in the baseline scenario), and Paris (only in the activated scenario) have been used to illustrate the effectiveness, usability, validity, and generalizability of the proposed methodological framework. The employed artificial intelligence models covered the settings of supervised, semi-supervised, and unsupervised learning; with the tasks of natural language processing, image recognition, and multi-modal machine learning; and with the architectures of machine learning algorithms, convolutional neural networks, attention-based recurrent neural networks, and graph neural networks. The obtained results suggested that User-Generated Content on social media platforms, with the aid of the methodological framework proposed in this research, has the ability and potential to function as the resourceful starting point for developing knowledge documentation tools to be applied globally at a large scale.

7.3 Reflection on the Research

7.3.1 Scientific Contribution

This dissertation is a cross-/inter-disciplinary study that applies cutting-edge Artificial Intelligence algorithms and models in the investigation of cultural significance for urban heritage. Many of the methods discovered and mentioned in Chapter 2 originally developed and used in various previously partially related disciplines were integrated into the methodological framework proposed in this dissertation. By combining methods from computer science, social sciences, heritage studies, and spatial analysis, this study offers possibilities to augment a research field that has been previously dominated by expert-based qualitative inspections with large-scale evidence that can be reproduced efficiently in other contexts. Such evidence could support scientific research on topics including digital humanities, people-centred heritage, heritage communities, participatory planning, collective memory, urban images, volunteered geographic information, and so on.

All the steps within the methodological framework proposed in this dissertation have been provided with adequate mathematical descriptions, allowing for generalized discussions and customization in different use cases. The collected datasets "WHOSe Heritage"[1], "Heri-Graphs"[2], and "HREs" together with the trained AI model checkpoints with these datasets have been or will be made open source, providing testing grounds and initial numerical benchmarks for the machine learning community as real-world datasets in tasks including text classification, multi-modal graph-node classification, semi-supervised learning, spatiotemporal clustering, and even federated learning on graphs (Zhang et al., 2023). The datasets developed in this dissertation have labels tailor-made about cultural significance, i.e., UNESCO World Heritage Outstanding Universal Value selection criteria and Heritage Attributes specifically in urban settings. To the best of the author's knowledge, they are among the very first open-access datasets that serve these purposes.

Additionally, the mathematical descriptions and workflows are also applicable beyond heritage studies. The essence of the methodological framework proposed in this study is to transform the unstructured multi-modal social media user-generated data into high-dimensional vectors embedded in a graph structure representing the spatiotemporal and social contexts of the posts. Since the components of generating heritage-related labels and maps are mostly modular, they can also be substituted with other human-generated information on spatial networks, applicable in other domains, such as in urban studies, social sciences, and policy analyses, in order to measure the safety, vitality, and popularity of urban spaces.

7.3.2 Societal Contribution

As part of a pan-European research and training network HERILAND[3] (Cultural Heritage and the Planning of European Landscapes), this dissertation, together with 14 other research projects, reflected on the societal challenges of heritage and landscape in the ever-changing 21st century (Burgers, 2021). The main challenges brought up by HERILAND, i.e., the Spatial Turn, Democratisation, Digital Transformations, Shifting Demographies and Contested Identities, and Changing Environments, are generally applicable both within the European context and in the globe. Being part of the Work Package of "Democratisation", this dissertation contributes as a tool to include and empower the citizens and get their voice heard during heritage management. This transition is promoted by inter-governmental institutions and doctrinal documents such as the Faro Convention (Council of Europe, 2005), the Recommendation on the Historic Urban Landscape (UNESCO, 2011), and the Sustainable Development Goals 11.3 "Inclusive and Sustainable Urbanization" and 11.4 "Protect the World Cultural and Natural Heritage" (Sachs et al., 2019; Vinuesa et al., 2020), and by non-governmental organizations and campaigns such

[1] https://github.com/zzbn12345/WHOSe_Heritage
[2] https://github.com/zzbn12345/Heri_Graphs
[3] https://www.heriland.eu, accessed 26 May 2023.

as Our World Heritage[4].

The methodological framework aided with Artificial Intelligence in this dissertation is shown to be able to collect information and map the knowledge of the community about the cultural significance of heritage. It fulfills the expectation and requirement of Historic Urban Landscape as knowledge documentation and civic engagement tools, useful and informative for future socially inclusive heritage management processes. It has the potential to be developed into a transparent heritage management and planning toolbox, systematically summarizing information from the online public. The approach can also eventually be checked by the citizens on how their voices are being heard and implemented in the decision-making process. Albeit not yet implemented in real-world design and planning practices, the methods and results have also been circulated and discussed with global and local heritage managers on formal and informal occasions, triggering much interest and attention[5]. Together with other studies in the same line of research with similar aims, the new insights and experiences obtained in different European and global cities can promote democratic and inclusive participation practices, especially when planning and cultural identity meet.

7.3.3 Restating the Major Limitations

Throughout the dissertation, only a small fraction of AI models are tested as a humble exploration. Much more possibilities and potentials still exist for applying AI in heritage management (Condorelli et al., 2020; Matrone et al., 2020; Yuan et al., 2022; Foroughi et al., 2023). In this dissertation, the models trained on datasets “WHOSe Heritage” and “Heri-Graphs” always suffered from a shortage of data in certain categories, e.g., the OUV selection criterion (v), and the Heritage Attribute category of “Artifact Product”, which imply that the models and datasets need to be further augmented and improved. However, even with a perfectly trained error-free model at a later stage after solving all the technological challenges and technical difficulties, careful inspections of the validity, reliability, and coherence of the models and interpretations of the derived results by humans with their expert knowledge are always needed, especially during policy decision-making on World Heritage for the social benefit. The AI-based models are on the one hand inherently less biased than a single human expert since the chances that this person adds implicit personal experiences on top of what is really there are sufficiently reduced. They are on the other hand still always biased based on the available training data reflecting possible unfairness and training methods restricted with all sorts of known or hidden assumptions, which can sometimes fall into sub-optimal solutions and even lead in wrong directions (Ntoutsi et al., 2020; Ferrer et al., 2021). The use of AI and social media data is never the “eternal solution” for mapping cultural significance, which

[4] https://www.ourworldheritage.org, accessed 26 May 2023

[5] An example of such conversations was made with Inez Weyermans from the heritage department of Amsterdam municipality in the Digital Citizen Engagement with Heritage | Future Making in the Anthropocene Podcast, accessed 26 May 2023.

could potentially create more new challenges and problems than it manages to solve. One of the biases that need special caution from the end of users is the so-called "automation bias", showing that people favour the results generated by automatic systems for decision-making processes (Parasuraman and Manzey, 2010). The privacy issue and data security are always worth noticing in this type of study.

Furthermore, the use of one or two specific social media platforms as the data source may have strong limitations to getting a comprehensive picture since there is always an unequal representation of users and non-users. As research demonstrates, despite the proliferation of digital technologies, a significant number of the population may still be disadvantaged in using digital platforms and tools, due to a lack of access to the internet, equipment, and difficulty with digital skills (Craglia et al., 2021). Some of these inequalities are related to age ranges, socioeconomic backgrounds, or spatial divides. Thus the outcomes of social media surveys may be considered unavoidably biased towards the users of digital platforms, implying a generational, socioeconomic, and/or spatiotemporal gap in its representation. These factors call for careful consideration at the early stage of applications and emphasize the need for integrated research and mixed analysis methods combining qualitative and quantitative knowledge. Social media may be helpful for setting the stage for planning and management through an initial data capture, but its limitations should also be balanced with other methods of data collection and analyses, as well as cross-sectoral integration of different data sources, such as official documents (Tarrafa Silva and Pereira Roders, 2010; Rosetti et al., 2022; Lin et al., 2023), archival maps (Potdar and Verbakel, 2022), design and planning practices (Fredholm et al., 2021; Castro de Azevedo, 2023), interviews (Li et al., 2021; Tarrafa Silva et al., 2023), surveys (Gonçalves et al., 2021; Ducci et al., 2023), behavioural data (Bai et al., 2023b), and participatory workshops (Pintossi et al., 2023; Zheng, 2023).

7.4 Recommendations for Future Research

This dissertation can be extended and continued in different directions.

First of all, the methodological framework can be applied in more case studies to document the collective knowledge of cultural significance in various cities of diverse cultural backgrounds, as already mentioned in Section 7.2. By doing so, the generalizability of the proposed framework could be tested. In case some parts of the methodological framework do not work, inductive error analyses and/or deductive reasoning could provide insights on why the system fails, and then the mathematical formulations and the AI-based computational models could be revised and updated accordingly. For example, more examples could be collected, annotated, or even generated as training data augmentation, and the categories of cultural significance, especially heritage attributes, could be upgraded so that they can be both universal and case-specific. After the validation, datasets could be ideally collected, processed, analysed, and visualized in all cities with urban areas inscribed in the UNESCO WHL and/or tentative properties. In this way, the scope goes beyond any specific case study and aims at a general rule or even a universal law about cultural significance perceived and expressed by people on social media. Heri-graphs of each participating city can be constructed, effectively transforming the central task from node classification to graph classification, requiring higher levels of abstraction and aggregation (Ma and Tang, 2021; Bai et al., 2022).

Second, as has been argued extensively in the official doctrines and scientific literature (ICOMOS, 2013; UNESCO, 2011; Tarrafa Silva and Pereira Roders, 2012; Veldpaus, 2015; Foroughi et al., 2022; Lin et al., 2023), heritage values (why) and heritage attributes (what) are two critical components of the high-level concept "cultural significance", which is represented as "Outstanding Universal Value (OUV)" in case of World Heritage. This dissertation covers one side of the story linking heritage attributes to OUV, transforming categories from both concepts into computational classification models. Future studies are encouraged to complete the other side of the story by first building up machine replicas on the classification of heritage value categories. This is a harder task as no structured annotated dataset is currently available. The challenge could be possibly tackled by combining computational workflows with an expert or crowd-sourcing evaluation (active learning), by exploring more advanced weakly supervised learning algorithms, and by integrating the prior expert knowledge into classification models (Settles, 2011; Shen et al., 2021).

Third, two separate variants of the methodological framework have been developed in this dissertation for the baseline and the activated scenarios. Both scenarios are still not explicitly compared together. In principle, there should always be significant "activated" spatiotemporal clusters in the long-term "baseline" datasets (Shen, 2018; Lai, 2019), since the periods of events were not deliberately removed from data collection. It could be an interesting extension to apply the variant of the methodological framework developed for the activated scenarios on the datasets collected for the baseline scenarios, and vice versa. Then the questions to be answered could be: how are the foci of the discussion on social media developed through the years and how can the emergent heritage values and attributes shown in social media posts during a radical event be mapped? As such, more systematic knowledge about the mechanisms and dynamics of cultural significance expressed online could be obtained.

Fourth, ways of integrating the proposed methodological framework in real-world urban planning and heritage management could be explored. It could start with organizing workshops among interested mayors and officers from the World Heritage cities with the help of UNESCO and other international or national networks (Rosetti et al., 2022). It is foreseen that digital literacy (the ability to understand what AI does and can do) and digital numeracy (the ability to write codes and realize AI algorithms) would be both necessary for the heritage managers and scholars in an

AI-intense near future[6]. Workshops could be a helpful introduction to such digital transformations, later to be followed by a curriculum change in the education of heritage and urban studies.

Finally, the effect of the methodological framework for social inclusion in heritage management needs to be verified through in-depth qualitative interviews and large-scale quantitative surveys with the stakeholders, e.g., the laypersons who actively use social media to express their opinions, feelings, and perceptions on the cultural heritage in their own city and other cities with urban areas inscribed in the UNESCO WHL. The degree of their perceived social inclusion and their willingness to further engage in heritage management through social media could be measured (Jansen et al., 2014; Taylor and Gibson, 2017). Only after such a process, one can confidently argue that the research has made a difference and the level of social inclusion has been increased.

All being said, this dissertation is only a modest starting point to explore many more possibilities. Hopefully, it can be a bridge among all the involved disciplines and get them to embrace each other eventually in the new future, all for the same aim of a smooth transformation towards inclusive heritage management and sustainability.

References

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In Aggarwal, C. C., editor, Social Network Data Analytics, chapter 1, pages 1–15. SPRINGER.

Bai, N., Ducci, M., Mirzikashvili, R., Nourian, P., and Pereira Roders, A. (2023a). Mapping urban heritage images with social media data and artificial intelligence, a case study in testaccio, rome. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023:139–146.

Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2022). Heri-graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. ISPRS International Journal of Geo-Information, 11(9).

Bai, N., Nourian, P., and Pereira Roders, A. (2021). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.

Bai, N., Nourian, P., Pereira Roders, A., Bunschoten, R., Huang, W., and Wang, L. (2023b). Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data. Environment and Planning B: Urban Analytics and City Science, 50(1):94–116.

Barabási, A.-L. et al. (2016). Network Science. Cambridge University Press.

Burgers, G.-J. (2021). Tourism, leisure and cultural heritage: The challenge of participatory planning and design. In Tourism and Regional Science: New Roads, pages 71–85. Springer.

Castro de Azevedo, A. (2023). Barriers to public participation in memorialization processes: Evidence from the dutch holocaust memorial of names. In Rodenberg, J., Wagenaar, P., and Burgers, G.-J., editors, Calling on the Community, Explorations in Heritage Studies, pages 162–188. Berghahn Books, United Kingdom.

Condorelli, F., Rinaudo, F., Salvadore, F., and Tagliaventi, S. (2020). A neural networks approach to detecting lost heritage in historical video. ISPRS International Journal of Geo-Information, 9(5):297.

[6] https://www.heriland.eu/2023/03/14/webinar-by-dr-sennay-ghebreab-ai-the-four-worlds-of-2050/, accessed 30 May 2023.

Council of Europe (2005). Convention on the value of cultural heritage for society (faro convention). Technical report, Council of Europe, Faro.
Craglia, M., Micheli, M., Hradec, J., Calzada, I., Luitjens, S., Ponti, M., Scholten, H. J., and Boter, J. (2021). Digitranscope: The governance of digitally-transformed society. Craglia, M., Scholten, HJ, Micheli, M., Hradec, J., Calzada, I., Luitjens, S., Ponti, M. and Boter, J., Digitranscope: The governance of digitally-transformed society, EUR, 30590.
Dubois, A. and Gadde, L.-E. (2002). Systematic combining: an abductive approach to case research. Journal of business research, 55(7):553–560.
Ducci, M., Janssen, R., Burgers, G.-J., and Rotondo, F. (2023). Mapping local perceptions for the planning of cultural landscapes. International Journal of E-Planning Research (IJEPR), 12(1):1–27.
Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., and Criado, N. (2021). Bias and discrimination in ai: a cross-disciplinary perspective. IEEE Technology and Society Magazine, 40(2):72–80.
Foroughi, M., de Andrade, B., and Pereira Roders, A. (2022). Peoples' values and feelings matter: Participatory heritage management using social media. In Muntañola, J., editor, Artificial Intelligence and Architectural Design, volume 33, pages 107–120. Oficina de Publicacions Acadèmiques Digitals de la UPC.
Foroughi, M., de Andrade, B., Roders, A. P., and Wang, T. (2023). Public participation and consensus-building in urban planning from the lens of heritage planning: A systematic literature review. Cities, 135:104235.
Fredholm, S., Dore, M., and Brorström, S. (2021). Strategic responses to wicked problems of heritage management: Experiences from the west link infrastructure project in gothenburg, sweden. Land, 10(10):1032.
Ginzarly, M., Pereira Roders, A., and Teller, J. (2019). Mapping historic urban landscape values through social media. Journal of Cultural Heritage, 36:1–11.
Gonçalves, J., Mateus, R., Silvestre, J. D., Roders, A. P., and Bragança, L. (2021). Attitudes matter: Measuring the intention-behaviour gap in built heritage conservation. Sustainable Cities and Society, 70:102913.
ICOMOS, A. (2013). The Burra Charter: The Australia ICOMOS charter for places of cultural significance 2013. Australia ICOMOS Incorporated.
Jansen, W. S., Otten, S., van der Zee, K. I., and Jans, L. (2014). Inclusion: Conceptualization and measurement. European journal of social psychology, 44(4):370–385.
Jokilehto, J. (2008). What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties. Technical report, ICOMOS, ICOMOS Berlin.
Lai, J. (2019). Urban Place Profiling Using Geo-Referenced Social Media Data. PhD thesis, UCL (University College London).
Li, J., Krishnamurthy, S., Roders, A. P., and van Wesemael, P. (2021). Imagine the old town of lijiang: Contextualising community participation for urban heritage management in china. Habitat International, 108:102321.
Lin, M., Pereira Roders, A., Nevzgodin, I., and de Jonge, W. (2023). Values and interventions: dynamic relationships in international doctrines. Journal of Cultural Heritage Management and Sustainable Development, ahead-of-print(ahead-of-print). Publisher: Emerald Publishing Limited.
Liu, P. and De Sabbata, S. (2021). A graph-based semi-supervised approach to classification learning in digital geographies. Computers, Environment and Urban Systems, 86:101583.
Ma, Y. and Tang, J. (2021). Deep learning on graphs. Cambridge University Press.
Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., and Remondino, F. (2020). Comparing machine and deep learning methods for large 3d heritage semantic segmentation. ISPRS International Journal of Geo-Information, 9(9):535.
Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., et al. (2020). Bias in data-driven artificial intelligence systems—an introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3):e1356.
Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359.
Parasuraman, R. and Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3):381–410.
Pereira Roders, A. (2019). The Historic Urban Landscape Approach in Action: Eight Years Later. In Reshaping Urban Conservation, pages 21–54. Springer.
Pintossi, N., Kaya, D. I., van Wesemael, P., and Roders, A. P. (2023). Challenges of cultural heritage adaptive reuse: A stakeholders-based comparative study in three european cities. Habitat International, 136:102807.
Potdar, K. and Verbakel, E. (2022). Eidetic mapping: An exploration for sustainability and resilience of historic urban landscapes. In The Twelfth International Convention of Asia Scholars (ICAS 12), volume 1, pages 547–559. Amsterdam University Press.
Rosetti, I., Bertrand Cabral, C., Pereira Roders, A., Jacobs, M., and Albuquerque, R. (2022). Heritage and sustainability: Regulating participation. Sustainability, 14(3):1674.
Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., and Rockström, J. (2019). Six transformations to achieve the sustainable development goals. Nature sustainability, 2(9):805–814.
Settles, B. (2011). From theories to queries: Active learning in practice. In Active learning and experimental design workshop in conjunction with AISTATS 2010, pages 1–18. JMLR Workshop and Conference Proceedings.

Shen, J. (2018). Profiling and Grouping Space-time Activity Patterns of Urban Individuals. PhD thesis, UCL (University College London).

Shen, J., Qiu, W., Meng, Y., Shang, J., Ren, X., and Han, J. (2021). Taxoclass: Hierarchical multi-label text classification using only class names. In NAACL-HLT 2021.

Tarrafa Silva, A. and Pereira Roders, A. (2010). The cultural significance of World Heritage cities : Portugal as case study. In Heritage and Sustainable Development, pages 255–263, Évora, Portugal.

Tarrafa Silva, A. and Pereira Roders, A. (2012). Cultural Heritage Management and Heritage (Impact) Assessments. In Joint CIB W070, W092 & TG72 International Conferenceon Facilities Management, Procurement Systems and Public Private Partnership.

Tarrafa Silva, A., Pereira Roders, A., Cunha Ferreira, T., and Nevzgodin, I. (2023). Critical analysis of policy integration degrees between heritage conservation and spatial planning in amsterdam and ballarat. Land, 12(5):1040.

Taylor, J. and Gibson, L. K. (2017). Digitisation, digital interaction and social media: embedded barriers to democratic heritage. International Journal of Heritage Studies, 23(5):408–420.

UNESCO (1972). Convention Concerning the Protection of the World Cultural and Natural Heritage. Technical Report november, UNESCO, Paris.

UNESCO (2008). Operational guidelines for the implementation of the world heritage convention. Technical Report July, UNESCO World Heritage Centre.

UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.

Veldpaus, L. (2015). Historic urban landscapes: framing the integration of urban and heritage planning in multilevel governance. PhD thesis, Technische Universiteit Eindhoven.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., and Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. Nature communications, 11(1):233.

Yuan, W., Yang, L., Yang, Q., Sheng, Y., and Wang, Z. (2022). Extracting spatio-temporal information from chinese archaeological site text. ISPRS International Journal of Geo-Information, 11(3):175.

Zhang, X., Hong, M., and Chen, J. (2023). Glasu: A communication-efficient algorithm for federated learning with vertically distributed graph data. arXiv preprint arXiv:2303.09531.

Zheng, N. (2023). Coming to grips with diverse voices in participatory heritage initiatives. PhD thesis, Vrije Universiteit Amsterdam.

Zimmermann, A., Lorenz, A., and Oppermann, R. (2007). An operational definition of context. In Modeling and Using Context: 6th International and Interdisciplinary Conference, CONTEXT 2007, Roskilde, Denmark, August 20-24, 2007. Proceedings 6, pages 558–571. Springer.

Bibliography

Aaker, J. L. (1997). Dimensions of brand personality. Journal of marketing research, 34(3):347–356.

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In Osdi, volume 16, pages 265–283. Savannah, GA, USA.

Abdel Tawab (2019). The Assessment of Historic Towns' Outstanding Universal Value Based on the Interchange of Human Values They Exhibit. Heritage, 2(3):1874–1891.

Abeysinghe, S., Manchanayake, I., Samarajeewa, C., Rathnayaka, P., Walpola, M. J., Nawaratne, R., Bandaragoda, T., and Alahakoon, D. (2018). Enhancing decision making capacity in tourism domain using social media analytics. In 2018 18th International Conference on Advances in ICT for Emerging Regions (ICTer), pages 369–375. IEEE.

Acheampong, F. A., Wenyu, C., and Nunoo-Mensah, H. (2020). Text-based emotion detection: Advances, challenges, and opportunities. Engineering Reports, 2(7):e12189.

Adamic, L. A. and Adar, E. (2003). Friends and neighbors on the web. Social networks, 25(3):211–230.

Adamic, L. A., Lento, T. M., Adar, E., and Ng, P. C. (2016). Information evolution in social networks. WSDM 2016 - Proceedings of the 9th ACM International Conference on Web Search and Data Mining, pages 473–482.

Adams, R. P. and MacKay, D. J. (2007). Bayesian online changepoint detection. arXiv preprint arXiv:0710.3742.

Adepeju, M. (2017). Modelling of sparse spatio-temporal point process (STPP): an application in predictive policing. PhD thesis, UCL (University College London).

Afyouni, I., Khan, A., and Aghbari, Z. A. (2022). Deep-eware: spatio-temporal social event detection using a hybrid learning model. Journal of Big Data, 9.

Afzaal, M., Usman, M., Fong, A. C., and Fong, S. (2019). Multiaspect-based opinion classification model for tourist reviews. Expert Systems, 36(2):e12371.

Aggarwal, C. C. (2011). An Introduction to Social Network Data Analytics. In Aggarwal, C. C., editor, Social Network Data Analytics, chapter 1, pages 1–15. SPRINGER.

Al-Sultany, G. A. and Abd Al-Ameer, A. A. (2019). Geotagged photos clustering using adapted density-based spatial clustering of applications with noise algorithm. Journal of Computational and Theoretical Nanoscience, 16(3):1056–1061.

Alaei, A. R., Becken, S., and Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. Journal of travel research, 58(2):175–191.

Albers, P. C. and James, W. R. (1988). Travel photography: A methodological approach. Annals of tourism research, 15(1):134–158.

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3):175–185.

Alviz-Meza, A., Vásquez-Coronado, M. H., Delgado-Caramutti, J. G., and Blanco-Victorio, D. J. (2022). Bibliometric analysis of fourth industrial revolution applied to heritage studies based on web of science and scopus databases from 2016 to 2021. Heritage Science, 10(1):189.

Aly, M. (2005). Survey on multiclass classification methods. Neural Netw, 19:1–9.

Amato, F., Cozzolino, G., Di Martino, S., Mazzeo, A., Moscato, V., Picariello, A., Romano, S., and Sperlí, G. (2016). Opinions analysis in social networks for cultural heritage applications. Smart Innovation, Systems and Technologies, 55:577–586.

Andrade, B. (2022). I can see through the waters eyes. covid-19 in heritage cities: Citizen participation and self-organization for greater conservation and sustainability: The case of venezia pulita (clean venice). In Living (World) Heritage Cities: Opportunities, challenges, and future perspectives of people-centered approaches in dynamic historic urban landscapes, pages 199–212. Sidestone Press.

Ankerst, M., Breunig, M. M., Kriegel, H.-P., and Sander, J. (1999). Optics: Ordering points to identify the clustering structure. ACM Sigmod record, 28(2):49–60.

Anselin, L. (1995). Local indicators of spatial association—lisa. Geographical analysis, 27(2):93–115.

Anselin, L. (2003). An introduction to spatial autocorrelation analysis with geoda. Spatial Analysis Laboratory, University of Illinois, Champagne-Urbana, Illinois.

Arcaute, E., Molinero, C., Hatna, E., Murcio, R., Vargas-Ruiz, C., Masucci, A. P., and Batty, M. (2016). Cities and regions in britain through hierarchical percolation. Royal Society open science, 3(4):150691.

Arjona, J. O. (2020). Analysis of the space-temporal patterns of events from twitter data: The case of madrid 2017 world pride. Estudios Geograficos, 81.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics surveys, 4:40–79.

Assmann, J. and Czaplicka, J. (1995). Collective memory and cultural identity. New german critique, pages 125–133.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007+ ASWC 2007, Busan, Korea, November 11-15, 2007. Proceedings, pages 722–735. Springer.
Australia ICOMOS (2013). The Burra Charter: The Australia ICOMOS Charter for Places of Cultural Significance (1999). Technical report, Australia ICOMOS.
Avila-Robinson, A. and Wakabayashi, N. (2018). Changes in the structures and directions of destination management and marketing research: A bibliometric mapping study, 2005–2016. Journal of Destination Marketing & Management, 10:101–111.
Bahdanau, D., Cho, K., and Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
Bai, N., Cheng, T., Nourian, P., and Pereira Roders, A. (2023a). An exploratory data analysis of the spatiotemporal patterns of heritage-related events on twitter. The 30th International Conference on Geoinformatics (CPGIS 2023), London, UK.
Bai, N., Ducci, M., Mirzikashvili, R., Nourian, P., and Pereira Roders, A. (2023b). Mapping urban heritage images with social media data and artificial intelligence, a case study in testaccio, rome. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023:139–146.
Bai, N., Luo, R., Nourian, P., and Pereira Roders, A. (2021a). WHOSe Heritage: Classification of UNESCO World Heritage statements of "Outstanding Universal Value" with soft labels. In Findings of the Association for Computational Linguistics: EMNLP 2021, pages 366–384, Punta Cana, Dominican Republic. Association for Computational Linguistics.
Bai, N., Nourian, P., Luo, R., Cheng, T., and Roders, A. P. (2023c). Screening the stones of venice: Mapping social perceptions of cultural significance through graph-based semi-supervised classification. ISPRS Journal of Photogrammetry and Remote Sensing, 203:135–164.
Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2021b). "What is OUV" revisited: A computational interpretation on the statements of Outstanding Universal Value. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-M-1-2021:25–32.
Bai, N., Nourian, P., Luo, R., and Pereira Roders, A. (2022). Heri-graphs: A dataset creation framework for multi-modal machine learning on graphs of heritage values and attributes with social media. ISPRS International Journal of Geo-Information, 11(9).
Bai, N., Nourian, P., and Pereira Roders, A. (2021c). Global citizens and world heritage: Social inclusion of online communities in heritage planning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021:23–30.
Bai, N., Nourian, P., Pereira Roders, A., Bunschoten, R., Huang, W., and Wang, L. (2023d). Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data. Environment and Planning B: Urban Analytics and City Science, 50(1):94–116.
Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. In Proceedings of the fourth ACM international conference on Web search and data mining, pages 65–74.
Baltrusaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2):423–443.
Bandarin, F. and Van Oers, R. (2012). The historic urban landscape: managing heritage in an urban century. John Wiley & Sons.
Barabási, A.-L. (2013). Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1987):20120375.
Barabási, A.-L. et al. (2016). Network Science. Cambridge University Press.
Barabasi, A.-L. and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. Nature reviews genetics, 5(2):101–113.
Barbagallo, D., Bruni, L., Francalanci, C., and Giacomazzi, P. (2012). An Empirical Study on the Relationship between Twitter Sentiment and Influence in the Tourism Domain. In Fuchs, M and Ricci, F and Cantoni, L., editor, INFORMATION AND COMMUNICATION TECHNOLOGIES IN TOURISM 2012, pages 506–516, SACHSENPLATZ 4-6, A-1201 VIENNA, AUSTRIA. SPRINGER-VERLAG WIEN.
Barbier, G. and Liu, H. (2011). Data mining in social media. Social network data analytics, pages 327–352.
Barros, C., Moya-Gómez, B., and Gutiérrez, J. (2020). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. Current Issues in Tourism, 23(10):1291–1310.
Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In Proceedings of the International AAAI Conference on Web and Social Media, volume 3, pages 361–362.
Battiato, S., Farinella, G. M., Milotta, F. L., Ortis, A., Addesso, L., Casella, A., D'Amico, V., and Torrisi, G. (2016). The social picture. In Proceedings of the 2016 ACM on international conference on multimedia retrieval, pages 397–400.
Batty, M. (1976). Urban modelling. Cambridge University Press Cambridge.
Batty, M. (2013). The new science of cities. MIT press.
Batty, M. (2023). A new kind of search. Environment and Planning B: Urban Analytics and City Science, 50(3):575–578.

Batty, M., Xie, Y., and Sun, Z. (1999). Modeling urban dynamics through gis-based cellular automata. Computers, environment and urban systems, 23(3):205–233.
Baumer, E., Elovic, E., Qin, Y., Polletta, F., and Gay, G. (2015). Testing and comparing computational approaches for identifying the language of framing in political news. In Proceedings of the 2015 conference of the North American chapter of the Association for Computational Linguistics: human language technologies, pages 1472–1482.
Bekker, R. (2020). Creating insights in tourism with flickr photography, visualizing and analysing spatial and temporal patterns in venice. Master's thesis, Rijksuniversiteit Groningen.
Bellens, A., Banc, N. V. L., Eloire, F., Grabar, N., Kergosien, E., and Severo, M. (2016). Social media and european cultural routes: Instagram networks on the via francigena. In Proceedings of the 8th International Conference on Management of Digital EcoSystems, pages 122–128.
Bello-Orgaz, G., Jung, J. J., and Camacho, D. (2016). Social big data: Recent achievements and new challenges. Information Fusion, 28:45–59.
Benzi, M. and Klymko, C. (2014). A matrix analysis of different centrality measures. arXiv preprint arXiv:1312.6722.
Berman, M. L., Åhlfeldt, J., and Wick, M. (2012). Historical gazetteer system integration: Chgis, regnum francorum, and geonames. Placing names: enriching and integrating gazetteers. Indiana University Press, Bloomington and Indianapolis.
Bernadou, D. (2017). Construire l'image touristique d'une région à travers les réseaux sociaux : le cas de l'Émilie-romagne en italie. Cybergeo, 2017.
Bertocchi, D. and Visentin, F. (2019). "the overwhelmed city": Physical and social over-capacities of global tourism in venice. Sustainability, 11(24):6937.
Bifulco, F., RUSSO SPENA, T., et al. (2016). The databenc experience: a smart innovation habitat. In Managing Cultural heritage, pages 46–63. McGraw-Hill.
Bigne, E., Ruiz, C., Cuenca, A., Perez, C., and Garcia, A. (2021). What drives the helpfulness of online reviews? a deep learning study of sentiment analysis, pictorial content and reviewer expertise for mature destinations. Journal of Destination Marketing & Management, 20:100570.
Bingham-Hall, J. and Law, S. (2015). Connected or informed?: Local twitter networking in a london neighbourhood. Big Data & Society, 2(2):2053951715597457.
Birant, D. and Kut, A. (2007). St-dbscan: An algorithm for clustering spatial–temporal data. Data & knowledge engineering, 60(1):208–221.
Bird, S., Klein, E., and Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".
Bishop, C. M. and Nasrabadi, N. M. (2006). Pattern recognition and machine learning, volume 4. Springer.
Blanchard, P. and Volchenkov, D. (2008). Mathematical analysis of urban spatial networks. Springer Science & Business Media.
Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022.
Blum, A. and Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In Proceedings of the eleventh annual conference on Computational learning theory, pages 92–100.
Boeing, G. (2017). Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. Computers, Environment and Urban Systems, 65:126–139.
Boland, A., Cherry, M. G., and Dickson, R. (2017). Doing a Systematic Review: A Student's Guide. SAGE PUBLICATIONS INC.
Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. Journal of mathematical sociology, 2(1):113–120.
Bonci, A., Clini, P., Martin, R., Pirani, M., Quattrini, R., and Raikov, A. (2018). Collaborative intelligence cyber-physical system for the valorization and re-use of cultural heritage. Journal of Information Technology in Construction, 23(1):305–323.
Bond, R. and Messing, S. (2015). Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook. American Political Science Review, 109(1):62–78.
Borg, I. and Groenen, P. J. (2005). Modern multidimensional scaling: Theory and applications. Springer Science & Business Media.
Boy, J. D. and Uitermark, J. (2017). Reassembling the city through instagram. Transactions of the Institute of British Geographers, 42(4):612–624.
Bozdag, E., Gao, Q., Houben, G.-J., and Warnier, M. (2014). Does offline political segregation affect the filter bubble? an empirical analysis of information diversity for dutch and turkish twitter users. Computers in human behavior, 41:405–415.
Breiman, L. (1996a). Bagging predictors. Machine learning, 24(2):123–140.
Breiman, L. (1996b). Stacked regressions. Machine learning, 24(1):49–64.
Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32.
Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901.

Bruns, A., Burgess, J. E., Crawford, K., and Shaw, F. (2012). # qldfloods and QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods. Technical Report Cci, ARC Centre of Excellence for Creative Industries and Innovation.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., et al. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712.

Budescu, D. V. and Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. Acta psychologica, 104(3):371–398.

Budescu, D. V. and Yu, H.-T. (2007). Aggregation of opinions based on correlated cues and advisors. Journal of Behavioral Decision Making, 20(2):153–177.

Burgers, G.-J. (2021). Tourism, leisure and cultural heritage: The challenge of participatory planning and design. In Tourism and Regional Science: New Roads, pages 71–85. Springer.

Burgess, S., Sellitto, C., Buultjens, J., and Cox, C. (2015). How australian smes engage with social media. In Proceedings of the European conference on e-learning, pages 45–51.

Calder, M., Craig, C., Culley, D., de Cani, R., Donnelly, C. A., Douglas, R., Edmonds, B., Gascoigne, J., Gilbert, N., Hargrove, C., Hinds, D., Lane, D. C., Mitchell, D., Pavey, G., Robertson, D., Rosewell, B., Sherwin, S., Walport, M., and Wilson, A. (2018). Computational modelling for decision-making: Where, why, what, who and how. Royal Society Open Science, 5(6).

Calvino, I. (1978). Invisible cities. Houghton Mifflin Harcourt.

Campillo-Alhama, C. and Martinez-Sala, A.-M. (2019). Events 2.0 in the transmedia branding strategy of World Cultural Heritage Sites. PROFESIONAL DE LA INFORMACION, 28(5).

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., and Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018), pages 67–74. IEEE.

Cao, R., Tu, W., Yang, C., Li, Q., Liu, J., Zhu, J., Zhang, Q., Li, Q., and Qiu, G. (2020). Deep learning-based remote and social sensing data fusion for urban region function recognition. ISPRS Journal of Photogrammetry and Remote Sensing, 163:82–97.

Card, D., Boydstun, A., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 438–444.

Cartwright, W. E. (2010). Addressing the value of art in cartographic communication. ISPRS Journal of Photogrammetry and Remote Sensing, 65(3):294–299.

Castiglione, A., Colace, F., Moscato, V., and Palmieri, F. (2018). Chis: A big data infrastructure to manage digital cultural items. Future Generation Computer Systems, 86:1134–1145.

Castro de Azevedo, A. (2023). Barriers to public participation in memorialization processes: Evidence from the dutch holocaust memorial of names. In Rodenberg, J., Wagenaar, P., and Burgers, G.-J., editors, Calling on the Community, Explorations in Heritage Studies, pages 162–188. Berghahn Books, United Kingdom.

Cavnar, W. B. and Trenkle, J. M. (1994). N-gram-based text categorization. In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, pages 161–175.

Chaabani, Y., Toujani, R., and Akaichi, J. (2018). Sentiment analysis method for tracking touristics reviews in social media network. Smart Innovation, Systems and Technologies, 76:299–310.

Chan, J. C.-W. and Paelinckx, D. (2008). Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. Remote Sensing of Environment, 112(6):2999–3011.

Chen, F.-W., Guevara Plaza, A., and Alarcón Urbistondo, P. (2017). Automatically extracting tourism-related opinion from chinese social media. Current Issues in Tourism, 20(10):1070–1087.

Chen, M., Wei, Z., Huang, Z., Ding, B., and Li, Y. (2020). Simple and deep graph convolutional networks. In International Conference on Machine Learning, pages 1725–1735. PMLR.

Chen, Y. (2021). An analytical process of spatial autocorrelation functions based on moran's index. PloS one, 16(4):e0249589.

Cheng, J., Adamic, L. A., Kleinberg, J., and Leskovec, J. (2016). Do cascades recur? 25th International World Wide Web Conference, WWW 2016, pages 671–681.

Cheng, T., Haworth, J., and Wang, J. (2012). Spatio-temporal autocorrelation of road network data. Journal of Geographical Systems, 14:389–413.

Cheng, T. and Wang, J. (2009). Accommodating spatial associations in drnn for space–time analysis. Computers, Environment and Urban Systems, 33(6):409–418.

Cheng, T. and Wicks, T. (2014). Event detection using twitter: A spatio-temporal approach. PloS one, 9(6):e97807.

Chianese, A., Marulli, F., and Piccialli, F. (2016). Cultural Heritage and Social Pulse: A Semantic Approach for CH Sensitivity Discovery in Social Media Data. In Proceedings - 2016 IEEE 10th International Conference on Semantic Computing, ICSC 2016, pages 459–464. Institute of Electrical and Electronics Engineers Inc.

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. Syntax, Semantics and Structure in Statistical Translation, page 103.

Cho, N., Kang, Y., Yoon, J., Park, S., and Kim, J. (2022). Classifying tourists' photos and exploring tourism destination image using a deep learning model. Journal of Quality Assurance in Hospitality & Tourism, pages 1–29.

Choi, C. and Hong, S.-Y. (2021). Mdst-dbscan: A density-based clustering method for multidimensional spatiotemporal data. ISPRS International Journal of Geo-Information, 10(6):391.

Chorowski, J. and Jaitly, N. (2017). Towards better decoding and language model integration in sequence to sequence models. In Proc. Interspeech 2017, pages 523–527.

Chua, T.-S., Tang, J., Hong, R., Li, H., Luo, Z., and Zheng, Y. (2009). Nus-wide: a real-world web image database from national university of singapore. In Proceedings of the ACM international conference on image and video retrieval, pages 1–9.

Church, K. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. Computational linguistics, 16(1):22–29.

Clark, K., Khandelwal, U., Levy, O., and Manning, C. D. (2019). What does BERT look at? an analysis of BERT's attention. In Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, pages 276–286, Florence, Italy. Association for Computational Linguistics.

Claster, W. B., Cooper, M., and Sallis, P. (2010). Thailand–tourism and conflict: Modeling sentiment from twitter tweets using naïve bayes and unsupervised artificial neural nets. In 2010 second international conference on computational intelligence, Modelling and Simulation, pages 89–94. IEEE.

Clemens, K. (2015). Geocoding with openstreetmap data. GEOProcessing 2015, page 10.

Clemente, P., Calvache, M., Antunes, P., Santos, R., Cerdeira, J. O., and Martins, M. J. (2019). Combining social media photographs and species distribution models to map cultural ecosystem services: The case of a natural park in portugal. Ecological indicators, 96:59–68.

Colace, F., De Santo, M., Greco, L., Amato, F., Moscato, V., and Picariello, A. (2014). Terminological ontology learning and population using latent dirichlet allocation. Journal of Visual Languages & Computing, 25(6):818–826.

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. Journal of machine learning research, 12(ARTICLE):2493–2537.

Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. IEEE Transactions on pattern analysis and machine intelligence, 24(5):603–619.

Condorelli, F., Rinaudo, F., Salvadore, F., and Tagliaventi, S. (2020). A neural networks approach to detecting lost heritage in historical video. ISPRS International Journal of Geo-Information, 9(5):297.

Cosgrove, D. (1982). The myth and the stones of venice: an historical geography of a symbolic landscape. Journal of Historical Geography, 8(2):145–169.

Costa, M. A. and Kulldorff, M. (2014). Maximum linkage space-time permutation scan statistics for disease outbreak detection. International journal of health geographics, 13(1):1–14.

Couclelis, H. (2005). "Where has the future gone?" Rethinking the role of integrated land-use models in spatial planning. Environment and Planning A, 37(8):1353–1371.

Council of Europe (2005). Convention on the value of cultural heritage for society (faro convention). Technical report, Council of Europe, Faro.

Craglia, M., Micheli, M., Hradec, J., Calzada, I., Luitjens, S., Ponti, M., Scholten, H. J., and Boter, J. (2021). Digitranscope: The governance of digitally-transformed society. Craglia, M., Scholten, HJ, Micheli, M., Hradec, J., Calzada, I., Luitjens, S., Ponti, M. and Boter, J., Digitranscope: The governance of digitally-transformed society, EUR, 30590.

Crandall, D., Backstrom, L., Huttenlocher, D., and Kleinberg, J. (2009). Mapping the world's photos. WWW'09 - Proceedings of the 18th International World Wide Web Conference, pages 761–770.

Cristelli, M., Batty, M., and Pietronero, L. (2012). There is more than a power law in zipf. Scientific reports, 2(1):1–7.

Cristianini, N., Shawe-Taylor, J., et al. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

Dahal, B., Kumar, S. A., and Li, Z. (2019). Topic modeling and sentiment analysis of global climate change tweets. Social network analysis and mining, 9:1–20.

Dai, P., Zhang, S., Chen, Z., Gong, Y., and Hou, H. (2019). Perceptions of cultural ecosystem services in urban parks based on social network data. Sustainability, 11(19):5386.

De Angelis, A., Gasparetti, F., Micarelli, A., and Sansonetti, G. (2017). A social cultural recommender based on linked open data. In Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, pages 329–332.

De Kleijn, M., van Aart, C. J., Van Manen, N., Burgers, G.-J., and Scholten, H. J. (2013). Testaccio, a digital cultural biography app. In UMAP Workshops.

de Souza, C. R., Redmiles, D., Cheng, L.-T., Millen, D., and Patterson, J. (2004). Sometimes you need to see through walls: a field study of application programming interfaces. In Proceedings of the 2004 ACM conference on Computer supported cooperative work, pages 63–71.

Deeksha, S., Ashrith, H., Bansode, R., and Kamath, S. (2015). A spatial clustering approach for efficient landmark discovery using geo-tagged photos. In 2015 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), pages 1–6. IEEE.

Del Vecchio, P., Mele, G., Ndou, V., and Secundo, G. (2018). Creating value from social big data: Implications for smart tourism destinations. Information Processing & Management, 54(5):847–860.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee.
Deng, Y., Wang, M., Yang, Y., and Yue, Y. (2022). Hd-ccsom: Hierarchical and dense collaborative continuous semantic occupancy mapping through label diffusion. In 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 2417–2422. IEEE.
Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
Dewulf, A. and Bouwen, R. (2012). Issue framing in conversations for change: Discursive interaction strategies for "doing differences". The Journal of Applied Behavioral Science, 48(2):168–193.
Dickinger, A. and Lalicic, L. (2016). An analysis of destination brand personality and emotions: a comparison study. Information Technology & Tourism, 15:317–340.
Dickinger, A., Lalicic, L., and Mazanec, J. (2017). Exploring the generalizability of discriminant word items and latent topics in online tourist reviews. International Journal of Contemporary Hospitality Management, 29(2):803–816.
Dong, W., Heller, K., and Pentland, A. (2012). Modeling infection with multi-agent dynamics. In Social Computing, Behavioral-Cultural Modeling and Prediction: 5th International Conference, SBP 2012, College Park, MD, USA, April 3-5, 2012. Proceedings 5, pages 172–179. Springer.
Dubois, A. and Gadde, L.-E. (2002). Systematic combining: an abductive approach to case research. Journal of business research, 55(7):553–560.
Ducci, M., Janssen, R., Burgers, G.-J., and Rotondo, F. (2023). Mapping local perceptions for the planning of cultural landscapes. International Journal of E-Planning Research (IJEPR), 12(1):1–27.
Easley, D. and Kleinberg, J. (2010). Networks, Crowds and Markets: Reasoning about a Highly Connected World. Cambridge University Press, Cambridge.
Eisenstein, J. (2018). Natural Language Processing. MIT Press.
Ekman, P. (1992). An argument for basic emotions. Cognition & emotion, 6(3-4):169–200.
Encalada, L., Boavida-Portugal, I., Cardoso Ferreira, C., and Rocha, J. (2017). Identifying tourist places of interest based on digital imprints: Towards a sustainable smart city. Sustainability, 9(12):2317.
Encalada-Abarca, L., Ferreira, C. C., and Rocha, J. (2022). Measuring tourism intensification in urban destinations: An approach based on fractal analysis. Journal of Travel Research, 61(2):394–413.
Eom, Y.-H. and Jo, H.-H. (2015). Tail-scope: Using friends to estimate heavy tails of degree distributions in large-scale complex networks. Scientific reports, 5(1):1–9.
Esch, T., Heldens, W., Hirner, A., Keil, M., Marconcini, M., Roth, A., Zeidler, J., Dech, S., and Strano, E. (2017). Breaking new ground in mapping human settlements from space–the global urban footprint. ISPRS Journal of Photogrammetry and Remote Sensing, 134:30–42.
Estellés-Arolas, E. and González-Ladrón-de Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information science, 38(2):189–200.
Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, page 226–231. AAAI Press.
Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In LREC, volume 6, pages 417–422.
Farnaghi, M., Ghaemi, Z., and Mansourian, A. (2020). Dynamic spatio-temporal tweet mining for event detection: a case study of hurricane florence. International Journal of Disaster Risk Science, 11:378–393.
Faruqui, M., Dodge, J., Jauhar, S. K., Dyer, C., Hovy, E., and Smith, N. A. (2015). Retrofitting word vectors to semantic lexicons. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 1606–1615.
Feizollah, A., Ainin, S., Anuar, N. B., Abdullah, N. A. B., and Hazim, M. (2019). Halal products on twitter: Data extraction and sentiment analysis using stack of deep learning algorithms. IEEE Access, 7:83354–83362.
Fejérdy, T. (2007). Evolution and possible enhancement of the concept of ouv. In Values and Criteria in Heritage Conservation, pages 323–328. Polistampa.
Ferrarin, C., Bajo, M., Benetazzo, A., Cavaleri, L., Chiggiato, J., Davison, S., Davolio, S., Lionello, P., Orlić, M., and Umgiesser, G. (2021). Local and large-scale controls of the exceptional venice floods of november 2019. Progress in Oceanography, 197:102628.
Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., and Criado, N. (2021). Bias and discrimination in ai: a cross-disciplinary perspective. IEEE Technology and Society Magazine, 40(2):72–80.
Févotte, C. and Idier, J. (2011). Algorithms for nonnegative matrix factorization with the β-divergence. Neural computation, 23(9):2421–2456.
Fey, M. and Lenssen, J. E. (2019). Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428.
Figueredo, M., Ribeiro, J., Cacho, N., Thome, A., Cacho, A., Lopes, F., and Araujo, V. (2018). From photos to travel itinerary: A tourism recommender system for smart tourism destination. In 2018 IEEE Fourth International Conference on Big Data Computing Service and Applications (BigDataService), pages 85–92. IEEE.

Fischer, G. (2009). Democratizing design: New challenges and opportunities for computer-supported collaborative learning. Computer Supported Collaborative Learning Practices, CSCL 2009 Conference Proceedings - 9th International Conference, pages 282–286.
Floridi, L. and Chiriatti, M. (2020). Gpt-3: Its nature, scope, limits, and consequences. Minds and Machines, 30:681–694.
Floris, R., Campagna, M., et al. (2014). Social media geographic information in tourism planning. TEMA, 257:417–430.
Floris, R. and Zoppi, C. (2015). Social media-related geographic information in the context of strategic environmental assessment of municipal masterplans: A case study concerning sardinia (italy). Future Internet, 7(3):276–293.
Foroughi, M., de Andrade, B., and Pereira Roders, A. (2022). Peoples' values and feelings matter: Participatory heritage management using social media. In Muntañola, J., editor, Artificial Intelligence and Architectural Design, volume 33, pages 107–120. Oficina de Publicacions Acadèmiques Digitals de la UPC.
Foroughi, M., de Andrade, B., Roders, A. P., and Wang, T. (2023). Public participation and consensus-building in urban planning from the lens of heritage planning: A systematic literature review. Cities, 135:104235.
Fostikov, A. (2023). First impressions on using ai powered chatbots, tools and search engines: Chatgpt, perplexity and other–possibilities and usage problems. Review of the National Center for Digitization 2023 (preprint).
Fredholm, S., Dore, M., and Brorström, S. (2021). Strategic responses to wicked problems of heritage management: Experiences from the west link infrastructure project in gothenburg, sweden. Land, 10(10):1032.
Freund, Y., Schapire, R. E., et al. (1996). Experiments with a new boosting algorithm. In icml, volume 96, pages 148–156. Citeseer.
Fukui, M. and Ohe, Y. (2019). Assessing the role of social media in tourism recovery in tsunami-hit coastal areas in Tohoku, Japan. Tourism Economics.
Gabrielli, L., Rinzivillo, S., Ronzano, F., and Villatoro, D. (2014). From Tweets to Semantic Trajectories: Mining Anomalous Urban Mobility Patterns. In Nin, J and Villatoro, D., editor, CITIZEN IN SENSOR NETWORKS, volume 8313 of Lecture Notes in Artificial Intelligence, pages 26–35, HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY. SPRINGER-VERLAG BERLIN.
Galesic, M., Bruine de Bruin, W., Dalege, J., Feld, S. L., Kreuter, F., Olsson, H., Prelec, D., Stein, D. L., and van Der Does, T. (2021). Human social sensing is an untapped resource for computational social science. Nature, 595(7866):214–222.
Galke, L. and Scherp, A. (2022). Bag-of-words vs. graph vs. sequence in text classification: Questioning the necessity of text-graphs and the surprising strength of a wide MLP. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 4038–4051, Dublin, Ireland. Association for Computational Linguistics.
Gao, H. and Ji, S. (2019). Graph u-nets. In international conference on machine learning, pages 2083–2092. PMLR.
García-Vega, M., Díaz-Galiano, M., García-Cumbreras, M., Del Arco, F., Montejo-Ráez, A., Jiménez-Zafra, S., Martínez Cámara, E., Aguilar, C., Cabezudo, M., Chiruzzo, L., et al. (2020). Overview of tass 2020: Introducing emotion detection. In Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) Co-Located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, pages 163–170.
Gardner, M. W. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric environment, 32(14-15):2627–2636.
Garduño Freeman, C. and Gonzalez Zarandona, J. A. (2021). Digital spectres: the notre-dame effect. International Journal of Heritage Studies, 27(12):1264–1277.
GeoMatt22 (2020). Similarity metrics for more than two vectors? Stack Exchange. (version: 2020-12-10) (access date: 2022-08-31).
George, Y., Karunasekera, S., Harwood, A., and Lim, K. H. (2021). Real-time spatio-temporal event detection on geotagged social media. Journal of Big Data, 8(1):91.
Gerrig, R. J., Zimbardo, P. G., Campbell, A. J., Cumming, S. R., and Wilkes, F. J. (2015). Psychology and life. Pearson Higher Education AU.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2019a). Machine learning and point of interests: typical tourist Italian cities. Current Issues in Tourism, 0(0):1–13.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2019b). Using social media to identify tourism attractiveness in six italian cities. Tourism management, 72:306–312.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2019c). Using social media to identify tourism attractiveness in six Italian cities. Tourism Management, 72:306–312.
Giglio, S., Bertacchini, F., Bilotta, E., and Pantano, P. (2020). Machine learning and points of interest: typical tourist italian cities. Current Issues in Tourism, 23(13):1646–1658.
Ginzarly, M., Pereira Roders, A., and Teller, J. (2019). Mapping historic urban landscape values through social media. Journal of Cultural Heritage, 36:1–11.
Ginzarly, M. and Srour, F. J. (2022). Cultural heritage through the lens of covid-19. Poetics, 92:101622.

Ginzarly, M., Srour, F. J., and Roders, A. P. (2022). The interplay of context, experience, and emotion at world heritage sites: a qualitative and machine learning approach. Tourism Culture & Communication, 22(4):321–340.

Gomez, R., Gomez, L., Gibert, J., and Karatzas, D. (2019). Learning from #barcelona instagram data what locals and tourists post about its neighbourhoods. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11134 LNCS:530–544.

Gonçalves, J., Mateus, R., Silvestre, J. D., Roders, A. P., and Bragança, L. (2021). Attitudes matter: Measuring the intention-behaviour gap in built heritage conservation. Sustainable Cities and Society, 70:102913.

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. nature, 453(7196):779–782.

Goodchild, M. and Longley, P. (1999). The future of gis and spatial analysis. In Geographical information systems: principles, techniques, management and applications, volume 1, pages 567–580. John Wiley New York.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. MIT press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

Gosal, A. S., Geijzendorffer, I. R., Václavík, T., Poulin, B., and Ziv, G. (2019). Using social media, machine learning and natural language processing to map multiple recreational beneficiaries. Ecosystem Services, 38:100958.

Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2021). Knowledge distillation: A survey. International Journal of Computer Vision, 129(6):1789–1819.

Gould, P. R. (1967). On the geographical interpretation of eigenvalues. Transactions of the Institute of British Geographers, pages 53–86.

Grandi, R. and Neri, F. (2014). Sentiment analysis and city branding. In New Trends in Databases and Information Systems: 17th East European Conference on Advances in Databases and Information Systems, pages 339–349. Springer.

Gravetter, F. J., Wallnau, L. B., Forzano, L.-A. B., and Witnauer, J. E. (2020). Essentials of statistics for the behavioral sciences. Cengage Learning.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv preprint arXiv:2203.05794.

Guo, T., Guo, B., Ouyang, Y., Yu, Z., Lam, J. C., and Li, V. O. (2018). Crowdtravel: scenic spot profiling by using heterogeneous crowdsourced data. Journal of Ambient Intelligence and Humanized Computing, 9:2051–2060.

Gustcoven, E. (2016). Attributes of world heritage cities, sustainability by management–a comparative study between the world heritage cities of amsterdam, edinburgh and querétaro. Master's thesis, KU Leuven.

Hagberg, A., Swart, P., and S Chult, D. (2008). Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Haklay, M. and Weber, P. (2008). Openstreetmap: User-generated street maps. IEEE Pervasive computing, 7(4):12–18.

Hamilton, J. D. (2020). Time series analysis. Princeton university press.

Hamilton, W., Ying, Z., and Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in neural information processing systems, 30.

Hashida, S., Tamura, K., and Sakai, T. (2018). Classifying sightseeing tweets using convolutional neural networks with multi-channel distributed representation. In 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pages 178–183. IEEE.

Hasnat, M. M. and Hasan, S. (2018). Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. Transportation Research Part C: Emerging Technologies, 96:38–54.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778.

He, Z., Deng, N., Li, X., and Gu, H. (2022). How to "read" a destination from images? machine learning and network methods for dmos' image projection and photo evaluation. Journal of Travel Research, 61(3):597–619.

Hillier, B. and Hanson, J. (1989). The Social Logic of Space. Cambridge University Press.

Hinton, G. E. (1990). Connectionist learning procedures. In Machine learning, pages 555–610. Elsevier.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8):1735–1780.

Holt, D. (2016). Branding in the age of social media. Harvard business review, 94(3):40–50.

Honkela, T. (1997). Self-organizing maps in natural language processing. PhD thesis, Citeseer.

Howard, J. and Gugger, S. (2020). Fastai: A layered api for deep learning. Information (Switzerland), 11(2):1–26.

Howard, J. and Ruder, S. (2018). Universal language model fine-tuning for text classification. In Gurevych, I. and Miyao, Y., editors, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers, pages 328–339. Association for Computational Linguistics.

Howley, P., Scott, M., and Redmond, D. (2009). Sustainability versus liveability: an investigation of neighbourhood satisfaction. Journal of environmental planning and management, 52(6):847–864.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., and Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. Computers, Environment and Urban Systems, 54:240–254.

Hu, Y., Wang, F., Guin, C., and Zhu, H. (2018). A spatio-temporal kernel density estimation framework for predictive crime hotspot mapping and evaluation. Applied geography, 99:89–97.

Hu, Z., Dong, Y., Wang, K., and Sun, Y. (2020). Heterogeneous graph transformer. In Proceedings of The Web Conference 2020, pages 2704–2710.

Huang, Q., He, H., Singh, A., Lim, S.-N., and Benson, A. R. (2020). Combining label propagation and simple models out-performs graph neural networks. arXiv preprint arXiv:2010.13993.

Huang, T. (1996). Computer vision: Evolution and promise. 1996 CERN SCHOOL OF COMPUTING, page 21.

Huang, W. and Li, S. (2016). Understanding human activity patterns based on space-time-semantics. ISPRS journal of photogrammetry and remote sensing, 121:1–10.

Huang, X., Wang, C., Li, Z., and Ning, H. (2019). A visual–textual fused approach to automated tagging of flood-related tweets during a flood event. International Journal of Digital Earth, 12(11):1248–1264.

Huang, Y., Li, Y., and Shan, J. (2018). Spatial-temporal event detection from geo-tagged tweets. ISPRS International Journal of Geo-Information, 7(4):150.

Huiskes, M. J. and Lew, M. S. (2008). The MIR Flickr retrieval evaluation. Proceedings of the 1st International ACM Conference on Multimedia Information Retrieval, MIR2008, Co-located with the 2008 ACM International Conference on Multimedia, MM'08, pages 39–43.

ICOMOS, A. (2013). The Burra Charter: The Australia ICOMOS charter for places of cultural significance 2013. Australia ICOMOS Incorporated.

IUCN, ICOMOS, ICROM, and WHC (2010). Guidance on the preparation of retrospective Statements of Outstanding Universal Value for World Heritage Properties. Technical report, IUCN, ICOMOS, ICROM and WHC.

Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. PloS one, 9(6):e98679.

Jansen, W. S., Otten, S., van der Zee, K. I., and Jans, L. (2014). Inclusion: Conceptualization and measurement. European journal of social psychology, 44(4):370–385.

Janssen, J., Luiten, E., Renes, H., and Stegmeijer, E. (2017). Heritage as sector, factor and vector: conceptualizing the shifting relationship between heritage management and spatial planning. European Planning Studies, 25(9):1654–1672.

Jia, Z., Nourian, P., Luscuere, P., and Wagenaar, C. (2023). Spatial decision support systems for hospital layout design: A review. Journal of Building Engineering, page 106042.

Jindal, N. and Liu, B. (2006). Identifying comparative sentences in text documents. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pages 244–251.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. In Nedellec, C. and Rouveirol, C., editors, Machine Learning: ECML-98, 10th European Conference on Machine Learning, Chemnitz, Germany, April 21-23, 1998, Proceedings, volume 1398 of Lecture Notes in Computer Science, pages 137–142. Springer.

Johansson, R. (2007). On case study methodology. Open house international, 32(3):48–54.

Johnson, R. and Zhang, T. (2017). Deep pyramid convolutional neural networks for text categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 562–570, Vancouver, Canada. Association for Computational Linguistics.

Jokilehto, J. (2006). World Heritage: Defining the Outstanding Universal Value. City & Time, 2(2):1–10.

Jokilehto, J. (2007). Aesthetics in the world heritage context. In Values and Criteria in Heritage Conservation, pages 183–194. Polistampa.

Jokilehto, J. (2008). What is OUV? Defining the Outstanding Universal Value of Cultural World Heritage Properties. Technical report, ICOMOS, ICOMOS Berlin.

Junker, C., Akbar, Z., and Cuquet, M. (2017). The network structure of visited locations according to geotagged social media photos. arXiv, pages 1–8.

Jurafsky, D. and Martin, J. H. (2020). Speech and language processing An Intoduction to Natural Language Processing, Computational Liguistics, and Speech Recognition. Third Edition Draft.

Kang, Y., Cho, N., Yoon, J., Park, S., and Kim, J. (2021). Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos. ISPRS International Journal of Geo-Information, 10(3):137.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. IEEE transactions on pattern analysis and machine intelligence, 24(7):881–892.

Kaplan, A. M. and Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. Business horizons, 53(1):59–68.

Karypis, G. and Kumar, V. (1995). Analysis of multilevel graph partitioning. In Supercomputing'95: Proceedings of the 1995 ACM/IEEE conference on Supercomputing, pages 29–29. IEEE.

Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika, 18(1):39–43.

Kennedy, L., Naaman, M., Ahern, S., Nair, R., and Rattenbury, T. (2007). How flickr helps us make sense of the world: context and content in community-contributed media collections. In Proceedings of the 15th ACM international conference on Multimedia, pages 631–640.

Kersten, J. and Klan, F. (2020). What happens where during disasters? a workflow for the multifaceted characterization of crisis events based on twitter data. Journal of Contingencies and Crisis Management, 28(3):262–280.
Kim, J. and Kang, Y. (2022). Automatic classification of photos by tourist attractions using deep learning model and image feature vector clustering. ISPRS International Journal of Geo-Information, 11(4):245.
Kim, W., Son, B., and Kim, I. (2021). Vilt: Vision-and-language transformer without convolution or region supervision. In International Conference on Machine Learning, pages 5583–5594. PMLR.
Kim, Y. (2014). Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.
Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.
Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
Kırmızı, Ö. and Karaman, A. (2021). A participatory planning model in the context of historic urban landscape: The case of kyrenia's historic port area. Land use policy, 102:105130.
Kisilevich, S., Mansmann, F., and Keim, D. (2010). P-dbscan: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. In Proceedings of the 1st international conference and exhibition on computing for geospatial research & application, pages 1–4.
Knyazev, B., Taylor, G. W., and Amer, M. (2019). Understanding attention and generalization in graph neural networks. Advances in neural information processing systems, 32.
Korakakis, M., Spyrou, E., Mylonas, P., and Perantonis, S. J. (2017). Exploiting social media information toward a context-aware recommendation system. Social Network Analysis and Mining, 7:1–20.
Kounadi, O., Lampoltshammer, T. J., Leitner, M., and Heistracher, T. (2013). Accuracy and privacy aspects in free online reverse geocoding services. Cartography and Geographic Information Science, 40(2):140–153.
Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. Social networks, 10(4):359–381.
Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.
Krothapalli, U. and Abbott, A. L. (2020). Adaptive label smoothing. arXiv preprint arXiv:2009.06432.
Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29(1):1–27.
Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R., and Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. PLoS medicine, 2(3):e59.
Kumar, P. (2019). Learning from the Past and Preparing for the Future: Cases and Tools for Cultural Heritage during Disasters. PhD thesis, IMT School for Advanced Studies Lucca.
Kumar, P. (2020a). Crowdsourcing to rescue cultural heritage during disasters: A case study of the 1966 florence flood. International Journal of Disaster Risk Reduction, 43:101371.
Kumar, P. (2020b). Twitter, disasters and cultural heritage: A case study of the 2015 nepal earthquake. Journal of Contingencies and Crisis Management, 28(4):453–465.
Kumar, P., Ofli, F., Imran, M., and Castillo, C. (2020). Detection of disaster-affected cultural heritage sites from social media images using deep learning techniques. Journal on Computing and Cultural Heritage (JOCCH), 13(3):1–31.
Lafon, S. and Lee, A. B. (2006). Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE transactions on pattern analysis and machine intelligence, 28(9):1393–1403.
Lai, J. (2019). Urban Place Profiling Using Geo-Referenced Social Media Data. PhD thesis, UCL (University College London).
Lai, J., Cheng, T., and Lansley, G. (2017). Improved targeted outdoor advertising based on geotagged social media data. Annals of GIS, 23(4):237–250.
Landauer, T. K., Foltz, P. W., and Laham, D. (1998). An introduction to latent semantic analysis. Discourse processes, 25(2-3):259–284.
Lansley, G. and Longley, P. A. (2016). The geography of twitter topics in london. Computers, Environment and Urban Systems, 58:85–96.
Latora, V., Nicosia, V., and Russo, G. (2017). Complex networks: principles, methods and applications. Cambridge University Press.
Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Social science. computational social science. Science (New York, NY), 323(5915):721–723.
LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. nature, 521(7553):436–444.
LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4):541–551.
Lee, D.-H. et al. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Workshop on challenges in representation learning, ICML, volume 3, page 896.
Lee, H. and Kang, Y. (2021). Mining tourists' destinations and preferences through lstm-based text classification and spatial clustering using flickr data. Spatial Information Research, 29(6):825–839.

Lee, J., Lee, I., and Kang, J. (2019). Self-attention graph pooling. In International conference on machine learning, pages 3734–3743. PMLR.
Lee, R., Wakamiya, S., and Sumiya, K. (2011). Discovery of unusual regional social activities using geo-tagged microblogs. World Wide Web, 14:321–349.
Lefebvre, H. (2014). The production of space (1991). In The people, place, and space reader, pages 323–327. Routledge.
Leung, D., Law, R., Van Hoof, H., and Buhalis, D. (2013). Social media in tourism and hospitality: A literature review. Journal of travel & tourism marketing, 30(1-2):3–22.
Leung, R., Vu, H. Q., and Rong, J. (2017). Understanding tourists' photo sharing and visit pattern at non-first tier attractions via geotagged photos. Information Technology & Tourism, 17:55–74.
Li, D., Zhou, X., and Wang, M. (2018a). Analyzing and visualizing the spatial interactions between tourists and locals: A flickr study in ten us cities. Cities, 74:249–258.
Li, J., Krishnamurthy, S., Roders, A. P., and Van Wesemael, P. (2020). Community participation in cultural heritage management: A systematic literature review comparing chinese and international practices. Cities, 96:102476.
Li, J., Krishnamurthy, S., Roders, A. P., and van Wesemael, P. (2021). Imagine the old town of lijiang: Contextualising community participation for urban heritage management in china. Habitat International, 108:102321.
Li, L.-J. and Fei-Fei, L. (2007). What, where and who? classifying events by scene and object recognition. In 2007 IEEE 11th international conference on computer vision, pages 1–8. IEEE.
Li, Q., Han, Z., and Wu, X.-M. (2018b). Deeper insights into graph convolutional networks for semi-supervised learning. In Thirty-Second AAAI conference on artificial intelligence, pages 1–8.
Li, Y., Tarlow, D., Brockschmidt, M., and Zemel, R. (2015). Gated graph sequence neural networks. arXiv preprint arXiv:1511.05493.
Liao, L., Chen, W., Xiao, J., Wang, Z., Lin, C.-W., and Satoh, S. (2022). Unsupervised foggy scene understanding via self spatial-temporal label diffusion. IEEE Transactions on Image Processing, 31:3525–3540.
Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., and Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Journal of clinical epidemiology, 62(10):e1–e34.
Lieber, R. (1980). On the organization of the lexicon. PhD thesis, Massachusetts Institute of Technology.
Liew, C. L. (2014). Participatory cultural heritage: A tale of two institutions' use of social media. D-Lib Magazine, 20(3-4):1–17.
Lin, M., Pereira Roders, A., Nevzgodin, I., and de Jonge, W. (2023). Values and interventions: dynamic relationships in international doctrines. Journal of Cultural Heritage Management and Sustainable Development, ahead-of-print(ahead-of-print). Publisher: Emerald Publishing Limited.
Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision, pages 740–755. Springer.
Lipizzi, C., Iandoli, L., and Marquez, J. E. R. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using twitter streams. International Journal of Information Management, 35(4):490–503.
Lipton, Z. C., Berkowitz, J., and Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019.
Liu, J. (2007). Qap: A unique method of measuring "relationships" in relational data. Chinese Journal of Sociology(in Chinese Version), 27(4):164–174.
Liu, J., Li, T., Xie, P., Du, S., Teng, F., and Yang, X. (2020). Urban big data fusion based on deep learning: An overview. Information Fusion, 53:123–133.
Liu, M., Liu, X., Li, Y., Chen, X., Hauptmann, A. G., and Shan, S. (2016). Exploiting feature hierarchies with convolutional neural networks for cultural event recognition. Proceedings of the IEEE International Conference on Computer Vision, 2016-Febru:274–279.
Liu, P. and De Sabbata, S. (2021). A graph-based semi-supervised approach to classification learning in digital geographies. Computers, Environment and Urban Systems, 86:101583.
Liu, S. B. (2011). Grassroots heritage: A multi-method investigation of how social media sustain the living heritage of historic crises. PhD thesis, University of Colorado at Boulder.
Liu, T., Butler, R. J., and Zhang, C. (2019). Evaluation of public perceptions of authenticity of urban heritage under the conservation paradigm of historic urban landscape—a case study of the five avenues historic district in tianjin, china. Journal of architectural conservation, 25(3):228–251.
Liu, X., Ji, K., Fu, Y., Tam, W., Du, Z., Yang, Z., and Tang, J. (2022). P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 61–68.
Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., and Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. Annals of the Association of American Geographers, 105(3):512–530.

Liu, Z., Shan, J., Glassey Balet, N., and Fang, G. (2017). Semantic social media analysis of chinese tourists in switzerland. Information technology & tourism, 17:183–202.
Loper, E. and Bird, S. (2002). Nltk: the natural language toolkit. In Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics-Volume 1, pages 63–70.
Lorini, V., Rufolo, P., and Castillo, C. (2022). Venice was flooding... one tweet at a time. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2):1–16.
Lu, W. and Stepchenkova, S. (2015a). User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. Journal of Hospitality Marketing & Management, 24(2):119–154.
Lu, W. and Stepchenkova, S. (2015b). User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. Journal of Hospitality Marketing and Management, 24(2):119–154.
Luo, J.-D. and Cheng, M.-Y. (2015). Guanxi circles' effect on organizational trust: Bringing power and vertical social exchanges into intraorganizational network analysis. American Behavioral Scientist, 59(8):1024–1037.
Luo, Y., Card, D., and Jurafsky, D. (2020). Detecting stance in media on global warming. arXiv preprint arXiv:2010.15149.
Lupo, B. M. (2021). Patrimônio cultural e cat´ ástrofe: Os concursos internacionais não-oficiais realizados para a notre dame de paris após o incêndio de 2019. Herança-Revista de História, Património e Cultura, 4(2):018–038.
Lynch, K. (1964). The image of the city. MIT press.
Ma, Y., Ren, Z., Jiang, Z., Tang, J., and Yin, D. (2018). Multi-dimensional network embedding with hierarchical structure. In Proceedings of the eleventh ACM international conference on web search and data mining, pages 387–395.
Ma, Y. and Tang, J. (2021). Deep learning on graphs. Cambridge University Press.
Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., and Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. International Journal of Geographical Information Science, 27(4):662–684.
Manning, C. and Schutze, H. (1999). Foundations of statistical natural language processing. MIT press.
Manning, C. D. (2009). An introduction to information retrieval. Cambridge university press.
Manríquez, R., Guerrero-Nancuante, C., Martínez, F., and Taramasco, C. (2021). Spread of epidemic disease on edge-weighted graphs from a database: A case study of covid-19. International Journal of Environmental Research and Public Health, 18(9):4432.
Mao, J., Lu, K., Li, G., and Yi, M. (2016). Profiling users with tag networks in diffusion-based personalized recommendation. Journal of Information Science, 42(5):711–722.
Mariani, M. M., Di Felice, M., and Mura, M. (2016). Facebook as a destination marketing tool: Evidence from italian regional destination management organizations. Tourism management, 54:321–343.
Marine-Roig, E. and Anton Clavé, S. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. Journal of Destination Marketing and Management, 4(3):162–172.
Marine-Roig, E., Martin-Fuentes, E., and Daries-Ramon, N. (2017). User-generated social media events in tourism. Sustainability, 9(12):2250.
Maron, M. E. (1961). Automatic indexing: An experimental inquiry. J. ACM, 8(3):404–417.
Martí, P., García-Mayor, C., and Serrano-Estrada, L. (2021). Taking the urban tourist activity pulse through digital footprints. Current Issues in Tourism, 24(2):157–176.
Martínez-López, B., Perez, A., and Sánchez-Vizcaíno, J. (2009). Combined application of social network and cluster detection analyses for temporal-spatial characterization of animal movements in salamanca, spain. Preventive veterinary medicine, 91(1):29–38.
Martínez-Sala, A.-M., Albeza, R., and Martínez Cano, F. J. (2018). Social networks of tourist destination marketing organizations as potential sources of ewom. Observatorio, 12:246–271.
Mascaro, R., Teixeira, L., and Chli, M. (2021). Diffuser: Multi-view 2d-to-3d label diffusion for semantic scene segmentation. In 2021 IEEE International Conference on Robotics and Automation (ICRA), pages 13589–13595. IEEE.
Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., and Remondino, F. (2020). Comparing machine and deep learning methods for large 3d heritage semantic segmentation. ISPRS International Journal of Geo-Information, 9(9):535.
Mazloom, M., Hendriks, B., and Worring, M. (2017). Multimodal context-aware recommender for post popularity prediction in social media. In Proceedings of the on Thematic Workshops of ACM Multimedia 2017, pages 236–244.
McInnes, L., Healy, J., and Astels, S. (2017). hdbscan: Hierarchical density based clustering. J. Open Source Softw., 2(11):205.
McMullen, M. (2020). 'pinning'tourist photographs: Analyzing the photographs shared on pinterest of heritage tourist destinations. Current Issues in Tourism, 23(3):376–387.
Meghraoui, M. and Sbeinati, R. (2023). Archeoseismology and the lost villages in northern syria, the impact of large earthquakes on cultural heritage. In Sustainable Conservation of UNESCO and Other Heritage Sites Through Proactive Geosciences, pages 445–461. Springer.

Mendieta, J., Suárez, S., Vaca, C., Ochoa, D., and Vergara, C. (2016). Geo-localized social media data to improve characterization of international travelers. In 2016 Third International Conference on eDemocracy & eGovernment (ICEDEG), pages 126–132. IEEE.

Miah, S. J., Vu, H. Q., Gammack, J., and McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. Information & Management, 54(6):771–785.

Micera, R. and Crispino, R. (2017). Destination web reputation as “smart tool” for image building: the case analysis of naples city-destination. International Journal of Tourism Cities.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. In Bengio, Y. and LeCun, Y., editors, 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26:3111–3119.

Milgram, S. (1967). The small world problem. Psychology today, 2(1):60–67.

Miller, G. A. (1995). Wordnet: a lexical database for english. Communications of the ACM, 38(11):39–41.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., and Grp, P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). Physical Therapy, 89(9):873–880.

Molina, S. O. and Molina, M. Á. O. (2021). Notre-dame de parís. dov’era e com’era: la réplica que habita en la ruina. Loggia, Arquitectura & Restauración, 34:8–27.

Monteiro, V., Henriques, R., Painho, M., and Vaz, E. (2014). Sensing World Heritage An Exploratory Study of Twitter as a Tool for Assessing Reputation. In Murgante, B and Misra, S and Rocha, AMAC and Torre, C and Rocha, JG and Falcao, MI and Taniar, D and Apduhan, BO and Gervasi, O., editor, COMPUTATIONAL SCIENCE AND ITS APPLICATIONS - ICCSA 2014, PT II, volume 8580 of Lecture Notes in Computer Science, pages 404–419. Univ Minho; Univ Perugia; Univ Basilicata; Monash Univ; Kyushu Sangyo Univ; Assoc Portuguesa Investigacao Operac.

Monti, L., Delnevo, G., Mirri, S., Salomoni, P., and Callegati, F. (2018). Digital invasions within cultural heritage: Social media and crowdsourcing. In Smart Objects and Technologies for Social Good: Third International Conference, GOODTECHS 2017, Pisa, Italy, November 29-30, 2017, Proceedings 3, pages 102–111. Springer.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. Biometrika, 37(1/2):17–23.

Moreno, A., Jabreel, M., and Huertas, A. (2015). Automatic analysis of the communication of tourist destination brands through social networks. In 2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), pages 546–553. IEEE.

Moreno-Sánchez, I., Font-Clos, F., and Corral, Á. (2016). Large-scale analysis of zipf’s law in english texts. PloS one, 11(1):e0147073.

Muangon, W., Muangprathub, J., saelee, J., Soonklang, T., Pongpinigpinyo, S., and Sitdhisanguan, K. (2018). An information retrieval system on thailand tourism community websites. In Proceedings of the 2018 10th International Conference on Information Management and Engineering, pages 101–105.

Müller, R., Kornblith, S., and Hinton, G. (2019). When does label smoothing help? In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Red Hook, NY, USA. Curran Associates Inc.

Munar, A. M. (2012). Social Media Strategies and Destination Management. SCANDINAVIAN JOURNAL OF HOSPITALITY AND TOURISM, 12(2):101–120.

Naramski, M., Szromek, A. R., Herman, K., and Polok, G. (2022). Assessment of the activities of european cultural heritage tourism sites during the covid-19 pandemic. Journal of Open Innovation: Technology, Market, and Complexity, 8(1):55.

Nenko, A. and Petrova, M. (2018). Emotional geography of st. petersburg: detecting emotional perception of the city space. In Digital Transformation and Global Society: Third International Conference, DTGS 2018, St. Petersburg, Russia, May 30–June 2, 2018, Revised Selected Papers, Part II 3, pages 95–110. Springer.

Newman, M. (2010). Networks: An Introduction. Oxford University Press.

Nguyen, D. (2018). Comparing automatic and human evaluation of local explanations for text classification. In Walker, M. A., Ji, H., and Stent, A., editors, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers), pages 1069–1078. Association for Computational Linguistics.

Nguyen, D. Q., Vu, T., and Nguyen, A. T. (2020). Bertweet: A pre-trained language model for english tweets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 9–14.

Nguyen, G. H., Lee, J. B., Rossi, R. A., Ahmed, N. K., Koh, E., and Kim, S. (2018). Continuous-time dynamic network embeddings. In Companion Proceedings of the The Web Conference 2018, pages 969–976.

Nigam, K., Lafferty, J., and McCallum, A. (1999). Using maximum entropy for text classification. In IJCAI-99 workshop on machine learning for information filtering, volume 1, pages 61–67. Stockholom, Sweden.

Nourian, P. (2016). Configraphics: Graph Theoretical Methods for Design and Analysis of Spatial Configurations. TU Delft.

Nourian, P., Rezvani, S., Sariyildiz, I., and van der Hoeven, F. (2016). Spectral modelling for spatial network analysis. In Proceedings of the Symposium on Simulation for Architecture and Urban Design (simAUD 2016), pages 103–110. SimAUD.

Nowak, S. and Rüger, S. (2010). How reliable are annotations via crowdsourcing. In Proceedings of the international conference on Multimedia information retrieval, pages 557–566.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., et al. (2020). Bias in data-driven artificial intelligence systems—an introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3):e1356.

Olsson, K. (2008). Citizen input in urban heritage management and planning: A quantitative approach to citizen participation. Town Planning Review, 79(4):371–395.

Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., and Plieninger, T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five european sites. Ecological Indicators, 94:74–86.

Padilha, R., Andaló, F. A., Lavi, B., Pereira, L. A., and Rocha, A. (2021a). Temporally sorting images from real-world events. Pattern Recognition Letters, 147:212–219.

Padilha, R., Andaló, F. A., Pereira, L. A., and Rocha, A. (2021b). Unraveling the notre-dame cathedral fire in space and time: an x-coherence approach. In Crime Science and Digital Forensics, pages 3–19. CRC Press.

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Pal, A., Selvakumar, M., and Sankarasubbu, M. (2020). MAGNET: multi-label text classification using attention-based graph neural network. In Rocha, A. P., Steels, L., and van den Herik, H. J., editors, Proceedings of the 12th International Conference on Agents and Artificial Intelligence, ICAART 2020, Volume 2, Valletta, Malta, February 22-24, 2020, pages 494–505. SCITEPRESS.

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., and González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. EPJ Data Science, 4(1):1–17.

Pan, J., Mou, N., and Liu, W. (2019). Emotion analysis of tourists based on domain ontology. In Proceedings of the 2019 International Conference on Data Mining and Machine Learning, pages 146–150.

Pan, S. J. and Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359.

Pang, Y., Zhao, Y., and Li, D. (2021). Graph pooling via coarsened graph infomax. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 2177–2181.

Pantano, E. and Dennis, C. (2019). Store buildings as tourist attractions: Mining retail meaning of store building pictures through a machine learning approach. Journal of Retailing and Consumer Services, 51:304–310.

Parasuraman, R. and Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3):381–410.

Park, D., Kim, W. G., and Choi, S. (2019). Application of social media analytics in tourism crisis communication. Current Issues in Tourism, 22(15):1810–1824.

Passaro, L. C., Bondielli, A., Dell'Oglio, P., Lenci, A., and Marcelloni, F. (2022). In-context annotation of topic-oriented datasets of fake news: A case study on the notre-dame fire event. Information Sciences, 615:657–677.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 32.

Patterson, G. and Hays, J. (2012). Sun attribute database: Discovering, annotating, and recognizing scene attributes. In 2012 IEEE Conference on Computer Vision and Pattern Recognition, pages 2751–2758. IEEE.

Patterson, G., Xu, C., Su, H., and Hays, J. (2014). The sun attribute database: Beyond categories for deeper scene understanding. International Journal of Computer Vision, 108(1):59–81.

Pearson, K. (1905). The problem of the random walk. Nature, 72(1865):294–294.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830.

Peng, X. and Huang, Z. (2017). A novel popular tourist attraction discovering approach based on geo-tagged social media big data. ISPRS International Journal of Geo-Information, 6(7):216.

Penn, A. (2003). Space syntax and spatial cognition: or why the axial line? Environment and behavior, 35(1):30–65.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In Moschitti, A., Pang, B., and Daelemans, W., editors, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, pages 1532–1543. ACL.

Pentland, A. (2015). Social Physics: How social networks can make us smarter. Penguin.

Pereira Roders, A. (2007). Re-architecture: lifespan rehabilitation of built heritage. PhD thesis, Technische Universiteit Eindhoven.

Pereira Roders, A. (2010). Revealing the World Heritage cities and their varied natures. In Heritage 2010: Heritage and Sustainable Development, Vols 1 and 2, chapter Heritage a, pages 245–253. Green Lines Institute.

Pereira Roders, A. (2019). The Historic Urban Landscape Approach in Action: Eight Years Later. In Reshaping Urban Conservation, pages 21–54. Springer.

Pereira Roders, A. and van Oers, R. (2011). World Heritage cities management. Facilities, 29(7):276–285.

Pérez, J. M., Furman, D. A., Alonso Alemany, L., and Luque, F. M. (2022). RoBERTuito: a pre-trained language model for social media text in Spanish. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 7235–7243, Marseille, France. European Language Resources Association.

Petzet, M. (2007). What is outstanding universal value. In Values and Criteria in Heritage Conservation, pages 315–322. Polistampa.

Pickering, C., Rossi, S. D., Hernando, A., and Barros, A. (2018). Current knowledge and future research directions for the monitoring and management of visitors in recreational and protected areas. Journal of Outdoor Recreation and Tourism, 21(November 2017):10–18.

Pinto, M. R., Viola, S., Fabbricatti, K., and Pacifico, M. G. (2020). Adaptive reuse process of the historic urban landscape post-covid-19. the potential of the inner areas for a “new normal”. VITRUVIO-International Journal of Architectural Technology and Sustainability, 5(2):87–105.

Pintossi, N., Ikiz Kaya, D., and Pereira Roders, A. (2019). Adaptive reuse of cultural heritage in amsterdam: Identifying challenges and solutions through the historic urban landscape approach. In Proceedings of the LDE Heritage Conference on Heritage and the Sustainable Development Goals: Proceedings, pages 304–314.

Pintossi, N., Kaya, D. I., van Wesemael, P., and Roders, A. P. (2023). Challenges of cultural heritage adaptive reuse: A stakeholders-based comparative study in three european cities. Habitat International, 136:102807.

Platt, J. et al. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers, 10(3):61–74.

Plummer, B. A., Wang, L., Cervantes, C. M., Caicedo, J. C., Hockenmaier, J., and Lazebnik, S. (2015). Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. In Proceedings of the IEEE international conference on computer vision, pages 2641–2649.

Potdar, K. and Verbakel, E. (2022). Eidetic mapping: An exploration for sustainability and resilience of historic urban landscapes. In The Twelfth International Convention of Asia Scholars (ICAS 12), volume 1, pages 547–559. Amsterdam University Press.

Praticò, Y., Ochsendorf, J., Holzer, S., and Flatt, R. J. (2020). Post-fire restoration of historic buildings and implications for notre-dame de paris. Nature Materials, 19(8):817–820.

Prince, M. (2004). Does active learning work? a review of the research. Journal of engineering education, 93(3):223–231.

Psarra, S. (2018). The Venice Variations: Tracing the Architectural Imagination. UCL press.

Pustejovsky, J. and Stubbs, A. (2012). Natural Language Annotation for Machine Learning: A guide to corpus-building for applications. " O'Reilly Media, Inc.".

Pérez, J. M., Giudici, J. C., and Luque, F. (2021). pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks.

QGIS Development Team (2023). QGIS Geographic Information System. Open Source Geospatial Foundation.

Qi, S., Wong, C. U. I., Chen, N., Rong, J., and Du, J. (2018). Profiling macau cultural tourists by using user-generated content from online social media. Information Technology & Tourism, 20:217–236.

Ragin, C. C. and Becker, H. S. (1992). What is a case?: exploring the foundations of social inquiry. Cambridge university press.

Rakic, T. and Chambers, D. (2008). World Heritage: Exploring the Tension Between the National and the ‘Universal’. Journal of Heritage Tourism, 2(3):145–155.

Ramanathan, V. and Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about oman tourism. In 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), pages 1–5. IEEE.

Rani, M. and Kaushal, S. (2022). Geoclust: Feature engineering based framework for location-sensitive disaster event detection using ahp-topsis. Expert Systems with Applications, 210.

Rao, D. and McMahan, B. (2019). Natural Language Processing with PyTorch - Build Intelligent Language Applications Using Deep Learning. O'Reilly Media, Inc.

Ratti, C. (2004). Space syntax: Some inconsistencies. Environment and Planning B: Planning and Design, 31(4):487–499.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.

Reiter, R. (1989). Towards a logical reconstruction of relational database theory. In Readings in Artificial Intelligence and Databases, pages 301–327. Elsevier.

Ren, Y., Cheng, T., and Zhang, Y. (2019). Deep spatio-temporal residual neural networks for road-network-based data modeling. International Journal of Geographical Information Science, 33(9):1894–1912.

Rey, S. J. and Anselin, L. (2007). PySAL: A Python Library of Spatial Analytical Methods. The Review of Regional Studies, 37(1):5–27.
Rish, I. et al. (2001). An empirical study of the naive bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence, volume 3, pages 41–46.
Rochon, T. R. (1998). Culture moves: Ideas, activism, and changing values. Princeton University Press.
Roders, A. P. and Van Oers, R. (2011). World heritage cities management. Facilities, 29(7/8):276–285.
Rodriguez, A. and Laio, A. (2014). Clustering by fast search and find of density peaks. science, 344(6191):1492–1496.
Rogerson, P. and Sun, Y. (2001). Spatial monitoring of geographic patterns: an application to crime analysis. Computers, Environment and Urban Systems, 25(6):539–556.
Rogerson, P. A. (2021). Spatial Statistical Methods for Geography. SAGE Publications Ltd.
Rojas-Padilla, E., Metze, T., and Termeer, K. (2022). Seeing the visual: A literature review on why and how policy scholars would do well to study influential visualizations. Policy Studies Yearbook, 12(1):103–136.
Rosetti, I., Bertrand Cabral, C., Pereira Roders, A., Jacobs, M., and Albuquerque, R. (2022). Heritage and sustainability: Regulating participation. Sustainability, 14(3):1674.
Roy, S. and Goldwasser, D. (2020). Weakly supervised learning of nuanced frames for analyzing polarization in news media. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 7698–7716.
Rubinstein, R. Y. and Kroese, D. P. (2013). The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning. Springer Science & Business Media.
Ruby, D. (2023). Social media users in the world — (2023 demographics). demandsage. (version: 2023-03-20) (access date: 2023-05-27).
Ruffino, P., Permadi, D., Gandino, E., Haron, A., Osello, A., and Wong, C. (2019). Digital technologies for inclusive cultural heritage: The case study of serralunga d'alba castle. In ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, volume 4, pages 141–147.
Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. nature, 323(6088):533–536.
Ruskin, J. (1879). The stones of Venice. Crowell.
Ruskin, J. and Quill, S. (2015). Ruskin's Venice: The Stones Revisited. Lund Humphries.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115:211–252.
Russell, S. J. and Norvig, P. (2010). Artificial Intelligence: A modern approach. Pearson Education, Inc.
Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., and Rockström, J. (2019). Six transformations to achieve the sustainable development goals. Nature sustainability, 2(9):805–814.
Safavian, S. R. and Landgrebe, D. (1990). Topics in inference and decision-making with partial knowledge. Technical report, Purdue University.
Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4):e1249.
Salur, M. U., Aydin, İ., and Alghrsi, S. A. (2019). Smartsenti: A twitter-based sentiment analysis system for the smart tourism in turkey. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), pages 1–5. IEEE.
Sánchez, M. L., Cabrera, A. T., and Del Pulgar, M. L. G. (2020). Guidelines from the heritage field for the integration of landscape and heritage planning: A systematic literature review. Landscape and Urban Planning, 204:103931.
Sansonetti, G., Gasparetti, F., Micarelli, A., Cena, F., and Gena, C. (2019). Enhancing cultural recommendations through social and linked open data. User Modeling and User-Adapted Interaction, 29:121–159.
Schapire, R. E. and Singer, Y. (1998). Improved boosting algorithms using confidence-rated predictions. In Proceedings of the eleventh annual conference on Computational learning theory, pages 80–91.
Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L., and Helbing, D. (2014). A network framework of cultural history. science, 345(6196):558–562.
Schirpke, U., Meisch, C., Marsoner, T., and Tappeiner, U. (2018). Revealing spatial and temporal patterns of outdoor recreation in the european alps and their surroundings. Ecosystem services, 31:336–350.
Schlichtkrull, M., Kipf, T. N., Bloem, P., Berg, R. v. d., Titov, I., and Welling, M. (2018). Modeling relational data with graph convolutional networks. In European semantic web conference, pages 593–607. Springer.
Schroeder, A., Pennington-Gray, L., Donohoe, H., and Kiousis, S. (2013). Using Social Media in Times of Crisis. Journal of Travel and Tourism Marketing, 30(1-2):126–143.
Schroff, F., Kalenichenko, D., and Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 815–823.
Schubert, E., Sander, J., Ester, M., Kriegel, H. P., and Xu, X. (2017). Dbscan revisited, revisited: why and how you should (still) use dbscan. ACM Transactions on Database Systems (TODS), 42(3):1–21.
Schuff, H. (2020). Explainable question answering beyond f1: metrics, models and human evaluation. Master's thesis, Universitaet Stuttgart.

Sebastiani, F. and Esuli, A. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of the 5th international conference on language resources and evaluation, pages 417–422. European Language Resources Association (ELRA) Genoa, Italy.
Serrano, S. and Smith, N. A. (2020). Is attention interpretable? ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, pages 2931–2951.
Settles, B. (2011). From theories to queries: Active learning in practice. In Active learning and experimental design workshop in conjunction with AISTATS 2010, pages 1–18. JMLR Workshop and Conference Proceedings.
Shah, K. (2015). Documentation and cultural heritage inventories case of the historic city of ahmadabad. In ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, volume 2, pages 271–278.
Shen, J. (2018). Profiling and Grouping Space-time Activity Patterns of Urban Individuals. PhD thesis, UCL (University College London).
Shen, J., Qiu, W., Meng, Y., Shang, J., Ren, X., and Han, J. (2021). Taxoclass: Hierarchical multi-label text classification using only class names. In NAACL-HLT 2021.
Shi, Z. and Pun-Cheng, L. S. (2019). Spatiotemporal data clustering: a survey of methods. ISPRS international journal of geo-information, 8(3):112.
Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In Bengio, Y. and LeCun, Y., editors, 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015.
Sofaer, J., Davenport, B., Sørensen, M. L. S., Gallou, E., and Uzzell, D. (2021). Heritage sites, value and wellbeing: learning from the covid-19 pandemic in england. International Journal of Heritage Studies, 27(11):1117–1132.
Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., and Li, C.-L. (2020). Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems, 33:596–608.
Song, S.-G. and Kim, D.-Y. (2016). A pictorial analysis of destination images on pinterest: The case of tokyo, kyoto, and osaka, japan. Journal of Travel & Tourism Marketing, 33(5):687–701.
Spoormans, L., Czischke, D., Pereira Roders, A., and de Jonge, W. (2023). “do i see what you see?”—differentiation of stakeholders in assessing heritage significance of neighbourhood attributes. Land, 12(3):712.
Stevens, T. M., Aarts, N., and Dewulf, A. (2020). Using emotions to frame issues and identities in conflict: farmer movements on social media. Negotiation and Conflict Management Research.
Stival, L., Pinto, A., Andrade, F. d. S. P. d., Santiago, P. R. P., Biermann, H., Torres, R. d. S., and Dias, U. (2023). Using machine learning pipeline to predict entry into the attack zone in football. PloS one, 18(1):e0265372.
Stone, M. (1961). The opinion pool. The Annals of Mathematical Statistics, pages 1339–1342.
Sun, C., Qiu, X., Xu, Y., and Huang, X. (2019). How to Fine-Tune BERT for Text Classification? Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11856 LNAI(2):194–206.
Sun, K., Lin, Z., and Zhu, Z. (2020). Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 5892–5899.
Sun, M., Zhang, F., Duarte, F., and Ratti, C. (2022). Understanding architecture age and style through deep learning. Cities, 128:103787.
Sun, X. and Lu, W. (2020). Understanding Attention for Text Classification. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3418–3428.
Sun, Y., Ma, H., and Chan, E. H. (2017). A model to measure tourist preference toward scenic spots based on social media data: A case of dapeng in china. Sustainability, 10(1):43.
Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H., and Hospedales, T. M. (2018). Learning to compare: Relation network for few-shot learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1199–1208.
Suzuki, M. and Yamamoto, Y. (2020). Analysis of relationship between confirmation bias and web search behavior. In Proceedings of the 22nd International Conference on Information Integration and Web-Based Applications & Services, pages 184–191.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1–9.
Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2818–2826.
Taecharungroj, V. and Mathayomchan, B. (2019). Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tourism Management, 75(July):550–568.
Tai, K. S., Socher, R., and Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of

the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers, pages 1556–1566. The Association for Computer Linguistics.

Tang, J. and Li, J. (2016). Spatial network of urban tourist flow in xi'an based on microblog big data. Journal of China Tourism Research, 12(1):5–23.

Tang, L. and Liu, H. (2009). Relational learning via latent social dimensions. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 817–826.

Tang, M., Gandhi, P., Kabir, M. A., Zou, C., Blakey, J., and Luo, X. (2019). Progress notes classification and keyword extraction using attention based deep learning models with BERT. arXiv.

Tao, Y., Zhang, F., Shi, C., and Chen, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. Sustainability, 11(18):5070.

Tarrafa Silva, A. and Pereira Roders, A. (2010). The cultural significance of World Heritage cities : Portugal as case study. In Heritage and Sustainable Development, pages 255–263, Évora, Portugal.

Tarrafa Silva, A. and Pereira Roders, A. (2012). Cultural Heritage Management and Heritage (Impact) Assessments. In Joint CIB W070, W092 & TG72 International Conferenceon Facilities Management, Procurement Systems and Public Private Partnership.

Tarrafa Silva, A., Pereira Roders, A., Cunha Ferreira, T., and Nevzgodin, I. (2023). Critical analysis of policy integration degrees between heritage conservation and spatial planning in amsterdam and ballarat. Land, 12(5):1040.

Taubenböck, H., Esch, T., Felbier, A., Wiesner, M., Roth, A., and Dech, S. (2012). Monitoring urbanization in mega cities from space. Remote sensing of Environment, 117:162–176.

Taylor, J. and Gibson, L. K. (2017). Digitisation, digital interaction and social media: embedded barriers to democratic heritage. International Journal of Heritage Studies, 23(5):408–420.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., and Toivonen, T. (2017). Instagram, flickr, or twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Scientific reports, 7(1):1–11.

Tenzer, M. (2022). Tweets in the peak: Twitter analysis-the impact of covid-19 on cultural landscapes. Internet Archaeology, 59.

Thakuriah, P. V., Sila-Nowicka, K., Hong, J., Boididou, C., Osborne, M., Lido, C., and McHugh, A. (2020). Integrated multimedia city data (imcd): a composite survey and sensing approach to understanding urban living and mobility. Computers, Environment and Urban Systems, 80:101427.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American society for information science and technology, 61(12):2544–2558.

Thorp, H. H. (2023). Chatgpt is fun, but not an author.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. Economic geography, 46(sup1):234–240.

Tsoumakas, G. and Katakis, I. (2007). Multi-label classification: An overview. International Journal of Data Warehousing and Mining (IJDWM), 3(3):1–13.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., and Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018).

Turner, A. (2007). From axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis. Environment and Planning B: planning and Design, 34(3):539–555.

Turner, M. (2008). Values of heritage in great religious and cultural areas: From existentialism to historicism-a view of the holy land and the sites of jesus and the apostles. In Values of Heritage in Great Religious and Cultural Areas, pages 1000–1007. Polistampa.

UNESCO (1972). Convention Concerning the Protection of the World Cultural and Natural Heritage. Technical Report november, UNESCO, Paris.

UNESCO (2008). Operational guidelines for the implementation of the world heritage convention. Technical Report July, UNESCO World Heritage Centre.

UNESCO (2011). Recommendation on the historic urban landscape. Technical report, UNESCO, Paris.

UNESCO (2020). Heritage in Urban Contexts: Impact of Development Projects on World Heritage properties in Cities. Technical Report January, UNESCO World Heritage Centre.

Urry, J. and Larsen, J. (2011). The tourist gaze 3.0. Sage.

Valese, M., Noardo, F., and Pereira Roders, A. (2020). World heritage mapping in a standard-based structured geographical information system. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B4-2020:81–88.

Vallat, R. (2018). Pingouin: statistics in python. J. Open Source Softw., 3(31):1026.

van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. Journal of machine learning research, 9(11).

van der Zee, E. and Bertocchi, D. (2018). Finding patterns in urban tourist behaviour: A social network analysis approach based on tripadvisor reviews. Information Technology & Tourism, 20(1-4):153–180.

van Dijck, J. (2011). Flickr and the culture of connectivity: Sharing views, experiences, memories. Memory Studies, 4(4):401–415.

van Eck, C. W., Mulder, B. C., and Dewulf, A. (2020). Online climate change polarization: Interactional framing analysis of climate change blog comments. Science Communication, 42(4):454–480.

van Eck, N. J. and Waltman, L. (2014). Visualizing bibliometric networks. Measuring scholarly impact: Methods and practice, pages 285–320.
van Weerdenburg, D., Scheider, S., Adams, B., Spierings, B., and van der Zee, E. (2019). Where to go and what to do: Extracting leisure activity potentials from web data on urban space. Computers, Environment and Urban Systems, 73:143–156.
VanderWeele, T. J. and Mathur, M. B. (2019). Some desirable properties of the bonferroni correction: is the bonferroni correction really so bad? American journal of epidemiology, 188(3):617–618.
Varnajot, A. (2019). Digital rovaniemi: contemporary and future arctic tourist experiences. Journal of Tourism Futures, 6(1):6–23.
Vassakis, K., Petrakis, E., Kopanakis, I., Makridis, J., and Mastorakis, G. (2019). Location-based social network data for tourism destinations. Springer Singapore.
Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A., Gouws, S., Jones, L., Kaiser, Ł., Kalchbrenner, N., Parmar, N., et al. (2018). Tensor2tensor for neural machine translation. In Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track), pages 193–199.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, pages 6000–6010.
Veldpaus, L. (2015). Historic urban landscapes: framing the integration of urban and heritage planning in multilevel governance. PhD thesis, Technische Universiteit Eindhoven.
Veldpaus, L. and Roders, A. P. (2014). Learning from a legacy: Venice to valletta. Change over time, 4(2):244–263.
Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2017). Graph attention networks. arXiv preprint arXiv:1710.10903.
Verbruggen, R., Pereira Roders, A., Stash, N., Leony, D., and De Bra, P. (2014). Protected urban planet: monitoring the evolution of protected urban areas worldwide. In Special Session'Real Spaces and Cyber Spaces: New Challenges in Regional Science'within ERSA 54th congress Regional Development & Globalisation: Best practices, Saint Petersburg, Russia.
Vig, J. (2019). A multiscale visualization of attention in the transformer model. In Costa-jussà, M. R. and Alfonseca, E., editors, Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 3: System Demonstrations, pages 37–42. Association for Computational Linguistics.
Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., and Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the sustainable development goals. Nature communications, 11(1):233.
von Droste, B. (2011). The concept of outstanding universal value and its application: "From the seven wonders of the ancient world to the 1,000 world heritage places today". Journal of Cultural Heritage Management and Sustainable Development, 1(1):26–41.
Wagner, J., Arora, P., Vaíllo, S. C., Barman, U., Bogdanova, D., Foster, J., and Tounsi, L. (2014). Dcu: Aspect-based polarity classification for semeval task 4. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 223–229.
Wallach, H. M. (2006). Topic modeling: beyond bag-of-words. In Proceedings of the 23rd international conference on Machine learning, pages 977–984.
Wang, D., Szymanski, B. K., Abdelzaher, T., Ji, H., and Kaplan, L. (2019). The age of social sensing. Computer, 52(1):36–45.
Wang, H. and Leskovec, J. (2021). Combining graph convolutional neural networks and label propagation. ACM Transactions on Information Systems (TOIS), 40(4):1–27.
Wang, L., Han, X., He, J., and Jung, T. (2022a). Measuring residents' perceptions of city streets to inform better street planning through deep learning and space syntax. ISPRS Journal of Photogrammetry and Remote Sensing, 190:215–230.
Wang, P., Luo, H., Obaidat, M. S., and Wu, T.-Y. (2018). The internet of things service recommendation based on tripartite graph with mass diffusion. In 2018 IEEE International Conference on Communications Workshops (ICC Workshops), pages 1–6. IEEE.
Wang, T.-C. and Phoa, F. K. H. (2016). A scanning method for detecting clustering pattern of both attribute and structure in social networks. Physica A: Statistical Mechanics and its Applications, 445:295–309.
Wang, Y., Jin, W., and Derr, T. (2022b). Graph neural networks: Self-supervised learning. In Wu, L., Cui, P., Pei, J., and Zhao, L., editors, Graph Neural Networks: Foundations, Frontiers, and Applications, pages 391–420. Springer Singapore, Singapore.
Wang, Y., Yao, Q., Kwok, J. T., and Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM computing surveys (csur), 53(3):1–34.
Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.
Waterton, E., Smith, L., and Campbell, G. (2006). The utility of discourse analysis to heritage studies: The burra charter and social inclusion. International journal of heritage studies, 12(4):339–355.
Watkins, J. (2007). Social media, participatory design and cultural engagement. In Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces, pages 161–166.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. nature, 393(6684):440–442.
Wilkinson, A. (2019). The historic urban landscape approach in edinburgh's old and new towns: Implementation of projects on the ground in a living capital city. Reshaping Urban Conservation: The Historic Urban Landscape Approach in Action, pages 223–233.
Williams, N. L., Inversini, A., Ferdinand, N., and Buhalis, D. (2017). Destination eWOM: A macro and meso network approach? Annals of Tourism Research, 64:87–101.
Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In Liu, Q. and Schlangen, D., editors, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020 - Demos, Online, November 16-20, 2020, pages 38–45.
Won, M., Murrieta-Flores, P., and Martins, B. (2018). ensemble named entity recognition (ner): evaluating ner tools in the identification of place names in historical corpora. Frontiers in Digital Humanities, 5:2.
Wooldridge, J. M. (2013). Introductory Econometrics: A Modern Approach. South-Western Cengage Learning, 5 edition.
Wu, L., Cui, P., Pei, J., and Zhao, L. (2022). Graph Neural Networks: Foundations, Frontiers, and Applications. Springer Singapore, Singapore.
Wu, X., Huang, Z., Peng, X., Chen, Y., and Liu, Y. (2018). Building a spatially-embedded network of tourism hotspots from geotagged social media data. IEEE Access, 6:21945–21955.
Xie, G., Li, J., Gu, G., Sun, Y., Lin, Z., Zhu, Y., and Wang, W. (2021). Bgmsdda: a bipartite graph diffusion algorithm with multiple similarity integration for drug–disease association prediction. Molecular Omics, 17(6):997–1011.
Xu, Y., Zhou, B., Jin, S., Xie, X., Chen, Z., Hu, S., and He, N. (2022). A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. Computers, Environment and Urban Systems, 95:101807.
Yanenko, O. and der Weberei, A. (2019). Introducing social distance to st-dbscan. In Proceedings of the 22nd AGILE Conference.
Yang, Z., Cohen, W., and Salakhudinov, R. (2016a). Revisiting semi-supervised learning with graph embeddings. In International conference on machine learning, pages 40–48. PMLR.
Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. J., and Hovy, E. H. (2016b). Hierarchical attention networks for document classification. In Knight, K., Nenkova, A., and Rambow, O., editors, NAACL HLT 2016, pages 1480–1489.
Yarza Pérez, A. J. and Verbakel, E. (2022). The role of adaptive reuse in historic urban landscapes towards cities of inclusion. the case of acre. Journal of Cultural Heritage Management and Sustainable Development.
Ying, Z., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. (2019). Gnnexplainer: Generating explanations for graph neural networks. Advances in neural information processing systems, 32.
Yuan, W., Yang, L., Yang, Q., Sheng, Y., and Wang, Z. (2022). Extracting spatio-temporal information from chinese archaeological site text. ISPRS International Journal of Geo-Information, 11(3):175.
Yuster, R. and Zwick, U. (2005). Fast sparse matrix multiplication. ACM Transactions On Algorithms (TALG), 1(1):2–13.
Zagato, L. et al. (2015). The notion of "heritage community" in the council of europe's faro convention. its impact on the european legal framework. Adell Nicolas, pages 141–168.
Zancheti, S. M. and Jokilehto, J. (1997). Values and urban conservation planning: some reflections on principles and definitions. Journal of architectural conservation, 3(1):37–51.
Zeng, B. and Gerritsen, R. (2014). What do we know about social media in tourism? a review. Tourism management perspectives, 10:27–36.
Zeng, H., Zhou, H., Srivastava, A., Kannan, R., and Prasanna, V. (2019). Graphsaint: Graph sampling based inductive learning method. In International Conference on Learning Representations.
Zhai, X., Luo, Q., and Wang, L. (2020). Why tourists engage in online collective actions in times of crisis: Exploring the role of group relative deprivation. Journal of Destination Marketing and Management, 16(August 2019).
Zhan, J., Gurung, S., and Parsa, S. P. K. (2017). Identification of top-k nodes in large networks using katz centrality. Journal of Big Data, 4(1):1–19.
Zhang, A., Lipton, Z. C., Li, M., and Smola, A. J. (2021). Dive into deep learning. arXiv preprint arXiv:2106.11342.
Zhang, C., Song, D., Huang, C., Swami, A., and Chawla, N. V. (2019a). Heterogeneous graph neural network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 793–803.
Zhang, C.-B., Jiang, P.-T., Hou, Q., Wei, Y., Han, Q., Li, Z., and Cheng, M.-M. (2020a). Delving deep into label smoothing. arXiv preprint arXiv:2011.12562.
Zhang, F., Zhou, B., Ratti, C., and Liu, Y. (2019b). Discovering place-informative scenes and objects using social media photos. Royal Society open science, 6(3):181375.
Zhang, J., Yang, Y., Tian, Q., Zhuo, L., and Liu, X. (2017). Personalized social image recommendation method based on user-image-tag model. IEEE Transactions on Multimedia, 19(11):2439–2449.

Zhang, M., Cui, Z., Neumann, M., and Chen, Y. (2018). An end-to-end deep learning architecture for graph classification. In Thirty-second AAAI conference on artificial intelligence, pages 4438–4445.
Zhang, M.-L. and Zhou, Z.-H. (2007). Ml-knn: A lazy learning approach to multi-label learning. Pattern recognition, 40(7):2038–2048.
Zhang, W. and Gelernter, J. (2014). Geocoding location expressions in twitter messages: A preference learning method. Journal of Spatial Information Science, 9:37–70.
Zhang, X., Hong, M., and Chen, J. (2023). Glasu: A communication-efficient algorithm for federated learning with vertically distributed graph data. arXiv preprint arXiv:2303.09531.
Zhang, Y. and Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events. Computers, Environment and Urban Systems, 79:101403.
Zhang, Y., Li, Y., Zhang, E., and Long, Y. (2022a). Revealing virtual visiting preference: Differentiating virtual and physical space with massive tiktok records in beijing. Cities, 130:103983.
Zhang, Y., Zhang, F., and Chen, N. (2022b). Migratable urban street scene sensing method based on vision language pre-trained model. International Journal of Applied Earth Observation and Geoinformation, 113:102989.
Zhang, Z., Cui, P., and Zhu, W. (2020b). Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering.
Zheng, N. (2023). Coming to grips with diverse voices in participatory heritage initiatives. PhD thesis, Vrije Universiteit Amsterdam.
Zhong, Q., Li, C., Zhang, Y., Sun, H., Yang, S., Xie, D., and Pu, S. (2016). Towards good practices for recognition & detection. In CVPR workshops, volume 1.
Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., and Torralba, A. (2017). Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6):1452–1464.
Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. (2014). Learning deep features for scene recognition using places database. Advances in neural information processing systems, 27.
Zhou, Y. and Long, Y. (2016). Sinogrids: a practice for open urban data in china. Cartography and Geographic Information Science, 43(5):379–392.
Zhou, Z.-H. (2012). Ensemble methods: foundations and algorithms. CRC press.
Zhou, Z.-H. (2021). Machine learning. Springer Nature.
Zhou, Z.-H. and Li, M. (2010). Semi-supervised learning by disagreement. Knowledge and Information Systems, 24(3):415–439.
Zhu, X. and Ghahramani, Z. (2002). Learning from labeled and unlabeled data with label propagation. Tech. Rep., Technical Report CMU-CALD-02–107, Carnegie Mellon University.
Zhu, X. and Goldberg, A. B. (2009). Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning, 3(1):1–130.
Zimmermann, A., Lorenz, A., and Oppermann, R. (2007). An operational definition of context. In Modeling and Using Context: 6th International and Interdisciplinary Conference, CONTEXT 2007, Roskilde, Denmark, August 20-24, 2007. Proceedings 6, pages 558–571. Springer.

APPENDIX A

Official Definitions for Cultural Significance of Heritage

UNESCO World Heritage | Outstanding Universal Value

Selection Criteria

According to the Operational Guidelines of UNESCO (UNESCO, 2008), Outstanding Universal Value means

> "cultural and/or natural significance which is so exceptional as to transcend national boundaries and to be of common importance for present and future generations of all humanity. As such, the permanent protection of this heritage is of the highest importance to the international community as a whole."

For any property to be inscribed in the World Heritage List, it must satisfy at least one of the ten Outstanding Universal Value (OUV) selection criteria and meet the conditions of integrity and/or authenticity. However, it is to be stressed that the definition of the selection criteria is regularly revised by the World Heritage Committee to reflect the evolution of World Heritage (WH) itself[1]. For example, cultural (criteria i-vi, also sometimes denoted as C1-C6 in this dissertation) and natural (criteria vii-x, also denoted as N7-N10) OUV used to be justified apart as two sets. Since 2004, the two sets are combined.

Although WH properties are usually justified with OUV from one category (cultural or

[1] http://whc.unesco.org/en/criteria/

natural), within the domain of mix heritage and cultural landscape, OUV from both categories can co-occur in one property (e.g., Mount Tai has all first seven OUV).

Table A.1 gives the original definition of the OUV selection criteria based on UNESCO.

TABLE APP. A.1 The definition for each UNESCO World Heritage OUV selection criterion and its main topic according to UNESCO (2008), Jokilehto (2008), and Bai et al. (2021a).

Criterion	Focus	Definition
(i)	Masterpiece	To represent a masterpiece of human creative genius;
(ii)	Values/Influence	To exhibit an important interchange of human values, over a span of time or within a cultural area of the world, on developments in architecture or technology, monumental arts, town-planning or landscape design;
(iii)	Testimony	To bear a unique or at least exceptional testimony to a cultural tradition or to a civilization which is living or which has disappeared;
(iv)	Typology	To be an outstanding example of a type of building, architectural or technological ensemble or landscape which illustrates (a) significant stage(s) in human history;
(v)	Land-Use	To be an outstanding example of a traditional human settlement, land-use, or sea-use which is representative of a culture (or cultures), or human interaction with the environment especially when it has become vulnerable under the impact of irreversible change;
(vi)	Associations	To be directly or tangibly associated with events or living traditions, with ideas, or with beliefs, with artistic and literary works of outstanding universal significance;
(vii)	Natural Beauty	To contain superlative natural phenomena or areas of exceptional natural beauty and aesthetic importance;
(viii)	Geological Process	To be outstanding examples representing major stages of earth's history, including the record of life, significant on-going geological processes in the development of landforms, or significant geomorphic or physiographic features;
(ix)	Ecological Process	To be outstanding examples representing significant on-going ecological and biological processes in the evolution and development of terrestrial, fresh water, coastal and marine ecosystems and communities of plants and animals;
(x)	Bio-diversity	To contain the most important and significant natural habitats for in-situ conservation of biological diversity, including those containing threatened species of outstanding universal value from the point of view of science or conservation.

Following are the official description and Statements of Outstanding Universal Value for five UNESCO World Heritage properties that are [partially] selected as case studies in different chapters to demonstrate the methodological framework of this dissertation. The documents are displayed following alphabetic order.

Classical Gardens of Suzhou

Date of Inscription: 1997
Significant modifications to the boundaries: 2000
Criteria: (i)(ii)(iii)(iv)(v)
Property: 11.922 ha
Buffer zone: 26.839 ha
Suzhou City, Jiangsu Province (N31 19 0 E120 27 0)

Classical Chinese garden design, which seeks to recreate natural landscapes in miniature, is nowhere better illustrated than in the nine gardens in the historic city of Suzhou. They are generally acknowledged to be masterpieces of the genre. Dating from the 11th-19th century, the gardens reflect the profound metaphysical importance of natural beauty in Chinese culture in their meticulous design.[2]

Criterion (i): The classical gardens of Suzhou that have been influenced by the traditional Chinese craftsmanship and artistry first introduced by the freehand brushwork of traditional Chinese paintings, embody the refined sophistication of traditional Chinese culture. This embodiment of artistic perfection has won them a reputation as the most creative gardening masterpieces of ancient China.

Criterion (ii): Within a time span of over 2,000 years, a unique but systematic form of landscaping for these particular types of gardens was formed. Its planning, design, construction techniques, as well as artistic effect have had a significant impact on the development of landscaping in China as well as the world.

Criterion (iii): The classical gardens of Suzhou first originated from the ancient Chinese intellectuals' desire to harmonize with nature while cultivating their temperament. They are the finest remnants of the wisdom and tradition of ancient Chinese intellectuals.

Criterion (iv): The classical gardens of Suzhou are the most vivid specimens of the culture expressed in landscape garden design from the East Yangtze Delta region in the 11th to 19th centuries. The underlying philosophy, literature, art, and craftsmanship shown in the architecture, gardening as well as the handcrafts reflect the monumental achievements of the social, cultural, scientific, and technological developments of this period.

Criterion (v): These classical Suzhou gardens are outstanding examples of the harmonious relationship achieved between traditional Chinese residences and artfully contrived nature. They showcase the life style, etiquette and customs of the East Yangtze Delta region during the 11th to 19th centuries.

Historic Centre of Rome, the Properties of the Holy See in that City Enjoying Extraterritorial Rights and San Paolo Fuori le Mura

Date of Inscription: 1980
Significant modifications to the boundaries: 1990
Minor boundary modification inscribed year: 2015
Criteria: (i)(ii)(iii)(iv)(vi)
Property: 1,430.8 ha
Province of Roma, Lazio region (IT) / Vatican City State (VA) (N41 53 24.8 E12 29 32.3)

Founded, according to legend, by Romulus and Remus in 753 BC, Rome was first the centre of the Roman Republic, then of the Roman Empire, and it became the capital of the Christian world in the 4th century. The World Heritage site, extended in 1990 to the walls of Urban VIII, includes some of the major monuments of antiquity such as the Forums, the Mausoleum of Augustus, the Mausoleum of Hadrian, the Pantheon, Trajan's Column and the Column of Marcus Aurelius, as well as the religious and public buildings of papal Rome.[3]

[2] https://whc.unesco.org/en/list/813, available under license CC-BY-SA IGO 3.0
[3] https://whc.unesco.org/en/list/91, available under license CC-BY-SA IGO 3.0

Criterion (i): The property includes a series of testimonies of incomparable artistic value produced over almost three millennia of history: monuments of antiquity (like the Colosseum, the Pantheon, the complex of the Roman and the Imperial Forums), fortifications built over the centuries (like the city walls and Castel Sant'Angelo), urban developments from the Renaissance and Baroque periods up to modern times (like Piazza Navona and the "Trident" marked out by Sixtus V (1585-1590) including Piazza del Popolo and Piazza di Spagna), civil and religious buildings, with sumptuous pictorial, mosaic and sculptural decorations (like the Capitoline Hill and the Farnese and Quirinale Palaces, the Ara Pacis, the Major Basilicas of Saint John Lateran, Saint Mary Major and Saint Paul's Outside the Walls), all created by some of the most renowned artists of all time.

Criterion (ii): Over the centuries, the works of art found in Rome have had a decisive influence on the development of urban planning, architecture, technology and the arts throughout the world. The achievements of ancient Rome in the fields of architecture, painting and sculpture served as a universal model not only in antiquity, but also in the Renaissance, Baroque and Neoclassical periods. The classical buildings and the churches, palaces and squares of Rome have been an unquestioned point of reference, together with the paintings and sculptures that enrich them. In a particular way, it was in Rome that Baroque art was born and then spread throughout Europe and to other continents.

Criterion (iii): The value of the archaeological sites of Rome, the centre of the civilization named after the city itself, is universally recognized. Rome has maintained an extraordinary number of monumental remains of antiquity which have always been visible and are still in excellent state of preservation. They bear unique witness to the various periods of development and styles of art, architecture and urban design, characterizing more than a millennium of history.

Criterion (iv): The historic centre of Rome as a whole, as well as its buildings, testifies to the uninterrupted sequence of three millennia of history. The specific characteristics of the site are the stratification of architectural languages, the wide range of building typologies and original developments in urban planning which are harmoniously integrated in the city's complex morphology. Worthy of mention are significant civil monuments such as the Forums, Baths, city walls and palaces; religious buildings, from the remarkable examples of the early Christian basilicas of Saint Mary Major, St John Lateran and St Paul's Outside the Walls to the Baroque churches; the water systems (drainage, aqueducts, the Renaissance and Baroque fountains, and the 19th-century flood walls along the Tiber). This evidently complex diversity of styles merges to make a unique ensemble, which continues to evolve in time.

Criterion (vi): For more than two thousand years, Rome has been both a secular and religious capital. As the centre of the Roman Empire which extended its power throughout the then known world, the city was the heart of a widespread civilization that found its highest expression in law, language and literature, and remains the basis of Western culture. Rome has also been directly associated with the history of the Christian faith since its origins. The Eternal City was for centuries, and remains today, a symbol and one of the most venerable goals of pilgrimages, thanks to the Tombs of Apostles, the Saints and Martyrs, and to the presence of the Pope.

Paris, Banks of the Seine

Date of Inscription: 1991
Criteria: (i)(ii)(iv)
Property: 365 ha
Ile de France (N48 51 55.8 E2 19 16.1)

From the Louvre to the Eiffel Tower, from the Place de la Concorde to the Grand and Petit Palais, the evolution of Paris and its history can be seen from the River Seine. The Cathedral of Notre-Dame and the Sainte Chapelle are architectural masterpieces while Haussmann's wide squares and boulevards influenced late 19th- and 20th-century town planning the world over.[4].

Criterion (i): The banks of the Seine are studded with a succession of architectural and urban masterpieces built from the Middle Ages to the 20th century, including the Cathedral of Notre-Dame and the Sainte Chapelle, the Louvre, the Palais de l'Institut, the Hôtel des Invalides, Place de la Concorde, Ecole Militaire, the Monnaie (the Mint), the Grand Palais of the Champs Elysées, the Eiffel Tower and the Palais de Chaillot.

Criterion (ii): Buildings along the Seine, such as Notre-Dame and the Sainte Chapelle, became the source of the spread of Gothic architecture, while the Place de la Concorde and the vista at the Invalides exerted influence on urban development of European capitals. Haussmann's urban planning, which marks the western part of the city, inspired the construction of the great cities of the New World, in particular in Latin America. Finally, the Eiffel Tower and the Grand and Petit Palais, the Pont Alexandre III and the Palais de Chaillot are the living testimony of the universal exhibitions, which were of such great importance in the 19th and 20th centuries.

Criterion (iv): United by a grandiose river landscape, the monuments, the architecture and the representative buildings along the banks of the Seine in Paris each illustrate with perfection, most of the styles, decorative arts and building methods employed over nearly eight centuries.

Seventeenth-Century Canal Ring Area of Amsterdam inside the Singelgracht

Date of Inscription: 2010
Criteria: (i)(ii)(iv)
Property: 198.2 ha
Noord Holland (N52 21 54 E4 53 16)

The historic urban ensemble of the canal district of Amsterdam was a project for a new 'port city' built at the end of the 16th and beginning of the 17th centuries. It comprises a network of canals to the west and south of the historic old town and the medieval port that encircled the old town and was accompanied by the repositioning inland of the city's fortified boundaries, the Singelgracht. This was a long-term programme that involved extending the city by draining the swampland, using a system of canals in concentric arcs and filling in the intermediate spaces. These spaces allowed the development of a homogeneous urban ensemble including gabled houses and numerous monuments. This urban extension was the largest and most homogeneous of its time. It was a model of large-scale town planning, and served as a reference throughout the world until the 19th century.[5]

Criterion (i): The Amsterdam Canal District is the design at the end of the 16th century and the construction in the 17th century of a new and entirely artificial 'port city.' It is a masterpiece of hydraulic engineering, town planning, and a rational programme of construction and bourgeois architecture. It is a unique and innovative, large-scale but homogeneous urban ensemble.

Criterion (ii): The Amsterdam Canal District bears witness to an exchange of considerable

[4] https://whc.unesco.org/en/list/600, available under license CC-BY-SA IGO 3.0
[5] https://whc.unesco.org/en/list/1349, available under license CC-BY-SA IGO 3.0

influences over almost two centuries, in terms not only of civil engineering, town planning, and architecture, but also of a series of technical, maritime, and cultural fields. In the 17th century Amsterdam was a crucial centre for international commercial trade and intellectual exchange, for the formation and the dissemination of humanist thought; it was the capital of the world-economy in its day.

Criterion (iv): The Amsterdam Canal District represents an outstanding example of a built urban ensemble that required and illustrates expertise in hydraulics, civil engineering, town planning, construction and architectural knowhow. In the 17th century, it established the model for the entirely artificial 'port city' as well as the type of Dutch single dwelling with its variety of façades and gables. The city is testimony, at the highest level, to a significant period in the history of the modern world.

Venice and Its Lagoon

Date of Inscription: 1987
Criteria: (i)(ii)(iii)(iv)(v)(vi)
Property: 70,176.4 ha
Province of Venezia, Veneto Region (N45 26 3.5 E12 20 20.2)

Founded in the 5th century and spread over 118 small islands, Venice became a major maritime power in the 10th century. The whole city is an extraordinary architectural masterpiece in which even the smallest building contains works by some of the world's greatest artists such as Giorgione, Titian, Tintoretto, Veronese and others.[6]

Criterion (i): Venice is a unique artistic achievement. The city is built on 118 small islands and seems to float on the waters of the lagoon, composing an unforgettable landscape whose imponderable beauty inspired Canaletto, Guardi, Turner and many other painters. The lagoon of Venice also has one of the highest concentrations of masterpieces in the world: from Torcello's Cathedral to the church of Santa Maria della Salute.The years of the Republic's extraordinary Golden Age are represented by monuments of incomparable beauty: San Marco, Palazzo Ducale, San Zanipolo, Scuola di San Marco, Frari and Scuola di San Rocco, San Giorgio Maggiore, etc.

Criterion (ii): The influence of Venice on the development of architecture and monumental arts is considerable; first through the Serenissima's fondachi or trading stations, along the Dalmatian coast, in Asia Minor and in Egypt, in the islands of the Ionian Sea, the Peloponnesus, Crete, and Cyprus, where the monuments were clearly built following Venetian models. But when it began to lose its power over the seas, Venice exerted its influence in a very different manner, thanks to its great painters. Bellini and Giorgione, then Tiziano, Tintoretto, Veronese and Tiepolo completely changed the perception of space, light and colour thus leaving a decisive mark on the development of painting and decorative arts in the whole of Europe.

Criterion (iii): With the unusualness of an archaeological site which still breathes life, Venice bears testimony unto itself. This mistress of the seas is a link between the East and the West, between Islam and Christianity and lives on through thousands of monuments and vestiges of a time gone by.

Criterion (iv): Venice possesses an incomparable series of architectural ensembles illustrating the hight of the Republic's splendour. From great monuments such as Piazza San Marco and

[6] https://whc.unesco.org/en/list/394, available under license CC-BY-SA IGO 3.0

Piazzetta (the cathedral, Palazzo Ducale, Marciana, Museo Correr Procuratie Vecchie), to the more modest residences in the calli and campi of its six quarters (Sestieri), including the 13th century Scuole hospitals and charitable or cooperative institutions, Venice presents a complete typology of medieval architecture, whose exemplary value goes hand-in-hand with the outstanding character of an urban setting which had to adapt to the special requirements of the site.

Criterion (v): In the Mediterranean area, the lagoon of Venice represents an outstanding example of a semi-lacustral habitat which has become vulnerable as a result of irreversible natural and climate changes. In this coherent ecosystem where the muddy shelves (alternately above and below water level) are as important as the islands, pile-dwellings, fishing villages and rice-fields need to be protected no less than the palazzi and churches.

Criterion (vi): Venice symbolizes the people’s victorious struggle against the elements as they managed to master a hostile nature. The city is also directly and tangibly associated with the history of humankind. The "Queen of the Seas”, heroically perched on her tiny islands, extended her horizon well beyond the lagoon, the Adriatic and the Mediterranean. It was from Venice that Marco Polo (1254-1324) set out in search of China, Annam, Tonkin, Sumatra, India and Persia. His tomb at San Lorenzo recalls the role of Venetian merchants in the discovery of the world - after the Arabs, but well before the Portuguese.

Historic Urban Landscape | Heritage Values and Attributes

Six Steps of HUL Approach

As proposed in UNESCO (2011) and further introduced by Pereira Roders (2019), six main steps are identified and proposed for the HUL Approach:

1. To undertake comprehensive surveys and mapping of the city's natural, cultural and human resources;
2. To reach consensus using participatory planning and stakeholder consultations on what values to protect for transmission to future generations and to determine the attributes that carry these values;
3. To assess vulnerability of these attributes to socio-economic stresses and impacts of climate change;
4. To integrate urban heritage values and their vulnerability status into a wider framework of city development, which shall provide indications of areas of heritage sensitivity that require careful attention to planning, design and implementation of development projects;
5. To prioritize actions for conservation and development;
6. To establish the appropriate partnerships and local management frameworks for each of the identified projects for conservation and development, as well as to develop mechanisms for the coordination of the various activities between different actors, both public and private.

A selection of the steps was discussed in Section 1.1.3 as the relevance of this dissertation for inclusive heritage management.

Heritage Values

Table A.2 gives the definition of the categories of heritage values according to previous scholars (Pereira Roders, 2007; Tarrafa Silva and Pereira Roders, 2012).

TABLE APP. A.2 The definition for heritage value category not directly applied in this dissertation.

Value	Sub-Type	Definition
SOCIAL	Spiritual	Beliefs, myths, religions (organized or not), legends, stories, testimonials of past generations
	Emotional (individual)	Memory and personal life experiences
	Emotional (collective)	Notions related with cultural identity, motivation and pride, sense of "place attachment" and communal value
	Allegorical	Objects/places representative of some social hierarchy/status
ECONOMIC	Use	The function and utility of the asset, original or attributed
	Non-use	The asset's expired function, which has its value in the past, and should be remained by its existence (of materials), option (to make use of it) and bequest value
	Entertainment	The role that it might be/have for the contemporaneous market, mainly for the tourism industry
	Allegorical	Oriented to publicizing financial property
POLITICAL	Educational	The education role that heritage assets may play, using it for political targets (e.g. Birth-nations myths, glorification of political leaders, etc.)
	Management	Made part of strategies and policies (past or present)
	Entertainment	It is part of strategies for dissemination of cultural awareness, explored for political targets
	Symbolic	Emblematic, power, authority and prosperous perceptions stem from the heritage asset
HISTORIC	Educational	Heritage asset as a potential to gain knowledge about the past in the future through
	Historic-Artistic	Quality of an object to be part of a few or unique testimonial of historic stylistic or artistic movements, which are now part of the history
	Historic-Conceptual	Quality of an object to be part of a few or unique testimonial that retains conceptual signs (architectural, urban planning, etc.), which are now part of history
	Symbolic	Fact that the object has been part/related with an important event in the past
	Archaeological	Connected with ancient civilizations
AESTHETICAL	Artistic	Original product of creativity and imagination
	Notable	Product of a creator, holding his signature
	Conceptual	Integral materialization of conceptual intentions (imply a conceptual background)
	Evidential	Authentic exemplar of a decade, part of the history of art or architecture
SCIENTIFIC	Workmanship	Original result of human labour, craftsmanship
	Technological	Skillfulness on techniques and materials, representing an outstanding quality of work
	Conceptual	Integral materialization of conceptual intentions (imply a conceptual background)
AGE	Workmanship	Craftsmanship value oriented towards the production period
	Existential	Piece of memory, reflecting the passage/lives of past generations
	Maturity	Marks of the time passage (patina) presents on the forms, components and materials
ECOLOGICAL	Spiritual	Harmony between the building and its environment (natural and artificial)
	Essential	Identification of ecological ideologies on its design and construction
	Existential	Manufactured resources which can either be reused, reprocessed or recycled

Heritage Attributes

Table A.3 gives the definition of the categories of heritage attributes mainly in urban settings according to previous scholars (Veldpaus, 2015; Gustcoven, 2016; Ginzarly et al., 2019).

TABLE APP. A.3 The definition for depicted scenery as heritage attribute category in this dissertation and its tangible/intangible type.

Attribute	Type	Definition
Monuments and Buildings	Tangible	The exterior of a whole building, structure, construction, edifice, or remains that host(ed) human activities, storage, shelter or other purpose;
Building Elements	Tangible	Specific elements, details, or parts of a building, which can be constructive, constitutive, or decorative;
Urban Form Elements	Tangible	Elements, parts, components, or aspects of/in the urban landscape, which can be a construction, structure, or space, being constructive, constitutive, or decorative;
Urban Scenery	Tangible	A district, a group of buildings, or specific urban ensemble or configuration in a wider (urban) landscape or a specific combination of cultural and/or natural elements;
Natural Features and Landscape Scenery	Tangible	Specific flora and/or fauna, such as water elements of/in the historic urban landscape produced by nature, which can be natural and/or designed;
Interior Scenery	Tangible/ Intangible	The interior space, structure, construction, or decoration that host(ed) human activity, showing a specific (typical, common, special) use or function of an interior place or environment;
People's Activity and Association	Intangible	Human associations with a place, element, location, or environment, which can be shown with the activities therein;
Gastronomy	Intangible	The (local) food-related practices, traditions, knowledge, or customs of a community or group, which may be associated with a community or society and/or their cultural identity or diversity;
Artifact Products	Intangible	The (local) artifact-related practices, traditions, knowledge, or customs of a community or group, which may be associated with a community or society and/or their cultural identity or diversity.

APPENDIX B
Supplementary Materials for Chapters

Supplementary Materials Chapter 2

The Coding Scheme for Systematic Literature Review

The following hierarchical scheme is used to code all the records included in the systematic literature review. The binary variables about whether or not a record fulfills a standard will be noted with a “(B)” after the parameter name. The scheme has been gradually formulated during the reviewing process, in a manner similar to grounded theory.

Research Metadata
Title, Journal, Keywords, Abstract, Tags, Research Area, Institutions, Language of the Record.

Research Context
- Geographical Distribution:

  The City of Research Institution, The City of Case Study, Continent of Case Study.
- Case Study Category:

  Explicitly Declaring Heritage (B), Explicitly Stating the Case Study (B), Study Level (Global, Local), Heritage Type (Cultural, Natural, Mixed), Case Count (Single, Two, Multiple), Case Name(s), UNESCO World Heritage ID.
- Data Collection:

  Social Media Platform, Data Gathering Method, Extraction Phase, Duration, Data Quantity, External Database being used other than Social Media.
- Type of Data:

Capture (B), Geo-locations (B), Interactions such as Retweet, Like, Mentions (B), Picture (B), Ratings (B), Tags (B), Timestamps (B), User-Information (B), Video (B).

- Aspect or Type of User-Generated Content

  Context (B), Content (B), Structure (B).

- Object being Studied:

  Building (B), Exhibition (B), Hotel (B), Product (B), Restaurant (B), Transportation (B), Urban Environment (B).

Research Content

- Research Scenarios:

  Everyday Baseline Scenario (B), Event-Triggered Activated Scenario (B).

- Main Objectives:

  Creating New Platforms (B), Describing Property (B), Explaining Mechanism (B), Exploring Usage of Social Media (B), Giving Recommendations (B), Predicting Progresses (B), Proposing Algorithms (B), Proposing Workflow (B), Simulating Dynamics (B), Suggesting Policies and Designs (B).

- Focus Group:

  Local Residents (B), Touring Visitors (B), Managerial Officials (B), Managers as Suppliers (B), Government (B), Humans as Demanders (B), Property as Destinations (B), Algorithms as Technology (B), Explicitly Stressing the Difference between Groups (B), and if the record explicitly excludes Data of Locals (B).

- Analytical Approach:

  Computational (B), Graph Theory (B), Mathematical (B), Qualitative (B), Spatial Analysis (B), and Statistics (B).

Research Methodology

- Graph Theory / Social Network Analysis:

  Tools used for Processing graph data, Directed or Undirected Graph, Weighted or Unweighted Graph, Random Graph or Scale-free Graph, Mono-partite or bipartite Graph, Meaning of Nodes, and Meaning of Links.

- Usage of Graph Statistics

  Assortivity (B), Betweenness Centrality (B), Clustering Coefficients (B), Components (B), Core-periphery Structure (B), Degrees (B), Degree Distribution (B), Density (B), Diameter (B), Edge Betweenness (B), Efficiency (B), Eigenvector Centrality (B), Reciprocity (B), Spreading Speed in Dynamics (B), Subgroup Cliques and Hubs (B), Weight Distribution (B).

- Natural Language Understanding

  Research Approach (Dictionary-based, Manual, Conventional Machine Learning, Deep Learning, Hybrid), Tools used, Models used, Algorithms used, Pre-processing Methods, Textual Features Used, Aspect-based Detection (B), Association among Textual Concepts (B), Classification of subjective/objective tunes (B), Context Recognition (B), Discussion Pathway Identification (B), Emotion Detection (B), Named Entity Extraction (B), Sentiment Detection (B), Topic Detection (B), and Performance on Tasks.

- Image Recognition

  Research Approach (Manual, Conventional Machine Learning, Deep Learning, Hybrid), Tools used, Library used, Models used, Image Context Detection (B), Object Detection (B), Topic Detection (B), Number of Classes in Classification Tasks, Performance on Tasks.

- General Machine Learning Techniques

  Research Approach (Manual, Conventional Machine Learning, Deep Learning, Semi-automated Workflow), Tools used, Models used, Goals of Machine Learning (Classification, Clustering, Regression), Type of Supervision (Supervised, Unsupervised, Both), Aim of using Machine Learning, Type of Output (Single-label, Multi-label), Performance on Tasks.

- Spatial Mapping and Analysis

  [GIS] Tools/Platforms used, [Python] Library used, Form of Data (Raster, Vector, Heatmap, etc), Level of Aggregation for Spatial Data, Mapping the Count of Data (B), Mapping Content (B), Mapping Sentiments (B), Mapping Temporal Information (B), Conducting Spatial Statistics (B), Comparing Different Maps (B).

Research Presentation

- Using of Abbreviations:

  CES - Cultural Ecosystem Services (B), DMO - Destination Marketing Organizations (B), eWoM - electronic Word of Mouth (B), HUL - Historic Urban Landscape (B), KPI - Key Performance Indicators (B), OUV - Outstanding Universal Value (B), POI - Point of Interest (B), SEA - Strategic Environmental Assessment (B), UGC - User-Generated Content (B), VGI - Volunteered Geographic Information (B).

- Reported Research Outcome Elements in the Paper:

  Algorithm (B), Chord Diagram (B), Data Structure (B), Definition (B), Descriptive Statistics (B), Formulas (B), Interface (B), Machine Learning Metrics (B), Map (B), Networks or Graphs (B), Statistical Tests (B), Table (B), Wordcloud (B), Workflow (B).

Supplementary Materials Chapter 3

Proof of Equivalence of Label Smoothing

Here we will show that the Vanilla Label Smoothing (LS) defined in Equations (3.4) and (3.5) is equivalent to the original LS assigning noise to all classes.

Proof. The LS defined in Szegedy et al. (2016):

$$q'(k) = (1-\epsilon)\boldsymbol{\delta}_{k,y} + \frac{\epsilon}{K} \tag{B.1}$$

could be rewritten as following to fit the context of mathematical notations in this paper:

$$\boldsymbol{y}^{O}_{i,j,k} = (1-\epsilon)\boldsymbol{y}_{i,j,k} + \frac{\epsilon}{K}\mathbf{1}, \tag{B.2}$$

where $\boldsymbol{y}_{i,j,k}$ is a one-hot vector of "ground-truth" label, K is the total number of classes (instead of $\kappa + 1$ in the paper for brevity and generality), ϵ is smoothing parameter as scalar, and $\mathbf{1}$ is a vector of 1s of size $K \times 1$.

On the other hand, the Vanilla LS proposed in this paper could be written as:

$$\boldsymbol{y}^{V}_{i,j,k} = \mathbf{f}(\boldsymbol{y}_{i,j,k} + \alpha\mathbf{1}) = \frac{e^{\boldsymbol{y}_{i,j,k}+\alpha\mathbf{1}} - \mathbf{1}}{e^{(\boldsymbol{y}_{i,j,k}+\alpha\mathbf{1})^T}\mathbf{1} - K}. \tag{B.3}$$

We will show that when

$$\epsilon = \frac{(e^{\alpha}-1)K}{e^{1+\alpha} + (K-1)e^{\alpha} - K}, \tag{B.4}$$

the vectors in Equations (B.2) and (B.3) are the same.

First, it is trivial that both the vectors are with the same shape of $\boldsymbol{y}_{i,j,k}$, i.e., $K \times 1$, and that the sums of all entries in both vectors are 1; e.g., observe that the denominator of the right-hand side of Equation (B.3) is equal to the vectorised summation of the values of the nominator.

Second, we assume, without loss of generality, that the "ground-truth" of the one-hot vector $\boldsymbol{y}_{i,j,k}$ is at its first entry, which means that $\boldsymbol{y}_{i,j,k} = [1, 0, ..., 0]_{K\times 1}$. Then both vectors could be rewritten as:

$$\boldsymbol{y}^{O}_{i,j,k} = \left[1 - \epsilon + \frac{\epsilon}{K}, \frac{\epsilon}{K}, ..., \frac{\epsilon}{K}\right]_{K\times 1}, \tag{B.5}$$

$$\boldsymbol{y}^{V}_{i,j,k} = \left[\frac{e^{1+\alpha}-1}{S}, \frac{e^{\alpha}-1}{S}, ..., \frac{e^{\alpha}-1}{S}\right]_{K\times 1}, \tag{B.6}$$

where $S := e^{1+\alpha} + (K-1)e^{\alpha} - K$.

Substituting Equation (B.4) into the entries in Equation (B.5), the first entry could be rewritten as $1 - \epsilon + \frac{\epsilon}{K} = 1 - \frac{(e^{\alpha}-1)K}{S} + \frac{e^{\alpha}-1}{S} = \frac{S-(e^{\alpha}-1)K+e^{\alpha}-1}{S} = \frac{e^{1+\alpha}+(K-1)e^{\alpha}-K-Ke^{\alpha}+K+e^{\alpha}-1}{S} = \frac{e^{1+\alpha}-1}{S}$. And the other entries could be rewritten as $\frac{\epsilon}{K} = \frac{e^{\alpha}-1}{S}$. Both types of entries are exactly the same as the ones shown in Equation (B.6).

Last, we will show that ϵ has a one-to-one relation with α based on Equation (B.4) when $\alpha \geq 0$. The partial derivative of ϵ with respect to α:

$$\frac{\partial \epsilon}{\partial \alpha} = \frac{Ke^{\alpha}(e-1)}{(e^{1+\alpha}+(K-1)e^{\alpha}-K)^2} > 0 \tag{B.7}$$

is non-negative, suggesting that the function is monotonic. Furthermore, $\epsilon = 0$ when $\alpha = 0$, and $\lim_{\alpha\to+\infty} \epsilon = \lim_{\alpha\to+\infty} \frac{K}{\frac{e^{\alpha}(e-1)}{e^{\alpha}-1}+K} = \frac{K}{e-1+K} > 0$ when $\alpha \to +\infty$, suggesting that it is incremental. This means that a unique $\epsilon \in \left[0, \frac{K}{e-1+K}\right)$ always exists for any non-negative α and vice versa. $\square$

Model Implementation Details

For all baselines, Adam (Kingma and Ba, 2017) is used as the optimizer with L2 regularization. Hyperparameter tuning is conducted as grid-search within a small range for each one being searched (and/or selected according to common experience if not mentioned), based on the top-k accuracy on validation split with an early-stopping criterion of 5 epochs, if not explicitly mentioned below. The models are implemented in PyTorch (Rao and McMahan, 2019) and experiments are performed on NVIDIA Tesla P100 GPU (N-gram, GRU+Attn, BERT) and Intel Core i7-8850H CPU (BoE, ULMFiT), respectively.

N-gram

The N-gram model used the TfidfVectorizer from Scikit-learn Python library to get an embedding vector of all 1-grams and 2-grams in the sample that appeared at least twice in the vocabulary. The embedding vectors are then fed in a 2-layer Multi-layer Perceptron (MLP) to get the model prediction. Hyperparameter tuning is performed on the size of the MLP hidden layer in {50, 100, 150, 200}, batch size in {64, 128, 256}, L2 in {0, 1e-5. 1e-4}, and dropout rate in {0.1, 0.2, 0.5} with 108 configurations. The best configuration applied in later experiments of Label Smoothing (LS) has a hidden dimension of 200, batch size of 128, L2 of 1e-5, learning rate of 2e-4, and dropout rate of 0.5.

BoE

The Bag-of-Embedding (BoE) model used the GloVe-6B-300d vectors[1] as initial embeddings, which are set to be tunable during training. Only words that have a higher frequency than a threshold in the full dataset will be kept, while the others will be transformed to a special `<UNK>` token. The word embeddings of all words in the sentence is averaged before being fed to a 2-layer MLP. Hyperparameter tuning is performed on the size of the MLP hidden layer in {50, 100, 150, 200}, batch size in {64, 128, 256}, and frequency threshold in {1, 3, 5} with 36 configurations. The best model has a hidden dimension of 200, batch size of 64, cut-off frequency of 1, L2 of 1e-5, learning rate of 5e-4, and dropout rate of 0.1.

GRU+Attn

The GRU+Attn model also used the GloVe-6B-300d as embeddings, which are frozen during the training. The embedding sequence is then fed into a GRU network. Word-level attention (Yang et al., 2016b) is applied to compute the sentence vector by a learned word context vector and the last hidden state of the GRU. The sentence vector is fed to a 1-layer feed-forward network for the output of the model. Hyperparameter tuning is performed on the size of the hidden layer in GRU in {64, 128, 256}, whether or not to use bi-directional GRU, batch size in {64, 128, 256}, L2 in {0, 1e-5, 1e-4}, learning rate in {1e-3. 5e-4. 2e-4}, and dropout rate in {0, 0.1, 0.2, 0.5} with 648 configurations. The best model is a uni-dimensional GRU with hidden dimension of 128, batch size of 256, L2 of 1e-5, learning rate of 1e-3, and dropout rate of 0.1.

ULMFiT

The ULMFiT model employs the idea of Universal Language Model Fine-tuning from a general-domain pretrained language model on Wikitext-103 with AWD-LSTM architecture (Howard and Ruder, 2018). A domain-specific language model is then fine-tuned with the full UNESCO WHL dataset including SD using fastai API (Howard and Gugger, 2020). One epoch is trained with a learning rate of 1e-2, with only the last layer unfrozen, reaching a perplexity of 46.71. Then the entire model is unfrozen and further trained for 10 epochs, with a learning rate of 1e-3, obtaining a fine-tuned WH domain-specific language model reaching a 30.78 perplexity. Some examples of the language model at this step are shown here, starting with the given phrases marked in bold:

> ***This site is unique because*** it is the only example of a complex of karst complexes that is clearly recognised as being of outstanding universal value. The island of zanzibar has been inscribed as a world heritage site in <num>. The inscriptions, which bear witness to the civilisation of...
>
> ***This architecture has a special layout***, especially in the form of the body of the building. The planet’s primary feature is the addition of the ideal island, which lies at an elevation of <num>m above the sea floor, and is home to some <num>...

[1] https://nlp.stanford.edu/projects/glove/

The encoder of the fine-tuned language model is loaded in PyTorch followed by a Pooling Linear Classifier[2] for classifier fine-tuning. Gradual unfreezing is applied in a simplified manner to prevent catastrophic forgetting: 1) for the 1st epoch, only the decoder is unfrozen and trained with a learning rate of 2e-2; 2) for the 2nd to 4th epoch, one more layer is unfrozen each time and trained with a learning rate of 1e-2, 1e-3, and 1e-4, respectively; 3) from the 5th epoch onward, the full model is unfrozen and trained with a learning rate of 2e-5. An early-stopping criterion of 3 is applied. No extensive hyperparameter tuning is performed since: 1) tuning ULMFiT is expensive on CPU; 2) the hyperparameter configuration from experience suggested by Howard and Ruder (2018) and Howard and Gugger (2020) already performs reasonably well; 3) the purpose of this study is not necessarily finding the best hyperparameter. The final model uses batch size of 64, L2 of 1e-5, and the default dropout rate for the decoder.

BERT

The BERT model uses the uncased base model using The Transformers library (Wolf et al., 2020). The pooler output processed from the last hidden-state of the `[CLS]` token during pretraining is fed into a 1-layer feed-forward network to fine-tune the classifier (Sun et al., 2019). An early-stopping criterion of 10 is applied. Hyperparameter tuning is performed on the batch size in {16, 24, 48, 64}, L2 in {0, 1e-5, 1e-4}, and dropout rate in {0, 0.1, 0.2} with 36 configurations. The best model uses batch size of 64, L2 of 1e-4, learning rate of 2e-5, and dropout rate of 0.2.

LS Configuration Tuning

A single random seed 1337 is used for hyperparameter tuning. Afterwards, ten random seeds in $\{0, 1, 2, 42, 100, 233, 1024, 1337, 2333, 4399\}$ are used to tune the LS configuration with $\alpha \in \{0, 0.01, 0.05, 0.1, 0.2, 0.5, 1\}$ for all three variants. The best LS configuration is selected based on the sum of the lower bound of 95% confidence interval on both top-1 and top-k accuracy. The best LS configuration is then used to evaluate the model performance on single seed 1337. The total runs on each baseline are, therefore, the sum of the number of hyperparameter configurations and random seeds experiments (which is 210).

Resource and Time

Table B.1 shows some further information on the model performance in terms of training resource utilization, model size, and inference time. Training processes are conducted on CPU or GPU, respectively, while inference is fully conducted with CPU.

It can be noted that the best-performing models ULMFiT and BERT also consume the most resources, in terms of training time and infrastructure usage, and have the largest model sizes. Though most time-consuming during training, ULMFiT takes a remarkably short time for inference on CPU compared to BERT. This suggests that ULMFiT might be an optimal choice for further development and application when

[2] https://fastai1.fast.ai/text.models.html

time is a critical matter.

TABLE APP. B.1 The model performance in terms of resource occupancy and inference time. The inference is conducted on Intel Core i7-8850H CPU. Inference time per Item shows the average time the model uses to make a prediction on one sentence. And Inference time for SD shows the total time the model needs to fully process and predict the independent Short Description (SD) test set.

Performance	N-gram	BoE	GRU+Attn	ULMFiT	BERT
Infrastructure	GPU	CPU	GPU	CPU	GPU $\times$ 4
Training Time per Item (s)	0.34	0.18	0.03	2.53	0.54
Training Time per Epoch (s)	12.69	3.18	1.97	213.61*	46.20
Early-Stopping Criteria	5	5	5	3	10
Training Epochs	32	20	15	7**	10
Trainable Parameters (M)	3.82	1.88	0.18	24.55	109.49
Inference Time per Item (s)	0.0031	0.0007	0.2245	0.0589	0.5542
Inference Time for SD (s)	6.92	1.44	4.44	151.75	1598.06

*1180.20 during language model fine-tuning.
**11 during language model fine-tuning.

Nomenclature

Tables B.2 gives an overview of the mathematical notations used in the Chapter 3.

TABLE APP. B.2 The nomenclature of mathematical notations used in Chapter 3 in alphabetic order.

Symbol	Data Type/Shape	Description
$\boldsymbol{A}$	Matrix of Integers $\boldsymbol{A} := [A_{k,l}]$, $k, l \in [1, \kappa]$	The co-occurrence matrix of all OUV selection criteria in the World Heritage properties $\mathcal{P}$, where the diagonal entries indicate the number of cases each criterion is used alone.
α	Scalar Value	The scalar leveraging the effect of Label Smoothing.
$\boldsymbol{\alpha}$	Vector of Floats $\boldsymbol{\alpha} = [\alpha_t]_{\frac{\kappa(\kappa-1)}{2} \times 1}$, $t \in [0, \frac{\kappa(\kappa-1)}{2})$	The "unrolled" upper triangular entries of the normalized co-occurrence matrix $\boldsymbol{A}$.
$\mathcal{B}_w$	Undirected weighted bipartite graph	The bipartite graph showing the relations of the OUV selection criteria and the vocabulary $\mathcal{V}$.
$\boldsymbol{\beta}^{(m,s)}$	Vector of Floats $\boldsymbol{\beta}^{(m,s)} = [\beta_t^{(m,s)}]_{\frac{\kappa(\kappa-1)}{2} \times 1}$, $t \in [0, \frac{\kappa(\kappa-1)}{2})$	The "unrolled" upper triangular entries of the normalized confusion matrices $\boldsymbol{C}^{(m,s)}$.
$\boldsymbol{\beta}$	Vector of Floats $\boldsymbol{\beta} = [\beta_t]_{\frac{\kappa(\kappa-1)}{2} \times 1}$, $t \in [0, \frac{\kappa(\kappa-1)}{2})$	The aggregated vector of $\boldsymbol{\beta}^{(m,s)}$ using dimensionality reduction algorithms.
$\boldsymbol{C}^{(m,s)}$	Matrices of Floats $\boldsymbol{C}^{(m,s)} = [C_{k,l}^{(m,s)}]_{\kappa \times \kappa}$, $k, l \in [0, \kappa)$, $s \in$ {train, val, test}	The confusion matrices by the model $\mathbf{m}_m$ in the s datasets (train, validation, or test).
$\boldsymbol{e}_k$	Vector of Booleans $\boldsymbol{e}_k \in \{0, 1\}^{\kappa \times 1}$	A one-hot vector with length of κ, where only the k_{th} entry is 1 and all other entries are 0.
ϵ	Scalar value	A small number.
$\mathbf{f}(\boldsymbol{z})$	A function returning a logit vector taking a non-negative float vector as input	The function transforming any non-negative vector to a logit vector that sums up to one, as a variant of softmax function.
$\boldsymbol{f}_k$	Vectors of Floats	The semantic representation of OUV selection criterion k as the average GloVe word embeddings of all words belonging to each set $\mathcal{W}_k$.
$\mathbf{g}(w_n)$	A function returning a Float vector taking a phrase as input	The function to look up the 300-dimensional GloVe embedding vectors of all words in the phrase w_n and take the sum of the vectors.
$\boldsymbol{\gamma}_i$	Vector of Floats $\boldsymbol{\gamma}_i := [\gamma_{i,k}]_{(\kappa+1) \times 1}$	The "parental" label of each World Heritage property marking the selection criteria it fulfils. A noise $\gamma_{i,\kappa+1} = 0.2$ is appended to the end of all Boolean vectors for the additional class "Others".
$\boldsymbol{\gamma}$	Vector of Floats $\boldsymbol{\gamma} = [\gamma_t]_{\frac{\kappa(\kappa-1)}{2} \times 1}$, $t \in [0, \frac{\kappa(\kappa-1)}{2})$	The "unrolled" upper triangular entries of the semantic similarity matrix $\boldsymbol{H}$.
$\mathcal{G}_\alpha, \mathcal{G}_\beta, \mathcal{G}_\gamma$	Undirected weighted unipartite graphs	The graph of the OUV selection criteria whose edge weights are represented respectively with $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$.
H	Scalar value	The statistics of Kruskal-Wallis H tests.
$\boldsymbol{H}$	Matrix of Floats $\boldsymbol{H} = [H_{k,l}]_{\kappa \times \kappa}$, $k, l \in [0, \kappa)$	The matrix showing the pair-wise cosine similarity of OUV selection criteria using the semantic representations $\boldsymbol{f}_k$.
i	Integer indices	The index of World Heritage properties in the set $\mathcal{P}$.
j	Integer indices	The index of sentences describing the criterion k possessed by the World Heritage property p_i.

TABLE APP. B.2 Cont.

Symbol	Data Type/Shape	Description
k, l	Integer indices $k, l \in [1, \kappa+1]$, $\kappa = 10$	The index of the ten OUV selection criteria, where $k = 11$ marks the negative class "Others".
$\lambda_0, \lambda_1, \lambda_2$	Scalar values $\lambda_0, \lambda_1 \geq 1 \in \mathbb{R}^+$, $\lambda_2 \in \mathbb{R}^+$	The parameters used to adjust respectively the weight of better models (λ_0), the weight of phrases with higher rankings (λ_1), and the threshold of weights to enter the final vocabulary (λ_2).
m	Integer indices	The index of Natural Language Processing models $\mathcal{M}$ used for classifying datasets.
$\mathcal{M}$	Set of models $\mathcal{M} = \{\mathfrak{m}_m \mid m = [0, 5)\}$	The Natural Language Processing models used for classifying datasets.
$\boldsymbol{\mu}_k$	Vector of non-negative Floats $\boldsymbol{\mu}_k := [\mu_{l,k}]_{(\kappa+1)\times 1}$	The k_{th} column of the column-normalized version of the co-occurrence matrix $\boldsymbol{A}$.
n	Integer indices	The index of phrases in the vocabulary $\mathcal{V}^{(0)}$.
N_0	Scalar integer	The maximum allowed number of phrases in the final lexicon.
o	Integer indices	The index of sentences in the short description $\boldsymbol{S}_i$.
$\boldsymbol{\omega}$	Vector of positive Floats $\boldsymbol{\omega} := [\omega_m]_{5\times 1} = [1, 1, 1, \lambda_0, \lambda_0]^{\mathsf{T}}$, $\lambda_0 \geq 1 \in \mathbb{R}^+$	The weighting vector to determine the importance of different models in $\mathcal{M}$.
p	Scalar value	The significance of statistical tests.
$\mathcal{P}$	Set of objects $p_i \in \mathcal{P}$	The set of all World Heritage properties (sites).
p_i	An object $p_i \in \mathcal{P}$	One example of World Heritage property.
$\boldsymbol{S}_i$	Array of raw texts	The paragraphs of texts shortly describing the site p_i fulfils.
r	Integer $r \in [1, 50)$	The rankings of phrases predicted by models.
r_p	Scalar value	Pearson correlation coefficients.
ρ	Scalar value	Spearman correlation coefficients.
$s_{i,o}$	The raw texts	The o_{th} sentence of the short description $\boldsymbol{S}_i$.
$\sigma_{\lvert \mathcal{W}'_k \rvert}$	Scalar value	The standard deviation of the sizes of sets $\mathcal{W}'_k$.
t	Integer indices $t \in [0, \frac{\kappa(\kappa-1)}{2})$	The index of the unrolled vectors $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ generated from matrices.
T	Scalar value	The statistics of student T-test.
U	Scalar value	The statistics of Mann-Whitney U tests.
$\boldsymbol{\Upsilon}$	A three-dimensional array of integers $\boldsymbol{\Upsilon} = [\upsilon_{n,k,m}]_{\lvert \mathcal{V}^{(0)} \rvert \times (\kappa+1) \times 5}$	The array recording the ranking r of the n_{th} phrase predicted with the model $\mathfrak{m}_m$ for the OUV selection criterion k.
$\boldsymbol{\Upsilon}'$	A two-dimensional array of integers $\boldsymbol{\Upsilon}' := [\upsilon'_{n,k}]_{\lvert \mathcal{V}_0 \rvert \times (\kappa+1)}$	The overall importance of each phrase w_n for the OUV selection criterion k.
$\upsilon_{n,k,m}$	A integer	The ranking r of the n_{th} phrase predicted with the model $\mathfrak{m}_m$ for the OUV selection criterion k.

TABLE APP. B.2 Cont.

Symbol	Data Type/Shape	Description
$\mathcal{V}^{(0)}$	A set of phrases $\mathcal{V}^{(0)} = \bigcup_{k=0}^{\kappa+1} \bigcup_{m=0}^{5} \{w \mid (w, *) \in \mathcal{W}_k^{(m)}\}$	The initial vocabulary containing all the phrases that entered the top-50 list of keywords by all models in $\mathcal{M}$, $*$ can refer to any ranking r.
$\mathcal{V}^{(1)}$	A set of phrases $\mathcal{V}^{(1)} = \bigcup_{k=0}^{\kappa+1} \{w \mid (w, *) \in \mathcal{W}'_k\}$	The vocabulary after filtering.
$\mathcal{V}$	A set of phrases $\mathcal{V} = \bigcup_{k=0}^{\kappa+1} \{w \mid (w, *) \in \mathcal{W}_k\}$	The final set of vocabulary.
$\mathcal{W}_k^{(m)}$	Ordered sets of tuples $\mathcal{W}_k^{(m)} = \{(w, r)\}$, $\lvert\mathcal{W}_k^{(m)}\rvert = 50, r \in [1, 50]$	The ordered set of the phrases w that belonged to the OUV selection criterion k with their ranking r of confidence predicted with model $\mathfrak{m}_m$.
$\mathcal{W}'_k$	Sets of tuples $\mathcal{W}'_k = \{(w_n, v'_{n,k}) \mid v'_{n,k} \geq \lambda_2\}$	The set of the filtered phrases w_n that belonged to the OUV selection criterion k with their relative importance $v'_{n,k}$.
$\mathcal{W}_k$	Sets of tuples	The final set of the filtered phrases w_n that belonged to the OUV selection criterion k as the OUV-related lexicon.
$\xi_\alpha, \xi_\beta, \xi_\gamma$	Scalar values	The thresholds determining the edge weights to be visualized in the graphs $\mathcal{G}_\alpha, \mathcal{G}_\beta, \mathcal{G}_\gamma$.
$\boldsymbol{X}_i$	Array of raw texts	The paragraphs of texts justifying all OUV selection criteria that the site p_i fulfils.
$x_{i,j,k}$	The raw texts	The j_{th} sentence in the paragraph $\boldsymbol{X}_i$ describing the OUV selection criterion k.
$\boldsymbol{y}_{i,j,k}$	Vector of Booleans $\boldsymbol{y}_{i,j,k} = \left[\boldsymbol{e}_k^\mathsf{T}, 0\right]^\mathsf{T} \in \{0,1\}^{(\kappa+1)\times 1}$	The one-hot "ground-truth" label of a sentence $x_{i,j,k}$ describing the OUV selection criterion k it fulfils.
$\hat{\boldsymbol{y}}_{i,j,k}$	Vector of Floats	The predicted label vector of the sentence $x_{i,j,k}$ as logit vector or probability distribution.
$\tilde{\boldsymbol{y}}_{i,j,k}$	Vector of Floats	The smoothed label of the sentence $x_{i,j,k}$ combining its parental label $\boldsymbol{\gamma}_i$ and ground-truth label $\boldsymbol{y}_{i,j,k}$.
$\boldsymbol{z}$	Vector of non-negative Floats	A generic non-negative vector that has the dimension of d.
$\boldsymbol{\zeta}$	Vector of non-negative Floats $\boldsymbol{\zeta} = [\zeta_r]_{51\times 1} = [0, \lambda_1^2, \ldots, \lambda_1^2, \lambda_1, \ldots, \lambda_1, 1, \ldots, 1]^\mathsf{T}$, $\lambda_1 \geq 1 \in \mathbb{R}^+$	The weighting vector to determine the importance of different rankings r.

Supplementary Materials Chapter 4

Model Implementation Details

A dataset with 902 sample images collected in Tripoli, Lebanon and classified with expert-based annotations presented in Ginzarly et al. (2019) was used to train several ML models to replicate the experts' behaviour on classifying depicted scenery. For each image, a unique class label among the 9 depicted scenes mentioned in Table A.3 was provided. In total, 10% of the images were separated and kept away during training as the test dataset and the remaining 812 images were used to train ML models with Scikit-learn python library (Pedregosa et al., 2011). Among the 812 data points, `train_test_split` method of the library was further used to split out a validation dataset with 203 samples (25%). The 512-dimensional visual representation introduced in Section 4.4.1 was generated from the images as the input of ML models, while the class label was used as categorical output of the multi-class single-label classification task.

For each of the selected ML models, `GridSearchCV` function with 10-fold cross-validation was used to wrap the model with a set of tunable parameters in a small range to be selected, while the average top-1 accuracy was used as the criterion for model selection. All 812 samples were input to the cross-validation to tune the hyper-parameters, after which the trained models with their optimal hyper-parameters were tested on the 203 validation data samples and the unseen test dataset with the remaining 90 samples. For the latter steps, the top-1 accuracy and macro-average F1 scores (harmonic average of the precision and recall scores) of all classes were used as the evaluation metrics. All experiments were conducted using a 12th Gen Intel(R) Core(TM) i7-12700KF CPU.

The implementation details of the models are as follows:

MLP

The model used L2 penalty of 1×10^{-4}, solver of stochastic gradient descent, adaptive learning rate and early stopping with maximum 300 iterations. It was tuned on the initial learning rate in $\{0.05, 0.1, 0.2\}$, and hidden sizes of one layer in $\{32, 64, 128, 256\}$ or two layers in $\{(256, 128), (256, 64), (256, 32), (128, 64), (128, 32)\}$. The best model had two hidden layers of $(256, 128)$ with a learning rate of 0.05.

KNN

The model was tuned on the number of neighbours in range $[3, 11] \subset \mathbb{N}$, and the weights of uniform, Manhattan distance, or Euclidean distance. The best model had six neighbours in Euclidean distance.

GNB

The model did not have a tunable hyper-parameter.

SVM

The model was tuned on the kernel type in $\{$linear, poly, rbf, sigmoid$\}$, regularization parameter C in range $[0.1, 2.0] \subset \mathbb{R}$, kernel coefficient gamma in $\{$scale, auto$\}$, and degree of the polynomial kernel function in range $[2, 4] \subset \mathbb{N}$. The best model used RBF kernel with scaled weights and regularization parameter of 1.8.

RF

The model did not restrict the maximum depth of the trees. It was tuned on the class weight in settings of uniform, balanced, and balanced over sub-samples, and the minimum samples required to split a tree node in $\{2, 7, 12, \ldots, 97\}$. The best model had a balanced class weight and a minimum of 17 samples to split a tree node.

Bagging

The model had 10 base estimators in the ensemble. It was tuned on the base estimator in SVM, Decision Tree, and KNN classifiers, and the proportion of maximum features used to train internal weak classifiers within the range $[0.1, 1.0] \subset \mathbb{R}$. The best model used maximum 50% of all features to fit SVM as internal base estimator.

Voting

The model took the first six aforementioned trained models as inputs in the ensemble to vote for the output and was tuned on the choice of hard (voting on top-1 prediction) and soft (voting on the averaged logits) voting mechanism. The best model used the soft voting mechanism.

Stacking

The model stacked the outputs of the first six aforementioned trained models in the ensemble followed by a final estimator and was tuned on the choice of final estimator among SVM and Logistic Regression. The best model used Logistic Regression as the final estimator.

Nomenclature

Tables B.3 and B.4 give an overview of the mathematical notations and functions used in the Chapter 4.

TABLE APP. B.3 The nomenclature of mathematical notations used in Chapter 4 in alphabetic order. All superscripts of matrices are merely tags, not to be confused with exponents and operations, with the exception of transpose operator $\square^{\mathsf{T}}$.

Symbol	Data Type/Shape	Description
$\boldsymbol{A}$	Matrix of Boolean $\boldsymbol{A} := (\boldsymbol{A}^{\mathsf{TEM}} > 0) \bigvee (\boldsymbol{A}^{\mathsf{SOC}} > 0) \bigvee (\boldsymbol{A}^{\mathsf{SPA}} > 0) \in \{0,1\}^{K \times K}$	The adjacency matrix of all post nodes in the set $\mathcal{V}$ that have at least one link connecting them as a composed simple graph.
$\boldsymbol{A}^{(*)}$	Matrix of Float $\boldsymbol{A}^{(*)} := [a^{(*)}_{i,i'}]_{K \times K} \in \mathbb{R}^{K \times K}$, $\boldsymbol{A}^{(*)} \in \{\boldsymbol{A}^{\mathsf{TEM}}, \boldsymbol{A}^{\mathsf{SOC}}, \boldsymbol{A}^{\mathsf{SPA}}\}$	The weighted adjacency matrix of each of the three sub-graphs $\mathcal{G}^{(*)}$ of the multi-graph $\mathcal{G}$, "(*)" represents one of the link types in {TEM, SOC, SPA}.
$\boldsymbol{A}^{\mathcal{U}}$	Matrix of Boolean $\boldsymbol{A}^{\mathcal{U}} := [a^{\mathcal{U}}_{j,j'}]_{\|\mathcal{U}\| \times \|\mathcal{U}\|} \in \{0,1\}^{\|\mathcal{U}\| \times \|\mathcal{U}\|}$	The adjacency matrix of all unique users $\mathcal{U}$ marking their direct friendship which also included the relationship among themselves.
$\boldsymbol{A}^{\mathcal{U}'}$	Matrix of Float $\boldsymbol{A}^{\mathcal{U}'} := [a^{\mathcal{U}'}_{j,j'}]_{\|\mathcal{U}\| \times \|\mathcal{U}\|} \in [0,1]^{\|\mathcal{U}\| \times \|\mathcal{U}\|}$	The weighted adjacency matrix of all unique users $\mathcal{U}$ marking their mutual interest in terms of the Jaccard Index of the public groups that they follow.
$\alpha_{\mathcal{T}}, \alpha^{(n)}_{\mathcal{U}}$	Float scalars $\alpha_{\mathcal{T}}, \alpha^{(1)}_{\mathcal{U}}, \alpha^{(2)}_{\mathcal{U}}, \alpha^{(3)}_{\mathcal{U}} \in [0,1]$	Parameters adjusting the weights of linear combination in relationship matrices $\mathfrak{T}$ and $\mathfrak{U}$.
$\beta_{\mathcal{U}}$	Float scalar $\beta_{\mathcal{U}} \in (0,1)$	The threshold to define mutual interest of two users as the Jaccard Index of public groups.
χ^2	Float Scalar	The Chi-square statistics of two distributions.
$\mathfrak{d}_i, \mathfrak{D}$	Object Tuples $\mathfrak{d}_i = (\mathfrak{I}_i, \mathcal{S}_i, \mathfrak{u}_i, \mathfrak{t}_i, \mathfrak{l}_i)$, $\mathfrak{d}_i \in \mathfrak{D} = \{\mathfrak{d}_1, \mathfrak{d}_2, \ldots, \mathfrak{d}_K\}$	The tuple of all raw data (image, sentences, user ID, timestamp, and geo-location) from one sample point.
D_{KL}	Float Scalar	The Kullback–Leibler (KL) divergence of two distributions.
ε	Float Scalar	An arbitrary small number to avoid zero-division.
$\boldsymbol{F}$	Matrix of Integers and Floats $\boldsymbol{F} = [\mathfrak{f}_i]_{3 \times K}, \mathfrak{f}_i = [\mathfrak{f}_{1,i}, \mathfrak{f}_{2,i}, \mathfrak{f}_{3,i}], \mathfrak{f}_{1,i} \in \mathbb{N}, \mathfrak{f}_{2,i}, \mathfrak{f}_{3,i} \in [0,1]$	The face recognition result of an image sample in terms of the number of faces detected $\mathfrak{f}_{1,i}$, the model confidence for the prediction $\mathfrak{f}_{2,i}$, and the proportion of total area of bounding boxes of detected faces to the total area of images $\mathfrak{f}_{3,i}$.
G_0	Undirected weighted graph $G_0 = (V_0, E_0, \boldsymbol{w}_0)$	The complete spatial network in a city weighted by the travel time with all sorts of transportation between spatial nodes.
G	Undirected weighted graph $G = (V, E, \boldsymbol{w}), V \subseteq V_0, E \subseteq E_0, \boldsymbol{w} \subseteq \boldsymbol{w}_0$	The spatial network in a city weighted by the travel time between spatial nodes (no more than 20 min) that have at least one sample posted near them.
$\mathcal{G}$	Weighted multi-graph $\mathcal{G} = (\mathcal{V}, \{\mathcal{E}^{\mathsf{TEM}}, \mathcal{E}^{\mathsf{SOC}}, \mathcal{E}^{\mathsf{SPA}}\}, \{\boldsymbol{w}^{\mathsf{TEM}}, \boldsymbol{w}^{\mathsf{SOC}}, \boldsymbol{w}^{\mathsf{SPA}}\})$	The graph including the temporal, social, and spatial links $\mathcal{E}^{(*)}$ among the post nodes from set $\mathcal{V}$, weighted by the respective connection strengths $\boldsymbol{w}^{(*)}$.
$\mathcal{G}^{(*)}$	Undirected weighted graph $\mathcal{G}^{(*)} = (\mathcal{V}, \mathcal{E}^{(*)}, \boldsymbol{w}^{(*)})$, $\mathcal{G}^{(*)} \in \{\mathcal{G}^{\mathsf{TEM}}, \mathcal{G}^{\mathsf{SOC}}, \mathcal{G}^{\mathsf{SPA}}\}$	The sub-graph of the multi-graph $\mathcal{G}$, while "(*)" represents one of the link types in {TEM, SOC, SPA}.
$\boldsymbol{H}^{\mathsf{B}}$	Matrix of Floats $\boldsymbol{H}^{\mathsf{B}} = [\boldsymbol{h}^{\mathsf{BERT}}_i]_{768 \times K}$	The last hidden layer for `[CLS]` token of BERT model pre-trained on WHOSe_Heritage.
$\boldsymbol{H}^{\mathsf{V}}$	Matrix of Floats $\boldsymbol{H}^{\mathsf{V}} = [\boldsymbol{h}^{\mathsf{V}}_i]_{512 \times K}$	The last hidden layer of ResNet-18 model pre-trained on Places365.

TABLE APP. B.3 Cont.

Symbol	Data Type/Shape	Description
i, i'	Integer Indices $i, i' \in \{1, 2, .., K\} \subset \mathbb{N}$	The index of samples in the dataset $\mathfrak{D}$ of one case city.
$\mathfrak{I}_i$	Tensor of Integers within $[0, 255] \in \mathbb{N}$ of size $150 \times 150 \times 3$ or $320 \times 240 \times 3$	The raw image data of one sample post with RGB channels.
$\boldsymbol{I}$	Matrix of Boolean $\boldsymbol{I} \in \{0, 1\}^{\lvert\mathcal{U}\rvert \times \lvert\mathcal{U}\rvert}$	The diagonal identity matrix marking the identity of unique users in $\mathcal{U}$.
j, j'	Integer Indices $j, j' \in \{1, 2, .., \lvert\mathcal{U}\rvert\} \subset \mathbb{N}$	The index of users in the ordered set $\mathcal{U}$ of all unique users from one case city.
k	Integer Indices $k \in \{1, 2, .., \lvert\mathcal{T}\rvert\} \subset \mathbb{N}$	The index of timestamps in the ordered set $\mathcal{T}$ of all unique timestamps from one case city.
K	Integer $K = \lvert\mathfrak{D}\rvert$	The sample size (number of posts) collected in one case city.
$\boldsymbol{K}^{\text{HA}}$	Matrix of Floats $\boldsymbol{K}^{\text{HA}} = [\boldsymbol{\kappa}_i^{\text{HA}}]_{2 \times K}$	The confidence indicator matrix for heritage attributes labels including the top-n confidence and agreement between VOTE and STACK models.
$\boldsymbol{K}^{\text{HV}}$	Matrix of Floats $\boldsymbol{K}^{\text{HV}} = [\boldsymbol{\kappa}_i^{\text{HV}}]_{2 \times K}$	The confidence indicator matrix for heritage values labels including the top-n confidence and agreement between BERT and ULMFiT models.
l, l'	Integer Indices $l, l' \in \{1, 2, .., \lvert V\rvert\} \subset \mathbb{N}$	The index of nodes in the ordered set V of all spatial nodes from one case city.
$\mathfrak{l}_i$	Tuple of Floats $\mathfrak{l}_i = (\mathfrak{x}_i, \mathfrak{y}_i)$	The geographical coordinate of latitude ($\mathfrak{y}_i$) and longitude ($\mathfrak{x}_i$) as location of one sample.
$\boldsymbol{L}^{\text{a}}$	Matrix of logit vectors $\boldsymbol{L}^{\text{a}} = [\boldsymbol{l}_i^{\text{a}}]_{102 \times K}$	The last softmax layer of ResNet-18 model pre-trained on SUN predicting scene attributes.
$\boldsymbol{L}^{\text{s}}$	Matrix of logit vectors $\boldsymbol{L}^{\text{s}} = [\boldsymbol{l}_i^{\text{s}}]_{365 \times K}$	The last softmax layer of ResNet-18 model pre-trained on Places365 predicting scene categories.
$\mathcal{M}$	A set of objects	The set of machine learning models used to train classifiers on Tripoli data.
$\boldsymbol{O}$	Matrix of Boolean $\boldsymbol{O} := [\mathfrak{o}_i] \in \{0, 1\}^{3 \times K}$	The language detection result of the original language appearance of the sentences in each sample, in terms of English $\mathfrak{o}_1$, local language $\mathfrak{o}_2$, and other languages $\mathfrak{o}_3$.
$\boldsymbol{R}, \boldsymbol{R}^{(*)}$	Matrix of Float $\boldsymbol{R}, \boldsymbol{R}^{(*)} \in \mathbb{R}^{N \times K}$, $\boldsymbol{R}^{(*)} \in \{\boldsymbol{R}^{\text{TEM}}, \boldsymbol{R}^{\text{SOC}}, \boldsymbol{R}^{\text{SPA}}\}$	The embedding matrices of each of the samples to a N-dimensional vector based on the general structure of the multi-graph $\mathcal{G}$ and the specific types of links.
$\mathcal{S}_i$	Set of Strings $\mathcal{S}_i = \{s_i^{(1)}, s_i^{(2)}, ..., s_i^{(\lvert\mathcal{S}_i\rvert)}\}$ or Empty Set $\mathcal{S}_i = \varnothing$	The processed textual data as a set of individual sentences that have a valid semantic meaning and have been translated into English.
$\boldsymbol{S}$	Boolean Matrix $\boldsymbol{S} := [s_{l,i}] \in \{0, 1\}^{\lvert V\rvert \times K}$	The one-hot embedding matrix of the samples corresponding to the geo-node set V.
$\mathfrak{S}$	Matrix of Float $\mathfrak{S} := [\mathfrak{s}_{l,l'}] \in [0, 1]^{\lvert V\rvert \times \lvert V\rvert}$	A matrix marking the spatial closeness of all the unique spatial nodes from set V that can be reached within 20 min.
$\mathcal{T}$	An ordered Set $\mathcal{T} = \{\tau_1, \tau_2, ..., \tau_{\lvert\mathcal{T}\rvert}\}$	The ordered set of all unique timestamps from one case city.

TABLE APP. B.3 Cont.

Symbol	Data Type/Shape	Description
τ_k	Timestamp $\tau_k \in \mathcal{T}$	A timestamp in the ordered set $\mathcal{T}$ of all unique timestamps.
$\mathfrak{t}_i$	Timestamp $\mathfrak{t}_i \in \mathcal{T}$	A timestamp indexed with sample ID in the ordered set $\mathcal{T}$ of all unique timestamps.
$\boldsymbol{T}$	Boolean Matrix $\boldsymbol{T} := [t_{k,i}] \in \{0,1\}^{\lvert\mathcal{T}\rvert \times K}$	The one-hot embedding matrix of the samples corresponding to the timestamp set $\mathcal{T}$.
$\mathfrak{T}$	Matrix of Float $\mathfrak{T} \in [0,1]^{\lvert\mathcal{T}\rvert \times \lvert\mathcal{T}\rvert}$	A matrix marking the temporal similarity of all the unique timestamps from set $\mathcal{T}$.
$\mathcal{U}$	An ordered Set $\mathcal{U} = \{\mu_1, \mu_2, \ldots, \mu_{\lvert\mathcal{U}\rvert}\}$	The ordered set of all unique users from one case study city.
μ_j	User ID Object $\mu_j \in \mathcal{U}$	An instance of user in the ordered set $\mathcal{U}$ of all unique users.
$\mathfrak{u}_i$	User ID Object $\mathfrak{u}_i \in \mathcal{U}$	An instance of user indexed with sample ID in the ordered set $\mathcal{U}$ of all unique users.
$\boldsymbol{U}$	Boolean Matrix $\boldsymbol{U} := [u_{j,i}] \in \{0,1\}^{\lvert\mathcal{U}\rvert \times K}$	The one-hot embedding matrix of the samples corresponding to the user set $\mathcal{U}$.
$\mathfrak{U}$	Matrix of Float $\mathfrak{U} \in [0,1]^{\lvert\mathcal{U}\rvert \times \lvert\mathcal{U}\rvert}$	A matrix marking the social similarity of all the unique users from set $\mathcal{U}$, as a linear combination of identity matrix $\boldsymbol{I}$ and adjacency matrices $\boldsymbol{A}^{\mathcal{U}}, \boldsymbol{A}^{\mathcal{U}'}$.
V	A set of nodes $V = \{v_1, v_2, \ldots, v_{\lvert V\rvert}\}$	The set of all the spatial nodes that have at least one sample posted near them.
v_l	Spatial node $v_l \in V$	A node in the set V of all spatial nodes that have at least one sample posted near them.
$\mathcal{V}$	A set of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_K\}$	The set of all nodes of posts in one case city.
v_i	Post/Sample node $v_i \in \mathcal{V}$	A node in the set $\mathcal{V}$ of all nodes of posts in one case city.
$\boldsymbol{w}, \boldsymbol{w}^{(*)}$	Vector of Float $\boldsymbol{w} := [w_e] \in [0,20]^{\lvert E\rvert}, \boldsymbol{w}^{(*)} := [w_e^{(*)}] \in \mathbb{R}^{\lvert\mathcal{E}\rvert}, \boldsymbol{w}^{(*)} \in \{\boldsymbol{w}^{\mathsf{TEM}}, \boldsymbol{w}^{\mathsf{SOC}}, \boldsymbol{w}^{\mathsf{SPA}}\}$	The weight vector of spatial network G and post graphs $\mathcal{G}^{\mathsf{TEM}}, \mathcal{G}^{\mathsf{SOC}}, \mathcal{G}^{\mathsf{SPA}}$, these weights are directly interchangeable with the adjacency matrices.
$\boldsymbol{X}^{\mathsf{vis}}$	Matrix of Floats and Integers $\boldsymbol{X}^{\mathsf{vis}}_{982 \times K} = \left[\boldsymbol{H}^{\mathsf{v}\mathsf{T}}, \boldsymbol{F}^{\mathsf{T}}, \boldsymbol{\sigma}^{(5)}(\boldsymbol{L}^{\mathsf{s}})^{\mathsf{T}}, \boldsymbol{\sigma}^{(10)}(\boldsymbol{L}^{\mathsf{a}})^{\mathsf{T}}\right]^{\mathsf{T}}$	The final visual feature concatenating the hidden layer $\boldsymbol{H}^{\mathsf{v}}$, the face detection results $\boldsymbol{F}$, the filtered top-5 scene prediction $\boldsymbol{\sigma}^{(5)}(\boldsymbol{L}^{\mathsf{s}})$, and the filtered top-10 attribute prediction $\boldsymbol{\sigma}^{(10)}(\boldsymbol{L}^{\mathsf{a}})$.
$\boldsymbol{X}^{\mathsf{tex}}$	Matrix of Floats and Integers $\boldsymbol{X}^{\mathsf{tex}}_{771 \times K} = \left[\boldsymbol{H}^{\mathsf{B}\mathsf{T}}, \boldsymbol{O}^{\mathsf{T}}\right]^{\mathsf{T}}$	The final textual feature concatenating the hidden layer $\boldsymbol{H}^{\mathsf{B}}$ of BERT on `[CLS]` token, and the original language detection results $\boldsymbol{O}$.
$\boldsymbol{Y}^{\mathsf{HA}}$	Matrix of Floats $\boldsymbol{Y}^{\mathsf{HA}} = [\boldsymbol{y}_i^{\mathsf{HA}}]_{9 \times K}$	The final generated label of heritage attributes on 9 depicted scenes, as the average of prediction from VOTE and STACK models.
$\boldsymbol{Y}^{\mathsf{HV}}$	Matrix of Floats $\boldsymbol{Y}^{\mathsf{HV}} = [\boldsymbol{y}_i^{\mathsf{HV}}]_{11 \times K}$	The final generated label of heritage values on 10 OUV selection criteria and an additional negative class, as the average of prediction from BERT and ULMFiT models.

TABLE APP. B.4 The nomenclature of functions defined and used in Chapter 4 in alphabetic order.

Symbol	Data Type/Shape	Description
$\text{argmx}(\boldsymbol{l}, n)$	Function outputting a set of floats or objects	The set of largest n elements of any float vector $\boldsymbol{l}$.
$\boldsymbol{f}_{\text{BERT}}(\mathcal{S}\mid\boldsymbol{\Theta}_{\text{BERT}})$	Function inputting a sentence/paragraph or a batch of sentences/paragraphs, outputting a vector or a matrix of vectors	The pre-trained uncased BERT model fine-tuned on WHOSe_Heritage with the model parameters $\boldsymbol{\Theta}_{\text{BERT}}$ that can process some textual inputs into the 768-dimensional hidden output vector $\boldsymbol{h}^{\text{BERT}}$ of the `[CLS]` token.
$\boldsymbol{f}_{\text{ResNet-18}}(\mathfrak{I}\mid\boldsymbol{\Theta}_{\text{ResNet-18}})$	Function inputting a tensor or a batch of tensors, outputting three vectors or three matrices of vectors	The ResNet-18 model pre-trained on Places365 dataset with the model parameters $\boldsymbol{\Theta}_{\text{ResNet-18}}$ that can process the image tensor $\mathfrak{I}$ into the predicted vectors of scenes $\boldsymbol{l}^{\text{s}}$, predicted vectors of attributes $\boldsymbol{l}^{\text{a}}$, and the last hidden layer $\boldsymbol{h}^{\text{v}}$.
$\boldsymbol{g}_{\text{BERT}}(\mathcal{S}\mid\boldsymbol{\Theta}_{\text{BERT}})$	Function inputting a sentence/paragraph or a batch of sentences/paragraphs, outputting a vector or a matrix of vectors	The end-to-end pre-trained uncased BERT model fine-tuned on WHOSe_Heritage with the model parameters $\boldsymbol{\Theta}_{\text{BERT}}$ together with the MLP classifiers that can process some textual inputs into the logit prediction vector $\boldsymbol{y}^{\text{BERT}}$ of 11 heritage value classes concerning OUV.
$\boldsymbol{g}_{\text{ULMFiT}}(\mathcal{S}\mid\boldsymbol{\Theta}_{\text{ULMFiT}})$	Function inputting a sentence/paragraph or a batch of sentences/paragraphs, outputting a vector or a matrix of vectors	The end-to-end pre-trained ULMFiT model fine-tuned on WHOSe_Heritage with the model parameters $\boldsymbol{\Theta}_{\text{ULMFiT}}$ together with the MLP classifiers that can process some textual inputs into the logit prediction vector $\boldsymbol{y}^{\text{ULMFiT}}$ of 11 heritage value classes concerning OUV.
$\boldsymbol{h}_{\text{VOTE}}(\boldsymbol{h}^{\text{v}}\mid\Theta_{\text{VOTE}}, \mathcal{M}, \boldsymbol{\Theta}_{\mathcal{M}})$	Function inputting a vector or a batch of vectors, outputting a vector or a matrix of vectors	The ensemble Voting Classifier with model parameter Θ_{VOTE} of machine learning models from $\mathcal{M}$ with their respective model parameters $\boldsymbol{\Theta}_{\mathcal{M}}$, which processes the visual feature vector $\boldsymbol{h}^{\text{v}}$ into the logit prediction vector $\boldsymbol{y}^{\text{VOTE}}$ of 9 heritage attribute classes concerning depicted scenes.
$\boldsymbol{h}_{\text{STACK}}(\boldsymbol{h}^{\text{v}}\mid\Theta_{\text{STACK}}, \mathcal{M}, \boldsymbol{\Theta}_{\mathcal{M}})$	Function inputting a vector or a batch of vectors, outputting a vector or a matrix of vectors	The ensemble Stacking Classifier with model parameter Θ_{STACK} of machine learning models from $\mathcal{M}$ with their respective model parameters $\boldsymbol{\Theta}_{\mathcal{M}}$, which processes the visual feature vector $\boldsymbol{h}^{\text{v}}$ into the logit prediction vector $\boldsymbol{y}^{\text{STACK}}$ of 9 heritage attribute classes concerning depicted scenes.
$\mathcal{I}(\mu_j)$	Function outputting an ordered set of objects	The set of public groups that are followed by user μ_j.
$\text{IoU}(\mathcal{A}, \mathcal{B})$	Function outputting a non-negative float	The Jaccard Index of any two sets $\mathcal{A}, \mathcal{B}$ as the cardinality of the intersection of the two sets over that of the union of them.
$\max(\boldsymbol{l}, n)$	Function outputting a float	The n_{th} largest element of any float vector $\boldsymbol{l}$.
$\boldsymbol{\sigma}^{(n)}(\boldsymbol{l})$	Function both inputting and outputting a logit vector	The activation filter to keep the top-n entries of any logit vector $\boldsymbol{l}$ and smooth all the others entries based on the total confidence (sum) of top-n entries.

Supplementary Materials Chapter 5

Proof of Equivalence of Label Diffusion

In this section, we will show that adding the last state of a node $\hat{\boldsymbol{\mathcal{Y}}}^{(t)}$ to the calculation of its current state during the diffusion process is equivalent to what has been proposed in Equations (5.12) and (5.13) for computing the steady-state $\boldsymbol{\mathcal{Y}}$.

Proof. By adding the term of the last state of a node itself, Equation (5.12) could be adapted as:

$$\hat{\mathbf{y}}_k^{(t+1)} = \alpha_1 \hat{\mathbf{y}}_k^{(t)} + \alpha_2 \hat{\mathbf{y}}_k + \alpha_3 \frac{\sum_{\nu_{k'} \in \mathcal{N}_G(\nu_k)} W_{k,k'} \hat{\mathbf{y}}_{k'}^{(t)}}{\sum_{\nu_{k'} \in \mathcal{N}_G(\nu_k)} W_{k,k'}}, \tag{B.8}$$

or in its matrix form:

$$\hat{\boldsymbol{\mathcal{Y}}}^{(t+1)} = \alpha_1 \hat{\boldsymbol{\mathcal{Y}}}^{(t)} + \alpha_2 \hat{\boldsymbol{\mathcal{Y}}} + \alpha_3 \hat{\boldsymbol{\mathcal{Y}}}^{(t)} \left(\boldsymbol{W}\boldsymbol{D}^{-1}\right), \tag{B.9}$$

where $\alpha_1, \alpha_2, \alpha_2 \in [0,1], \alpha_1 + \alpha_2 + \alpha_3 = 1$ are parameters balancing the importance of the last state of a node, the initial state of a node, and the last state of its neighbouring nodes. Then the steady state could be written as:

$$\boldsymbol{\mathcal{Y}} = \alpha_1 \boldsymbol{\mathcal{Y}} + \alpha_2 \hat{\boldsymbol{\mathcal{Y}}} + \alpha_3 \boldsymbol{\mathcal{Y}} \left(\boldsymbol{W}\boldsymbol{D}^{-1}\right), \tag{B.10}$$

$$\boldsymbol{\mathcal{Y}} \left((1-\alpha_1)\boldsymbol{I} - \alpha_3 \boldsymbol{W}\boldsymbol{D}^{-1}\right) = \alpha_2 \hat{\boldsymbol{\mathcal{Y}}}, \tag{B.11}$$

$$\text{therefore, } \boldsymbol{\mathcal{Y}} = \frac{\alpha_2}{\alpha_2+\alpha_3} \hat{\boldsymbol{\mathcal{Y}}} \left(\boldsymbol{I} - \frac{\alpha_3}{\alpha_2+\alpha_3} \boldsymbol{W}\boldsymbol{D}^{-1}\right)^{-1}, \tag{B.12}$$

substituting the number $\alpha_3/(\alpha_2+\alpha_3) \in (0,1]$ with another parameter $\alpha_0 \in (0,1]$, then Equation (B.12) could be written as:

$$\boldsymbol{\mathcal{Y}} = (1-\alpha_0) \hat{\boldsymbol{\mathcal{Y}}} \left(\boldsymbol{I} - \alpha_0 \boldsymbol{W}\boldsymbol{D}^{-1}\right)^{-1}, \tag{B.13}$$

exactly the same as Equation (5.13). Here the parameter α_0 represents the relative importance of the last state of the neighbouring nodes of a node and its initial state, conceptually consistent with the original α mentioned in Section 5.3.3. $\square$

It is worth noting that the diffusion chain presented here employs a Markov transition probability matrix but it is not a Markov Chain in its entirety because it is not a memory-less machine; in fact, the initial state contributes to the direction of the steady state vectors. Note by putting α_2 equal to zero we can turn this chain into a Markov Chain, in which case the **y** ends up being an eigenvector centrality array.

Model Implementation Details

For all models, Adam (Kingma and Ba, 2017) with L2 regularization of 2e-4 is used as the optimizer. The hyper-parameter tuning, model training, and inference on VEN are performed on NVIDIA GeForce RTX 3060 GPU, and the inference on VEN-XL is performed on Intel Core i7-12700KF CPU since it is too large to fit in GPU. Hyper-parameter tuning is performed in a small range with grid-search. The detail of training, the resource occupancy, and the inference time are given respectively in the following sections and in Table B.5.

TABLE APP. B.5 The training resource occupancy, the model checkpoint size, and inference time (per each mini-batch) of each type of models.

Model	Number of Epochs at Early-Stopping	Model Size	Training Time	Inference Time GPU (*VEN*)	Inference Time CPU (*VEN-XL*)
MLP	126/300	2.1 MB	0.02s	0.02s	0.33s
GCN-KNN	207/500	115.2 MB	0.02s	0.01s	0.05s
GAT	442/1000	6.0 MB	0.05s	0.03s	4.18s
GSA	170/300	13.6 MB	0.09s	0.06s	13.54s
HGSA	300/300	1.6 MB	0.03s	0.03s	3.39s
HGT	300/300	0.6 MB	0.04s	0.02s	1.33s

RDC

No hyperparameter is tuned for the random classifier. The random choice function of Numpy library is used to generate top-3 OUV and top-1 HA predictions for each data sample based on the initial prior distribution of classes.

MLP

The training takes 300 epochs with early-stopping criterion of 30 epochs. The hyper-parameters being tuned include learning rate in {.01, .001, .0005}, drop out rate in {.1, .2, .5}, number of hidden layers in {2, 3, 5}, and the size of hidden layers in {32, 64, 128, 256, 512}. The final selected model has a learning rate of .001, dropout rate of .1, and 3 hidden layers each with a size of 256.

GCN

The training takes 500 epochs with early-stopping criterion of 100 epochs. The models use the initial residual connection alpha of 0.5, parameter to compute the strength of identity mapping theta of 1.0, and do not enable shared weights between the smoothed representation and the initial residuals. The hyper-parameters being tuned include learning rate in {.01, .001, .0001}, drop out rate in {.1, .2, .5}, number of hidden layers in {3, 6, 9}, and the size of hidden layers in {128, 256, 512, 1024, 2048}. The final selected model has a learning rate of .0001, dropout rate of .1, and 3 hidden layers each with a size of 2048. Furthermore, it turned out that the models using KNN links rather than the original graph structure perform better, therefore the same searched hyper-parameters are used to re-train a model checkpoint with KNN links as the final model.

GAT

The training takes 1000 epochs with early-stopping criterion of 100 epochs. The models have two hidden GAT layers while the second one only has one attention head. The output of a linear hidden layer is concatenated with output of GAT filters before the final output layer. The hyper-parameters being tuned include learning rate in {.01, .001, .0001}, drop out rate in {.1, .3, .6}, number of attention heads for the first GAT layer in {2, 5, 8}, and the size of hidden layers in {32, 64, 128, 256, 512}. The final selected model has a learning rate of .0001, dropout rate of .1, 2 attention heads, and hidden layer size of 256.

GSA

The training takes 300 epochs with early-stopping criterion of 30 epochs. The hyper-parameters being tuned include learning rate in {.01, .001, .0001}, drop out rate in {.1, .3, .5}, number of hidden layers in {2, 3, 5}, and the size of hidden layers in {32, 64, 128, 256, 512}. The final selected model has a learning rate of .0001, dropout rate of .1, and 5 hidden layers each with a size of 512.

HGSA

The training takes 300 epochs with early-stopping criterion of 100 epochs. The output of a linear hidden layer is concatenated with output of Hetero GSA filters before the final output layer. The hyper-parameters being tuned include learning rate in {.01, .001, .0001}, number of hidden layers in {2, 3, 5}, and the size of hidden layers in {32, 64, 128, 256, 512}. The final selected model has a learning rate of .0001, and 3 hidden layers each with a size of 32.

HGT

The training takes 300 epochs with early-stopping criterion of 100 epochs. The output of a linear hidden layer is concatenated with output of HGT before the final output layer. The hyper-parameters being tuned include learning rate in {.01, .001, .0005, .0001}, number of attention heads in {2, 4}, way of grouping attention heads in {sum, mean}, number of hidden layers in {2, 3, 5}, and the size of hidden layers in {32, 64, 128, 256}. The final selected model has a learning rate of .0005, 2 attention heads, grouping method of mean, and 3 hidden layers each with a size of 32.

Extended Results

Figure B.1 shows the effect of α on the distribution of OUV and HA categories in the final diffused spatial label arrays $\boldsymbol{\mathcal{Y}}$. As α gets larger and closer to its theoretical maximum of $\min(1, 1/\lambda)$, the spatial labels get more to the extreme where all the labels are dominated only by the large classes. This is similar to the problem of "over-smoothing" in GNN literature (Li et al., 2018b).

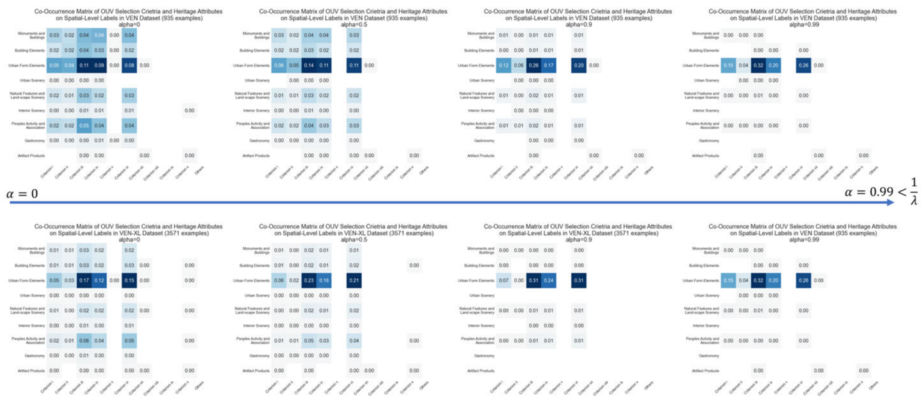


FIG. APP. B.1 The change of normalised co-occurrence matrices $\boldsymbol{O}$ of the OUV and HA categories in spatial level label array $\mathscr{Y}$ in both VEN and VEN-XL datasets, as the scaling parameter α changes.

Computing the relative importance of all features while classifying each OUV/HA category using GNNExplainer will generate a soft mask vector for each node. Figure B.2 plots all the 10-quantile values (similar to the median at the 50% partition, yet showing all values at the 10%, 20%, ..., 90% partitions) of the soft mask values of each feature among all considered nodes, respectively using trained GAT and GSA as the base model. The distribution of the features shows that the relative importance computed by GNNExplainer on the explainable features is far less than that on the hidden features. How to explain and/or interpret those "non-explainable" hidden features would be an interesting future research direction. Inspecting the visualized distributions, that of GAT is slightly different from GSA in the sense that the hidden visual features (with the indices of 0-511, i.e., the left part of the images) are given higher relative importance in GSA. Furthermore, the red lines indicating the threshold of entering the top-250 entries for all the nodes imply that the two models work very differently using the information of all features. GAT has a lower top-250 threshold with a far wider confidence interval than GSA, suggesting that GAT uses very different numbers of features to predict the nodes, while the thresholds and thus the number of features being used in GSA are relatively more stable.

Figure B.3 demonstrates a similar change pattern of Moran's I as in Figure 5.11 with conventional definition of weight matrix:

$$I_C = \frac{(c - \bar{c}\mathbf{1})^\top \tilde{\boldsymbol{W}} (c - \bar{c}\mathbf{1})}{(c - \bar{c}\mathbf{1})^\top (c - \bar{c}\mathbf{1})}, \tag{B.14}$$

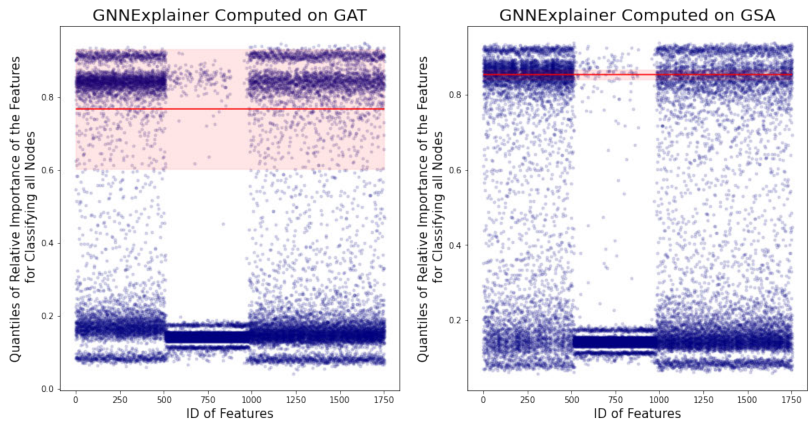


FIG. APP. B.2 The scatter plots of all the 10-quantile values for the relative importance of all visual and textual features while classifying each node in $\mathcal{V}_{\text{train}}, \mathcal{V}_{\text{val}}, \mathcal{V}_{\text{test}}$ in GAT and GSA models, computed with GNNExplainer. The explainable visual features are with the indices of 512-981. The red lines and their shadows mark the means and standard deviations of the relative importance by the top-250_{th} feature.

where the diagonal entries of $\tilde{\boldsymbol{W}}$ are all 0 and the row-sums of the matrix are all 1. Since a few spatial nodes in V (20 in VEN and 27 in VEN-XL) were isolated without any neighbours, rendering the row-standardization operation invalid, these nodes are omitted from the computation.

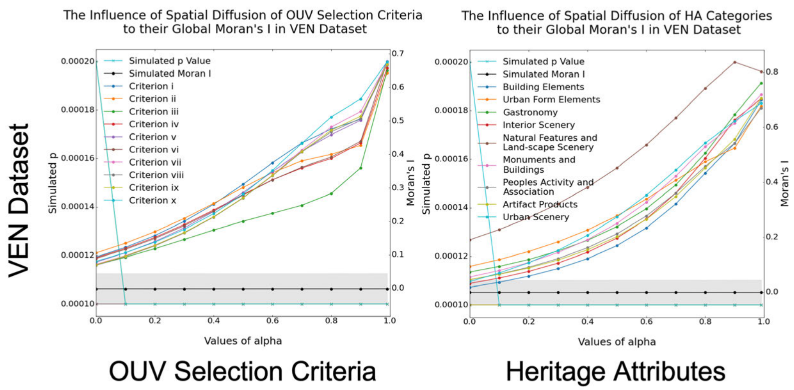


FIG. APP. B.3 The change of global Moran's I in VEN with conventional row-standardized weight matrix only having zero diagonal entries. The Moran's I are generally smaller than in Figure 5.11 since the self-correlations are not considered. For most categories, the spatial correlation is already significant without diffusion. For smaller α, the deviation of Moran's I is also smaller while significantly dropping the p values. Note the expected I value gets to the conventional scale of $-1/(N-1)$.

FIG. APP. B.4 Comparison of the geographical distribution of post-level and spatial-level OUV node labels in VEN and spatial-level labels in VEN-XL. Post-level labels are accompanied by a kernel-density heatmap.

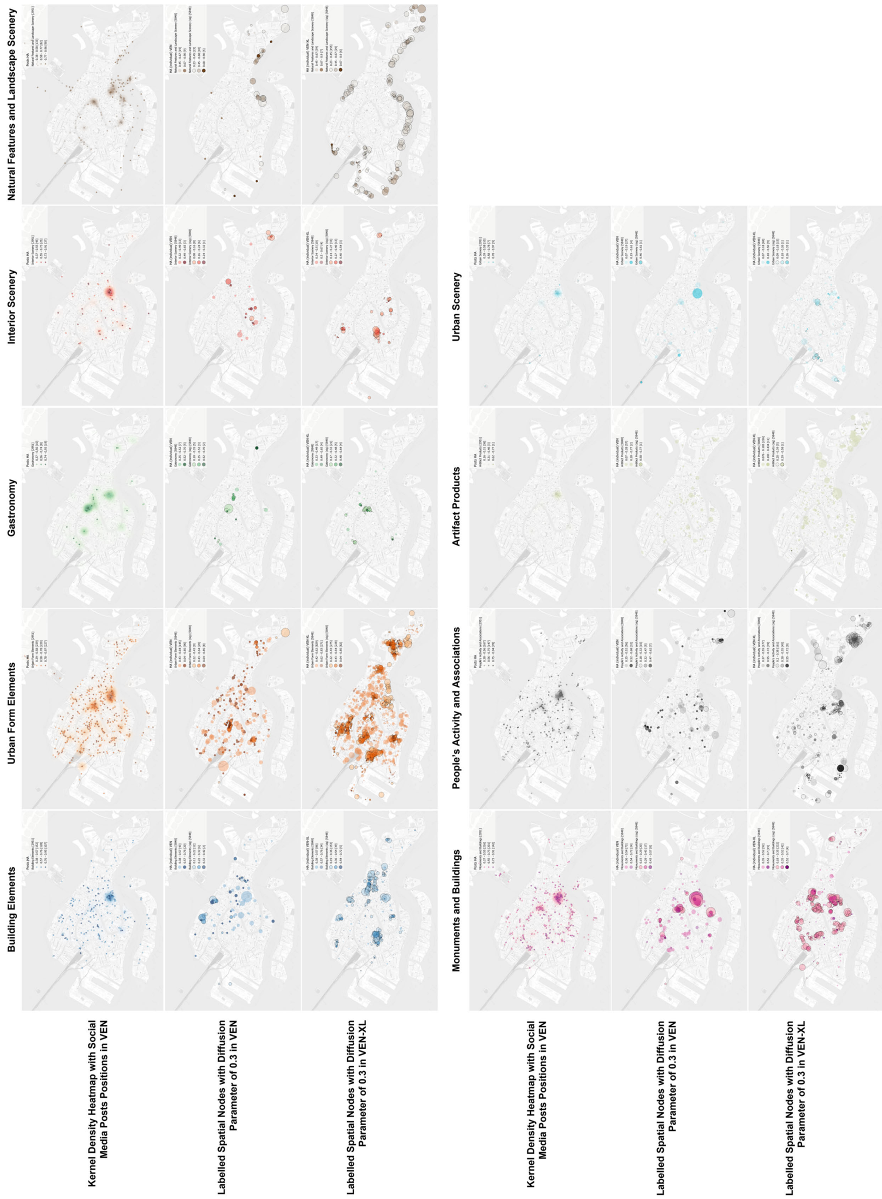


FIG. APP. B.5 Comparison of the geographical distribution of post-level and spatial-level HA node labels in VEN and spatial-level labels in VEN-XL. Post-level labels are accompanied by a kernel-density heatmap.

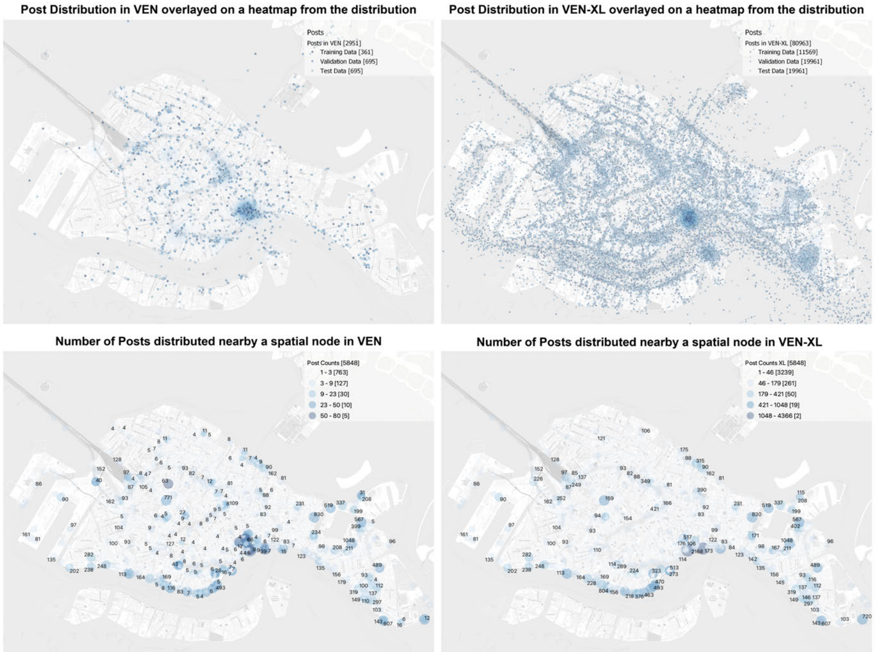


FIG. APP. B.6 Top: the dis-aggregated distribution of all the geo-tagged posts in both VEN and VEN-XL datasets; Bottom: the number of posts distributed nearby each spatial node.

Figure B.4 and B.5 respectively plot the distribution of high values on spatial nodes level for each OUV and HA category in VEN and VEN-XL datasets, and the high values on post levels overlapping with a kernel-density heatmap in VEN dataset only. A relatively stable pattern could be observed in the sense that the “hotspots” in VEN are generally detectable in VEN-XL, but not vice versa. In a few cases such as the HA category of Interior Scene, some significant clusters in VEN are diluted and no longer visible in VEN-XL with possibly more diverse post topics concerning OUV and HA. In general, the distribution in VEN-XL with more posts as data samples can be regarded as more reliable.

Note the methodology proposed in this study can be seen as an alternative and/or supplement to the conventional kernel-density heatmap weighted by the value in each channel. Figure B.4 and B.5 also show the similarity and difference between the two methods in the case of VEN dataset. Generally, the hotspots are distributed in similar locations with both methods, since a spatial node can only be assigned high values when nearby posts also have high values consistently. However, the method proposed also considers confidence and agreement as crucial weighting parameters, preventing the risk in heatmaps that a very large number of medium-low values will also result in an overall hotspot in almost all categories, which is obvious in the case of San Marco square. Another benefit of the proposed method is that it is more specific and discretized than the kernel-density heatmap, yet more general and aggregated than mapping individual posts. The former is beneficial since it can point to certain places (street intersections) instead of only a broad region while tracing the posts as demonstrated in Figures 5.13 and 5.14, easier for targeting useful information. The latter is beneficial since the method will not be too sensitive to individual posts while losing the main points. Furthermore, the proposed method performs aggregation on a fixed number of spatial nodes (a maximum of 5848 in Venice), easier for human comprehension, especially when the number of posts at hand grows to a larger scale, as demonstrated in Figure B.6 where the top-right subplot mapping all the posts collected in Venice gets too crowded with points. However, Figure B.6 also showcases another drawback of the dataset provided by Bai et al. (2022), that the spatial nodes only consisted of the ones on the main island and omitted places such as Giudecca island and San Giorgio Maggiore, pulling the posts on those places as well as on the canals to their nearest walkable spatial nodes on the southern harbour areas. This may have partially influenced the results of spatial distribution of categories such as OUV Criterion (vi) about Association and HA Natural Features and Landscape Scenery. This issue could be fixed in future studies by updating the assignment matrix $\boldsymbol{B}$ and spatial weight matrix $\boldsymbol{W}$.

Nomenclature

Tables B.6 gives an overview of the mathematical notations used in the Chapter 5.

TABLE APP. B.6 The nomenclature of mathematical notations used in Chapter 5 in alphabetic order.

Symbol	Data Type/Shape	Description
$\boldsymbol{A}$	Matrix of Boolean $\boldsymbol{A} := (\boldsymbol{A}^{\text{TEM}} > 0) \bigvee (\boldsymbol{A}^{\text{SPA}} > 0) \bigvee (\boldsymbol{A}^{\text{SOC}} > 0) \in \{0,1\}^{K \times K}$	The adjacency matrix of all post nodes in the set $\mathcal{V}$ that have at least one link connecting them as a composed simple graph.
$\boldsymbol{A}^{(*)}$	Matrix of Boolean $\boldsymbol{A}^{(*)} := \left[A^{(*)}_{i,i'}\right]_{K \times K} \in \{0,1\}^{K \times K}$, $\boldsymbol{A}^{(*)} \in \{\boldsymbol{A}^{\text{TEM}}, \boldsymbol{A}^{\text{SPA}}, \boldsymbol{A}^{\text{SOC}}\}$	The adjacency matrix of each of the three sub-graphs $\mathcal{G}^{(*)}$ of the multi-graph $\mathcal{G}$, "(*)" represents one of the link types in {TEM, SPA, SOC}.
$\boldsymbol{A}_s, \boldsymbol{A}_s^{(*)}$	Matrix of Boolean $\boldsymbol{A}_s, \boldsymbol{A}_s^{(*)} \in \{0,1\}^{\|\mathcal{V}_s\| \times \|\mathcal{V}_s\|}$	The sampled adjacency matrix in sub-graph $\mathcal{G}_s$ for model training and inference.
$\boldsymbol{A}^{\text{KNN}}$	Matrix of Boolean $\boldsymbol{A}^{\text{KNN}} := \left[A^{\text{KNN}}_{i,i'}\right] \in \{0,1\}^{K \times K}$	The adjacency matrix of the k-Nearest Neighbour graph computed with visual features of posts.
$\alpha, \alpha_1, \alpha_2, \alpha_3$	Scalar Values $\alpha, \alpha_1, \alpha_2, \alpha_3 \in [0,1]$	The parameters adjusting the relative importance of neighbours in diffusion process.
$\mathcal{B}$	Bipartite Graph $\mathcal{B} = (\mathcal{V}, V, \mathscr{E}, \boldsymbol{B})$	The bipartite graph of postal nodes $\mathcal{V}$ and spatial nodes V with matrix $\boldsymbol{B}$ and edges $\mathscr{E}$.
$\boldsymbol{B}$	Matrix of Boolean $\boldsymbol{B} := [B_{i,k}] \in \{0,1\}^{K \times \|V\|}$	The bi-adjacency matrix of postal nodes $\mathcal{V}$ and spatial nodes V.
β	Scalar Value	The attenuation parameter for the computation of Katz centrality.
C	Integer Indices $C \in \{1, 2, \ldots, 20\} \subset \mathbb{N}$	The index of the OUV and HA label category channels.
$\boldsymbol{D}$	Matrix of Floats $\boldsymbol{D} \in \mathbb{R}_+^{\|V\| \times \|V\|}$	A diagonal matrix where each entry records the weighted degree of graph G.
$\boldsymbol{e}_C$	1D Array of Boolean $\boldsymbol{e}_C \in \{0,1\}^{20 \times 1}$	A one-hot unit vector marking the C_{th} entry as 1.
$\mathcal{F}$	A set of objects $\mathcal{F} = \{\mathbf{f}_j\}, j \in [0, \|\mathcal{F}\|)$	The set of candidate MLP or GNN models to be trained.
$\mathcal{G}$	Multi-Graph $\mathcal{G} = (\mathcal{V}, \{\mathcal{E}^{\text{TEM}}, \mathcal{E}^{\text{SPA}}, \mathcal{E}^{\text{SOC}}\})$	The graph with temporal, spatial, and social links $\mathcal{E}^{(*)}$ among post nodes set $\mathcal{V}$.
$\mathcal{G}'$	Undirected Simple Graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E})$	The simple composed graph of the multi-graph $\mathcal{G}$ with the same node set $\mathcal{V}$.
$\mathcal{G}_s$	Undirected Multi-Graph or Simple Graph, $\mathcal{G}_s = (\mathcal{V}_s, \{\mathcal{E}_s^{\text{TEM}}, \mathcal{E}_s^{\text{SPA}}, \mathcal{E}_s^{\text{SOC}}\})$ or $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$	The sub-graphs sampled from the original graph $\mathcal{G}$ or $\mathcal{G}'$ to train the models and make inference.
G	Undirected Weighted Graph $G = (V, E, \boldsymbol{W})$	The backend geographical representation of the city as a spatial network.
γ, ϕ	Scalar parameters $\gamma, \phi \in \mathbb{R}$	The parameters to adjust the relative contribution of agreement and confidence scores in the computation of attention values $\boldsymbol{S}$.
i, i'	Integer Indices $i, i' \in \{0, 1, 2, .., K-1\} \subset \mathbb{N}$	The index of samples in the dataset.
I_C	Scalar Value of Float	The global Moran's I computed for the C_{th} label channel.
$\boldsymbol{I}_C$	1D Array of Float $\boldsymbol{I}_C \in \mathbb{R}^{\|V\| \times 1}$	The local Moran's I on all spatial nodes computed for the C_{th} label channel.
j	Integer Indices $j \in \{0, 1, 2, .., \|\mathcal{F}\| - 1\} \subset \mathbb{N}$	The index of candidate models to be trained.
k, k'	Integer Indices $j \in \{0, 1, 2, .., \|V\| - 1\} \subset \mathbb{N}$	The index of spatial nodes in the spatial network.

TABLE APP. B.6 Cont.

Symbol	Data Type/Shape	Description
K	Integer	The sample size (number of posts).
$\boldsymbol{\kappa}^{\text{con}}, \boldsymbol{\kappa}^{\text{agr}}$	1D Array of Floats $\boldsymbol{\kappa}^{\text{con}}, \boldsymbol{\kappa}^{\text{agr}} \in [0, 1]^{K \times 1}$	The prediction confidence and agreement value of the models in $\mathcal{F}$ for all the posts.
$\ell_{\text{OUV}}, \ell_{\text{HA}}$	Function returning Scalar Values	Topic-specific evaluation metrics for OUV and HA classification tasks.
$\mathcal{L}_{\text{train}}, \mathcal{L}_{\text{val}}^{\text{V/A}}$	Function returning Scalar Values	The loss function of a training batch and the entire validation sets.
λ	Scalar Value	The largest eigenvalue of the matrix $\boldsymbol{W}\boldsymbol{D}^{-1}$
$\mathcal{N}_{\mathcal{B}}, \mathcal{N}_{G}$	Function returning a set of nodes	The function returning the neighbours of a spatial node ν_k in either the bipartite graph $\mathcal{B}$ as a set of postal nodes or the spatial network G as a set of spatial nodes.
$\omega_{V/A}$	Scalar parameter	The relative importance of OUV and HA performance during training.
$p_j, p_{*,j}^{\text{V/A}(\cdot)}$	Scalar Values $p_j = p_{\text{val,j}}^{\text{OUV(nJ)}} + p_{\text{val,j}}^{\text{HA(1)}} + p_{\text{test,j}}^{\text{OUV(nJ)}} + p_{\text{test,j}}^{\text{HA(1)}} \in \mathbb{R}^{+}, p_{*,j}^{\text{OUV(1)}}, p_{*,j}^{\text{OUV(n)}}, p_{*,j}^{\text{OUV(nJ)}}, p_{*,j}^{\text{HA(1)}} \in [0, 1]$	The value of a specific evaluation metric (top-1 accuracy, top-n accuracy, order-n Jaccard Index) in the validation or test set for OUV or HA categories by the model $\mathbf{f}_j$.
$\boldsymbol{s}_C$	1D Array of Floats $\boldsymbol{s}_C \in [0, 1]^{K \times 1}$	The vector of attention values of all post nodes in $\mathcal{V}$ of the label channel C.
$\boldsymbol{S}$	2D Array of Floats $\boldsymbol{S} \in [0, 1]^{20 \times K}$	The matrix of attention values of all post nodes in $\mathcal{V}$ of all label channels.
$\sigma_{\boldsymbol{Z}_i, 1}$	Scalar Value	The first singular value computed with SVD on the matrix $\boldsymbol{Z}_i$.
$\boldsymbol{\Theta}_j$	Array of Floats	The model parameter by the candidate model $\mathbf{f}_j$.
$\mathcal{V}$	A set of nodes $\mathcal{V} = \{v_i\}, i \in [0, K)$	The set of all nodes of posts in the graph $\mathcal{G}$.
$\mathcal{V}_{\text{batch}}$	A set of nodes $\mathcal{V}_{\text{batch}} \subset \mathcal{V}_{\text{train}}, \mathcal{V}_{\text{val}}, \mathcal{V}_{\text{test}}$	The set of post nodes as mini-batches used for model training and inference.
$\mathcal{V}_{\text{tex}\pm}$	A set of nodes $\mathcal{V}_{\text{tex+}}, \mathcal{V}_{\text{tex-}} \subset \mathcal{V}$	The set of post nodes with or without textual features.
$\mathcal{V}_{\text{train}}, \mathcal{V}_{\text{val}}, \mathcal{V}_{\text{test}}, \mathcal{V}_{\text{unlab}}$	A set of nodes $\mathcal{V}_{\text{train}}, \mathcal{V}_{\text{val}}, \mathcal{V}_{\text{test}}, \mathcal{V}_{\text{unlab}} \subset \mathcal{V}$	The set of post nodes respectively in the training set, validation set, test set, or unlabelled set.
$\mathcal{V}_{\text{V}\pm,\text{A}\pm}$	A set of nodes $\mathcal{V}_{\text{V+,A+}}, \mathcal{V}_{\text{V+,A-}}, \mathcal{V}_{\text{V-,A+}}, \mathcal{V}_{\text{V-,A-}} \subset \mathcal{V}$	The set of post nodes respectively with or without OUV or HA labels initially.
V	A set of nodes $V = \{\nu_k\}, k \in [0, \|V\|)$	The set of all spatial nodes of street intersections in the spatial network G.
$\boldsymbol{W}$	Matrix of Float $\boldsymbol{W} := [W_{k,k'}] \in [0, 1]^{\|V\| \times \|V\|}$	The weighted adjacency matrix marking the temporal closeness of spatial nodes.
$\boldsymbol{X}$	2D Array of Floats $\boldsymbol{X} := [\boldsymbol{x}_i]_{i \in [0, K)} \in \mathbb{R}^{1753 \times K}$	The visual and textual representation features of a post.
$\boldsymbol{X}_s$	2D Array of Floats $\boldsymbol{X}_s \in \mathbb{R}^{1753 \times \|\mathcal{V}_s\|}$	The sampled input visual and textual features of nodes in sub-graph $\mathcal{G}_s$ used for model training and inference.
$\boldsymbol{X}^{\text{tex}}$	2D Array of Floats $\boldsymbol{X}^{\text{tex}} \in \mathbb{R}^{771 \times K}$	The textual representation features of a post.
$\boldsymbol{X}^{\text{vis}}$	2D Array of Floats $\boldsymbol{X}^{\text{vis}} \in \mathbb{R}^{982 \times K}$	The visual representation features of a post.

TABLE APP. B.6 Cont.

Symbol	Data Type/Shape	Description
$\boldsymbol{y}_i^{\mathsf{HA}}, \boldsymbol{y}_i^{\mathsf{OUV}}$	1D Arrays of Floats $\boldsymbol{y}_i^{\mathsf{HA}} \in [0,1]^{9\times 1}, \boldsymbol{y}_i^{\mathsf{OUV}} \in [0,1]^{11\times 1}$	The HA and OUV labels of the node v_i if not empty
$\hat{\boldsymbol{y}}_{j,i}, \hat{\boldsymbol{y}}_{j,i}^{\mathsf{HA}}, \hat{\boldsymbol{y}}_{j,i}^{\mathsf{OUV}}$	1D Arrays of Floats $\hat{\boldsymbol{y}}_{j,i} \in [0,1]^{20\times 1} \hat{\boldsymbol{y}}_{j,i}^{\mathsf{HA}} \in [0,1]^{9\times 1}, \hat{\boldsymbol{y}}_{j,i}^{\mathsf{OUV}} \in [0,1]^{11\times 1}$	The predicted HA and OUV labels of the node v_i by the candidate model $\mathbf{f}_j$
$\hat{\boldsymbol{y}}_C$	1D Array of Floats $\hat{\boldsymbol{y}}_C := \hat{\boldsymbol{Y}}^{\mathsf{T}} \boldsymbol{e}_C \in [0,1]^{K\times 1}$	The labels of all post nodes in $\mathcal{V}$ for the C_{th} label channel.
$\boldsymbol{Y}_{\mathsf{V}\pm,\mathsf{A}\pm}$	2D Arrays of Floats or Empty Array	The "ground-truth" soft label arrays of post nodes respectively with or without OUV or HA labels initially.
$\hat{\boldsymbol{Y}}$	2D Array of Floats $\hat{\boldsymbol{Y}} := [\hat{\boldsymbol{y}}_i]_{v_i\in\mathcal{V}} \in [0,1]^{20\times K}$	The aggregated label array from $\hat{\boldsymbol{Y}}_j$ for all the posts by all the models in $\mathcal{F}$.
$\hat{\boldsymbol{Y}}_i$	2D Array of Floats $\hat{\boldsymbol{Y}}_i := \left[\hat{\boldsymbol{y}}_{j,i}\right]_{\mathbf{f}_i\in\mathcal{F}} \in [0,1]^{20\times\lvert\mathcal{F}\rvert}$	The predicted label array for the post v_i by all the models in $\mathcal{F}$.
$\hat{\boldsymbol{Y}}_j$	2D Array of Floats $\hat{\boldsymbol{Y}}_j := \left[\hat{\boldsymbol{y}}_{j,i}\right]_{v_i\in\mathcal{V}} \in [0,1]^{20\times K}$	The predicted label array for all the posts in $\mathcal{V}$ by the model $\mathbf{f}_j$.
$\hat{\mathbf{y}}_C$	1D Array of Floats $\hat{\mathbf{y}}_C := \hat{\boldsymbol{\mathfrak{Y}}}^{\mathsf{T}} \boldsymbol{e}_C \in [0,1]^{\lvert V\rvert\times 1}$	The initial soft label value on all spatial nodes in the C_{th} label channel.
$\mathbf{y}_C$	1D Array of Floats $\mathbf{y}_C \in [0,1]^{\lvert V\rvert\times 1}$	The final soft label value on all spatial nodes in the C_{th} label channel after diffusion.
$\hat{\boldsymbol{\mathfrak{Y}}}$	2D Array of Floats $\hat{\boldsymbol{\mathfrak{Y}}} := [\hat{\mathbf{y}}_k] \in [0,1]^{20\times\lvert V\rvert}$	The aggregated spatial label array for spatial nodes from their nearby posts.
$\hat{\boldsymbol{\mathfrak{Y}}}^{(t)}$	2D Array of Floats $\hat{\boldsymbol{\mathfrak{Y}}}^{(t)} := \left[\hat{\mathbf{y}}_k^{(t)}\right] \in [0,1]^{20\times\lvert V\rvert}$	The diffused spatial label array for spatial nodes from their neighbours at the t_{th} iteration, where $\hat{\boldsymbol{\mathfrak{Y}}}^{(0)} = \hat{\boldsymbol{\mathfrak{Y}}}$.
$\boldsymbol{\mathfrak{Y}}$	2D Array of Floats $\boldsymbol{\mathfrak{Y}} := [\mathbf{y}_k] \in [0,1]^{20\times\lvert V\rvert}$	The diffused final spatial label array for spatial nodes from their spatial neighbours.
$\boldsymbol{z}_{j,i}^{\mathsf{HA}}, \boldsymbol{z}_{j,i}^{\mathsf{OUV}}$	1D Arrays of Floats $\boldsymbol{z}_{j,i}^{\mathsf{HA}} \in \mathbb{R}^{9\times 1}, \boldsymbol{z}_{j,i}^{\mathsf{OUV}} \in \mathbb{R}^{11\times 1}$	The hidden layer outputs by model $\mathbf{f}_j$ corresponding to HA and OUV label channels
$\boldsymbol{Z}_i$	Matrix of Floats $\boldsymbol{Z}_i \in [-1,1]^{20\times\lvert\mathcal{F}\rvert}$	The centred and normalised label matrix calculated from $\hat{\boldsymbol{Y}}_i$ for SVD computation.

Supplementary Materials Chapter 6

Model Implementation Details

To obtain the key topics as semantic information of tweets, the BERTopic python library was used (Grootendorst, 2022). The inputs of the topic models were the translated and normalized tweets, as mentioned in Section 6.2.3, where each tweet was regarded as an individual document. Within the six main modules of the BERTopic library, the configurations were respectively as follows:

1. The Sentence-Transformer model "all-MiniLM-L6-v2" (Reimers and Gurevych, 2019) trained in English was used as the embedding model for the input texts.
2. The UMAP (Uniform Manifold Approximation and Projection) model with default parameter configurations was used as the dimensionality reduction algorithm, where the option of low memory was selected to prevent large datasets such as Notre-Dame fire from running out of memory.
3. The default HDBSCAN model was used to cluster the vector representations with reduced dimensions with a minimum topic size of 45 in the case of the Notre-Dame fire and 25 in the Venice flood.
4. For each cluster, the CountVectorizer tool from the Scikit-Learn Python library was used as the vectorizer model to obtain a bag-of-words matrix, where both single words and 2-grams (two consecutive words) were counted and the stop words lists provided by NLTK Python library for English and the local language (French or Italian) were excluded. Note that the stop words were only excluded here at a later stage for generating the verbal description (representation) of the cluster, but not before the sentence embedding step, since Transformer-based BERT models prefer to view words in their semantic contexts.
5. The adjusted version of TF-IDF, the c-TF-IDF (class-Term Frequency - Inverse Document Frequency) was used on the level of clusters by combining the bad-of-words matrices of all tweets belonging to each cluster. Specifically, the importance of very frequent words after removing the stop words was further reduced by taking the square root of all term frequencies. The initially obtained clusters were then automatically reduced by another round of HDBSCAN clustering on the c-TF-IDF cluster representations, resulting in the final detected topics.
6. To further improve the quality of the obtained topic representations, the algorithm of Maximal Marginal Relevance was used to decrease the redundancy of keywords and increase the diversity of keywords for each topic.

The matrix showing the probabilities of each tweet belonging to each obtained topic was calculated and saved. All the topics together with their keywords representation were checked manually to select the ones that might be relevant and interesting for this research, and clustered into six themes: emotions (emoji), heritage, incidence, actions, other sites, and politics, as already described in Section 6.4.3. All the other topics that were not selected were ignored for further analyses in this research.

Extended Results

List of Interesting Topics for the Notre-Dame Fire

Figure B.7 shows the complete timelines of all semantic categories of cultural significance, emotions, and interesting topics detected in the Notre-Dame fire dataset. A selection has been previously illustrated in Figure 6.9.

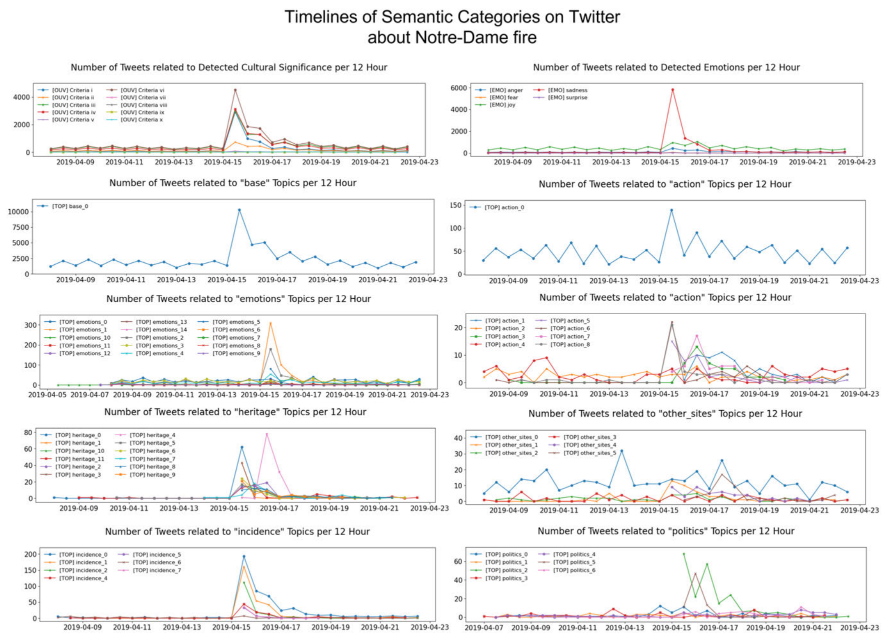


FIG. APP. B.7 The complete timelines showing the temporal development of semantic information along with the HREs in Notre-Dame fire.

The keywords associated with each detected interesting topic under each theme are listed below, note the emojis are transformed into verbal descriptions:

- **Base**
- [TOP] base 0: church, heritage, dame paris, dame cathedral, via, burning, dame fire, rebuild, notredamedeparis, may

- **Emotions**

- [TOP] emotions 0: face_with_tears_of_joy face_with_tears_of_joy, user face_with_tears_of_joy, face_with_tears_of_joy httpurl, ça face_with_tears_of_joy, plus face_with_tears_of_joy, aussi face_with_tears_of_joy, oui face_with_tears_of_joy, face_with_tears_of_joy vtep, grave face_with_tears_of_joy, know face_with_tears_of_joy
- [TOP] emotions 1: notredame notredame, httpurl notredame, notredame sad, crying notredame, httpurl sad, believe notredame, cry notredame, notredamedeparis notredame, attack notredame, awful notredame
- [TOP] emotions 2: loudly_crying_face loudly_crying_face, loudly_crying_face httpurl, loudly_crying_face red_heart, loudly_crying_face dame, loudly_crying_face face_with_tears_of_joy, loudly_crying_face crying_face, crying_face loudly_crying_face, loudly_crying_face broken_heart, loudly_crying_face paris, baby loudly_crying_face
- [TOP] emotions 3: eyes smiling_face_with_heart, smiling_face_with_3_hearts, smiling_face_with_3_hearts smiling_face_with_3_hearts, smiling_face_with_3_hearts user, beaming_face_with_smiling_eyes beaming_face_with_smiling_eyes, smiling_face_with_3_hearts httpurl, user beaming_face_with_smiling_eyes, smiling_face_with_sunglasses smiling_face_with_sunglasses, smiling_face_with_smiling_eyes smiling_face_with_smiling_eyes, user smiling_face_with_smiling_eyes
- [TOP] emotions 4: red_heart red_heart, yellow_heart, love red_heart, user blue_heart, green_heart, purple_heart purple_heart, red_heart notredame, blue_heart httpurl, red_heart thank, merci red_heart
- [TOP] emotions 5: broken_heart notredame, loudly_crying_face notredame, crying_face notredame, notredame crying_face, notredame loudly_crying_face, face_screaming_in_fear loudly_crying_face, sad_but_relieved_face notredame, face_screaming_in_fear notredame, crying_face notredamedeparis, broken_heart loudly_crying_face
- [TOP] emotions 6: grinning_face_with_sweat grinning_face_with_sweat, thinking_face grinning_face_with_sweat, merci grinning_face_with_sweat, grinning_face_with_sweat ouf, jew optimistic, way grinning_face_with_sweat, grimacing_face grinning_face_with_sweat, go grinning_face_with_sweat, grinning_face_with_sweat red_heart, grinning_face_with_sweat virgintonic
- [TOP] emotions 7: face_screaming_in_fear face_screaming_in_fear, tired_face face_screaming_in_fear, face_screaming_in_fear face_with_monocle, face_screaming_in_fear juvaja, understand face_screaming_in_fear, face_screaming_in_fear loudly_crying_face, flushed_face face_screaming_in_fear, face_screaming_in_fear cold_face, face_screaming_in_fear heritage, speechless face_screaming_in_fear
- [TOP] emotions 8: anxious_face_with_sweat, anxious_face_with_sweat anxious_face_with_sweat, anxious_face_with_sweat user, user anxious_face_with_sweat, httpurl anxious_face_with_sweat, anxious_face_with_sweat loudly_crying_face, anxious_face_with_sweat pensive_face, non anxious_face_with_sweat, hot_face anxious_face_with_sweat, 징글징글 anxious_face_with_sweat
- [TOP] emotions 9: face_vomiting face_vomiting, nauseated_face face_vomiting, nauseated_face nauseated_face, face_vomiting angry_face, angry_face face_vomiting, user face_vomiting, face_vomiting nauseated_face, reading face_vomiting, islamophobia like, innocuous survey
- [TOP] emotions 10: face_screaming_in_fear, httpurl face_screaming_in_fear, face_screaming_in_fear face_screaming_in_fear, face_screaming_in_fear crying_face, user face_screaming_in_fear, grinning_face face_screaming_in_fear, omg face_screaming_in_fear, words face_screaming_in_fear, quality diversity, provider face_screaming_in_fear
- [TOP] emotions 11: broken_heart user, broken_heart, broken_heart broken_heart, heartbroken broken_heart, confounded_face broken_heart, sad broken_heart, miskina broken_heart, pain meditation, misha tweet, mothers suicides
- [TOP] emotions 12: hug ganchita, hugs user, giant hug, ganchita thank, need hug, hug great, hug tds, hug tilda, hug viet, hug jesus
- [TOP] emotions 13: pleading_face, pleading_face pleading_face, wsh pleading_face, pleading_face damn, like pleading_face, pleading_face pensive_face, st pleading_face, thank pleading_face, time pleading_face, jsuis pleading_face
- [TOP] emotions 14: shocked user, surprise user, user shock, shock user, know shocked, shock, recal box, jui shocked, part surprise, policeman charge

- **Heritage**

- [TOP] heritage 0: notredame paris, paris notredame, httpurl notredame, paris httpurl, notredame symbol, france notredame, day france, symbol france, httpurl sad, today paris
- [TOP] heritage 1: spire collapsed, spire collapses, spire collapse, fire spire, cathedral collapses, paris collapsed, collapsed fire, collapse dame, spire cathedral, spire roof
- [TOP] heritage 2: rose window, rose windows, stained glass, glass windows, windows survived, window dame, window spared, rosettes, survived fire, three rose

- [TOP] heritage 3: notredame spire, collapsed notredame, spire fell, two towers, moment notredame, roof notredame, towers fire, collapse spire, roof collapsed, towers notredame
- [TOP] heritage 4: important artefacts, saved brave, fire positive, full important, braveheroes, positive signs, firefighters braveheroes, rebuilt restored, iconic building, notredame rebuilt
- [TOP] heritage 5: monuments, historical monuments, heritage professionals, monument like, monument user, people built, magnificent monument, conservators archaeologists, avoid facelift, historians kills
- [TOP] heritage 6: cathédrale dame, cathédrale, user cathedral, paris cathedral, cathedral httpurl, broken_heart dame, broken_heart cathédrale, cathedral dame, dame red_heart, black_heart church
- [TOP] heritage 7: cross, palm sunday, holyweek, notredame cross, cross stands, cross christ, arms cross, latin_cross, httpurl jesus, joseph
- [TOP] heritage 8: discussing historic, significance dame, homage dame, stone paper, place saint, pays homage, witness httpurl, paris work, dame paris, marseille leans
- [TOP] heritage 9: church notredame, notredame catholic, catholics, church building, catholic church, notredame owned, notredame much, place worship, catholic religion, notredame church
- [TOP] heritage 10: fire notre_dame, fdny, fire symbol, dame fire, historic houses, invaluable places, legacy fire, library mention, life dozens, history cry
- [TOP] heritage 11: paris cathedral, user dame, cathedral httpurl, garde photo, user cathédrale, dame paris, cathédrale dame, millefeuille dame, chapelle onze, kapellekerk

- **Incidence**

- [TOP] incidence 0: dame user, dame dame, dame fire, dame burning, dame burns, like dame, gothic, dame symbol, history dame, dame cathedral
- [TOP] incidence 1: paris cathedral, cathedral paris, cathedral dame, cathedral fire, user cathedral, dame paris, fire paris, fire breaks, cathedral notredame, httpurl cathedral
- [TOP] incidence 2: notredame fire, fire notredame, notredame burning, notredame notredamecathedralfire, fire notredamedeparis, httpurl notredamecathedralfire, fire paris, notredame paris, flames notredame, notredamecathedralfire notredame
- [TOP] incidence 4: courage firefighters, firefighters notredame, notredame firefighters, congratulations firefighters, firefighters mobilized, hope firefighters, heroes, fire firefighters, yubari, dear firefighters
- [TOP] incidence 5: paris fire, fire paris, dame paris, depths laments, laments stéphane, dame fire, ee, france affected, paris homework, video fire
- [TOP] incidence 6: fire user, fire fire, user fire, sub rogue, gros fire, ignites user, spontaneously ignites, smart plug, anything fire, firecatchesfire user
- [TOP] incidence 7: reduced ashes, sad fire, ashes fire, cathedral burnt, cathedral burning, cathedral burned, cathedral fire, france stfu, flames survided, fire mum

- **Actions**

- [TOP] action 0: helped turn, accounts helped, verified accounts, user trndnl, topic user, accounts, love user, user merci, know user, awful user
- [TOP] action 1: donations, donate, rebuild dame, donations dame, donated, donating, french billionaires, money rebuild, millionaires, pledges
- [TOP] action 2: thumbs_up thumbs_up, user thumbs_up, thumbs_down, thumbs_down thumbs_down, pretty scallop, ideas rainbow, httpurl thumbs_up, kiss_mark like, thumbs_up collection, thumbs_down httpurl
- [TOP] action 3: donations, donations notredame, taxes taxes, reconstruction notredame, notredame donations, million euros, notredamedesriches, donate notredame, millionaires, notredame billion
- [TOP] action 4: thinking_face thinking_face, rapport thinking_face, turn thinking_face, investigation thinking_face, talking thinking_face, answer thinking_face, paris thinking_face, eyes thinking_face, coincidence thinking_face, something thinking_face
- [TOP] action 5: rebuild notredame, notredame years, notredame rebuilt, deadline rebuild, accomplished five, rebuilding notredame, reconstruction notredame, saying rebuild, notredame rebuild, rebuilt years
- [TOP] action 6: arrow notredame, arrow collapsed, arrow fall, rebuild arrow, new arrow, identically modernize, arrow magnificent, eyes arrow, httpurl arrows, arrow adapted
- [TOP] action 7: 3d, andrew tallon, historian laser, helping rebuild, scans dame, architectural historian, laser scanners, used lasers, worked laser, historian andrew
- [TOP] action 8: user rebuilt, less complexity, local management, square construction, include gdf, lift petticoats, indeed facilities, irreplaceable ok, irl workflow, maintained cleaning

- **Other Sites**

- [TOP] other sites 0: louvre, httpurl paris, île france, paris île, picasso, musée, streetphotography, paris paris, seine river, gallery
- [TOP] other sites 1: victor hugo, hugo dame, 1831, hugo hunchback, hugo novel, hugo wrote, dame victor, empty skeleton, miserables, novel dame
- [TOP] other sites 2: 875 875, priests, xvi, pedophilia, churches france, pope benedict, churches attacked, pedophilia church, france vandalized, 875 churches
- [TOP] other sites 3: eiffel, eiffel tower, tower paris, floor eiffel, tower every, see eiffel, francissantamaria eiffel, restaurant eiffel, towereiffel, top eiffel
- [TOP] other sites 4: vatican, catholic church, say vatican, church donors, richest institutions, vatican sitting, much vatican, vatican give, church afford, user vatican
- [TOP] other sites 5: national museum, brazilian billionaire, 88 million, billionaire donated, donated 10, brazilians, brazilians donate, dame give, brazilian woman, find brazilian

- **Politics**

- [TOP] politics 0: emmanuel macron, macron20h, user macron, macron elected, debate emmanuel, macron want, speech, macron speak, macron20h httpurl, president
- [TOP] politics 1: vote, politicians, senate, elected, electoral, president republic, elected officials, republic user, voters, prime minister
- [TOP] politics 2: emmanuel macron, president macron, french president, five, macron dame, macron promises, years httpurl, macron notredame, macron rebuild, rebuild cathedral
- [TOP] politics 3: algeria, sudan, rwanda, genocide, algerians, ottoman, egypt, tunisia, african hemicycles, arab world
- [TOP] politics 4: yellowvests, yellow vests, yellow vest, yellowvests paris, yellowvests acte23, 20 yellowvests, ultimatum2, protest, paris protest, yellowvests actexxii
- [TOP] politics 5: pinault family, henri pinault, million euros, arnault family, bernard arnault, donation, francois, renounces tax, paris pinault, billionaire
- [TOP] politics 6: yellow vests, vests dame, vest movement, paris yellow, vest protesters, rebel yellow, levavasseur calls, funds march, paris protests, vests protesting

List of Interesting Topics for the Venice Flood

Figure B.8 shows the complete timelines of all semantic categories of cultural significance, emotions, and interesting topics detected in the Venice flood dataset. A selection has been previously illustrated in Figure 6.10.

The keywords associated with each detected interesting topic under each theme are listed below, note the emojis are transformed into verbal descriptions:

- **Base**

- [TOP] base 0: water, italy, venice, see, marco, venezia, day, city, san marco, user venice

- **Emotions**

- [TOP] emotions 0: httpurl httpurl, user httpurl, httpurl user, httpurl fuck, httpurl love, httpurl oh, httpurl understand, httpurl new, httpurl god, httpurl excuse
- [TOP] emotions 1: face_with_tears_of_joy, face_with_tears_of_joy face_with_tears_of_joy, face_with_tears_of_joy user, user face_with_tears_of_joy, face_with_tears_of_joy httpurl, loudly_crying_face loudly_crying_face, loudly_crying_face user, loudly_crying_face face_with_tears_of_joy, face_with_rolling_eyes face_with_rolling_eyes, fearful_face fearful_face
- [TOP] emotions 2: red_heart red_heart, africa gem_stone, aristoflownetwork copyright, rè gem_stone, registered ã^{n}d, level zero, beating_heart rè, gem_stone beating_heart, hätè gem_stone, prohibited gem_stone
- [TOP] emotions 3: beaming_face_with_smiling_eyes, eyes smiling_face_with_heart, smiling_face_with_smiling_eyes smiling_face_with_smiling_eyes, smiling_face_with_3_hearts, beaming_face_with_smiling_eyes beaming_face_with_smiling_eyes, beaming_face_with_smiling_eyes user, beaming_face_with_smiling_eyes httpurl, smiling_face_with_3_hearts httpurl, grinning_face_with_smiling_eyes grinning_face_with_smiling_eyes, beaming_face_with_smiling_eyes red_heart

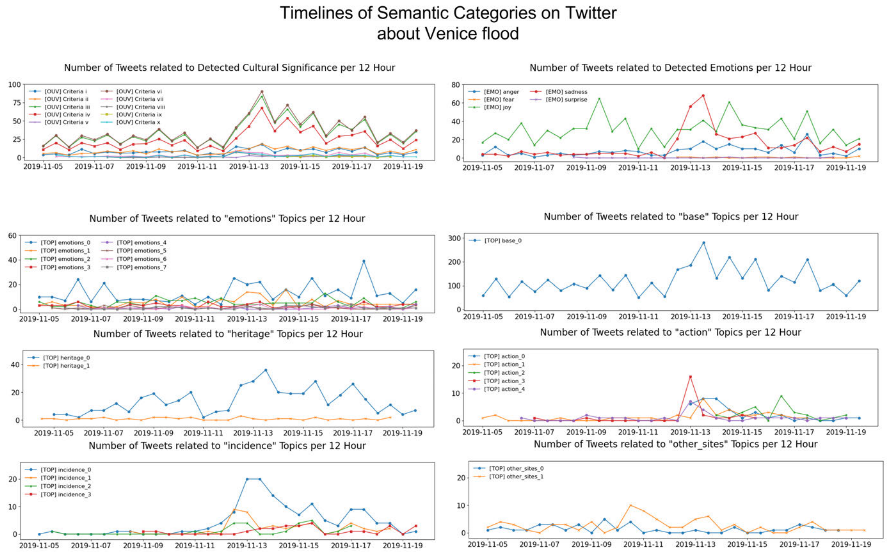


FIG. APP. B.8 The complete timelines showing the temporal development of semantic information along with the HREs in Venice flood.

- [TOP] emotions 4: pleading_face, pleading_face httpurl, pleading_face pleading_face, pleading_face growing_heart, expressionless_face, face_with_monocle face_with_monocle, cazzie video, growing_heart httpurl, exploding_head exploding_head, blue_heart blue_heart
- [TOP] emotions 5: cry httpurl, cried, cried httpurl, facetime, hysterical help, im bitches, inside crying, jessie death, mado raga, lot mom
- [TOP] emotions 6: rolling_on_the_floor_laughing, rolling_on_the_floor_laughing rolling_on_the_floor_laughing, rolling_on_the_floor_laughing face_with_tears_of_joy, rolling_on_the_floor_laughing httpurl, face_with_tears_of_joy rolling_on_the_floor_laughing, venice rolling_on_the_floor_laughing, rolling_on_the_floor_laughing grinning_squinting_face, face_screaming_in_fear rolling_on_the_floor_laughing, rolling_on_the_floor_laughing tagadala7, face_savoring_food face_savoring_food
- [TOP] emotions 7: clapping_hands, clapping_hands clapping_hands, clapping_hands user, user clapping_hands, thumbs_up clapping_hands, clapping_hands flexed_biceps, clapping_hands ok_hand, clapping_hands top_arrow, gesture, clapping_hands httpurl

- **Heritage**

- [TOP] heritage 0: salvini, venice italy, italia, league, venezia italia, venice, venetians, venice matera, venice acquaalta, venezia venice
- [TOP] heritage 1: holiness, shamrock cherry_blossom, sins, shamrock, bright_button shamrock, cherry_blossom bright_button, allah, graduation, deceased, prayer

- **Incidence**

- [TOP] incidence 0: climate, climate change, flooding, floods, flood, worst flooding, flooding venice, flooding 50, venice flooding, global warming
- [TOP] incidence 1: high tide, highest tide, tide 50, centimeters, httpurl small_orange_diamond, flooded highest, city hit, exceptional tide, tide hits, hit highest
- [TOP] incidence 2: sirens, siren, siren sounded, sirens sounded, four tones, sirens venice, alarm siren, httpurl sirens, water_wave water_wave, hear sirens
- [TOP] incidence 3: bookstore, alta bookshop, library venice, bertoni bookshop, acquaalta bookstore, books destroyed, many books, disappointed_face disappointed_face, person_raising_hand person_raising_hand, calle

- **Actions**

- [TOP] action 0: bribes, mose work, operation, project, veneto region, billion mose, venice mose, commissioners, euros spent, galan zaia
- [TOP] action 1: venezia httpurl, httpurl venise, acquaaltaavenezia backhand_index_pointing_down, receives support, blame political, numbers people, read history, companies involved, venessiamia httpurl, unloading
- [TOP] action 2: donate, magna, savevenice oneeuroforoneselfie, shareit helpvenice, helpvenice veneziaacquaalta, salviamovenezia comunedivenezia, euro could, oneeuroforoneselfie salviamovenezia, help city, million
- [TOP] action 3: sanservolo say, mose sanservolo, venezia mose, warning warning, acquaaltaavenezia httpurl, mose acquaalta, senator morra, impeachmenthearings fight_fight_against_cyber_violence, shit pile_of_poo, stikstofcrisis togetherforwonho
- [TOP] action 4: folded_hands, folded_hands folded_hands, venice folded_hands, user prayers, user folded_hands, writing_hand frasinliberta, person_raising_hand person_raising_hand, backhand_index_pointing_down backhand_index_pointing_down, speechless, palms_up_together

- **Other Sites**

- [TOP] other sites 0: biennale, venice biennale, user biennale, biennalearte2019, biennale arte, art gardens, biennale httpurl, biennial contemporary, biennale venezia, biennalearte2019 user
- [TOP] other sites 1: beach, venice beach, beach httpurl, beach boardwalk, venicebeach california, california sunset, caminomasqueunloco losangeles, beach bordwalk, sunset venice, los

Nomenclature

Tables B.7 gives an overview of the mathematical notations used in the Chapter 6.

TABLE APP. B.7 The nomenclature of mathematical notations used in Chapter 6 in alphabetic order.

Symbol	Data Type/Shape	Description
$\mathcal{A}, \mathcal{B}$	Sets of objects	Generic sets.
$\boldsymbol{c}$	Vector of non-negative integers $\boldsymbol{c} := [c_j]_{\lvert\mathcal{C}\rvert\times 1} \in \mathbb{N}^{\lvert\mathcal{C}\rvert\times 1}, c_j = \lvert\{\mathfrak{d}_i \lvert \mathfrak{c}_i = \zeta_j\}\rvert$	The number of tweets that are posted in the cities from the set $\mathcal{C}$.
$\boldsymbol{c}_{\mathsf{B}}, \boldsymbol{c}_{\mathsf{D}}, \boldsymbol{c}_{\mathsf{A}}$	Vectors of non-negative integers $\boldsymbol{c}_{\mathsf{B}}, \boldsymbol{c}_{\mathsf{D}}, \boldsymbol{c}_{\mathsf{A}} \in \mathbb{N}^{\lvert\mathcal{C}\rvert\times 1}, \boldsymbol{c}_{\mathsf{B}} + \boldsymbol{c}_{\mathsf{D}} + \boldsymbol{c}_{\mathsf{A}} = \boldsymbol{c}$	The number of tweets that are posted in each city before, during, and after the event.
$\mathcal{C}$	A set of objects $\mathcal{C} = \{\zeta_0, \zeta_1, ..., \zeta_{\lvert\mathcal{C}\rvert-1}\}$	The unique names of the cities in the dataset.
$\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2$	Sets of objects $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2 \subset \mathcal{C}, \mathcal{C}_0 = \{\zeta_0\}$	The cities that are the ones where the events happened ($\mathcal{C}_0$), from the same country ($\mathcal{C}_1$), or from far beyond ($\mathcal{C}_2$).
χ^2	Scalar value	The Chi-square statistics of two distributions.
$\boldsymbol{d}$	Vector of non-negative floats $\boldsymbol{d} := [d_i]_{K\times 1} \in \mathbb{R}^{K\times 1}$	The geodesic distances of the cities to the city where the event happened (ζ_0).
df	Scalar value	The degree of freedom.
$df*$	Scalar value	The minimum of the number of rows or columns minus 1 for a two-level Chi Square test.
$\mathfrak{d}_i, \mathfrak{D}$	Object Tuples $\mathfrak{d}_i = (\mathcal{S}_i, \mathcal{O}_i, \mathfrak{u}_i, \mathfrak{t}_i, \mathfrak{l}_i)$, $\mathfrak{d}_i \in \mathfrak{D} = \{\mathfrak{d}_0, \mathfrak{d}_1, ..., \mathfrak{d}_{K-1}\}$	The tuple of all raw data (sentences, ID of other associated tweets, user ID, timestamp, and geo-location) from one sample point.
$\mathcal{E}$	A Set of tuples $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$	The link sets denoting all links among the tweets.
$\mathcal{E}^{\mathsf{CONV}}, \mathcal{E}^{\mathsf{USER}}$	Sets of tuples $\mathcal{E} = \mathcal{E}^{\mathsf{CONV}} \bigcup \mathcal{E}^{\mathsf{USER}}$	The link sets denoting respectively the conversational links and the user links among the tweets.
$\mathcal{G}$	Simple directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$	The graph including the the tweets connected with all association types.
$\mathcal{G}^{\mathsf{MULT}}$	Directed multi-graph $\mathcal{G} = (\mathcal{V}, \{\mathcal{E}^{\mathsf{CONV}}, \mathcal{E}^{\mathsf{USER}}\})$	The graph including the conversational links and user links among the tweets.
$\boldsymbol{g}_{\mathsf{BERT}}, \boldsymbol{g}_{\mathsf{ULMFiT}}$	Models as end-to-end functions returning logit vectors	The pre-trained BERT and ULMFiT models on WHOSe Heritage datasets.
H	Scalar value	The statistics of the Kruskal-Wallis H tests.
$\mathsf{OUV}(\mathcal{A}, \mathcal{B})$	A function returning a scalar with sets as inputs	The function calculating the Intersection over Union of two sets.
i, i'	Integer Indices $i, i' \in \{0, 1, 2, .., K-1\} \subset \mathbb{N}$	The index of samples in the dataset $\mathfrak{D}$ of one case heritage-related event.
j	Integer Indices $j \in \{0, 1, 2, .., \lvert\mathcal{C}\rvert - 1\} \subset \mathbb{N}$	The index of cities in the set $\mathcal{C}$ of all unique names of the cities.
k	Integer Indices $k \in \{0, 1, 2, .., \lvert\mathcal{T}\rvert - 1\} \subset \mathbb{N}$	The index of timestamps in the ordered set $\mathcal{T}$ of all unique hours from one case city.

TABLE APP. B.7 Cont.

Symbol	Data Type/Shape	Description
$\boldsymbol{K}^{\mathsf{OUV}}$	Matrix of Floats $\boldsymbol{K}^{\mathsf{OUV}} = [\boldsymbol{\kappa}_i^{\mathsf{OUV}}]_{2\times\lvert\mathcal{V}\rvert}$	The confidence indicator matrix for OUV labels including the top-n confidence and agreement between BERT and ULMFiT models.
$\boldsymbol{k}^{\mathsf{EMS}}$	Vector of Boolean's $\boldsymbol{k}^{\mathsf{EMS}} = [\kappa_i^{\mathsf{EMS}}]_{1\times\lvert\mathcal{V}\rvert}$, $\kappa_i^{\mathsf{EMS}} = \kappa_i^{\mathsf{EM}} \wedge \kappa_i^{\mathsf{SE}}, \kappa_i^{\mathsf{EMS}}, \kappa_i^{\mathsf{EM}}, \kappa_i^{\mathsf{SE}} \in \{0,1\}$	The confidence indicator vector of emotion labels that shows both a consistent emotion prediction ($\kappa_i^{\mathsf{EM}} = 1$) and a similar sentiment prediction ($\kappa_i^{\mathsf{SE}} = 1$) with different models.
K	Integer $K = \lvert\mathfrak{D}\rvert$	The sample size (number of posts) collected in one case event.
$\boldsymbol{l}$	Vector of Floats	A generic vector.
$\mathfrak{l}_i$	Tuple of Floats $\mathfrak{l}_i = (\mathfrak{x}_i, \mathfrak{y}_i, \mathfrak{c}_i)$	The geographical coordinate of latitude ($\mathfrak{y}_i$) and longitude ($\mathfrak{x}_i$) and city name ($\mathfrak{c}_i$) as location of one sample.
m	Integer Indices $m \in \{0, 1, 2, .., \lvert\mathcal{Z}\rvert - 1\} \subset \mathbb{N}$	The index of generated topics $\mathcal{Z}$ from topic modelling.
$\mathrm{max}(\boldsymbol{l}, n)$	Function returning a float with a vector and an integer as inputs	The function returning the value of the n_{th} largest element of a vector $\boldsymbol{l}$.
n	Scalar value	The sample size in a statistical test.
$\boldsymbol{n}$	Vector of integers $\boldsymbol{n} = [1, 2, 3..., \lvert\mathcal{C}\rvert]^{\mathsf{T}}$	The ranking vector of the ordered set $\mathcal{C}$.
$\mathcal{O}_i$	A set of tuples or an empty set $\mathcal{O}_i = \{\mathfrak{d}_{i'} \mid \mathfrak{d}_{i'} \in \mathcal{D}\}$ or $\mathcal{O}_i = \varnothing$	All the tweets that are associated with the tweet $\mathfrak{d}_i$.
p	Scalar value	The significance of a statistical test.
$\mathcal{S}_i$	Set of Strings $\mathcal{S}_i = \{s_i^{(1)}, s_i^{(2)}, ..., s_i^{(\lvert\mathcal{S}_i\rvert)}\}$	The processed textual tweet data as a set of individual sentences that have been translated into English.
$\mathcal{T}$	An ordered Set $\mathcal{T} = \{\tau_0, \tau_1, ..., \tau_{\lvert\mathcal{T}\rvert-1}\}$	The ordered set of all unique timestamps from one case event.
$\mathcal{T}_{\mathsf{B}}, \mathcal{T}_{\mathsf{D}}, \mathcal{T}_{\mathsf{A}}$	Ordered subsets $\mathcal{T}_{\mathsf{B}}, \mathcal{T}_{\mathsf{D}}, \mathcal{T}_{\mathsf{A}} \subset \mathcal{T}$	The ordered set of all unique timestamps before, during, and after the event.
$\mathrm{top\text{-}n}(\boldsymbol{l}, n)$	Function returning a set with a vector and an integer as inputs	The function returning the index set of the largest n elements in the vector $\boldsymbol{l}$.
τ_k	Timestamp $\tau_k \in \mathcal{T}$	A timestamp in the ordered set $\mathcal{T}$ of all unique timestamps.
$\mathfrak{t}_i$	Timestamp $\mathfrak{t}_i \in \mathcal{T}$	A timestamp indexed with sample ID in the ordered set $\mathcal{T}$ of all unique timestamps.
$\boldsymbol{t}$	Vector of non-negative integers $\boldsymbol{t} := [t_k]_{\lvert\mathcal{T}\rvert\times 1} \in \mathbb{N}^{\lvert\mathcal{T}\rvert\times 1}, t_k = \lvert\{\mathfrak{d}_i \mid \mathfrak{t}_i = \tau_k\}\rvert$	The number of tweets that are posted at each unique timestamp.
$\boldsymbol{\Theta}_{\mathsf{BERT}}, \boldsymbol{\Theta}_{\mathsf{ULMFiT}}$	Parameters	Model paramters for the BERT and ULMFiT models.
$\mathcal{U}$	An ordered Set $\mathcal{U} = \{\mu_0, \mu_1, ..., \mu_{\lvert\mathcal{U}\rvert-1}\}$	The ordered set of all unique users from one case event.
$\mathfrak{u}_i$	User ID Object $\mathfrak{u}_i \in \mathcal{U}$	An instance of user indexed with sample ID in the ordered set $\mathcal{U}$ of all unique users.

TABLE APP. B.7 Cont.

Symbol	Data Type/Shape	Description
U	Scalar value	The statistics of the Mann-Whitney U tests.
V	Scalar value	The Cramer's V as effect size for Chi Square tests.
$\mathcal{V}$	A set of nodes $\mathfrak{d}_i \in \mathcal{V}, \mathcal{V} \subset \mathcal{D}$	The set of all nodes of tweets in a case event that are not isolated.
$\mathcal{V}^{\text{OUV}}, \mathcal{V}^{\text{EMS}}, \mathcal{V}^{\text{TOP}}$	Sets of nodes $\mathcal{V}^{\text{OUV}}, \mathcal{V}^{\text{EMS}}, \mathcal{V}^{\text{TOP}} \subset \mathcal{V} \subset \mathcal{D}$	The sets of filtered tweets that are found to give valid predictions on OUV, emotion, and topic labels.
$(\mathfrak{x}_i, \mathfrak{y}_i)$	Geographical Coordinates	The latitude and longitude of the tweet $\mathfrak{d}_i$.
(x_0, y_0)	Geographical Coordinates	The latitude and longitude of the city ζ_0 where the event happened.
$\boldsymbol{Y}^{\text{OUV}}$	Matrix of Floats $\boldsymbol{Y}^{\text{OUV}} = [\boldsymbol{y}_i^{\text{OUV}}]_{11 \times K}$	The OUV labels of tweets as probability distributions on 10 OUV selection criteria and an additional negative class, as the average of prediction from BERT and ULMFiT models.
$\boldsymbol{y}_i^{\text{BERT}}, \boldsymbol{y}_i^{\text{UMLFiT}}$	Logit vector of Floats $\boldsymbol{y}_i^{\text{BERT}} \in [0,1]^{11 \times 1}, \boldsymbol{y}_i^{\text{ULMFiT}} \in [0,1]^{11 \times 1}$	Predicted OUV labels for the tweet $\mathfrak{d}_i$ by BERT and ULMFiT models
$\boldsymbol{y}_i^{\text{EM(0)}}, \boldsymbol{y}_i^{\text{EM(1)}}$	Logit vector of Floats $\boldsymbol{y}_i^{\text{EM(0)}} \in [0,1]^{7 \times 1}, \boldsymbol{y}_i^{\text{EM(1)}} \in [0,1]^{6 \times 1}$	Predicted emotion labels for the tweet $\mathfrak{d}_i$ by pysentimiento and BERTweet emotion models.
$\boldsymbol{y}_i^{\text{SE(0)}}, \boldsymbol{y}_i^{\text{SE(1)}}$	Logit vector of Floats $\boldsymbol{y}_i^{\text{SE(0)}} \in [0,1]^{3 \times 1}, \boldsymbol{y}_i^{\text{SE(1)}} \in [0,1]^{3 \times 1}$	Predicted sentiment labels for the tweet $\mathfrak{d}_i$ by pysentimiento and BERTweet sentiment models.
$\boldsymbol{y}_i^{\text{TOP}}$	Logit vector of Floats $\boldsymbol{y}_i^{\text{TOP}} = \left[y_{i,m}^{\text{TOP}}\right]_{\lvert\mathcal{Z}\rvert \times 1} \in [0,1]^{\lvert\mathcal{Z}\rvert \times 1}$	Predicted topic labels for the tweet $\mathfrak{d}_i$ with topic modelling from BERTopic.
$\boldsymbol{\mathcal{Y}}^{\text{EMS}}$	Array of sets $\boldsymbol{\mathcal{Y}}^{\text{EMS}} = \left[\mathcal{Y}_i^{\text{EMS}}\right]$	The array of final emotion labels for all the tweets in $\mathcal{V}$, containing the top-1 emotions and top-1 sentiments if the prediction is valid, otherwise empty.
$\boldsymbol{\mathcal{Y}}^{\text{OUV}}$	Array of sets $\boldsymbol{\mathcal{Y}}^{\text{OUV}} = \left[\mathcal{Y}_i^{\text{OUV}}\right]$	The array of final OUV labels for all the tweets in $\mathcal{V}$, containing the top-3 OUV selection criteria if the prediction is valid, otherwise empty.
$\boldsymbol{\mathcal{Y}}^{\text{TOP}}$	Array of sets $\boldsymbol{\mathcal{Y}}^{\text{TOP}} = \left[\mathcal{Y}_i^{\text{TOP}}\right]$	The array of final emotion labels for all the tweets in $\mathcal{V}$, containing the topic name that has a higher probability than 0.5 if the prediction is valid, otherwise empty.
$\mathcal{Z}$	A set of objects $\mathcal{Z} = \{z_m \mid m = 0, 1, \ldots, \lvert\mathcal{Z}\rvert - 1\}$	The set of the generated topics obtained with BERTopic topic modelling.
$\mathcal{Z}^{\text{S}}$	A subset of objects $\mathcal{Z}^{\text{S}} \in \mathcal{Z}$	A subset of the generated topics obtained with BERTopic topic modelling that are interesting and informative for heritage management.
ζ_0	An object $\zeta_0 \in \mathcal{C}, \mathcal{C}_0 = \{\zeta_0\}$	The name of the city where the event happened.
ζ_j	An object $\zeta_j \in \mathcal{C}$	The name of a city that is one instance of the set $\mathcal{C}$.

Publications

Journal Articles

Bai, N, Nourian P, Luo R, Cheng T, Pereira Roders, A. (2023). Screening the Stones of Venice: Mapping Social Perceptions of Cultural Significance through Graph-based Semi-supervised Classification. ISPRS Journal of Photogrammetry and Remote Sensing. 203, 135-164. https://doi.org/10.1016/j.isprsjprs.2023.07.018

Bai, N, Nourian, P, Pereira Roders, A, Bunschoten, R, Huang, W, Wang, L. (2023). Investigating rural public spaces with cultural significance using morphological, cognitive and behavioural data. Environment and Planning B: Urban Analytics and City Science. 50(1), 94-116. https://doi.org/10.1177/23998083211064290

Bai N, Nourian P, Luo R, Pereira Roders A. (2022). Heri-Graphs: A Dataset Creation Framework for Multi-Modal Machine Learning on Graphs of Heritage Values and Attributes with Social Media. ISPRS International Journal of Geo-Information. 11(9): 469. https://doi.org/10.3390/ijgi11090469

Bai N. (2018). Form and Content driven by Ideology, a brief Analysis on the National Art School of Cuba (in Chinese). Huazhong Architecture. (05), p. 41-50. http://dx.doi.org/10.13942/j.cnki.hzjz.2018.05.005

Bai N, Wang L, Sun P. (2017). The Influence of Architectural Environment and Educational Background on Spatial Judgment Preference (in Chinese). Journal of Human Settlements in West China. 32(01). p. 49-56. http://dx.doi.org/10.13791/j.cnki.hsfwest.20170108

Nourian P, Azadi S, **Bai N**, de Andrade B, Abu Zaid N, Rezvani S, Pereira Roders A. EquiCity Game: A mathematical serious game for participatory design of spatial configurations. (Under Review).

Bai, N, Nourian P, Pereira Roders, A. Mapping the User-Generated Content for Researching Cultural Heritage - A Systematic Literature Review. (Under Preparation)

Bai, N, Nourian P, Cheng T, Pereira Roders, A. Network-based Spatiotemporal Mapping of Heritage-Related Events Detected on Social Media. (Under Preparation).

Conference Papers

Bai N, Ducci M, Mirzikashvili R, Nourian P, Pereira Roders, A. (2023). Mapping Urban Heritage Images with Social Media Data and Artificial Intelligence, A Case Study in Testaccio, Rome. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-M-2-2023. p. 139–146.
10.5194/isprs-archives-XLVIII-M-2-2023-139-2023

Bai N. Luo R, Nourian P, Pereira Roders, A. (2021). WHOSe Heritage: Classification of UNESCO World Heritage" Outstanding Universal Value" Documents with Soft Labels. In Findings of the Association for Computational Linguistics: EMNLP 2021. p. 366-384. Association for Computational Linguistics.
http://dx.doi.org/10.18653/v1/2021.findings-emnlp.34

Bai N, Nourian P, Luo R, Pereira Roders A. (2021). "What is OUV" Revisited: A Computational Interpretation on the Statements of Outstanding Universal Value. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, VIII-M-1-2021. p. 25–32.
https://doi.org/10.5194/isprs-annals-VIII-M-1-2021-25-2021

Bai N, Nourian P, Pereira Roders A. (2021). Global Citizens and World Heritage: Social Inclusion of Online Communities in Heritage Planning. In The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-M-1-2021. p. 23–30.
https://doi.org/10.5194/isprs-archives-XLVI-M-1-2021-23-2021

Bai N, Azadi S, Nourian P, Pereira Roders A. (2020). Decision-Making as a Social Choice Game: Gamifying an urban redevelopment process in search for consensus. In Anthropologic – Architecture and Fabrication in the cognitive age - Proceedings of the 38th eCAADe. Vol. 2. p. 555-564.
https://doi.org/10.52842/conf.ecaade.2020.2.555

Bai N, Nourian P, Xie A, Roders AP. (2020). Towards a finer heritage management: Evaluating the tourism carrying capacity using an agent-based model. In RE: Anthropocene, Design in the Age of Humans - Proceedings of the 25th CAADRIA. Vol. 1. p. 305-314.
https://doi.org/10.52842/conf.caadria.2020.1.305

Bai N, Wang L. (2019). Village Heritage Conservation in the New Data Age: Rural Information Modelling in the Context of Rural Vitalization in China. In Intelligent and Informed - Proceedings of the 24th CAADRIA. Vol. 2. p. 41-50.
https://doi.org/10.52842/conf.caadria.2019.2.041

Xu W, Ma H, **Bai N**. (2019). The Influence of Spatial Vitality around Subway Stations in Beijing on Pedestrians' Emotion. In Differences and Integration of Urban and Rural Environment Proceedings of the 13th International Symposium for

Environment-Behavior Studies. p. 268-274

Bai N, Ye W, Li J, Ding H, Pienaru MI, Bunschoten R. (2018). Customised Collaborative Urban Design – A Collective User-based Urban Information System through Gaming. In Computing for a better tomorrow - Proceedings of the 36th eCAADe. Vol. 1. p. 419- 428. https://doi.org/10.52842/conf.ecaade.2018.1.419

Bai N, Huang W. (2018). Quantitative Analysis on Architects Using Culturomics – Pattern Study of Pritzker Winners Based on Google N-gram Data. In Learning, Adapting and Prototyping - Proceedings of the 23rd CAADRIA. Vol. 2. p. 257-266. https://doi.org/10.52842/conf.caadria.2018.2.257

Book Chapters

Bai N, Pereira Roders, A, & Corten, J. P (2021). Roundtable IV: Heritage, Digitalization and Sustainability. In U. Pottgiesser, S. Fatoric, C. Hein, E. de Maaker, & A. Pereira Roders (Eds.), LDE Heritage Conference on Heritage and the Sustainable Development Goals: Proceedings (pp. 525-527). TU Delft Open. https://books.bk.tudelft.nl/index.php/press/catalog/book/781

Nourian P, **Bai N**. (2021). Reuse: On Evaluating the Fitness of Spatial Configurations Before & After Retrofitting for Reuse of Architectural Heritage. In A. Pereira Roders (Ed.), Mastermind Crash: A method to reveal the impact of architectural redesign (pp. 23-72). TU Delft Open. https://pure.tudelft.nl/ws/portalfiles/portal/101426392/

Nourian P, Azadi S, Uijtendaal R, **Bai, N**. Augmented Computational Design. In N. Abbasabadi & M Ashayeri (Eds.), Artificial Intelligence in Performance-driven Design: Theories, Methods, and Tools Towards Sustainability. Wiley. (Accepted).

Castro A, Pereira Roders A, Burgers G, Pendlebury J, Swideerski M, **Bai N**, Manen N, Wagenaar P, Nourian P, Meparishvili T. Work Package 2: Democratization. In Heriland Report | Heritage Planning Handbook (Under Review).

Popular Writings

Bai N. (2023). The Dusts of Venezia: A Story of Kai, the Great Scavenger. (Under Review).

Curriculum Vitæ

Nan BAI 白楠

03-08-1993	Born in Lanzhou, China

Education

2012–2016 Bachelor of Architecture
Tsinghua University, Beijing, China

Thesis: The Colosseum for the Young: Designing an Interactive Playground for Primary School Pupils

Supervisor: Prof. dr. L. Zhang

2014–2016 Bachelor of Science in Psychology (dual degree)
Tsinghua University, Beijing, China

Thesis: The Influence of Architectural Environment and Educational Background on Spatial Judgment Preference

Supervisor: Prof. dr. P. Sun

2014–2015 Exchange Student in Architecture
Kalsruher Institut für Technologie, Karlsruhe, Germany

2016–2019 Master of Architecture
Tsinghua University, Beijing, China

2017–2019 Master of Science in Architektur (dual degree)
Die Technische Universität Berlin, Berlin, Germany

Thesis: Morphology, Cognition and Behavior: the Image of the Villages and the Rural Information Modelling

Supervisors: Prof. dr. L. Wang, Prof. dr. R. Bunschoten

2019 Doctor of Philosophy
Technische Universiteit Delft

Thesis: Sensing the Cultural Significance with AI for Social Inclusion

Promotors: Prof. dr. A. Pereira Roders, Dr. P. Nourian

Experience

2022–2023 Visiting Researcher
SpaceTimeLab, University College London, London, United Kingdom

Conducting a research project on spatial-temporal big data mining of Twitter data during heritage-related events

Giving Guest lectures MSc Course “Spatial-Temporal Data Analysis and Data Mining”

2020–2023 Teaching Assistant
Technische Universiteit Delft

Developing and lecturing MSc Course “MASTERMIND – Heritage and Values”

2019 Research Intern
Palace Museum, Beijing, China

Conducting a research project on tourism carrying capacity in Yangxin Dian using Agent-based Models and Simulation

2019 Teaching Assistant
Tsinghua University, Beijing, China

Lecturing BArch Course “Chinese Vernacular Architecture”

2018-2019 Student Researcher
Tsinghua University, Beijing, China

Involved in National Key Research and Development Program of China (NO. 2018YFD1100303) for big data monitoring of rural vitalization

Awards

2023 Best Paper Award, 29th CIPA Symposium on Heritage Documentation

2020	Young CAADRIA Award, 25th International CAADRIA Conference
2019	Marie Sklodowska-Curie Scholarship, European Commission
2018	National Graduate Scholarship in China, Tsinghua University
2015	Ni Tianzeng Scholarship, Tsinghua University
2014	Baden-Württemberg Scholarship, Karlsruhe, Germany
2012	Second prize of Freshman Scholarship, Tsinghua University

Supervision

2023 — MSc Xingyu Fang (As 2nd Supervisor)
University College London.

Thesis: Semi-Supervised Learning-Based UNESCO World Cultural Heritage News Text Classification and Spatio-Temporal Analysis.

Supervisors: Prof. dr. T. Cheng, **N. Bai**

2023 — MSc Chunyu Jin (As 2nd Supervisor)
University College London.

Thesis: Spatio-Temporal-Semantic Event Detection in World Heritage News Data: A Hybrid Approach using BERTopic and MDST-DBSCAN.

Supervisors: Prof. dr. T. Cheng, **N. Bai**

2022-2023 — MSc Wen-Yu Chen. (As 5th Examiner)
Leiden University and Delft University of Technology.

Thesis: To Redefine, Not Reinforce – A Spatial Decision Support System with Generative Design Model for Exploring Optimal Improvements to Existing Street Networks for Enhancing Equity of Accessibility.

Supervisors: Dr.ir. T. Verma, Dr. P. Nourian, Dr. J. E. Goncalves, R. Nelson, **N. Bai**

Public Talks

2023.07.20	**Bai N**, Cheng T, Nourian P, Pereira Roders A. An Exploratory Data Analysis of the Spatiotemporal Patterns of Heritage-Related Events on Twitter. In The 30th International Conference on Geoinformatics (CPGIS 2023). July 19-21, 2023, University College London, London, the UK.
2023.05.05	**Bai N**. Heri-Graphs – Using Social Media Data to Construct Graph Datasets for Heritage Studies and Graph-based Multi-modal Machine Learning. In "AI and Ethics in Academic Research" Symposium. May 05, 2023, the University of York, the UK.
2023.04.21	**Bai N**. HeriGraph – Using Social Media Data to Construct Graph Datasets for Heritage Studies and Graph-based Multi-modal Machine Learning. In Open Research Meeting at Eindhoven University of Technology. April 21, 2023, Eindhoven University of Technology, the Netherlands.
2023.03.30	Gherbreab S, **Bai N**, Tenzer M, van Manen N. Panel Discussions for Heriland Webinar AI & The Four Worlds of 2050. March 30, 2023.
2022.11.30	**Bai N**, Weyermans I, van den Ende T. Digital Citizen Engagement with Heritage | Future Making. In the Anthropocene Podcast. European Heritage Tribune, November 30, 2022.
2022.10.20	**Bai N**. Collective Perception of Cultural Heritage Values on Social Media. In AET Bites, TU Delft, October 20, 2022.
2022.09.06	**Bai N**, Nourian P, Luo R, Pereira Roders A. HERI-GRAPHS: Constructing Semi-supervised Machine Learning Datasets of Heritage Values and Attributes for Sustainable Urban Heritage Management using Social Media Data. In Session Space Technology and Big Data Facilitate the Sustainability of Cultural Heritage of 2022 International Forum on Big Data for Sustainable Development Goals (FBAS 2022). September 6, 2022, Beijing, China.
2021.10.30	**Bai N**, Nourian P, Luo R, Pereira Roders A. "What is OUV" Revisited: A Computational Interpretation on Statements of Outstanding Universal Value. In City+ International Conference 2021, October 30, 2021, Politecnico di Milano, Italy.
2021.04.10	**Bai N**. Social Inclusion in Cultural Heritage Planning: Mapping Perceived Heritage Values using User-Generated Content in Social Media Platforms. In Our World Heritage- Transformational Impact of Information Technology Globinar2.0. April 10, 2021.
2021.02.24	**Bai N**, Nourian P. Small World by Concerned Citizens. In Symposium on Future Landscape. February 24, 2021, Vrije University Amsterdam, the Netherlands.

Languages

Chinese (Native), English (C1-C2), Dutch (B2-C1), German (B1-B2), Italian (A1-A2), French (A1-A2)

23#17

Sensing the Cultural Significance with AI for Social Inclusion

A Computational Spatiotemporal Network-based Framework of Heritage Knowledge Documentation using User-Generated Content

Nan Bai

Social Inclusion has been growing as a goal in heritage management. Whereas the 2011 UNESCO Recommendation on the Historic Urban Landscape (HUL) called for tools of knowledge documentation, social media already functions as a platform for online communities to actively involve themselves in heritage-related discussions. Such discussions happen both in “baseline scenarios” when people calmly share their experiences about the cities they live in or travel to, and in “activated scenarios” when radical events trigger their emotions. To organize, process, and analyse the massive unstructured multi-modal (mainly images and texts) user-generated data from social media efficiently and systematically, Artificial Intelligence (AI) is shown to be indispensable. This thesis explores the use of AI in a methodological framework to include the contribution of a larger and more diverse group of participants with user-generated data. It is an interdisciplinary study integrating methods and knowledge from heritage studies, computer science, social sciences, network science, and spatial analysis. AI models were applied, nurtured, and tested, helping to analyse the massive information content to derive the knowledge of cultural significance perceived by online communities. The framework was tested in case study cities including Venice, Paris, Suzhou, Amsterdam, and Rome for the baseline and/or activated scenarios. The AI-based methodological framework proposed in this thesis is shown to be able to collect information in cities and map the knowledge of the communities about cultural significance, fulfilling the expectation and requirement of HUL, useful and informative for future socially inclusive heritage management processes.

A+BE | Architecture and the Built Environment | TU Delft BK