Using social media data to understand the urban green space use before and after a pandemic

Nan Cui

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Geography

May, 2023

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The following chapters contain jointly authored manuscripts where Nan Cui is the lead author:

The work in Chapter 3 of the thesis has appeared in the publication as:
Cui, N., Malleson, N., Houlden, V. and Comber, A., 2021. Using VGI and social media data to understand urban green space: a narrative literature review. *ISPRS International Journal of Geo-Information*, *10*(7), p.425.
Nan Cui was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Alexis Comber, Nick Malleson and Vikki Houlden are the supervisors and supervised this study. They provided conceptual direction, advised on study design, analyses and discussion, and made comments and edits on the manuscript throughout.

The work in Chapter 4 of the thesis has appeared in the publication as:
Cui, N., Malleson, N., Houlden, V. and Comber, A., 2022. Using social media data to understand the impact of the COVID-19 pandemic on urban green space use. *Urban Forestry & Urban Greening*, 74, p.127677.
Nan Cui was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Alexis Comber, Nick Malleson and Vikki Houlden are the supervisors and provided conceptual direction, advised on study design, analyses and discussion, and made comments and edits on the manuscript from draft to proofs.

The work in Chapter 5 of the thesis has appeared in the publication as:
Cui, N., Malleson, N., Houlden, V. and Comber, A., Urban green space topics based on structural topic modelling during the COVID-19 pandemic. Submitted to a peer-reviewed journal.
Nan Cui was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Alexis Comber, Nick Malleson and Vikki Houlden

are the supervisors and supervised this study. They provided conceptual direction, advised on study design, analyses and discussion, and made comments and edits on the manuscript from draft to proofs.

Acknowledgements

I am deeply grateful for the unwavering support and guidance of my three supervisors Lex Comber, Nick Malleson, and Vikki Houlden, each of whom has contributed significantly to my academic and personal growth during my PhD journey.

Firstly, I would like to express my sincere appreciation to my main supervisor, Lex Comber. His expertise in GIS, R, and space-time analysis has been instrumental in shaping the technical direction and quality of my research. His patient guidance, insightful feedback, and dedication to my success have been invaluable, and I have learned so much from him about conducting spatial analysis using R. I feel fortunate to have had the opportunity to work under his supervision. Secondly, I would like to thank my second supervisor, Nick, for his critical thinking and review skills. His guidance in reviewing journal articles and providing constructive feedback has been invaluable in shaping the structure of my thesis. His insights have been an important factor in developing my critical thinking skills and have greatly benefited my research. Finally, I would like to express my gratitude to Vikki. Her passion, kindness, and patience have been instrumental in improving my writing skills. Her detailed comments and edits on my manuscripts have been a valuable resource, helping me to refine my writing and communication skills. I appreciate her dedicated support and encouragement throughout my PhD journey.

I feel blessed to have had such wonderful supervisors who have been so supportive and invested in my success. I am grateful for the experience of working with such knowledgeable and inspiring mentors, and I will always treasure their guidance and advice.

Thanks to the School of Geography, which is an extremely friendly and encouraging place to work and I would like to thank everyone in the school who helped me in this endeavour from application to completion. Thanks to the China Scholarship Council (No. 201906390033) and The University of Leeds for funding this PhD project. Huge thanks to my family and friends, for bearing with me and sharing with me the joy and frustration of completing this study.

Abstract

Urban green spaces (UGSs) are essential components of urban ecosystems that provide considerable benefits to residents, including recreational opportunities, improved air and water quality, and mental and physical health benefits. The COVID-19 pandemic and related restriction measures have affected people's daily lives in numerous ways, such as remote working and learning, online shopping, social distancing, travel restrictions, and outdoor activities. During the COVID-19 pandemic, UGSs have become the main places for outdoor activities. Understanding human-environment interactions in UGSs is an important research field that has broad implications for improving policies in response to a social crisis and informing urban planning strategies.

The main challenges of investigating human-environment interactions lie in effectively collecting research datasets that can reflect or reveal human behaviour patterns within UGSs. Volunteered Geographical Information (VGI) and social media can provide better information about real-time perceptions, attitudes and behaviours than traditional datasets such as surveys and questionnaires. This provides great opportunities to investigate human-environment interactions in UGS in real-time. Additionally, Twitter is one of the most popular social networks, and it can provide more comprehensive and unbiased datasets through a new academic research Application Programming Interface (API).

The overall aim of this thesis is to evaluate the contributions of UGS to human well-being, during a time of crisis, by investigating the characteristics and spatial-temporal patterns of UGS use across three periods: pre-, during- and after the COVID-19 pandemic. The thesis will document the process of examining spatial-temporal changes in UGS use associated with COVID-19 related pandemic, by using Twitter datasets incorporating approaches including text mining, topic modelling and spatial-temporal analysis. This is the first study to examine social media data over consistent time period before, during and after the lockdown in relation to UGS. The results show that the findings and method can potentially inform policy makers in their management and planning of UGS, especially in a period of social crisis like the COVID-19 pandemic. This research has great potential to help improve urban green space planning and management in urban areas.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Study Background

Urban green space (UGS) refers to urban areas covered by vegetation such as grass, trees, woods, shrubs, or other kinds of plants (Derkzen et al., 2015). UGS can range from small parks and gardens, cemeteries and green street corridors, to large parks and nature reserves (Breuste et al., 2013). Definitions vary, for example, in Poland (Feltynowski et al., 2018), UGS also includes municipal forests and zoological / botanical gardens and, definitions may vary depending on specific local objectives. OpenStreetMap for example defines 17 categories of UGS (https://www.openstreetmap.org), the European Union Urban Atlas defined 3 in 2006 and 8 in 2012 (Kabisch et al., 2016; Chałka et al., 2011; Feltynowski et al., 2018). In the U.K., the national mapping agency (Ordnance survey) defines 10 categories within the OS MasterMap Greenspace Layer (Dennis et al., 2018). Because of this variation it is important to define the what is meant by 'UGS' and to identify an appropriate UGS dataset before undertaking a specific analysis. In this thesis, UGS is defined as all publicly owned and publicly accessible open space that is covered by vegetation or sports facilities, e.g., parks, gardens, woodlands, nature areas, sports courts, cemetery, and other green space within the city boundary area (Schipperijn et al., 2013; Ordnance Survey, 2021).

UGS provides a number of environmental, social, and economic benefits to citizens and plays an essential role in maintaining and enhancing a sustainable city. It provides a number of ecosystem services, for instance, improving air quality, enhancing urban biodiversity, reducing noise and urban heat island effect (Liu, O.Y. and Russo, 2021; Burgess et al., 1988). UGS has been found to provide numerous benefits to public health and human wellbeing, including improving mental health, reducing stress, promoting physical activity, and enhancing social cohesion and community engagement (Gascon et al., 2015; Ronchi et al., 2020). An increasing number of studies have investigated the benefits of the interactions between humans and UGS (Kabisch et al., 2015; Houlden et al., 2019), confirming the relationship between UGS with physical, mental, and social well-being (Campbell et al., 2016). UGS is therefore recognised as component of urban sustainability and for enhancing the quality of life for citizens (Kim and Jin, 2018).

As the proportion of population that is urban increases (from 50 % in 2010 to nearly 70 % by 2050 (Sachs et al., 2019), the need for UGS will also increase requiring the planning and management of UGS provision (Haaland et al., 2015), in order to make public places more liveable and sustainable (Kashef, 2016). The contribution to human well-being of UGS has the potential to inform future urban planning and decisions about how to incorporate green spaces into the built environment. One way of quantifying the positive effect of UGS is to examine UGS visitation and the activities undertaken within UGSs.

This study examines UGS visitation in London by examining the content of geo-located microblogging posts. London, located in south-east England, is the capital city and the largest city of the United Kingdom, with a population of around 9 million (Office for National Statistics, 2022). As one of the world's major global cities, London exerts a strong influence on its arts, entertainment, fashion, commerce and finance, education, health care, media, science and technology, tourism, and transport and communications. London’s diverse cultures encompass over 300 languages, made it Europe third-most populous city.

In 2019, the National Park City Foundation confirmed London as the world's first National Park City (Townshend et al., 2018; National Park City Foundation, 2023). The aims of the London National Park City were to encourage more people to enjoy the outdoors and make the city greener, healthier and wilder (We are London, 2023). London is therefore selected as study area because of various UGSs in urban areas, high density population and large number of social media users from diverse cultural, educational and professional backgrounds. However, this thesis solely focuses on London as the study area, without considering other cities. As a result, the research findings can only shed light on UGS visitation patterns within London, which may not fully represent other global cities with varying populations, diverse cultures, and different racial demographics. To comprehensively capture the global impact of the COVID-19 pandemic on UGS visitation, comparative studies involving multiple cities around world are required.

Previous studies have investigated UGS interactions via visitation frequencies (Heikinheimo et al., 2020), attitudes to and perceptions of UGS (Plunz et al., 2019), UGS activities (Zhang, W. et al., 2015), and factors associated with UGS visitation (Zhang, S. and Zhou, 2018). These studies mostly used traditional methods of UGS data collection including survey questionnaires and observation (Brown, Greg et al., 2018; Hamstead et al., 2018; Roberts, H.V., 2017; Wang et al., 2018; Cohen et al., 2010; Heather E. Wright Wendel, 2012). However, such approaches are time consuming, often have low response participant rates causing potential uncertainties and biases (Evenson et al., 2013), and may generate site-specific, results that are difficult to generalise (Kovacs-Györi et al., 2018), reducing their utility in UGS planning (Evenson et al., 2013).

Social media such as Facebook, Twitter, Instagram, Flickr provide popular platforms where users can post and share their views, opinions, feelings, emotions and communicate when they visit UGSs (Liu, H. et al., 2017). The use of social media data has become increasingly popular in the field of UGS research, as it potentially provides great opportunities to investigate human behaviours and attitudes around UGS in real-time (Al-Kodmany, 2019). Applications of social media in the analysis of UGS include assessments of park use, the factors affecting the park visitation (Gu et al., 2016; Ladle et al., 2018; Liu, H. et al., 2017), exposure to and accessibility of UGS (Guo et al., 2019; Song et al., 2018; Tao et al., 2019), identification of UGS recreational resources and service areas (Upton et al., 2015; Zhai et al., 2018; Zhen et al., 2017), urban forest management (Brown, Gregory et al., 2014; Foster et al., 2015; Steenberg et al., 2019) and human health (Liu, K. et al., 2016). Social media data can be used to analyse UGS as the data (1) provides much useful information containing users' thoughts, perceptions, experience, which reflect their connection between users and their surroundings (Guerrero et al., 2016; Song et al., 2018); (2) allows users to post their content with a georeferencing and a timestamp allowing the spatial and temporal patterns of UGS uses to be examined (Roberts, H. et al., 2017; Sonter et al., 2016; Wood et al., 2013), and (3) can be treated as a method to supplement traditional methods which are spatially and temporally-limited (Sim et al., 2019).

Twitter is an online social media and social networking service (Kwak, 2010). Twitter was created in March 2006 and launched in July of that year. People can register as Twitter users can post limitation of 280 characters in theirs posts and they can post geo-tagged Tweets providing rich information including text content, coordinates, and timestamps. In April 2022, Twitter made the ability to add and view captions globally available. Twitter provides free apps for all of the mobile phones including iPhone, iPad, and Android, which enable users to post Tweets anywhere at any time, especially in urban areas where there are excellent internet and Wi-Fi services. The total number of Twitter active users in the UK is 19.05 million as of October 2021. In January 2020, male users accounted for 58.5% and female accounted for 41.5% of UK Twitter users (Strugar, 2022).

At least 500 million Tweets are sent on a daily basis, meaning there are plenty of conversations to join.

The thesis aims to examine spatial-temporal changes in UGS use in London associated with COVID-19 related lockdowns by using social media data. Twitter will be selected as data sources with advantages of data accessibility, high spatial-temporal resolution, and large number of Twitter users in London (Roberts, H.V., 2017; Brown, Greg et al., 2018). Additionally, Twitter research academic API (application programming interface) can provide full historical geo-referenced Tweets, which allow researchers to conduct comprehensive analyses (Pfeffer et al., 2022; Barrie and Ho, 2021). Tweet metadata typically provide information such as (1) the content of the Tweet, (2) timestamp, and (3) geographical information, with the latter being voluntarily added by Twitter users. Only Tweets that were posted within UGS areas will be selected for the analysis, these Tweets are refers to as UGS Tweets in this thesis.

The coordinates information and timestamps of Twitter data will be used to investigate the spatial-temporal patterns of UGS visitation. In detail, coordinates information will be employed for identifying and examining the popularity of different types of UGS in London, and exploring and comparing the characteristics of spatial patterns of UGS visitation across the three years. The timestamps of Tweets will be used to group Tweets into daily and hourly patterns, which enable us to generate heat maps to display the evolution of the temporal patterns of UGS visitation. The Tweet content will be used to detect the topics and related activities. In this part of analysis, structural topic model (STM) will be used to investigate the dynamics of UGS-related topics pre-, during- and after the COVID-19 outbreaks. In STM, the content covariate of the COVID-19 will be added into the modelling which enables us to deeply understand how UGS related topics were influenced by the pandemics.

However, it is important to acknowledge that this analysis has some limitations. First of all, issues of data bias and its representativeness should be noted. For example, Twitter users tend to be younger and have relatively higher levels of education compared to the general population. Besides, English is dominant language on Twitter, which may lead to language bias and exclusion of non-English-speakers. Thirdly, the analysis of Tweets might only capture the active Twitter users' UGS visitation patterns and opinions, but overlook those who are less active or do not use Twitter when they visited UGS.

Additionally, this thesis only utilizes Twitter data as its data source, thereby capturing the UGS use patterns solely from Twitter users, rather than providing insights into the broader population of UGS users in London. Thus additional research is warranted to attain a comprehensive understanding of how COVID-19 influenced UGS use.

In early 2020, a new coronavirus known as COVID-19 was identified in Wuhan China (Holmes et al., 2021) and declared as a pandemic by the WHO on 11March 2020 (Tison et al., 2020; World Health Organization, 2020). This led to significant changes in human mobility and behaviours as governments around the world implemented restrictions including the closure of businesses, cancelation of public events and schooling, social distancing, limitations on the size of social gatherings, and travel restrictions. While previous UGS studies have assessed UGS benefits and visitation during normal times, little research has examined the contributions of UGS during times of crisis such as a pandemic. The COVID-19 pandemic affected people's daily lives in numerous ways such as remote working and learning, online shopping, social distancing, travel restrictions, and outdoor activities (Chakraborty and Maity, 2020; Haleem et al., 2020; Theberath et al., 2022; Jackson et al., 2021; Pierce et al., 2020) and during the COVID-19 pandemic, UGS has become the main place for outdoor activities. A number of studies have shown that access to parks and gardens was beneficial to people’s health and wellbeing, especially during this period (Poortinga et al., 2021; Lesser and Nienhuis, 2020). In this context, this project uses social media data to investigate the impact of the COVID-19 pandemic on UGS use and the potential contribution of UGS to human well-being, particularly during a time of crisis.

1.2 Research aims and objectives

The overall aim of this work is to evaluate the contributions of UGS to human wellbeing, during a time of crisis, by investigating the characteristics spatial-temporal patterns of UGS use across three periods: pre-, during- and after the COVID-19 pandemic. The thesis first synthesises existing knowledge on the use of social media data in the field of UGS research. This allowed a number of future research directions to be identified from the perspectives of data availability, quality, analysis methods, through a literature review with meta-analysis. The thesis also investigates the spatial-temporal patterns of UGS visitation to assess how contributed to UGS users’ daily lives, and to better understand the dynamics in UGS activities, a series of text mining incorporating topic analysis and spatial modelling were undertaken.

These aims and desired outcomes were accomplished by breaking down the research into a number of objectives, as summarised in Table 1.1.

Table 1. 1 Research objectives of the thesis.

No.	Research objectives	Corresponding chapter(s)
I.	1) Obtain a comprehensive understanding of the use of social media data in the context of UGS visitation, including data availability, data collection, dataset quality, and analysis methods. 2) Summarise the research topics in previous studies that have explored VGI and social media in relation to UGS use, and examined the availability, quality, and analysis methods of VGI and social media data for UGS research. 3) Propose the guidelines for the future analysis of use social media data in the field of UGS research, including during a time of social crisis.	Chapter 3
II.	4) Design an appropriate strategy for collecting, pre-processing, storing and managing social media datasets, and demonstrate the utility of social media data for UGS planning and management. 5) Examine the impact of the COVID-19 pandemic on UGS use before-, during- and after the first lockdown period. 6) Investigate the characteristics of spatial and temporal patterns of UGS visitation in the context of COVID-19 lockdown period. Provide potential recommendations for future urban planning and management in response to the complexity of urban environments and emergencies such as COVID-19 pandemic.	Chapter 4
III.	7) Use structural topic modelling (STM) to comprehensively investigate the dynamics of the topics discussed by social media users in UGSs, to potentially reveal the contributions of UGS to human wellbeing, particularly in a time of crisis. 8) Explore the topics about, and attitudes to UGS expressed through Tweets, and the evolution of topic content and topic prevalence, with the context of a COVID-19 pandemic. 9) Investigate the spatial-temporal dynamics of the important topics in relation to the impact of a pandemic on UGS activities, and provide suggestions and guidance for future UGS policies, especially.	Chapter 5

1.3 Thesis structure

This thesis is presented in the alternative format as described by the Faculty of Environment, University of Leeds. The thesis consists of six chapters. Chapters 3, 4 and

5 take the form of a jointly-authored manuscripts where Nan Cui is the lead author, all of which have been published or submitted to peer-reviewed journals.

Chapter 2 provides a research context of the UGS related studies including during the COVID-19 pandemic. The methodologies for understanding the interactions between human and UGS by using social media data and for using meta-analysis to conduct a literature review were introduced. This chapter highlights limitations and current research gaps and synthesises existing knowledge of the social media data analysis, within the context of UGS research. It identifies that a dynamic analysis of UGS visitation could help understand the contributions of UGS to human wellbeing before-, during- and after a pandemic. The findings of this chapter motivate the direction of the research in the following chapters.

The work in Chapter 3 of the thesis has appeared in the publication as:

Cui, N., Malleson, N., Houlden, V. and Comber, A., 2021. Using VGI and social media data to understand urban green space: a narrative literature review. *ISPRS International Journal of Geo-Information, 10*(7), p.425. doi: https://doi.org/10.3390/ijgi10070425.

Chapter 3 employs meta-analysis and bibliometric approaches to conduct a literature review of the use of VGI and social media data in research examining UGS. The research priorities focus on investigating factors related to activities in UGS areas, assessing urban park visitation and accessibility, integrating the use of multiple social media platforms, and, where appropriate, optimizing the use of personal information. In addition, analysis approaches can be extended to examine the network suggested by social media posts that are shared, re-posted or reacted to and by being combined with textual, image and geographical data to extract more representative information for UGS analysis. The study has identified a number of opportunities for future research.

The work in Chapter 4 of the thesis has appeared in the publication as:

Cui, N., Malleson, N., Houlden, V. and Comber, A., 2022. Using social media data to understand the impact of the COVID-19 pandemic on urban green space use. *Urban*

Forestry & Urban Greening, 74, p.127677. doi: https://doi.org/10.1016/j.ufug.2022.127677.

Chapter 4 examines spatial-temporal changes in UGS use in London associated with COVID-19 related lockdowns by using social media data (Twitter). This is one of the first studies to examine social media data over consistent time period before, during and after the COVID-19 lockdown in relation to UGS use. The results show that the findings and methods can potentially inform policy makers in UGS management and planning, especially in a period of social crisis like the COVID-19 pandemic.

The work in Chapter 5 of the thesis has appeared in the paper as:

Cui, N., Malleson, N., Houlden, V. and Comber, A., 2023. Urban green space topics based on structural topic modelling during the COVID-19 pandemic. *Submitted to a peer-reviewed journal.*

Chapter 5 uses machine learning model (STM) and spatial model (IDE) to identify, analyse, and visualise dynamic topics in UGS by incorporating text mining techniques. The chapter examines the effectiveness of STM on investigating the changes in the topic content and topic prevalence in relation to UGS, and also explore the spatial-temporal patterns of UGS use pre-, during- and after the COVID-19 outbreaks. The results reflects UGS users' attitudes to and perceptions of restriction measures, which potentially provide suggestions and guidance for future UGS policies, especially in times of crisis.

Chapter 6 is the conclusion of this thesis. The major findings of the research are consolidated, and this chapter summarises the contribution to the literature in the broader use of social media analysis and UGS research. It is then followed by the critical discussion of limitations and outlooks.

References

Al-Kodmany, K. 2019. Improving understanding of city spaces for tourism applications. *Buildings*. ***9***(8), p.187.

Barrie, C. and Ho, J.C.-t. 2021. academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software.* ***6***(62), p3272.

Breuste, J., Haase, D. and Elmqvist, T. 2013. Urban landscapes and ecosystem services. *Ecosystem services in agricultural and urban landscapes*, pp.83-104.

Brown, G., Kelly, M. and Whitall, D. 2014. Which 'public'? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. *Journal of Environmental Planning and Management*. ***57***(2), pp.190-214.

Brown, G., Rhodes, J. and Dade, M. 2018. An evaluation of participatory mapping methods to assess urban park benefits. *Landscape and Urban Planning*. ***178***, pp.18-31.

Burgess, J., Harrison, C.M. and Limb, M. 1988. People, parks and the urban green: a study of popular meanings and values for open spaces in the city. *Urban studies.* ***25***(6), pp.455-473.

Campbell, L.K., Svendsen, E.S., Sonti, N.F. and Johnson, M.L. 2016. A social assessment of urban parkland: Analyzing park use and meaning to inform management and resilience planning. *Environmental Science & Policy*. ***62***, pp.34-44.

Chakraborty, I. and Maity, P. 2020. COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Science of the total environment.* ***728***, p138882.

Chałka, K., Olszewski, R. and Zieliński, J. 2011. Bazy Danych Obiektów Topograficznych i Ogólnogeograficznych-zakres merytoryczny i techniczny opracowywanego projektu rozporządzenia MSWIA. *Roczniki Geomatyki-Annals of Geomatics.* ***9***(6 (50)), pp.98-102.

Cohen, D.A., Marsh, T., Williamson, S., Derose, K.P., Martinez, H., Setodji, C. and McKenzie, T.L., 2010. Parks and physical activity: why are some parks used more than others?. Preventive medicine, ***50***, pp.S9-S12.

Dennis, M., Barlow, D., Cavan, G., Cook, P.A., Gilchrist, A., Handley, J., James, P., Thompson, J., Tzoulas, K. and Wheater, C.P. 2018. Mapping urban green infrastructure: A novel landscape-based approach to incorporating land use and land cover in the mapping of human-dominated systems. *Land.* ***7***(1), p17.

Derkzen, M.L., van Teeffelen, A.J. and Verburg, P.H. 2015. Quantifying urban ecosystem services based on high-resolution data of urban green space: an assessment for Rotterdam, the Netherlands. *Journal of Applied Ecology.* ***52***(4), pp.1020-1032.

Evenson, K.R., Wen, F., Golinelli, D., Rodríguez, D.A., Cohen, D.A.J.E. and behavior. 2013. Measurement properties of a park use questionnaire. ***45***(4), pp.526-547.

Feltynowski, M., Kronenberg, J., Bergier, T., Kabisch, N., Łaszkiewicz, E. and Strohbach, M.W. 2018. Challenges of urban green space management in the face of using inadequate data. *Urban forestry & Urban greening.* ***31***, pp.56-66.

Foster, A. and Dunham, I.M. 2015. Volunteered geographic information, urban forests, & environmental justice. *Computers, Environment and Urban Systems*. ***53***, pp.65-75.

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Forns, J., Plasència, A. and Nieuwenhuijsen, M.J. 2015. Mental health benefits of long-term exposure to residential green and blue spaces: a systematic review. *International journal of environmental research and public health.* ***12***(4), pp.4354-4379.

Gu, Z., Zhang, Y., Chen, Y. and Chang, X. 2016. Analysis of attraction features of tourism destinations in a mega-city based on check-in data mining—A case study of ShenZhen, China. *ISPRS International Journal of Geo-Information*. ***5***(11), p.210.

Guerrero, P., Møller, M.S., Olafsson, A.S. and Snizek, B. 2016. Revealing cultural ecosystem services through Instagram images: The potential of social media volunteered geographic information for urban green infrastructure planning and governance. *Urban Planning*. ***1***(2), pp.1-17.

Guo, S., Song, C., Pei, T., Liu, Y., Ma, T., Du, Y., Chen, J., Fan, Z., Tang, X., Peng, Y. and Wang, Y. 2019. Accessibility to urban parks for elderly residents: Perspectives from mobile phone data. *Landscape and urban planning*. ***191***, p.103642.

Haaland, C. and van Den Bosch, C.K. 2015. Challenges and strategies for urban green-space planning in cities undergoing densification: A review. *Urban forestry & urban greening*. ***14***(4), pp.760-771.

Haleem, A., Javaid, M. and Vaishya, R. 2020. Effects of COVID-19 pandemic in daily life. *Current medicine research and practice.* ***10***(2), p78.

Hamstead, Z.A., Fisher, D., Ilieva, R.T., Wood, S.A., McPhearson, T. and Kremer, P. 2018. Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*. ***72***, pp.38-50.

Heather E. Wright Wendel, R.K.Z., James R. Mihelcic. 2012. Accessibility and usability: Green space preferences, perceptions, and barriers in a rapidly urbanizing city in Latin America. *Landscape and Urban Planning.* ***107***, pp.272-282.

Heikinheimo, V., Tenkanen, H., Bergroth, C., Järv, O., Hiippala, T. and Toivonen, T. 2020. Understanding the use of urban green spaces from user-generated geographic information. *Landscape and Urban Planning*. ***201***, p103845.

Holmes, E.C., Goldstein, S.A., Rasmussen, A.L., Robertson, D.L., Crits-Christoph, A., Wertheim, J.O., Anthony, S.J., Barclay, W.S., Boni, M.F. and Doherty, P.C. 2021. The origins of SARS-CoV-2: A critical review. *Cell*. ***184***(19), pp.4848-4856.

Houlden, V., de Albuquerque, J.P., Weich, S. and Jarvis, S. 2019. A spatial analysis of proximate greenspace and mental wellbeing in London. *Applied Geography*. ***109***, p102036.

Jackson, S.B., Stevenson, K.T., Larson, L.R., Peterson, M.N. and Seekamp, E. 2021. Outdoor activity participation improves adolescents' mental health and well-being during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*. ***18***(5), p2506.

Kabisch, N., Qureshi, S. and Haase, D. 2015. Human–environment interactions in urban green spaces—A systematic review of contemporary issues and prospects for future research. *Environmental Impact assessment review*. ***50***, pp.25-34.

Kabisch, N., Strohbach, M., Haase, D. and Kronenberg, J. 2016. Urban green space availability in European cities. *Ecological indicators*. ***70***, pp.586-596.

Kashef, M. 2016. Urban livability across disciplinary and professional boundaries. *Frontiers of Architectural Research*. ***5***(2), pp.239-253.

Kim, D. and Jin, J. 2018. Does happiness data say urban parks are worth it? *Landscape and Urban Planning*. ***178***, pp.1-11.

Kolcsar, R., Resch, B., Crivellari, A. and Blaschke, T. 2018. Beyond spatial proximity—classifying parks and their visitors in London based on spatiotemporal and sentiment analysis of Twitter data. *ISPRS International Journal of Geo-Information*. ***7***(9), p.378.

Ladle, A., Galpern, P. and Doyle-Baker, P. 2018. Measuring the use of green space with urban resource selection functions: An application using smartphone GPS locations. *Landscape and Urban Planning*. ***179***, pp.107-115.

Lesser, I.A. and Nienhuis, C.P. 2020. The impact of COVID-19 on physical activity behavior and well-being of Canadians. *International journal of environmental research and public health*. ***17***(11), p3899.

Liu, H., Li, F., Xu, L. and Han, B. 2017. The impact of socio-demographic, environmental, and individual factors on urban park visitation in Beijing, China. *Journal of Cleaner Production*. ***163***, pp.S181-S188.

Liu, K., Siu, K.W.M., Gong, X.Y., Gao, Y. and Lu, D. 2016. Where do networks really work? The effects of the Shenzhen greenway network on supporting physical activities. *Landscape and Urban Planning*. ***152***, pp.49-58.

Liu, O.Y. and Russo, A. 2021. Assessing the contribution of urban green spaces in green infrastructure strategy planning for urban ecosystem conditions and services. *Sustainable Cities and Society*. ***68***, p102772.

National Park City Foundation. 2023. National park cities are grassroots movements for people making their cities greener, healthier and wilder. [Online]. [Accessed 16 March]. Available from: https://www.nationalparkcity.org/

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G. and Fuller, G., 2019. Sustainable Development Report 2019. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN). *Search in*.

Ordnance Survey. 2021. *OS Open Greenspace*. [Online]. [Accessed 25 October]. Available from: https://beta.ordnancesurvey.co.uk/products/os-open-greenspace

Pfeffer, J., Mooseder, A., Hammer, L., Stritzel, O. and Garcia, D. 2022. This Sample seems to be good enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API. *arXiv preprint arXiv*.***2204***,02290.

Pierce, M., Hope, H., Ford, T., Hatch, S., Hotopf, M., John, A., Kontopantelis, E., Webb, R., Wessely, S. and McManus, S. 2020. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. *The Lancet Psychiatry*. ***7***(10), pp.883-892.

Plunz, R.A., Zhou, Y., Vintimilla, M.I.C., Mckeown, K., Yu, T., Uguccioni, L. and Sutto, M.P. 2019. Twitter sentiment in New York City parks as measure of well-being. *Landscape and Urban Planning*. ***189***, pp.235-246.

Poortinga, W., Bird, N., Hallingberg, B., Phillips, R. and Williams, D. 2021. The role of perceived public and private green space in subjective health and wellbeing during and after the first peak of the COVID-19 outbreak. *Landscape and Urban Planning*. ***211***, p104092.

Roberts, H., Sadler, J. and Chapman, L. 2017. Using Twitter to investigate seasonal variation in physical activity in urban green space. *Geo: Geography and Environment*. ***4***(2), p.e00041.

Roberts, H.V. 2017. Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography*. ***81***, pp.13-20.

Ronchi, S., Arcidiacono, A. and Pogliani, L. 2020. Integrating green infrastructure into spatial planning regulations to improve the performance of urban ecosystems. Insights from an Italian case study. *Sustainable Cities and Society*. ***53***, p101907.

Schipperijn, J., Bentsen, P., Troelsen, J., Toftager, M. and Stigsdotter, U.K. 2013. Associations between physical activity and characteristics of urban green space. *Urban forestry & urban greening*. ***12***(1), pp.109-116.

Sim, J. and Miller, P. 2019. Understanding an urban park through big data. *International journal of environmental research and public health*. ***16***(20), p.3816.

Song, Y., Huang, B., Cai, J. and Chen, B. 2018. Dynamic assessments of population exposure to urban greenspace using multi-source big data. *Science of the Total Environment*. *634*, pp.1315-1325.

Sonter, L.J., Watson, K.B., Wood, S.A. and Ricketts, T.H. 2016. Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. *PLoS one*. ***11***(9), p.e0162372.

Steenberg, J.W., Robinson, P.J. and Duinker, P.N. 2019. A spatio-temporal analysis of the relationship between housing renovation, socioeconomic status, and urban forest ecosystems. *Environment and Planning B: Urban Analytics and City Science*. ***46***(6), pp.1115-1131.

Tao, Z. and Cheng, Y. 2019. Modelling the spatial accessibility of the elderly to healthcare services in Beijing, China. *Environment and Planning B: Urban Analytics and City Science*. ***46***(6), pp.1132-1147.

Theberath, M., Bauer, D., Chen, W., Salinas, M., Mohabbat, A.B., Yang, J., Chon, T.Y., Bauer, B.A. and Wahner-Roedler, D.L. 2022. Effects of COVID-19 pandemic on mental health of children and adolescents: A systematic review of survey studies. *SAGE open medicine*. ***10***, p.20503121221086712.

Tison, G.H., Avram, R., Kuhar, P., Abreau, S., Marcus, G.M., Pletcher, M.J. and Olgin, J.E. 2020. Worldwide effect of COVID-19 on physical activity: a descriptive study. *Annals of internal medicine*. ***173***(9), pp.767-770.

Townshend, T., Roe, M., Davies, C. and Qin, Q. 2018. National park city: Salutogenic city? In: *SDP 2018*.

Upton, V., Ryan, M., O'Donoghue, C. and Dhubhain, A.N. 2015. Combining conventional and volunteered geographic information to identify and model forest recreational resources. *Applied Geography*. ***60***, pp.69-76.

Wang, Z., Jin, Y., Liu, Y., Li, D. and Zhang, B. 2018. Comparing social media data and survey data in assessing the attractiveness of Beijing Olympic Forest Park. *Sustainability*. ***10***(2), p.382.

We are London. 2023. *London National Park City*. [Online]. [Accessed 16 Feburary]. Available from: https://www.london.gov.uk/programmes-strategies/environment-and-climate-change/parks-green-spaces-and-biodiversity

Wood, S.A., Guerry, A.D., Silver, J.M. and Lacayo, M. 2013. Using social media to quantify nature-based tourism and recreation. *Scientific reports*. ***3***(1), p.2976.

World Health Organization. 2020. *Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected: interim guidance, 13 March 2020.* World Health Organization.

Zhai, Y., Wu, H., Fan, H. and Wang, D. 2018. Using mobile signaling data to exam urban park service radius in Shanghai: methods and limitations. *Computers, Environment and Urban Systems*. ***71***, pp.27-40.

Zhang, S. and Zhou, W. 2018. Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. *Landscape and Urban Planning*. ***180***, pp.27-35.

Zhang, W., Yang, J., Ma, L. and Huang, C. 2015. Factors affecting the use of urban green spaces for physical activities: Views of young urban residents in Beijing. *Urban Forestry & Urban Greening*. ***14***(4), pp.851-857.

Zhen, F., Cao, Y., Qin, X. and Wang, B. 2017. Delineation of an urban agglomeration boundary based on Sina Weibo microblog ‘check-in’data: A case study of the Yangtze River Delta. *Cities*. ***60***, pp.180-191.

Chapter 2 Research context

2.1 Urban green space

2.1.1 Terminology and typology of urban green space

UGS provides numerous benefits for urban residents across multiple disciplines and research contexts, such as ecology, environmental and social sciences, urban design and planning, and public health, in the past decades (Kabisch et al., 2015). An increasing number of studies suggest that maintaining and developing UGS plays an essential role in creating a sustainable city. However, because UGS is studied from various objectives and disciplines, there can be differences in the meanings and results generated from different studies. For example, the terminology used to describe UGS can vary across different studies and disciplines. The terms "greenspace," "green space," "public greenspace," "urban greenspace," and "urban green space" have been used in previous studies associated with UGS. Additionally, some of the studies use more specific terms according to the research aims and projects. For example, terms such as "garden," "ecological garden," and "urban habitat" have been used in studies focusing on biological, earth, and environmental sciences. Terms such as "green environments," "green network," "green infrastructure," "greening project," "productive greenspace," and "working greenspace" have been used in studies focusing on architecture, urban environment, and building. In addition, terms such as "vegetated area," "water bodies," "woodland," "sky garden," "urban farm landscape," and "urban trees" have been frequently used in studies related to social sciences, behavioral sciences, and policy and political sciences.

The variation of terms with UGS indicated that the definition of UGS varied across different research projects, which may lead to issues in the lack of common terms for multidisciplinary research or in performing comparative studies. Providing a clear and considered definition of key terms is critical for researchers, otherwise, they risk relying on idiosyncratic personal interpretations of generalized terms. A review article that examined 125 published journal articles about green space suggested that the key term ‘green space’ was the main term used in all disciplines in the past fifty years, indicating that the term ‘green space’ is more commonly used than ‘greenspace’ (Taylor and Hochuli, 2017). Another review article that focused on the interactions between humans and environment used the term ‘urban green space’, ‘urban green’, and ‘urban park’ as basic search key terms during the extraction stage (Kabisch et al., 2015). In addition, a study that focused on the relations between green space and urban services also used the term 'urban green space' as the main keywords (Belmeziti et al., 2018). The thesis aims to answer the questions of how the UGS provide benefits during a pandemic period. So the terms should be provided in a common sense when take into account of public health into UGS research. For example, terms 'green space'; 'public green space' and 'park' were used in a study to explore the value of urban green space for health and wellbeing (Lee et al., 2015). The overall aim of this thesis are to investigate the urban green space visitation by using social media data, thus the term 'urban green space' was used throughout the thesis.

The typology of UGS also varied based on the research goals, research scale and methods. The types of UGS in a certain study were defined according to specific classification criteria (Belmeziti et al., 2018; Panduro and Veie, 2013). For example, the studies focusing on physical activities, accessibility of UGS, and environmental benefits such as reducing urban heat island classified UGS according to the size of UGS areas (Heather E. Wright Wendel, 2012; Schipperijn et al., 2023; Xiao et al., 2018). These studies aimed to quantify the area of urban green space, and link the types of urban green space and the size of the area. Other studies classified UGS based on their functions such as recreation, relaxation, and sports grounds (Shackleton and Blair, 2013), and the influences of UGS on housing prices (Panduro and Veie, 2013). Studies focusing on ecosystem services classified UGS according to what types of ecosystem services that UGS could provide to city residents (Luederitz et al., 2015; Belmeziti et al., 2018). The above studies suggest that the typology of UGS is essential in conducting comprehensive study results, when dealing with the specific research questions.

This thesis will focuses on the contributions of UGS to human wellbeing, concerning the social and environmental benefits of UGS to whole of society, especially during times of crisis. Specifically, UGS can provide places for outdoor physical activities, such as walking, running, and cycling, and for social interactions and activities; Spending time in UGS can improve mood, increase self-esteem, and reduce symptoms of mental illness. An increasing number of studies focus on the environmental and social benefits related to green space as being important to mitigating urbanisation-induced environmental effects and increasing the quality of life of citizens. These studies define UGS as a range of parks, street trees, urban agriculture, residential lawns, and roof gardens (Breuste et al., 2013). In this analysis, UGS is defined as all publicly owned and publicly accessible open space that is covered by vegetation or sports facilities, e.g., parks, gardens, woodlands, nature areas, sports facilities, cemetery, and other green space within the city boundary area (Schipperijn et al., 2013; Ordnance Survey, 2021a).

2.1.2 Urban green space use during normal and pandemic time

An increasing number of UGS studies have been providing information about how people interact with UGS. During normal time (pre-COVID-19), researches on UGS and its visitation mostly focused on the aspects of: 1) the general use of urban green space, the perceptions and general satisfaction, 2) the direct and indirect health effects of UGS, 3) issues about equal provision and access UGS by different social groups, 4) the development, planning, and management of UGS, and 5) economic value analyses focusing on the monetisation of UGS values using different methods. During the COVID-19 pandemics, researches on the relationship between the pandemic and UGS use mainly focused on the benefits of UGS to public health (Xie et al., 2020), the effects of the COVID-19 pandemic on the use and perceptions of UGS (Ugolini et al., 2020; Da Schio et al., 2021), and visitation rate to UGS before and during the COVID-19 lockdown pandemic (Grima et al., 2020).

Previous studies used questionnaire surveys (Johnson et al., 2019; Pipitone and Jović, 2021; Owczarczak-Garstecka et al., 2021), activity diaries (Theodorou et al., 2021), and GPS-based campaigns (Jankowska et al., 2015) to detect the UGS use. These methodologies and relevant datasets have been proven to potentially provide in-depth understanding about UGS use, but they are often limited in duration and number of participants because data collection that involves active participation can be time- and energy consuming. (Heikinheimo et al., 2020). This might hinder the effectiveness of these methods (Shi et al., 2017; Hamstead et al., 2018; Jing-Huei et al., 2022).

Social media data opens a new window for investigating the UGS use by using advanced methods such as big data analysis, topic modelling approaches, and sentiment analysis, particularly in the research of UGS use. For example, Heikinheimo et al. (2020) and Lim et al. (2018) used Twitter to investigate the patterns of leisure activities, and to detect the effects of green spaces on sentiments. Li, D. et al. (2018) used Flickr data to detect the spatial interactions between human and environments. Zhang and Zhou (2018) and Norman and Pickering (2019) used geo-tagged social media data to investigate the factors that affecting UGS use. These studies aimed to provide information for planning a more sustainable and socially equal cities with the growing of urban population and density (Heikinheimo et al., 2020; Fan et al., 2016), but failed to consider the situation in a time of pandemic or social crisis, which could potentially provide guidelines for future urban planning and management.

Previous studies on the usage of UGS were conducted during normal time, when there is no global epidemic such as COVID-19, which has changed human mobility and behaviour patterns in many countries. However, it is still not clear about how UGS was used during a pandemic period, and how UGS can make contributions to the whole society during a time of crisis. Thus there is a need for evidence-based decision making and urban planners and managers base decisions when adapting policies in response to the rapid change in demand and preferences during the pandemic (Jing-Huei et al., 2022). Overall, in the light of this challenge, this thesis will investigate the impact of a pandemic on the usage of UGS, and to inform UGS management during disruptions when use is altered and the needs of the public may be changing. Additionally, because of restriction measures such as stay-at-home and social distancing, it can be challenging to conduct traditional methods of data collection such as survey questionnaires and observations. Social media such as Twitter, Instagram, and Facebook are become a valuable tool for analysing UGS during times of crisis, such as the COVID-19 pandemic.

2.1.3 The review of UGS research

Most recently, an increasing number of studies began to explore UGS use associated to COVID-19 pandemic. For example, Wang and Li (2022) evaluated the impacts of the COVID-19 lockdown on the use and perceptions of UGS in Xuzhou, China, and identified the differences across groups through an online survey of 376 respondents. The study selected a middle level city in China and the time period from March to April 2022, the results may cannot represent the situations in big cities such as London, Beijing, New York, and Hong Kong. It is needed to investigate the UGS in such metropolitans as the results could be more representative of most big cities around the world. Jing-Huei et al. (2022) used Twitter data to investigate the how public sentiments and perceptions associated with NYC parks has shifted from pre-COVID to during-COVID, the time period of the analysis was from March 2019 to February 2020, and March 2020 to February 2021, separately. This study only selected four big parks in NYC (i.e. at least 500 acres) as a case study rather than the whole urban areas, which lost the knowledge about the use of small parks such as pocket parks and community parks.

Additionally, the lack of post COVID-19 period in 2021 is one of the limitations of this study, as no comparison was conducted in post-COVID-19 period. Grzyb et al. (2021) used Instagram data to compare the UGS visitation patterns during COVID-19 pandemic with the same period for the previous year of 2019. The results showed that the wilder the area, the lower visitation levels. However, this study has some limitations of lower representativeness and the study area only covered formally-recognized UGS with a recreational function rather than the whole of urban areas. Geng et al. (2021) analysed the impacts of COVID-19 pandemic on park use but at global, regional and national levels, which failed to provide the knowledge of UGS use at a specific urban area. There is a need for investigating how UGS use was influenced in a whole urban area, which could potentially inform more comprehensively urban UGS management during disruptions.

Overall, it is needed obtain the knowledge about how UGS was used before- during- and after the COVID-19 pandemic, by conducting a comprehensive investigation using social media data and advanced analysis methods. The results will potentially provide insights for future urban planning and management.

2.2 Urban green space in London

2.2.1 London is the first National Park City

London was officially declared the world's first National Park City in July 2019, 50% of Greater London is covered by green and blue space with 3000 of parks, front gardens, allotments, rivers, canals, and ponds (National Park City Foundation, 2023; We are London, 2023b). Green space in central London consists of five of the capital's eight Royal Parks (The Royal Parks, 2023), supplemented by a number of small garden squares scattered throughout the city centre. Open space in the rest of the region is dominated by the remaining three Royal Parks and many other parks and open spaces of a range of sizes, run mainly by the local London boroughs, although other owners include the National Trust and the City of London Corporation.

The parks in London are free to visit and everyone can enjoy parks at every day of the year. There are many programmes conducted by communities continuously improve green and blue spaces in London. For example, London’s Green Spaces Commission (We are London, 2023a) aims to identify new models for the delivery and management of London’s green spaces to help support boroughs transform how their parks services are managed and funded in the context of substantial and ongoing constraints imposed on public sector funding. The London Rewilding Taskforce (We are London, 2023c) aims to explore opportunities for rewilding in London to support nature recovery and enhance biodiversity, while bringing benefits to Londoners and addressing the climate and ecological emergency. The London Urban Forest Partnership (We are London, 2023d) aims to protect, manage and enhance the capital's trees and woodlands, and promote the benefits of the city's urban forest. Additionally, there are many funding sources for supporting the urban green space. For example, the Greener City Fund was the Mayor's programme for supporting green space projects during the 2016-2021 mayoral term, lots of projects have been conducted to plant trees, and help make the city greener. In addition, Londoners are encouraged to participate in a range of community and public engagement programmes, which not only can allow residents to interact with nature environments but also benefit to people’s human wellbeing.

National Park City Rangers connect and support thousands of people leading change in their own neighbourhoods, they share and celebrate the work of hundreds of organisations and individuals across London. Additionally, as the days start to get lighter and longer in spring season, people can enjoy parks in various ways. For example, people can see the blossoms and flowers such as daffodils, cherry trees, and bluebells; discover nature setting with families; explore the incredible heritage in parks; and spot beautiful bird species. Chinmoy Sarkar et al. (2015) used questionnaires to examine the effects of London urban green space upon walking behaviour patterns and physical activities in UGS.

2.2.1 Social media analysis related to UGS in London

Social media data have been used in previous studies to detect the London UGS use, which provided insights for urban planning and management. For example, Lansley and Longley (2016) used Twitter to investigate the topics in London, the analysis results suggested that people engaged UGS related activities such as photography and sightseeing in parks and gardens. Kovacs-Györi et al. (2018) used Twitter data to analyse the spatial-temporal park visiting behaviour of more than 4000 users for almost 1700 parks in London, the study found that the people tweeted mostly in parks 3–4 km away from their centre of activity and they were more positive than elsewhere while doing so. However, these studies used public streaming API to collect Twitter data which only extract 1–2% of the complete Tweets. This means that there is a risk that the study results may not accurately represent the broader population being studied due to potential biases or limitations in the sample selection. Additionally, the studies only selected one single research period, thus there is a lack of study that focuses on dynamics of spatial-temporal patterns of UGS use in a consecutive period, particularly during a time of crisis such as COVID-19 pandemic. The issues related to data volume, study period, and analysis methods need to be appropriately addressed when conducting analyses of social media data.

2.3 Volunteered geographic information and social media data

Urban green space management and planning are inspired from the information about UGS use and the interactions between human and environments. However, the data for such analysis are often scarce and laborious to collect (Heikinheimo et al., 2020). The questionnaire survey is one of the main data resources for the research in relation to UGS visitation, but this method is often limited by the low number of responses, time consuming and high labour cost (Tenkanen et al., 2017; Heikinheimo et al., 2017).

2.3.1 Evolution of datasets for UGS research

The data sources that can be used to investigate the usage of UGS has been transferring from professional domains (i.e. questionnaire survey) to public contents (social media platforms), this shift has occurred as a result of significant technological advances during the past two decades (Heikinheimo et al., 2020). The mobile devices can record the geographical information using GPS function, which could be used to create geo-referenced information at anyplace and anytime. Additionally, the proliferation of various mobile applications enable users to create and share volunteered information to the publics as much as they want. This open a new window for the UGS research.

The user generated data with geo-information and timestamps were called as volunteered geographic information (VGI) data. The definition of VGI evolved in the past decades. First coined by (Goodchild, 2007), VGI is defined as “the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals”. In 2013, Cinnamon and Schuurman (2013) argued that crowdsourcing implies a kind of consensus-producing process and the assumption that several people will provide information about the same thing so it will be more accurate than VGI. VGI, on the other hand, is produced by individuals without any such opportunity for convergence. Elwood et al. (2012) defined VGI as spatial information that is voluntarily made available, with an aim to provide information about the world.

2.3.2 Types of VGI and social media data

Previous studies have compared four types of volunteered geographic information (VGI) datasets: social media, sports tracking, mobile phone operator and public participation geographic information systems (PPGIS) in understanding the use of UGS (Heikinheimo et al., 2020). The study examined the ability of these data in providing the information about where, when and how users interact with UGS. The results showed that social media platforms can provide information about the patterns of leisure activities in UGS. Social media refers to web-based services that allow people to create and share content in online communities (McCay-Peet and Quan-Haase, 2017).

Social media data provides massive information within a short period of time, it can support researchers to analyse the data from in a greater detail and more systematic fashion than other methods such as interviews and surveys. Moreover, the methods for collecting the digital footprints are instantaneous and nonintrusive (Lomborg and Bechmann, 2014). Finally, social media data could be collected from multiple platforms. Specifically, Twitter and Weibo provide text-based VGI data, Instagram and Flickr provide image-based VGI data, and OpenStreetMap and MapMyFitness provide map-based VGI data (Senaratne et al., 2017). All these types of VGI data enable researchers to investigate the use of UGS from different perspectives (Guo et al., 2019).

2.3.3 Data opportunities for UGS research

Social media platforms are used by millions of people from different countries throughout the world (Martí et al., 2019). As the number of social network users grows, the quantity, quality, and potential utility of social media data continue to expand rapidly. In 2023, the number of social media users worldwide reached 4.76 billion (Chaffey, 2023). In the UK, there are 57.10 million social media users, accounting for 84.4% of the total population in January 2023. There are massive information posted every day, which offer great opportunities to detect UGS use by using these datasets. Social media user-generated data offer a valuable means of studying UGS, particularly when such data is georeferenced. The ability to reference the location of the posts related to UGS allows for the analysis of specific phenomena within a particular UGS area, providing added value for urban research. This georeferenced information enables researchers to identify patterns of UGS use and explore the relationship between UGS and social or environmental factors at a more precise and localized level. Twitter and Flickr are the most frequently used social media data sources in UGS and environmental studies because of the high accessibility of data and low cost of time and energy (Tenkanen et al., 2017). The Chapter 4 will further investigate the usage of various VGI data in relation to UGS research.

2.4 The imperative importance of conducting a literature review

The widespread use of social media data and analysis approaches allow the researcher to examine the interactions between UGS and human from the wide ways, various types of VGI datasets have been used in relation to UGS research. However, there is a lack of comprehensive literature review that synthesizes existing knowledge on the use of social media data in UGS research, making it difficult to establish a clear research direction for future studies in terms of dataset selection, analysis methods, and research directions. Therefore, this thesis aims to fill this gap by providing a comprehensive literature review of previous studies utilizing social media data for UGS research. Through this review, this thesis seeks to identify trends, challenges, and opportunities associated with the use of social media data in UGS research, as well as provide recommendations for future research directions.

2.4.1 Categories of literature review

Literature review plays an essential role in academic research by gather existing knowledge and to examine the state of a research area (Kunisch et al., 2018). Literature review allows researchers to identify research gaps, evaluate research methods, and interpret research results relevant to a specifical research question or research area.

The two main methods for conducting a literature review are narrative review and systematic review. A narrative review provides a descriptive overview of a specific subject, and the selection of studies is based on availability and the authors' discretion. In contrast, a systematic review is a structured research methodology that involves a comprehensive and rigorous analysis of existing literature on a specific research question or topic. This method typically includes a meta-analysis component, which utilizes statistical techniques to synthesize data from collected studies into a single quantitative estimate or summary effect size. It is a way to synthesize and summarize the results of multiple studies to provide an overall understanding of the topic. Systematic reviews aim to present a fair evaluation of a research topic by using a trustworthy, rigorous, and auditable methodology.

There are three main differences between the two types of review as follows.

(1) In terms of search method, a narrative review does not pre-define search strategies. The related studies are included or excluded based on the authors' research experience, which may introduce subjective selection bias. The search strategies for a systematic review would be pre-defined before searching and the related studies were selected based on the predefined selection criteria, which are much more objective. Additionally, the search engines or databases for a narrative review are selected by author's experiences, while a systematic review selects multiple search engines/databases such as Medline, EMBASE, CINAHL, Sport DISCUS and PubMed (Hunter et al., 2015), and Web of science, IEEE, and Scopus (Linnenluecke et al., 2020).

(2) In synthesising data, a narrative review generally give an overall description of the each study, and mainly focuses on several certain studies based on author's selection, the research results might be limited due to author's selection and research interests. A systematic review uses a continuous or categorical statistical values to analyse the collected data, which can generate the relative objective results based on the pre-defined data extraction and synthesis guidelines.

(3) In terms of the interpretation, a narrative review interprets the research findings by author's subjective intention, which could result in a number of bias of the research results and conclusions. A systematic review interprets the research results based on the data statistics and empirical evidence, which can proved the evidence-based summaries concerning a specific research topics such as UGS visitation.

Overall, in a narrative review, the absence of objective and systematic selection criteria in a review method can result in several methodological shortcomings, potentially leading to biases in the author's interpretation and conclusions. Recently, many journals have modified their acceptance policy for review papers, giving priority to systematic reviews as regular review articles while excluding narrative reviews. This change aimed at providing the most reliable evidence for all basic and clinical questions, as well as to generate new hypotheses for further research (Pae, 2015). In a conclusion, it is suggested that a systematic review can provide a more objective, comprehensive, reliable and instructive research results for a particular research question or research topic.

2.5 Challenges in studying UGS use via social media data

The analysis of social media data on UGS use has received considerable attention as a promising method for applied research. However, there are several challenges associated with using social media data for UGS research that have sparked debates about its validity. Despite these challenges, social media data offers opportunities such as large sample sizes, cost-effectiveness, and access to rich data, making it a promising avenue for UGS research if used appropriately and with careful consideration of its limitations. This section identifies the challenges, opportunities and methodological issues associated with using social media data in UGS research.

2.5.1 Opportunities and challenges of datasets

Social media platforms are used by millions of people from different countries throughout the world (Martí et al., 2019). As the number of social network users grows, the quantity, quality, and potential utility of social media data continue to expand rapidly. In 2023, the number of social media users worldwide reached 4.76 billion (Chaffey, 2023). In the UK, there are 57.10 million social media users, accounting for 84.4% of the total population in January 2023. There are massive information posted every day, which offer great opportunities to detect UGS use by using these datasets. Social media user-generated data offer a valuable means of studying UGS, particularly when such data is georeferenced. The ability to reference the location of the posts related to UGS allows for the analysis of specific phenomena within a particular UGS area, providing added value for urban research. This georeferenced information enables researchers to identify patterns of UGS use and explore the relationship between UGS and social or environmental factors at a more precise and localized level.

However, the location information retrieved by built-in GPS receivers might be inaccurate due to different effects (Ajao et al., 2015; Mahmud et al., 2014). It is important to note that the quality of geo-referenced social media data can be influenced by various intrinsic and extrinsic factors. Intrinsic factors may include issues related to mobile device characteristics, such as poor GPS signal strength or inconsistent data capture. Extrinsic factors may include aspects of the built environment such as tall buildings or canyons, and quality of internet or Wi-Fi, which can impede GPS signals and therefore impact data accuracy. Thus, researchers need to consider these factors when analysing geo-referenced social media data in UGS research (Ajao et al., 2015). Furthermore, users can individually choose to add their precise location to a Tweet or just a general attached location information (such as a city or district). This might result in imprecise and coarse location information of data.

Social media allows people to express their perceptions, interests, events, needs and behaviours by posting texts when they visit UGS. Additionally, users are generally not constrained when generating information, which can reflect real status of UGS users and their perceptions. This is an advantage because according to the Hawthorne effect, subjects may alter their behaviour when they realize they are being observed (McCarney et al., 2007). However, advanced analyse methods such as machine learning and NLP techniques are still required to effectively and precisely generate reliable results.

2.5.2 Data availability and quality

Social media networking sites provide various of public APIs (application programming interfaces) such as Flickr API (www.flickr.com/services/api/), Instagram API (www.instagram.com/developer/), and Twitter API (developer.twitter.com), which allow researchers to collect data for various types of analyses. APIs enable researchers to collect data from different social media websites (Lomborg and Bechmann, 2014). APIs support people to build a computer program which can send requests to social media platforms such as Twitter and then obtain required data (Lomborg and Bechmann, 2014). For example, researchers can collect data from APIs about location, timestamps, content information, videos and images, and so on, which facilitates understanding the interactions between people and UGS. APIs are therefore provide a means for researchers to access and collect data from social media platforms. Although this approach provides opportunities for investigating UGS, researches are often limited by issues of data accessibility, volume, and data quality, which may result in under or over representative results. Research that relies on APIs for data collection suffers from a lack of transparency regarding the data output and quality. This lack of transparency can weaken the research and limit its reliability, as researchers may not have control over the quality and quantity of the data provided by the API (Stieglitz et al., 2018).

Previous studies have employed different types of API, which may vary in terms of the quantity, quality, and representativeness of the datasets. For example, Twitter streaming API (Benhardus and Kalita, 2013; Martí et al., 2019) has been compared with the Twitter Firehose (Morstatter et al., 2013), it was found that in particular smaller samples collected from the streaming API tend to misrepresent the volume of hash-tags when compared to the Firehose(Figure 3.2). Indeed, APIs of social media platforms appear to offer restricted access to their underlying databases without providing thorough documentation of how the API filters the database (Morstatter et al., 2013). In summary, it is crucial to select a suitable and accessible API for acquiring high-quality datasets when conducting analyses.

Most recently, Twitter released an API v2, which significantly improved the access level of datasets. However, there are three types of access levels to Twitter API v2, which are Essential access, Elevated access, and Academic research access. The differences in retrieving data volume per month, number of project per account, and access levels to Twitter API v1 (standard, premium, and enterprise) still limit the representativeness of the collected data. This can impact the results and reliability of the analysis, especially in studies related to UGS that only occupies a small portion of the whole of urban area.

Twitter released a new version of the standard search Tweets endpoint as part of Twitter Application Programming Interface (API) in 2020 (Pfeffer et al., 2022). This includes many new features, such as the ability to access the full history of Tweets, allowing researchers to programmatically access public Tweets from the complete archive dating back to March 2006. In this way, the new API provides more complete, and unbiased data than previous Twitter APIs (Pfeffer et al., 2022).

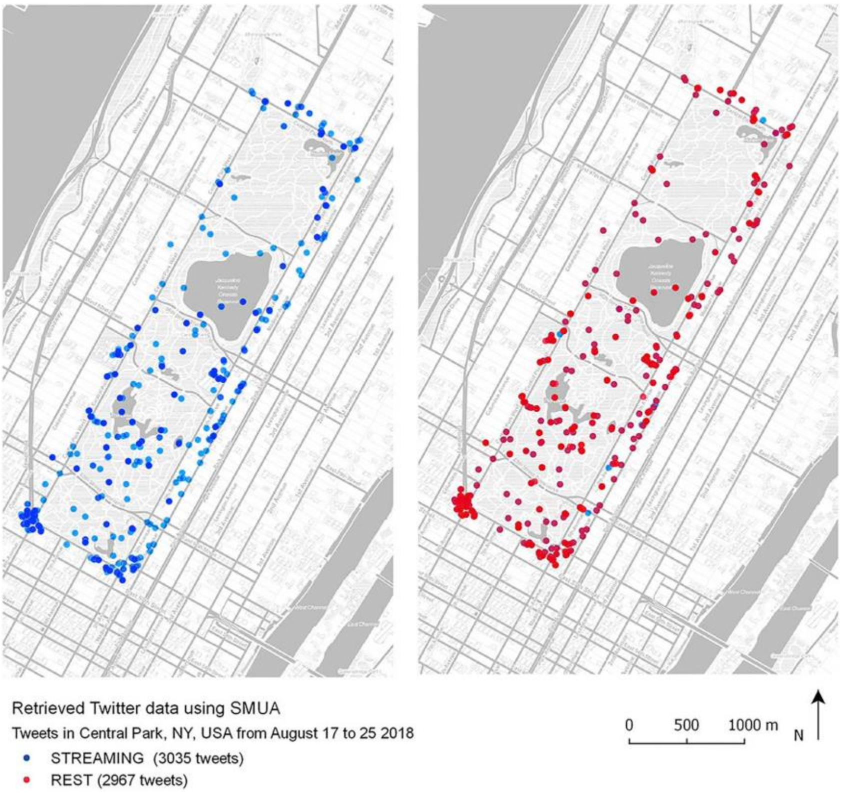

Figure 3. 1 Comparison between Twitter datasets obtained via Streaming and Rest API methods (Martí et al., 2019).

2.5.3 Methodological challenges of using social media data for UGS research

Challenges of using social media data have received considerable attention (Ahmed et al., 2017; Stieglitz et al., 2018; Steiger et al., 2015; Martí et al., 2019) in terms of analysis methods including spatial-temporal and semantic analysis. In spatial-temporal analysis methods, previous studies on UGS research used various analysis methods to detect the spatial patterns of UGS use. For example, KDE, K-means, Mean-Shift and DBSCAN algorithms were commonly used to assess the spatial patterns (Ullah et al., 2020; Ghermandi and Sinclair, 2019; Hasnat and Hasan, 2018). The approaches that combine different spatial analysis methods should therefore be developed in future works related to UGS research using VGI data. In temporal analysis, different time scales have been used in previous studies that mainly focused on daily, weekly, and monthly visitation patterns. The combination of spatial analysis and temporal analysis could be undertaken at the individual level. The semantic information was analysed by using a series of text analysis models. For example, text-mining, sentiment analysis, machine learning algorithms such as topic modelling and SVM were used by previous studies on detecting the perceptions of UGS. It is crucial to select a suitable text analysis methods when analyse social media datasets. Further methodological challenges and opportunities of using social media data will be discussed in chapter 3.

Another key methodological issue concerns sampling and associated questions of representativeness of social media data, there is a lack of differentiated statistical evidence of the social characteristics of users of distinct social media platforms (Blank and Lutz, 2017), particularly in UGS users. Evidently, the users who are most likely to generate most of these data in the social media are hardly representative of the entire population of users of social media (Lansley and Longley, 2016), thus it is hard to estimate the quality and representativeness of a sample if the researchers don’t know the population from which the sample is drawn. Research using APIs should always include a critical assessment of the sample and the in-built limitations to generalization when reporting findings. The explicit address of basic sampling biases creates transparency and thereby enhances the credibility of the empirical study. Finally, it should be noted that the accuracy of mobile devices can also influence research results to some extent.

2.5.4 Study period duration and extent

Despite an increasing number of studies have used social media data to investigate UGS visitation, many existing research papers are isolated case studies (Roberts, H.V., 2017;

Tenkanen et al., 2017) that collect a large data set during a specific time period on a specific subject and analyse it quantitatively. For example, Kovacs-Györi et al. (2018) used Twitter data to investigate the sentiment and emotion levels of park users. However, the results were represented in a static status throughout the year 2012, without considering the impact of seasonal changes on UGS visitation, which may result in over representative results especially in winter season. Lyu and Zhang (2019) used multi-source VGI data to understand the UGS use in Wuhan, China. One of the limitations in this study is the lack of consistency in the time periods of two datasets, one (Weibo check-in data) was collected from the whole year of 2017, while another one (Baidu heat map data) was collected from only 3 days in February 2017. The results might be influenced by the two different time periods.

It is important to consider the data collection period when study the UGS use because outdoor activities always have great correlations with weather and temperature (Cohen et al., 2010; Lu, 2019), especially when detecting the dynamics of UGS use in terms of spatial and temporal patterns. This guides us to mitigate the influences of seasonal variations while conducting research on the impact of COVID-19 on UGS use.

Chapter 3 Data and Methods

3.1 Datasets used in this thesis

The datasets in this thesis encompass literature metadata, Twitter datasets, and UGS datasets. Specifically, literature metadata will be collected from sources including Web of Science, Scopus, IEEE Xplore and Google search. Twitter datasets will be collected via Twitter research academic API. UGS datasets will be collected from Ordnance Survey(OS) and London boundaries datasets will be collected from the UK Data Service websites.

Literature metadata comprises information such as author names, article titles, journal names, key words, volume numbers, publication years, ISSN numbers, and DOI links. These information allow researchers to understand the research trends within the related field over past decades, to identify the hot research topics and research gaps, and to obtain information about which journal and institutes have made substantial contributions to the field. In this thesis, these information can effectively help us to understand the research question such as what types of social media websites or platforms were generally selected in these studies by investigating the key words in articles. However, this type of datasets have some limitations. For example, the limited search key words and search engines may not capture the whole related articles, affecting the accuracy of topic identification; the literature metadata may not include unpublished or non-peer-reviewed research, leading to potential bias in the analysis; and conference proceedings may not be adequately represented in metadata, impacting the comprehensiveness of the analysis. Overall, it should be acknowledged that there are still some limitations in the literature metadata, future analysis could addressed these issues when conducting literature review.

Twitter dataset can provide massive information for UGS related research. In detail, Tweet textual content can be used to detect UGS related topics and public perceptions of UGS. Geographical data can be used to investigate the spatial patterns of UGS visitation, especially during the COVID-19 pandemic period when there were challenges in obtaining observation and survey datasets. The timestamp data enable researchers to explore the temporal patterns of UGS visitation, providing valuable insights into how COVID-19-related restriction measures have impacted UGS use and individuals' daily routines. In the current thesis, tweet textual content, geographical and timestamps data will be employed to detect the impact COVID-19 on UGS use.

However, there is still other information in tweet datasets that has the potential to provide insights for UGS research. For example, user profile information can be used to investigate the relationship between UGS visitation and user characteristics, such as gender and age groups, with the approval of an ethics review. Future analysis could employed these information to enrich UGS-related research.

In the thesis, the Twitter datasets were collected via the Twitter academic research API. The API call selected geo-referenced Tweets located in London. They covered a three-month period (23rd March to 23rd Jun) for three consecutive years: 2019, 2020, and 2021. To begin with, the changes in the spatial distributions of geo-located Tweets for three individual years were analysed and the most frequently visited UGSs were identified in each year. Activities in UGS were also identified using text mining (Salloum et al., 2017), and their spatio-temporal changes were investigated. The Twitter data were analysed over hours and days to identify changes in temporal patterns. A paired sample t-test was used to significant differences between the use of UGS over the three years. All the analyses were undertaken using R software (Ihaka and Gentleman, 1996). The further details of each part of datasets will be introduced in the following sections.

3.1.1 Study area (Greater London and Urban green space)

Greater London is an administrative area in England governed by the Greater London Authority. It is divided into 33 local government districts including 32 London boroughs and the City of London. The Greater London region covers 1572 km^2 and has a population of 9 million. The 2021 census recorded that about 3.58 million people were foreign-born making it the city with the second largest immigrant population after New York (Catney et al., 2023). About 69% of children born in London in 2015 had at least one parent who was born abroad (Jolly et al., 2020). The multicultural and language characteristics make the analysis of UGS use achieve a better representativeness compare to other cities that with lower international levels, which will provide a comprehensive guidance for future international modern city planning and management. However, it should be acknowledged that only English Tweets will be select as data sources in the analysis, which may potentially limits the representation of non-English-speaking social media users.

The datasets of Greater London boundaries and district boundaries could be collected through the UK Data Service (https://ukdataservice.ac.uk/), which provides researchers with precise and freely accessible datasets. In this analysis, the datasets were collected and saved in shapefile format. A shapefile is a vector data storage format that stores the location, shape, and attributes of geographic features with the same geometry type (such as point, line, and polygon) and the same spatial reference. In this analysis, the datasets were gathered and stored in the shapefile format. A shapefile is a vector data storage format that retains the spatial information, geometry, and attributes of geographic features sharing the same geometry type (e.g., point, line, or polygon) and spatial reference.

3.1.2 Urban green space layers

Ordnance Survey (OS) is an authoritative mapping agency who provide the most accurate and up-to-date source for geospatial data in Great Britain (Shoari et al., 2021). OS provides Open Greenspace which depicts the location and extent of spaces such as parks and sports facilities that are likely to be accessible to the public (Ordnance Survey, 2021a), it is the most comprehensive Open dataset of greenspace which provides the foundation for people to help create greener and healthier communities. This dataset contains data for publicly accessible urban greenspaces in GB (Robinson et al., 2022). Its primary purpose is to enable UGS visitors to find and access greenspaces near them for exercise and recreation. Thus the Open Greenspace layers were suitable for investigating the UGS use. The Open Greenspace includes 10 types, i.e., *Allotments Or Community Growing Spaces, Bowling Green, Cemetery, Religious Grounds, Golf Course, Other Sports Facility, Play Space, Playing Field, Public Park Or Garden, and Tennis Court*. Table 2.1 gives a description of each type of Open Greenspace.

OS Open Greenspace is supplied as an shapefile which is freely downloaded. An shapefile is an open file format to store geometry and attribute information about spatial features. It is a plug-in and play format, allowing users to drag and drop the files directly into the analysis software such as GIS, and R Studio.

Table 2. 1 The description of each type of Open Greenspace (Ordnance Survey, 2021b).

Open Greenspace	Description
Allotments or Community Growing Spaces	Areas of land for growing fruit, vegetables and other plants, either in individual allotments or as a community activity. Produce is for the grower's own consumption and not primarily for commercial activity.
Bowling Green	A specially prepared area intended for playing bowls.
Cemetery	Areas of land associated with burial areas.

Religious Grounds	Areas of land associated with churches and other places of worship. Only included where there are significant areas of greenspace (over 500m2 of natural space - identified as surfaces that are not manmade, such as grass, woodland and bare earth).
Golf Course	A specially prepared area intended for playing golf.
Other Sports Facility	Land used for sports not specifically described by other categories. Includes those facilities where participation in sport is the primary use of the area.
Play Space	A specially prepared area intended for children's play, usually linked to housing areas or parks and containing purpose-built equipment. Not captured if within schools or paid-for tourist attractions.
Playing Field	Large, flat areas of grass or specially designed surfaces, generally with marked pitches, used primarily for outdoor sports, i.e. football, rugby, cricket.
Public Park or Garden	Areas of land designed, constructed, managed and maintained as a public park or garden. These normally have a defined perimeter and free public access, and generally sit within or close to urban areas. Access is granted for a wide range of uses and not usually restricted to paths or tracks within the area. May include areas with managed facilities such as benches and flowerbeds, and more natural areas.
Tennis Court	A specially prepared area intended for playing tennis.

OS Open Greenspace layer has been used in a series of studies, which proved that the datasets support scientific research very well. For example, Liao et al. (2021) used OS greenspace dataset to evaluate the accuracy of open datasets such as OSM for greenspace mapping in London. Robinson et al. (2022) used OS Open Greenspace data to determine greenspace presence, type, number of greenspaces in Great Britain urban centre. All of these studies proved that the OS Open Greenspace dataset greatly contributed to UGS research. Thus this dataset has been select in the thesis.

3.1.3 Twitter datasets collection and its application

Twitter is an online social media and social networking service (Kwak, 2010). Twitter was created in March 2006 and launched in July of that year. People can register as Twitter users and they can post or reply texts, images, and videos known as Tweets. Within Twitter users can post Tweets, like Tweets, re-Tweet Tweets. In addition, Twitter allow users to add the geospatial information and timestamps when post Tweets, the geo-tagged Tweets provide rich information including text content, coordinates, and timestamps. In 2017, Twitter allowed users to post limitation of 280 characters in theirs posts. Twitter users can also upload a photo with attached description in a Tweet. In April 2022, Twitter made the ability to add and view captions globally available. Descriptions can be added to any uploaded image with a limit of 1000 characters.

Twitter provides free apps for all of the mobile phones including iPhone, iPad, and Android, which enable users to post Tweets anywhere at any time, especially in urban areas where there are excellent internet and Wi-Fi services. The total number of Twitter active users in the UK is 19.05 million as of October 2021. The Twitter penetration rate among internet users in the UK in 2022 was 45%. In January 2020, men accounted for 58.5% and women for 41.5% of UK Twitter users (Strugar, 2022). At least 500 million Tweets are sent on a daily basis, meaning there are plenty of conversations to join. Twitter users are more likely than the general public to have both a college degree and higher income than the average adult worldwide, the most common age group is 25-34 years old. It was found that 10% of users who are most active on Twitter are responsible for 80% of all tweets, focusing mainly on the topics of politics and women (Wojcik and Hughes, 2019).

Twitter data has been proved to provide much useful data for UGS research, Twitter data can be collected through Twitter API, which can be used to retrieve and analyse Twitter data, as well as build for the conversation on Twitter. Over the years, the Twitter API has grown by adding additional levels of access for developers and academic researchers to be able to scale their access to enhance and research the public conversation. However, different levels of access may lead to different results. Twitter datasets have been used in previous studies related to UGS research. For example, Donahue et al. (2018) used Twitter data to understand the drivers of urban park visitation and the factors that affecting the use of UGS. Chapman et al. (2018) and Plunz et al. (2019) used Twitter data to investigate the sentiment variation and emotional responses of users to UGS. These studies demonstrated how Twitter data can be used in UGS research including sentiment analysis, spatial-temporal analysis, and topic analysis. However, there is a lack of study focusing on using the consecutive datasets to track the changes in UGS use. Wang and Li (2022) detected the changes of the use and perception of UGS during the pandemic period to explore the impact of COVID-19 lockdown, this study employed the online questionnaire surveys methods and sent the online questionnaire , which might cannot provide a real time analysis. In addition, the study focused on the period from March to April 2022, rather than the first lockdown period in 2020, which might lost the initial feeling of the lockdown. In order to reduce the influence of the weather and climates changes, the same period from march to June in three consecutive years in 2019, 2020, and 2021 were chooses as the study period, which could reflect the reliable study results and provide real time analysis.

In 2021, Twitter provided a new API for academic research which allowed researchers to access Twitter's real-time and historical public data with additional features and functionality that support collecting more precise, complete, and unbiased datasets (Pfeffer et al., 2022). This enable researchers to detect the changes in UGS use, especially during the time of crisis such as Covid-19 pandemics. The research academic API will be used to collect the datasets in the thesis, in which way the research results will be reliable and have good enough representativeness.

In the thesis, the Twitter datasets were collected via the Twitter academic research API. The API call selected geo-referenced Tweets located in London. They covered a three-month period (23rd March to 23rd Jun) for three consecutive years: 2019, 2020, and 2021. To begin with, the changes in the spatial distributions of geo-located Tweets for three individual years were analysed and the most frequently visited UGSs were identified in each year. Activities in UGS were also identified using text mining (Salloum et al., 2017), and their spatio-temporal changes were investigated. The Twitter data were analyzed over hours and days to identify changes in temporal patterns. A paired sample t-test was used to significant differences between the use of UGS over the three years. All the analyses were undertaken using R software (Ihaka and Gentleman, 1996).

3.2 Methodology

3.2.1 The methodology of undertaking a review with meta-analysis

To conduct a systematic literature review, the clear methodological steps should be pre-defined and the appropriate analysis tool should be selected, the details descriptions were shown as follows.

(1) Define the research questions

The first step is to pre-define the research questions under a specific research field, which aims to summarize the knowledge of existing studies and identify future research directions. A good systematic review will have a clear research question and focus on evidence that has been published on a topic or fields (Linnenluecke et al., 2020).

Additionally, this step is crucial because the pre-defined research questions would be the basis for the next steps of making search strategy and defining selection criteria. For example, Hunter et al. (2015) reviewed the impact of interventions to promote physical activity in UGS, and focused on assessing the effectiveness of interventions to encourage physical activity in urban green space. Such a clear question focusing on interventions and outcomes in a specific context is advantageous for a review, as it delineates clear boundaries. Stock (2018) reviewed the existing literatures about the approaches that were used to extract and analyse geographic data from social media, motivated by this goal, three specific research questions were selected which guided them to select the appropriate databases and make the suitable search strategies.

(2) Make a search strategy

A variety of databases can be used to search for relevant articles. A suitable search strategy should consider the factors such as scientific databases, searching time periods, keywords for related articles, and languages of articles. For example, Hunter et al. (2015) choose five databases and set the keywords of ‘physical activity’, ‘urban green space’ and ‘intervention’ in searching stage, and the time period was published before July 2014. Stock (2018) searched the candidate papers through four databases including EBSCO Discover, Web of Science, Scopus, and Google Scholar, by using the key words of ‘social media’, ‘geographic OR geospatial OR spatiotemporal’, and ‘mining OR extract’. The collection period ended to 19 July 2017.

(3) Eligibility criteria

After executing the search strategy across all selected databases, a number of potentially relevant studies are typically returned. Then it is necessary to establish inclusion and exclusion criteria for data cleaning. These criteria should be pre-defined according to the research scope to ensure that only relevant studies are included in the analysis. For example, only English articles, full-text available articles, and published in peer reviewed academic journal articles were included (Hunter et al., 2015; Stock, 2018). Additionally, review articles, duplicates, reports and articles that did not meet the specific research goals were excluded. Generally, systematic review papers should provide the details of inclusion and exclusion criteria, which allow the readers to understand how did the articles were selected and removed from the databases. Furthermore, during the screening process, the reviewers should carefully examine the title, abstract, and keywords of each record to assess their relevance to the research question. However, in some cases, it might be necessary to refer to the full publication text to determine the suitability of the publication for inclusion in the review.

(4) Synthesise and present results

After the data cleaning process, the selected articles were subjected to various approaches for synthesis and analysis. The appropriate analysis methods and visualisation techniques should be selected based on the number of articles that would be analysed. For example, a review that comprised only a small number of studies was not suitable for meta-analysis, while for a review that contained larger samples of papers, meta-analytic regression analysis could be used to extract findings and summarise knowledges (Linnenluecke et al., 2020).

There are numerous ways to present the results of a systematic literature review. Bibliometric analysis was one of the popular methods in conducting a systematic review. A central part of this technique is the production of a bibliographic map of the topic of interest for visualising the intellectual origins of that topic and the structure of the literature over time. Previous systematic review articles used various bibliographic tools such as *HistCite* and *CiteSpace (Tho et al., 2017), ResGap (http://resgap.com), Burstiness,* and *R packages including Revtools (Westgate, 2019), Litsearchr (Koffler et al., 2021), Metagear (Lajeunesse, 2016),* and *bibliometrix* (Aria and Cuccurullo, 2017), of which the *Bibliometrix* is by far the most popular R package which was used in an increasing number of papers (Rodríguez-Soler et al., 2020; Abbas et al., 2022).

Bibliometrix allows R users to import a bibliography database from SCOPUS or the Web of Science, stored either as a Bibtex (.bib) or Plain Text (.txt) file. The package has simple functions which allow for descriptive analyses, for example, in terms of most relevant authors by the number of publications, or the most cited documents sorted by GCS. Systematic reviews generally involve analysis of the conceptual structure among the articles analysed. Bibliometrix has many additional functions, such as thematic maps or the thematic evolution of concepts. The package also facilitates various network analyses, including co-citation analysis, coupling analysis, collaboration analysis or co-occurrence analysis.

Overall, the current section 2.4 introduces the concepts of literature review for academic research and the methods of conducting a literature review by using meta-data and bibliometric techniques. The importance of systematic review was emphasised because it can provide a more objective, comprehensive, reliable and instructive research results for a particular research question or research topic. Additionally, the *Bibliometrix* R package was introduced as it was increasingly used to conduct systematic review.

3.2.2 Methods for Text-based data analysis

Social media data provides a rich source of information, including texts, images, coordinates, and timestamps, that can enable researchers to explore the use of urban green spaces from various perspectives. Texts, in particular, can provide valuable insights into users' perceptions of UGS, their activities and events within UGS, and the topics and sentiments related to UGS. However, analysing text data from social media can be challenging because it is often unstructured and noisy. To address this challenge, researchers have developed various techniques for handling and analysing short texts, such as Tweets. These techniques include natural language processing (NLP) methods, such as sentiment analysis, topic modeling, and entity recognition, as well as machine learning algorithms, such as classification and topic modelling. This sub-section reviews the studies related to the techniques for handling and analysing short text like Tweets, which is actually a kind of unstructured data.

Text mining and natural language processing (NLP) are two techniques which can support researchers to extract and retrieve meaningful information and knowledge from textual context in documents. Specifically, text mining refers to the process of extracting high-quality information that is useful for particular purposes (Feldman and Dagan, 1995). Text is unstructured, amorphous, and complex to deal with. However, despite its challenges, text remains the most common medium for the formal exchange of information. (Gharehchopogh and Khalifelu, 2011). The notion of NLP was developed in the 1950s as the sub-field of artificial intelligence and linguistics (Nadkarni et al., 2011; Gharehchopogh and Khalifelu, 2011; Liddy, 2001). NLP aims to use computers to understand natural language, originally different from textual information retrieval (Manning and Schutze, 1999). Generally, text mining is extraction of interesting and useful patterns in text data but NLP technologies is as information discovery and related techniques supply text classification, text categorization, document clustering, finding groups of similar documents, information extraction, summarization and etc. Text mining techniques are dedicated to information extraction from unstructured textual data and NLP can then be seen as an interesting tool for the enhancement of information extraction procedures (Rajman and Besançon, 1998).

Text mining and NLP can be used to clean, pre-process, categorize, cluster, summarize, and extract insights from unstructured text data, numerous techniques and algorithms were widely used in social media text data analysis. This subsection provides an overview of the methods and applications of text mining and NLP techniques for processing and analysing social media text data. The brief introductions of text mining were given as follows.

(1) Text data cleaning, which aims to get rid of any extraneous information, like stopwords, punctuation, HTML entities, tags, non-alphabets, and any other kind of characters which might not be a part of the language (Lansley and Longley, 2016; Rajman and Besançon, 1998; Feldman and Dagan, 1995).

Stopwords refer to the most common words in any language, and these words do not add much information to the text, hence stopwords are generally removed before processing the text datasets (Sarica and Luo, 2021). Examples of a few stopwords in English are "the", "a", "an", "so", "what". There are two approaches for removing the stopwords, one is to count the frequency of all the words across the whole datasets, and then remove the words that larger than threshold value, another way is to select a pre-defined stopwords list (Sarica and Luo, 2021). There are many libraries of stopwords such as the Natural Language Toolkit (NLTK) (Bird, 2006) and Gensim (Řehůřek and Sojka, 2011) packages in Python; and the tm (Feinerer, 2013) and tidytext (Silge and Robinson, 2016) packages in R, which can be used for the removal of English stop words. These pre-defined stopwords enable researchers to efficiently work on text cleaning.

It should be noted that the selection of stopwords is important because it depends on the specific research tasks and purposes, which might result in different results. For example, in applications like document search engines and document classification, where keywords are more important than general terms, removing stopwords can be a good idea, but if there's some application about, for instance, lyrics search, or search specific quotes, stopwords can be important. Consider some examples like "To be, or not to be", "Look what you made me do" etc. Stopwords in such phrases actually play an important role, and hence, should not be dropped. Social media text data, such as Tweets, is often considered short text. Tweets are typically composed of incomplete, noisy, and poorly structured sentences, irregular expressions, ill-formed words, and non-dictionary terms, which can make processing and analyzing this data challenging (Jianqiang and Xiaolin, 2017). Thus, the text cleaning is essential to the next steps of analysis. After stopwords cleaning, then the next step is to remove tags, punctuations, numbers, and URL links which do not provide much useful information in text contents, especially in sentiment analysis and topic modelling (Jianqiang and Xiaolin, 2017; Sarica and Luo, 2021).

(2) Spammer, bot and fake account detection and removal (Data cleaning)

The increase of social media platforms such as Facebook, Twitter, and Instagram, have attracted large number of users around the world, which allow researchers to detect the UGS use by analysing the social media datasets. However, the results can be influenced by huge amount of fake accounts that do not belong to real humans, which can spread fake news, misleading web rating, and spam (Erşahin et al., 2017). For example, a coffee shop within UGS areas could create a fake account and post large numbers of duplicate Tweets and advertisements. Identifying and removing these fake accounts is crucial to ensure the accuracy and reliability of the research findings.

A wide variety of strategies and tools are available online that are said to be useful in identifying fake or spam accounts. For example, Lansley and Longley (2016) have identified the fake accounts by counting identical messages that occurred more than three times in the datasets, and also remove the accounts that have posted more than 3000 Tweets within one year. The approach in (Erşahin et al., 2017) was used to detect the fake account by using Naïve Bayes classification algorithm. Additionally, the user accounts need to be further separated according to specific research questions. For example, in the analysis of evaluating the sentiments of tourists or urban residents, researchers would firstly identify the tourists or remove them among large number of users. For example, Plunz et al. (2019) identified the users who have at least a 15-day gap between their first and last geolocated Tweets in the datasets as urban residents in New York.

(3) Text pre-processing is putting the cleaned text data into a form that text mining algorithms can quickly and simply evaluate. In this step, NLP techniques including tokenization, stemming, and lemmatization are all part of this process, which help computers understand, interpret and manipulate and respond to human in their natural language (Hodorog et al., 2022). Specifically, tokenization is the process breaking

complex data like paragraphs into simple units called tokens. There are two types of tokenization including sentence tokenization and word tokenization, the former refers to split a paragraph into list of sentences, and the later one refers to split a sentence into list of words (Webster and Kit, 1992). Word tokenization was popular in NLP especially in social media text data analysis (Pinto et al., 2016). Stemming is a normalization technique where list of tokenized words are converted into shorten root words to remove redundancy (Singh, J. and Gupta, 2017). For example, a stemming program transform the words like fishing, fished, and fisher to the stem fish. One of the drawbacks of stemming is that this progress produces intermediate representation of word, which may or may not return meaningful word. Lemmatization is therefore preferred because it considers morphological analysis of the words and returns meaningful word in proper form (Balakrishnan and Lloyd-Yemoh, 2014).

(4) After text cleaning and pre-processing, it is necessary to transform text data into numerical data such as a vector space model, this transformation task is generally called feature extraction of document data (Neethu and Rajasree, 2013). The results of feature extraction can be input for further analysis such as machine learning which can only understand numbers and doesn’t understand text data. There are various types of feature extraction techniques such as word frequency, n-grams, term frequency-inverse document frequency are typical features (TF-IDF), and word embedding.

(5) Finally, modelling aims to extract insights from the text data, collected features are subjected to text mining algorithms, in which the most frequently used were machine learning algorithms which will be further introduced in section 2.2.2.2 Overall, text mining and NLP techniques are crucial for advancing UGS management and planning because it is needed to transform relevant information locked in text into structured data that can be used by computer processes aimed at understanding the interactions between human and UGS.

3.2.3 Machine learning and topic modelling

Machine learning (ML) is a field of computer science and artificial intelligence that focuses on the development of algorithms and statistical models that can learn from and make predictions or decisions based on data, without being explicitly programmed (Mahesh, 2020). The purpose of ML is to make machines learn by themselves from the data and then improve the efficiency of handling, exploring and analysing the data

(Mahesh, 2020). Especially with the abundance of social media datasets available, the demand for machine learning is in rise (Balaji et al., 2021). ML consists of supervised ML and unsupervised ML. Supervised ML algorithms refers to the algorithms which needs external assistance (Singh, A. et al., 2016). Which means that the input dataset is divided into train and test dataset. The train dataset has output variable which needs to be predicted or classified. All algorithms learn some kind of patterns from the training dataset and apply them to the test dataset for prediction or classification. The commonly used supervised ML algorithms include Decision Tree, Navie Bayes, Support Vector Machine. Unsupervised ML algorithms mean that there is neither correct answers nor teachers for training the algorithms. Unsupervised algorithms are left to their own devices to discover and present the interesting structure in the data. The unsupervised learning algorithms learn few features from the data. When new data is introduced, it uses the previously learned features to recognize the class of the data. It is mainly used for clustering and feature reduction. The most commonly used unsupervised ML include K-means clustering, Principal Component Analysis, and topic modelling.

In terms of the application of ML in the field short texts analysis, the typical text classification techniques such as sentiment analysis and topic modelling have been widely used for clustering social media texts (Garcia and Berton, 2021; Rodrigues and Chiplunkar, 2019; Dahal et al., 2019). Topic modelling is gaining an increasing attention by scholars in exploring hidden information from large volumes of textual data. For example, Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) have been used in event detection by using Twitter data (Suri and Roy, 2017), and the LDA has been proved to be more semantically interpretable. Latent Dirichlet Allocation (LDA) is one of the topic modelling approaches and it is a powerful method which can detect the topics from textual data such as articles, newspapers, and social media posts such as Tweets (Jelodar et al., 2019; Blei et al., 2003; Lansley and Longley, 2016).

LDA is a generative probabilistic model largely used for topic modelling and is able to identify the probability of a given document being in a given topic through a "bag-of-words" interpretation of its contents (Blei et al., 2003). LDA is a three-level hierarchical Bayesian model, within which every item of a corpus is modelled as a finite mixture over an underlying set of topics. Each topic is then modelled as an infinite mixture over an underlying set of topic probabilities. There topic possibilities offer a specific illustration of a document (Blei et al., 2003). LDA represents documents as a mixture of topics that contain words with certain probabilities of occurrence. Given a collection of documents, some fixed number of topics to find, LDA learns the topic representation of each document and therefore the words associated to each topic via an iterative procedure. LDA then tries to backtrack from the documents to seek out a collection of topics that are likely to have generated the collection.

In the context of modern city like London and New York, Lansley and Longley (2016) have employed LDA to classify geo-tagged Tweets from Inner London to identify the 20 topic groups across the whole city, this study proved that the LDA could be effectively used to identify the topics in short text datasets such as Tweets. Additionally, in the analysis of interactions between human and UGS, Sim et al. (2020) have used LDA to investigate the activities within UGS in New York City, which proved that LDA could be used to detect the UGS activities via geo-tagged Tweets. During the covid-19 pandemics, Dorostkar and Najarsadeghi (2022) have used LDA to explore the emotions on Twitter social media in Tehran, the results shows that the emotions created by social media can positively or negatively affect the community, and the published tweets have a strong reaction to the COVID-19.

However, the themes in a document collection evolve over time, and it is of interest to explicitly model the dynamics of the underlying topics. In order to track the temporal changes of topics, a temporal component to topic modelling, referred to as dynamic topic model (DTM) introduced by (David M and John D, 2006). This model can capture the evolution of topics in a sequentially organized corpus of documents, this model evaluates the changes of topic proportions and word probability in a corpus of documents along a time series text data.

Recently, building upon earlier approaches to topic modelling such as LDA, Roberts, M.E. et al. (2014) developed structural topic modelling (STM) that allows researchers to make greater use of observed structural variables that characterize documents. STM modelling assumes that topics can be correlated with one another and that X covariates structure the distribution of topics over documents (Roberts, M.E. et al., 2014). The key innovation of STM is that it allows researchers to incorporate arbitrary metadata including information of each document into the topic model. The goal of the STM is to allow researchers to discover topics and estimate their relationship to document metadata, as well how topics change along the timeline.

STM is also a generative model of word counts. STM defines a data generating process for each document and then use the data to find the most likely values for the parameters within the model. Figure 3.1 provides a graphical representation of the model. The generative model begins at the top, with document-topic and topic-word distributions generating documents that have metadata associated with them (Xd, where d indexes the documents). Within this framework (which is the same as other topic models like LDA absent the metadata), a topic is defined as a mixture over words where each word has a probability of belonging to a topic. And a document is a mixture over topics, meaning that a single document can be composed of multiple topics. As such, the sum of the topic proportions across all topics for a document is one, and the sum word probabilities for a given topic is one.

Figure 3.1 and the statement below of the document generative process highlight the case where topical prevalence and topical content can be a function of document metadata. Topical prevalence refers to how much of a document is associated with a topic (described on the left hand side) and topical content refers to the words used within a topic (described on the right hand side). Hence metadata that explain topical prevalence are referred to as topical prevalence covariates, and variables that explain topical content are referred to as topical content covariates. It is important to note, however, that the model allows using topical prevalence covariates, a topical content covariate, both, or neither. In the case of no covariates, the model reduces to a (fast) implementation of the Correlated Topic Model (Blei and Lafferty, 2007). In the case of no covariates and point estimates for β, the model reduces to a (fast) implementation of the Correlated Topic Model (Blei and Lafferty, 2007).

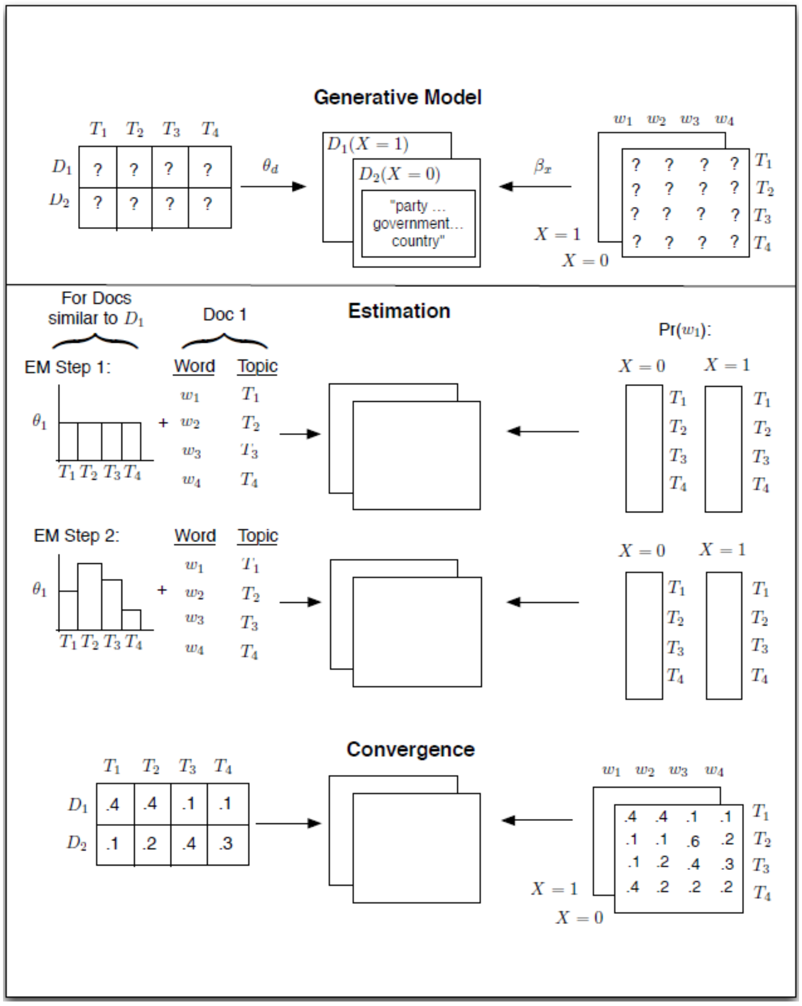

Figure 3. 2 Heuristic description of generative process and estimation of the STM. Source: (Roberts, M.E. et al., 2019).

The generative process for each document (indexed by d) with vocabulary of size V for a STM model with K topics can be summarized as:

1. Draw the document-level attention to each topic from a logistic-normal generalized linear model based on a vector of document covariates X_d.

$$\vec{\theta}_d | X_{d\gamma}, \Sigma \sim LogisticNormal(\mu = X_{d\gamma}, \Sigma) \quad (1)$$

where X_d is a 1-by-p vector, γ is a p-by-K −1 matrix of coefficients and ∑ is K −1-by-K − 1 covariance matrix.

2. Given a document-level content covariate yd, form the document-specific distribution over words representing each topic (k) using the baseline word distribution (m), the topic specific deviation $k_k^{(t)}$, the covariate group deviation $k_{yd}^{(c)}$ and the interaction between the two $k_{yd,k}^{(i)}$.

$$\beta_{d,k} \propto exp\ (m + k_k^{(t)} + k_{yd}^{(c)} + k_{yd,k}^{(i)}) \quad (2)$$

m, and each $k_k^{(t)}$k, $k_{yd}^{(c)}$, $k_{yd,k}^{(i)}$ are V -length vectors containing one entry per word in the vocabulary. When no convent covariate is present β can be formed as $\beta_{d,k} \propto \exp\left(m + k_k^{(t)}\right)$ or simply point estimated (this latter behavior is the default).

3. For each word in the document, $(n \in 1, \dots, N_d)$:

- Draw word's topic assignment based on the document-specific distribution over topics.

$$Z_{d,n}|\vec{\theta}_d \sim Multinomial(\vec{\theta}_d) \quad (3)$$

- Conditional on the topic chosen, draw an observed word from that topic.

$$\omega_{d,n}|Z_{d,n}, \beta_{d,k=Z_{d,n}} \sim Multinomial(\beta_{d,k=Z_{d,n}}) \quad (4)$$

STM has been widely used in previous studies, for example, Chen et al. (2020) used STM to investigate the evaluation of topics in computers and education academic publications over 40 years. Sachdeva et al. (2017) used STM to detect the topics of Tweets in order to investigate the air quality impacts of wildfire events in the United States, the results suggested that social media can be a valuable tool for providing insight into the socio-psychological dimensions of fire and smoke and their impact on people residing in affected areas. These studies demonstrated that STM modelling enabled the creation of tools for investigating the topic prevalence and relevant trends in the research. However, to the best of our knowledge, there are limited studies have use STM to investigate the changes in the topics expressed by Twitter users when visited UGS, in the context of social crisis such as COVID-19 pandemics.

3.2.4 Spatial-temporal analysis of social media data

Geotagged social media data can provide high resolution of time and space information, which allows researchers to investigate the spatial and temporal patterns of social media users. This has the potential to provide valuable information about the location and time of emergency events, as well as the needs and concerns of affected communities within city areas. Previous studies have used various techniques to quantify the spatial and temporal patterns of users, and this subsection provide an overview of these relevant techniques.

3.2.4.1 Inverse distance weight analysis

Various spatial analysis methods were implied in analysing the spatial patterns of point datasets. For example, Norman and Pickering (2017) collected volunteered geographic information (VGI) through three platforms to map the spatial distributions of urban park visitors. They employed 'points to line' and 'line density' tools in ArcMap (Institute,

2011) to evaluated the line densities of the park users, the spatial techniques in the study have helped authors to understand the spatial variations in park visit patterns. Tenkanen et al. (2017) and Girardin et al. (2009) used point density techniques to show the spatial distributions of social media users across the study area.

There are various spatial interpolation methods (Li, J. and Heap, 2011) of which inverse distance weighting (IDW) is one of the most popular or frequently used method. IDW is a kind of deterministic interpolation method that creates a continuous surface of values based on point data, where the values at any given location are determined by the weighted average of nearby points. The weight assigned to each point is inversely proportional to the distance from that point to the location being estimated.

IDW has been developed for and applied in many disciplines (Li, J. and Heap, 2014), such as environmental sciences analysis including air pollution and water quality, epidemiology, and agriculture analysis (Gu et al., 2021; Fatima et al., 2022; Pereira et al., 2022). However, previous studies rarely employed IDW to assess the spatial patterns of topics, particularly in UGSs. In the current thesis, IDW approach will be used to explore the spatial variations of potential topics across the whole of London over all time periods. The computational implementation of the IDW was performed using the *gstat* package (Gräler et al., 2016) from R software.

3.2.4.2 Temporal analysis of social media data

In the context of social media data associated to UGS, temporal analysis can potentially provide information about how temporal patterns of UGS use and topics evolved over time. This includes identifying trends or patterns in the frequency and timing of visitation related to specific topics or activities. Such analyses can offer valuable information for UGS research and decision-making.

The aggregation of social media data to an appropriate scale is crucial for accurate and meaningful analysis, depending on the specific research purpose. For example, Tenkanen et al. (2017) aggregated visitors from social media by summing the number of active social media users per day. If a user posted 5 Tweets within a single day, the user is treated as true human user. Additionally, the number of users were further grouped to a monthly level as the official statistics from the parks were reported on a monthly basis, which enable the results could be compared with official statistics. In this thesis, multiple temporal patterns were analysed by aggregating the data into daily, weekly and yearly patterns, which can potentially provide comprehensive information about the dynamic influence of COVID-19 pandemic on UGS visitation and the relevant topics and activities.

References

Abbas, A.F., Jusoh, A., Mas' od, A., Alsharif, A.H. and Ali, J. 2022. Bibliometrix analysis of information sharing in social media. *Cogent Business & Management.* ***9***(1), p2016556.

Ahmed, W., Bath, P.A. and Demartini, G. 2017. Using Twitter as a data source: An overview of ethical, legal, and methodological challenges. *The Ethics of Online Research.* ***2***, pp.79-107.

Ajao, O., Hong, J. and Liu, W., 2015. A survey of location inference techniques on Twitter. *Journal of Information Science,* ***41***(6), pp.855-864.

Aria, M. and Cuccurullo, C. 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics.* ***11***(4), pp.959-975.

Balaji, T., Annavarapu, C.S.R. and Bablani, A. 2021. Machine learning algorithms for social media analysis: A survey. *Computer Science Review.* ***40***, p100395.

Balakrishnan, V. and Lloyd-Yemoh, E. 2014. Stemming and lemmatization: A comparison of retrieval performances.

Belmeziti, A., Cherqui, F. and Kaufmann, B. 2018. Improving the multi-functionality of urban green spaces: Relations between components of green spaces and urban services. *Sustainable cities and society.* ***43***, pp.1-10.

Benhardus, J. and Kalita, J. 2013. Streaming trend detection in twitter. *International Journal of Web Based Communities.* ***9***(1), pp.122-139.

Bird, S. 2006. NLTK: the natural language toolkit. In: *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions,* pp.69-72.

Blank, G. and Lutz, C. 2017. Representativeness of social media in great britain: investigating Facebook, Linkedin, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist.* ***61***(7), pp.741-756.

Blei, D.M. and Lafferty, J.D. 2007. A correlated topic model of science.

Blei, D.M., Ng, A.Y. and Jordan, M. 2003. Latent dirichlet allocation. *Journal of machine Learning research.* ***3***(01), pp.993-1022.

Breuste, J., Haase, D. and Elmqvist, T. 2013. Urban landscapes and ecosystem services. *Ecosystem services in agricultural and urban landscapes,* pp.83-104.

Catney, G., Lloyd, C.D., Ellis, M., Wright, R., Finney, N., Jivraj, S. and Manley, D. 2023. Ethnic diversification and neighbourhood mixing: A rapid response analysis of the 2021 Census of England and Wales. *The Geographical Journal.*

Chaffey, D. 2023. *Global social media statistics research summary 2023.* [Online]. [Accessed 6 March]. Available from: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/

Chapman, L., Resch, B., Sadler, J., Zimmer, S., Roberts, H. and Petutschnig, A., 2018. Investigating the emotional responses of individuals to urban green space using twitter data: A critical comparison of three different methods of sentiment analysis. *Urban Planning**.** **3***(1), pp.21-33.

Chen, X., Zou, D., Cheng, G. and Xie, H. 2020. Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education. *Computers & Education.* ***151***, p103855.

Chinmoy Sarkar, Chris Webster, Matthew Pryor, Dorothy Tang, Scott Melbourne, Xiaohu Zhang and Jianzheng, L. 2015. Exploring associations between urban green, street design and walking: Results from the Greater London boroughs. *Landscape and Urban Planning.* ***143***, pp.112-125.

Cinnamon, J. and Schuurman, N. 2013. Confronting the data-divide in a time of spatial turns and volunteered geographic information. *GeoJournal.* ***78***, pp.657-674.

Cohen, D.A., Marsh, T., Williamson, S., Derose, K.P., Martinez, H., Setodji, C. and McKenzie, T.L. 2010. Parks and physical activity: why are some parks used more than others? *Preventive medicine.* ***50***, pp.S9-S12.

Da Schio, N., Phillips, A., Fransen, K., Wolff, M., Haase, D., Ostoić, S.K., Živojinović, I., Vuletić, D., Derks, J. and Davies, C. 2021. The impact of the COVID-19 pandemic on the use of and attitudes towards urban forests and green spaces: Exploring the instigators of change in Belgium. *Urban Forestry & Urban Greening.* ***65***, p127305.

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***, pp.1-20.

David M, B. and John D, L. 2006. Dynamic topic models. In: *Proceedings of the 23rd international conference on Machine learning*, pp.113–120.

Donahue, M.L., Keeler, B.L., Wood, S.A., Fisher, D.M., Hamstead, Z.A. and McPhearson, T. 2018. Using social media to understand drivers of urban park visitation in the Twin Cities, MN. *Landscape and Urban Planning.* ***175***, pp.1-10.

Dorostkar, E. and Najarsadeghi, M. 2022. How to evaluate urban emotions using twitter social media? *Cities*. ***127***, p103713.

Elwood, S., Goodchild, M.F. and Sui, D.Z. 2012. Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the association of American geographers*. ***102***(3), pp.571-590.

Erşahin, B., Aktaş, Ö., Kılınç, D. and Akyol, C. 2017. Twitter fake account detection. In: *2017 International Conference on Computer Science and Engineering (UBMK)*, Antalya, Turkey, pp. 388-392.

Fan, P., Xu, L., Yue, W., Chen, J.J.L. and Planning, U. 2016. Accessibility of public urban green space in an urban periphery: The case of Shanghai. pS0169204616302432.

Fatima, M., Butt, I. and Arshad, S. 2022. Geospatial clustering and hot spot detection of malaria incidence in Bahawalpur district of Pakistan. *GeoJournal*. ***87***(6), pp.4791-4806.

Feinerer, I. 2013. Introduction to the tm Package Text Mining in R. *Accessible en ligne: http://cran. r-project. org/web/packages/tm/vignettes/tm. pdf*.

Feldman, R. and Dagan, I. 1995. Knowledge Discovery in Textual Databases (KDT). In: *KDD*, pp.112-117.

Garcia, K. and Berton, L. 2021. Topic detection and sentiment analysis in Twitter content related to COVID-19 from Brazil and the USA. *Applied soft computing*. ***101***, p107057.

Geng, D., Innes, J., Wu, W. and Wang, G. 2021. Impacts of COVID-19 pandemic on urban park visitation: a global analysis. *Journal of forestry research*. ***32***(2), pp.553-567.

Gharehchopogh, F.S. and Khalifelu, Z.A. 2011. Analysis and evaluation of unstructured data: text mining versus natural language processing. In: *2011 5th International Conference on Application of Information and Communication Technologies (AICT)*, Baku, Azerbaijan, 2011, pp. 1-4.

Ghermandi, A. and Sinclair, M. 2019. Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change*. ***55***, pp.36-47.

Girardin, F., Vaccari, A., Gerber, A., Biderman, A. and Ratti, C. 2009. Quantifying urban attractiveness from the distribution and density of digital footprints. *International Journal of Spatial Data Infrastructures Research*. ***4***, 175-200.

Goodchild, M.F. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*. ***69***, pp.211-221.

Gräler, B., Pebesma, E.J. and Heuvelink, G.B. 2016. Spatio-temporal interpolation using gstat. *The Royal Journal*. ***8***(1), p204.

Grima, N., Corcoran, W., Hill-James, C., Langton, B., Sommer, H. and Fisher, B. 2020. The importance of urban natural areas and urban ecosystem services during the COVID-19 pandemic. *PloS one*. ***15***(12), pe0243344.

Grzyb, T., Kulczyk, S., Derek, M. and Woźniak, E. 2021. Using social media to assess recreation across urban green spaces in times of abrupt change. *Ecosystem Services*. ***49***, p101297.

Gu, K., Zhou, Y., Sun, H., Dong, F. and Zhao, L. 2021. Spatial distribution and determinants of PM 2.5 in China's cities: Fresh evidence from IDW and GWR. *Environmental monitoring and assessment*. ***193***, pp.1-22.

Guo, S., Yang, G., Pei, T., Ma, T., Song, C., Shu, H., Du, Y. and Zhou, C. 2019. Analysis of factors affecting urban park service area in Beijing: Perspectives from multi-source geographic data. *Landscape and Urban Planning*. ***181***, pp.103-117.

Hamstead, Z.A., Fisher, D., Ilieva, R.T., Wood, S.A., McPhearson, T. and Kremer, P. 2018. Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*. ***72***, pp.38-50.

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***, pp.38-54.

Heather E. Wright Wendel, R.K.Z., James R. Mihelcic,. 2012. Accessibility and usability: Green space preferences, perceptions, and barriers in a rapidly urbanizing city in Latin America. *Landscape and Urban Planning*. ***107***, pp.272-282.

Heikinheimo, V., Minin, E.D., Tenkanen, H., Hausmann, A., Erkkonen, J. and Toivonen, T. 2017. User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*. ***6***(3), p85.

Heikinheimo, V., Tenkanen, H., Bergroth, C., Järv, O., Hiippala, T. and Toivonen, T. 2020. Understanding the use of urban green spaces from user-generated geographic information. *Landscape and Urban Planning*. ***201***, p103845.

Hodorog, A., Petri, I. and Rezgui, Y. 2022. Machine learning and Natural Language Processing of social media data for event detection in smart cities. *Sustainable Cities and Society*. ***85***, p104026.

Hunter, R.F., Christian, H., Veitch, J., Astell-Burt, T., Hipp, J.A. and Schipperijn, J. 2015. The impact of interventions to promote physical activity in urban green space: a systematic review and recommendations for future research. *Social science & medicine*. ***124***, pp.246-256.

Institute, E.S.R. 2011. *ArcGIS desktop: release 10*. Environmental Systems Research Institute Redlands, CA.

Jankowska, M.M., Schipperijn, J. and Kerr, J. 2015. A framework for using GPS data in physical activity and sedentary behavior studies. *Exercise and sport sciences reviews*. ***43***(1), p48.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. and Zhao, L. 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools Applications*. ***78***(11), pp.15169-15211.

Jianqiang, Z. and Xiaolin, G. 2017. Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE access*. ***5***, pp.2870-2879.

Jing-Huei, H., Myron F., F., Laura G., T. and J., A.H. 2022. Exploring public values through Twitter data associated with urban parks pre- and post- COVID-19. *Landscape and Urban Planning*. ***227***, p104517.

Johnson, M.L., Campbell, L.K., Svendsen, E.S. and McMillen, H.L. 2019. Mapping urban park cultural ecosystem services: A comparison of twitter and semi-structured interview methods. *Sustainability*. ***11****(21)*, p6137.

Jolly, A., Thomas, S. and Stanyer, J. 2020. London’s children and young people who are not British citizens: A profile.

Kabisch, N., Qureshi, S. and Haase, D. 2015. Human–environment interactions in urban green spaces — A systematic. *Environmental Impact assessment review*. ***50***, pp.25-34.

Koffler, S., Barbiéri, C., Ghilardi-Lopes, N.P., Leocadio, J.N., Albertini, B., Francoy, T.M. and Saraiva, A.M. 2021. A buzz for sustainability and conservation: the growing potential of citizen science studies on bees. *Sustainability*. ***13***(2), p959.

Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A. and Blaschke, T. 2018. Beyond spatial proximity-classifying parks and their visitors in london based on spatiotemporal and sentiment analysis of twitter data. *ISPRS International Journal of Geo-Information.* ***7***(9), p378.

Kunisch, S., Menz, M., Bartunek, J.M., Cardinal, L.B. and Denyer, D. 2018. Feature topic at organizational research methods: how to conduct rigorous and impactful literature reviews? *Organizational Research Methods.* ***21***(3), pp.519-523.

Kwak, H. 2010. What is Twitter, a social network or a news media? *Proc. 19th Int. Conf. on World Wide Web, 2010.*

Lajeunesse, M.J. 2016. Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R. *Methods in Ecology and Evolution.* ***7***(3), pp.323-330.

Lansley, G. and Longley, P. 2016. The geography of Twitter topics in London. *Computers, Environment and Urban Systems.* ***58***, pp.85-96.

Lee, A.C.K., Jordan, H.C. and Horsley, J. 2015. Value of urban green spaces in promoting healthy living and wellbeing: prospects for planning. *Risk management and healthcare policy.* pp.131-137.

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***, pp.249-258.

Li, J. and Heap, A.D. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics.* ***6***(3-4), pp.228-241.

Li, J. and Heap, A.D. 2014. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software.* ***53***, pp.173-189.

Liao, Y., Zhou, Q. and Jing, X. 2021. A comparison of global and regional open datasets for urban greenspace mapping. *Urban Forestry & Urban Greening.* ***62***, p127132.

Liddy, E.D. 2001. Natural language processing.

Lim, K.H., Lee, K.E., Kendal, D., Rashidi, L., Naghizade, E., Winter, S. and Vasardani, M. 2018. The grass is greener on the other side: Understanding the effects of green spaces on Twitter user sentiments. In: *Companion Proceedings of the The Web Conference* ***2018***, pp.275-282.

Linnenluecke, M.K., Marrone, M. and Singh, A.K. 2020. Conducting systematic literature reviews and bibliometric analyses. *Australian Journal of Management.* ***45***(2), pp.175-194.

Lomborg, S. and Bechmann, A. 2014. Using APIs for data collection on social media. *The Information Society.* ***30***(4), pp.256-265.

Lu, Y. 2019. Using Google Street View to investigate the association between street greenery and physical activity. *Landscape Urban Planning.* ***191***, p103435.

Luederitz, C., Brink, E., Gralla, F., Hermelingmeier, V., Meyer, M., Niven, L., Panzer, L., Partelow, S., Rau, A.-L. and Sasaki, R. 2015. A review of urban ecosystem services: six key challenges for future research. *Ecosystem services.* ***14***, pp.98-112.

Lyu, F. and Zhang, L. 2019. Using multi-source big data to understand the factors affecting urban park use in Wuhan. *Urban Forestry & Urban Greening.* ***43***, p126367.

Mahesh, B. 2020. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet].* ***9***, pp.381-386.

Mahmud, J., Nichols, J. and Drews, C. 2014. *Home Location Identification of Twitter Users*. ACM. pp.1-21.

Manning, C. and Schutze, H. 1999. *Foundations of statistical natural language processing.* MIT press.

Martí, P., Serrano-Estrada, L. and Nolasco-Cirugeda, A. 2019. Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems.* ***74***, pp.161-174.

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M. and Fisher, P. 2007. The Hawthorne Effect: a randomised, controlled trial. *BMC medical research methodology.* ***7***(1), pp.1-8.

McCay-Peet, L. and Quan-Haase, A. 2017. What is social media and what questions can social media research help us answer. *The SAGE handbook of social media research methods.* pp.13-26.

Morstatter, F., Pfeffer, J., Liu, H. and Carley, K. 2013. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In: *Proceedings of the international AAAI conference on web and social media*. ***7***(1), pp.400-408.

Nadkarni, P.M., Ohno-Machado, L. and Chapman, W. 2011. Natural language processing: an introduction. *Journal of the American Medical Informatics Association.* ***18***(5), pp.544-551.

National Park City Foundation. 2023. *National Park Cities Are Grassroots Movements For People Making Their Cities Greener, Healthier And Wilder.* [Online]. [Accessed 16 March]. Available from: https://www.nationalparkcity.org/

Neethu, M. and Rajasree, R. 2013. Sentiment analysis in twitter using machine learning techniques. In: *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)*, Tiruchengode, India, pp.1-5.

Norman, P. and Pickering, C.M. 2017. Using volunteered geographic information to assess park visitation: Comparing three on-line platforms. *Applied Geography*. ***89***, pp.163-172.

Norman, P. and Pickering, C.M. 2019. Factors influencing park popularity for mountain bikers, walkers and runners as indicated by social media route data. *Journal of Environmental Management.* ***249***, p109413.

Office for National Statistics. 2022. *Population and household estimates, England and Wales: Census 2021*. ONS Newport, UK.

Ordnance Survey. 2021a. *OS Open Greenspace.* [Online]. [Accessed 25 October]. Available from: https://beta.ordnancesurvey.co.uk/products/os-open-greenspace

Ordnance Survey. 2021b. *OS Open Greenspace support: Technical information.* [Online]. [Accessed 26.02]. Available from: https://www.ordnancesurvey.co.uk/business-government/tools-support/open-map-greenspace-support#technicalInformation

Owczarczak-Garstecka, S.C., Graham, T.M., Archer, D.C. and Westgarth, C. 2021. Dog walking before and during the COVID-19 pandemic lockdown: experiences of UK dog owners. *International Journal of Environmental Research and Public Health.* ***18***(12), p6315.

Pae, C.-U. 2015. Why systematic review rather than narrative review? *Psychiatry investigation.* ***12***(3), p417.

Panduro, T.E. and Veie, K.L. 2013. Classification and valuation of urban green spaces—A hedonic house price valuation. *Landscape and Urban Planning.* ***120***, pp.119-128.

Pereira, L.C., dos Santos, G.R., Marques, E.A.G., Pires, J.D. and Renó, R. 2022. Construction of multidimensional geomechanical models with IDW and using R language. *Journal of South American Earth Sciences.* ***116***, p103775.

Pfeffer, J., Mooseder, A., Hammer, L., Stritzel, O. and Garcia, D. 2022. This Sample seems to be good enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API. *arXiv preprint arXiv*. ***2204***, 02290.

Pinto, A., Gonçalo Oliveira, H. and Oliveira Alves, A. 2016. Comparing the performance of different NLP toolkits in formal and social media text. In: *5th Symposium on Languages, Applications and Technologies (SLATE'16)*: Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Pipitone, J.M. and Jović, S. 2021. Urban green equity and COVID-19: Effects on park use and sense of belonging in New York City. *Urban Forestry & Urban Greening.* ***65***, p127338.

Plunz, R.A., Zhou, Y., Vintimilla, M.I.C., Mckeown, K., Yu, T., Uguccioni, L. and Sutto, M.P. 2019. Twitter sentiment in New York City parks as measure of well-being. *Landscape and Urban Planning*. ***189***, pp.235-246.

Rajman, M. and Besançon, R. 1998. Text mining: natural language techniques and text mining applications. In: *Data Mining and Reverse Engineering: Searching for semantics. IFIP TC2 WG2. 6 IFIP Seventh Conference on Database Semantics (DS-7) 7–10 October 1997, Leysin, Switzerland*: Springer, pp.50-64.

Řehůřek, R. and Sojka, P. 2011. Gensim—statistical semantics in python. *Retrieved from genism. org.*

Roberts, H.V. 2017. Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography*. ***81***, pp.13-20.

Roberts, M.E., Stewart, B.M. and Tingley, D. 2019. Stm: An R package for structural topic models. *Journal of Statistical Software*. ***91***, pp.1-40.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G. 2014. Structural topic models for open-ended survey responses. *American journal of political science*. ***58***(4), pp.1064-1082.

Robinson, J.M., Mavoa, S., Robinson, K. and Brindley, P. 2022. Urban centre green metrics in Great Britain: A geospatial and socioecological study. *Plos one*. ***17***(11), pe0276962.

Rodrigues, A.P. and Chiplunkar, N.N. 2019. A new big data approach for topic classification and sentiment analysis of Twitter data. *Evolutionary Intelligence*. pp.1-11.

Rodríguez-Soler, R., Uribe-Toril, J. and Valenciano, J.D.P. 2020. Worldwide trends in the scientific production on rural depopulation, a bibliometric analysis using bibliometrix R-tool. *Land Use Policy*. ***97***, p104787.

Sachdeva, S., McCaffrey, S. and Locke, D. 2017. Social media approaches to modeling wildfire smoke dispersion: Spatiotemporal and social scientific investigations. *Information, Communication & Society*. ***20***(8), pp.1146-1161.

Sarica, S. and Luo, J. 2021. Stopwords in technical language processing. *Plos one*. ***16***(8), pe0254937.

Schipperijn, J., Bentsen, P., Troelsen, J., Toftager, M. and Stigsdotter, U.K. 2013. Associations between physical activity and characteristics of urban green space. *Urban forestry & urban greening*. ***12***(1), pp.109-116.

Schipperijn, J., Bentsen, P., Troelsen, J., Toftager, M. and Stigsdotter, U.K. 2023. Associations between physical activity and characteristics of urban green space. *Urban Forestry & Urban Greening.* ***12***(1), pp.109-116.

Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C. and Haklay, M. 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*. ***31***(1), pp.139-167.

Shackleton, C.M. and Blair, A. 2013. Perceptions and use of public green space is influenced by its relative abundance in two small towns in South Africa. *Landscape and Urban Planning.* ***113***, pp.104-112.

Shi, B., Zhao, J. and Chen, P.J. 2017. Exploring urban tourism crowding in Shanghai via crowdsourcing geospatial data. *Current Issues in Tourism.* ***20***(11), pp.1186-1209.

Shoari, N., Ezzati, M., Doyle, Y.G., Wolfe, I., Brauer, M., Bennett, J. and Fecht, D. 2021. Nowhere to play: available open and green space in Greater London schools. *Journal of Urban Health.* ***98***, pp.375-384.

Silge, J. and Robinson, D. 2016. tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software.* ***1***(3), p37.

Sim, J., Miller, P. and Swarup, S. 2020. Tweeting the High Line life: A social media lens on urban green spaces. *Sustainability.* ***12***(21), p8895.

Singh, A., Thakur, N. and Sharma, A. 2016. A review of supervised machine learning algorithms. In: *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, pp.1310-1315.

Singh, J. and Gupta, V. 2017. A systematic review of text stemming techniques. *Artificial Intelligence Review*. ***48***, pp.157-217.

Steiger, E., Westerholt, R., Resch, B. and Zipf, A. 2015. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems.* ***54***, pp.255-265.

Stieglitz, S., Mirbabaie, M., Ross, B. and Neuberger, C. 2018. Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International journal of information management.* ***39***, pp.156-168.

Stock, K. 2018. Mining location from social media: A systematic review. *Computers, Environment and Urban Systems.* ***71***, pp.209-240.

Strugar, M. 2022. *Top 25 Surprising Twitter Statistics UK Edition [2022], Cyber Crew*.

Suri, P. and Roy, N.R. 2017. Comparison between LDA & NMF for event-detection from large text stream data. In: *2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT)*: IEEE, pp.1-5.

Taylor, L. and Hochuli, D.F. 2017. Defining greenspace: Multiple uses across multiple disciplines. *Landscape and urban planning.* ***158***, pp.25-38.

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), p17615.

The Royal Parks. 2023. *London's Royal Parks.* [Online]. [Accessed 12 March]. Available from: https://www.royalparks.org.uk/parks

Theodorou, A., Panno, A., Carrus, G., Carbone, G.A., Massullo, C. and Imperatori, C. 2021. Stay home, stay safe, stay green: The role of gardening activities on mental health during the Covid-19 home confinement. *Urban Forestry & Urban Greening.* ***61***, p127091.

Tho, S.W., Yeung, Y.Y., Wei, R., Chan, K.W. and So, W.W.M. 2017. A systematic review of remote laboratory work in science education with the support of visualizing its structure through the HistCite and CiteSpace software. *International journal of science and mathematics education*. ***15***, pp.1217-1236.

Ugolini, F., Massetti, L., Calaza-Martínez, P., Cariñanos, P., Dobbs, C., Ostoić, S.K., Marin, A.M., Pearlmutter, D., Saaroni, H. and Šaulienė, I. 2020. Effects of the COVID-19 pandemic on the use and perceptions of urban green space: An international exploratory study. *Urban forestry & urban greening.* ***56***, p126888.

Ullah, H., Wan, W., Haidery, S.A., Khan, N.U., Ebrahimpour, Z. and Muzahid, A.A.M. 2020. Spatiotemporal patterns of visitors in urban green parks by mining social media big data based upon WHO reports. *IEEE Access*. ***8***, pp.39197-39211.

Wang, S. and Li, A. 2022. Impacts of COVID-19 Lockdown on Use and Perception of Urban Green Spaces and Demographic Group Differences. *Land.* ***11***(11), p2005.

We are London. 2023a. *London Green Spaces Commission.* [Online]. [Accessed 14 March]. Available from: https://www.london.gov.uk/programmes-and-strategies/environment-and-climate-change/parks-green-spaces-and-biodiversity/london-green-spaces-commission

We are London. 2023b. *London National Park City.* [Online]. [Accessed 16 Feburary]. Available from: https://www.london.gov.uk/programmes-strategies/environment-and-climate-change/parks-green-spaces-and-biodiversity

We are London. 2023c. *London Rewilding Taskforce.* [Online]. [Accessed 06 January]. Available from: https://www.london.gov.uk/programmes-strategies/environment-and-climate-change/parks-green-spaces-and-biodiversity/london-rewilding-taskforce

We are London. 2023d. *The London Urban Forest Partnership.* [Online]. [Accessed 23 Feburary]. Available from: https://www.london.gov.uk/programmes-and-strategies/environment-and-climate-change/parks-green-spaces-and-biodiversity/london-urban-forest-partnership

Webster, J.J. and Kit, C. 1992. Tokenization as the initial phase in NLP. In: *COLING 1992 volume 4: The 14th international conference on computational linguistics*.

Westgate, M.J. 2019. revtools: An R package to support article screening for evidence synthesis. *Research synthesis methods.* ***10***(4), pp.606-614.

Wojcik, S. and Hughes, A. 2019. Sizing up Twitter users. *PEW research center*. ***24***, pp.1-23.

Xiao, X.D., Dong, L., Yan, H., Yang, N. and Xiong, Y. 2018. The influence of the spatial characteristics of urban green space on the urban heat island effect in Suzhou Industrial Park. *Sustainable Cities and Society*. ***40***, pp.428-439.

Xie, J., Luo, S., Furuya, K. and Sun, D. 2020. Urban parks as green buffers during the COVID-19 pandemic. *Sustainability*. ***12***(17), p6751.

Zhang, S. and Zhou, W. 2018. Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. *Landscape and Urban Planning*. ***180***, pp.27-35.

Chapter 4 aims to synthesize the existing literatures related to the utilization of social media data to examine UGS. Meta-analysis and bibliometric approaches were used to identify the research themes. This chapter provides a comprehensive overview of the characteristics of different social media platforms used for UGS research. It also outlines and compares the approaches on data pre-processing, data quality enhancement, data analysis and modelling. The results show that Twitter was the most frequently used platform in UGS studies, followed by Flickr, Instagram, Weibo and OpenStreetMap. This chapter provides guidance for next steps of analysis where Twitter will be selected as data sources. Additionally, it should be noted that the datasets have inherent bias and issues of representativeness which may can't reflect the real usage of UGS use. A number of limitations of using VGI and social media datasets have been identified in this chapter. For example, social media datasets can't represent the whole UGS users as many of them do not use social networks when they visit UGS. Social networks cannot replace traditional methods such as surveys, questionnaires, and observations in the analysis of UGS usage. Nevertheless, they can serve as valuable supplementary tools for UGS research. Researchers should keep this in mind when conducting social media analyses.

Chapter 4 Using VGI and social media data to understand urban green space: A narrative literature review

Abstract

Volunteered Geographical Information (VGI) and social media can provide information about real-time perceptions, attitudes and behaviours in urban green space (UGS). This paper reviews the use of VGI and social media data in research examining UGS. The current state of the art is described through the analysis of 177 papers to 1) summarize the characteristics and usage of data from different platforms, 2) provide an overview of the research topics using such data sources, and 3) characterize the research approaches based on data pre-processing, data quality assessment and improvement, data analysis and modelling. A number of important limitations and priorities for future research are identified. The limitations include issues of data acquisition and representativeness, data quality, as well as differences across social media platforms in different study areas such as urban and rural areas. The research priorities include a focus on investigating factors related to physical activities in UGS areas, urban park use and accessibility, the use of data from multiple sources and, where appropriate, making more effective use of personal

information. In addition, analysis approaches can be extended to examine the network suggested by social media posts that are shared, re-posted or reacted and by being combined with textual, image and geographical data to extract more representative information for UGS analysis.

Keywords: Urban green space; Volunteered geographical information; Social media data

4.1 Introduction

Urban green space (UGS) refers to urban land covered by vegetation (Niemelä and Conservation, 1999). It is an essential component of urban environmental systems and plays a critical role in sustaining urban natural environments as well as the social systems that use these spaces (Chiesura and planning, 2004). An increasing number of studies have examined the various benefits of UGS to humans via the interactions between humans and UGS (Kabisch et al., 2015). These include studies of the ecosystem services of UGS (Wolch et al., 2014), the events and physical activities that occur in UGS areas (Cohen et al., 2013; Zhang, W. et al., 2015), the benefits to mental health (Chiesura and planning, 2004; Roe et al., 2013), and the accessibility of UGS (Comber et al., 2008; Fan et al., 2017). These studies have confirmed that city residents largely rely on parks and green spaces for physical, mental, and social well-being (Campbell et al., 2016; Grose and planning, 2009). UGS is therefore recognized as one of the key features supporting urban sustainability and enhancing the quality of life of urban residents (Kim et al., 2018).

Worldwide, the proportion of people living in urban will increase from 50 % in 2010 to nearly 70 % by 2050 (Nations, 2019). Hence the demand for UGS is rapidly increasing in the context of urbanization, especially in metropolitan areas. This means that the planning and management of UGS is critical in order to satisfy the needs of urban residents (Haaland et al., 2015), requiring urban planners to make public places more liveable and sustainable (Kashef, 2016). The interactions between humans and UGS, in particular, play a fundamental role in UGS planning (Roberts and Victoria, 2017). For example, researchers have investigated the interactions between UGS and humans and their impacts on visitors' perception, as well as the benefits to residents’ well-being (Larson et al., 2016; Tsai et al., 2018).

Social media are internet-based applications that enable people to communicate and share resources (Taylor et al., 2012). These technologies allow the public to voluntarily produce geographic information which can be considered as Volunteered Geographic Information

(VGI). The georeferenced data provided by social media can be considered as VGI and social media as VGI sources. Examples of this are geotagged Tweets from Twitter, geotagged photographs from Flickr and Instagram, etc. (See et al., 2017). VGI is defined as user-generated digital geographical data, including both text and multimedia (See et al., 2016a), enabled through the use of a range of technologies to create, assemble, and disseminate geographic information. VGI can be used to support understanding and exploration of the socio-economic and environmental conditions of a place through analysis of different resources such as geo-tagged Tweets and photos (Ghermandi and Sinclair, 2019; Mitchell et al., 2013), check-in data (Lyu, F. et al., 2019), OpenStreetMap (Hennig, 2017) and so on. The widespread use of popular social media technologies such as Twitter, Facebook, Instagram and Flickr where users post and share their views, opinions, feelings and emotions provides a resource to examine UGS visits, behaviours and use (Liu et al., 2017). For example, studies have investigated perceptions of green environment quality by analysing park visit frequency through Point-of-Interest (PoI) check-ins (Chen, W. et al., 2018; Cohen et al., 2010), mapping cultural service areas (Paracchini et al., 2014; Figueroa-Alfaro and Tang, 2017) and investigating tourism patterns (Shi et al., 2017; Wood et al., 2013). Such data potentially provide opportunities for researchers to quickly obtain a large amount of useful information for scientific research (Al-Kodmany, 2019).

This review covers the use of major social media data platforms in urban green space research and examines data collection methods, the advantages and disadvantages of different social media VGI and highlights a number of research gaps. It does by considering the following questions:

- What were the research aims and the research topics in studies that explored VGI in relation to urban green space?
- What types of social media websites or platforms were generally selected in these studies?
- What were the methods used in collecting data, processing data and analysing data?
- What were the potential challenges and problems not yet resolved and researched?

The reason for this review now, focussed in this way, is because previous reviews about the application of VGI data in urban studies have mainly focused on smart city planning and management (Hao et al., 2015; Nitoslawski et al., 2019), data acquisition and quality issues (Basiri et al., 2019), data mining approaches and techniques (Stock and Systems, 2018; See et al., 2016b), and human mobility in urban areas (Wang, A. et al., 2021), with a focus on the broader context of urban management and planning (Martí et al., 2019; Hecht and Stephens, 2014). However, in the domain of UGS and VGI data application, few reviews have summarised the application of VGI data in the context of UGS planning.

4.2 Materials and Methods

In this study a bibliometric analysis of published research was undertaken in order to support investigation of the characteristics of previous studies. Then the key research areas (themes) are examined as well as the methods used (including data pre-processing as well spatial, temporal and semantic analysis) before highlighting a number of data quality issues and key areas for methodological improvement.

4.2.1 Bibliometric literature search

A bibliometric analysis was undertaken using 4 steps (Figure 4.1) based on established guidelines for conducting a systematic literature review (Kabisch et al., 2015; Petticrew, 2001). The aims of this analysis were to first establish the degree to which UGS analyses are increasingly using different forms of social media to understand UGS user attitudes and preferences, and then to determine the how they were being used (for example, in support of specific objectives such as tourism or ecosystem services benefits). This review examined articles published between 1 January 2010 and 1 December 2019 in English. First, the search terms were determined based on a number of keywords, which can be classified into two groups. One group is composed of words related to “urban green space” (Konijnendijk et al., 2013). The other group referred to “social media” or “volunteered geographic information”. The search terms are described in Table 1 and relate to two themes: topic (e.g., Urban green space) and data sources (e.g., Social media data). These were adapted for each database to ensure appropriate syntax. The search terms in this review were selected based on the authors’ knowledge and previous studies examining methods to conduct a systematic review (Kabisch et al., 2015; Konijnendijk et al., 2013). The search engines Web of Science, Scopus, IEEE Xplore and Google search were used as they cover a range of discipline areas, with the aim of capturing all relevant literature in this domain. The search terms were used to find matches in ‘title, abstract, and keywords’ for Scopus and ‘Topic’ for Web of Science. A final step was to synthesize the data and to extract relevant information (Table 4.1).

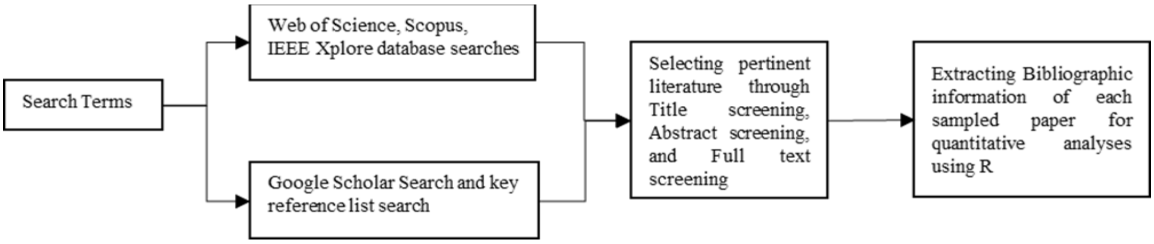

Figure 4. 1 Outline of search strategy.

Table 4. 1 Summary of literature search terms and their use in the search query.

UGS			Data	
Urban	AND	Green space/ Greenspace	AND	Social media
	OR	Green infrastructure	OR	Volunteer geographic information/ VGI
	OR	Park	OR	Crowd sourced geographic information
	OR	Recreation area	OR	Crowd source/ Crowdsource/ Crowdsourcing
	OR	Garden	OR	Citizen science/ Citizen contributed science
	OR	Playing field	OR	Flikr/ Twitter/ Weibo/ Foursquare/ Instagram
			OR	WeChat/ WhatsApp/ Facebook

The query for paper selection by key words was (TITLE-ABS-KEY (“urban”) AND (“green space” OR “greenspace” OR “green infrastructure” OR “park” OR “recreation area” OR “Garden” OR” playing field ”) AND TITLE-ABS-KEY(“Social media” OR “Volunteer geographic information” OR “VGI” OR “crowd source” OR “citizen science” OR “Flickr” OR “Twitter” OR “Weibo” OR “Foursquare” OR “Instagram”)) AND PUBYEAR > 2009 AND PUBYEAR < 2020.

After entering the search terms in each database, the papers were screened and some of them were excluded according to the content of the title or abstract. This was to remove articles that were not related or only marginally related to the objectives of the review. For example, articles examining the use of social media data without urban green space visitation were excluded. In addition, the literature considered in this review was restricted to publications in international, peer-reviewed journals articles and conference proceedings. The remaining papers were further screened for the exclusion criteria in Table 4.2. In addition, papers that did not appeare in the initial search results but were referenced within the identified papers were included if they related to the review aims (Konijnendijk et al., 2013). Finally, the bibliographic information of each paper was extracted for quantitative analyses, including trend detection, text and topic mining, and citation analysis. A final manual check of the papers was undertaken to ensure a minimum equal evaluation of topics and themes and as little assessment bias as possible (Kabisch et al., 2015).

Table 4. 2 Literature screening exclusion criteria

No.	Exclusion Criteria	Examples
1	Studies not written in English	(Sheng et al., 2018)
2	Studies concerned with Intelligent parking systems	(Blancaflor et al., 2019; Sadhukhan, 2017)
3	Studies concerned with apps information monitoring	(Sprake et al., 2014)
4	Surveillance of health by using web data	(Jung et al., 2019)
5	Studies not related to green pace	(Ben-Harush et al., 2012)
6	Studies selected the industrial park as study areas	(Zhe et al., 2016)
7	Studies concerned with disaster detection	(Weiler et al., 2016)
9	Studies concerned with emergency situation	(Barros et al., 2015)

4.2.2 Data processing

Bibliometric methods allow researchers to examine, organise, and analyse huge amounts of information to find hidden patterns (Broadus, 1987). Many bibliometric tools use information about authors, affiliations and citations to identify and explore patterns in conceptual maps, co-citation analyses, cluster and factor analyses (Daim et al., 2006). The “bibliometrix” R package (http://www.bibliometrix.org) (Aria and Cuccurullo, 2017) , an open-source tool for scientometric and bibliometric research, was used for quantitative analysis and for topic mining of the bibliographic data in R 4.0.3 (https://cran.r-project.org/bin/windows/base/old/4.0.3/). This package includes all major bibliometric analysis methods, with rapid analysis speeds and the use of data matrices for co-citation, coupling, collaborative analysis, and co-word analysis. In this study, bibliometric analysis methods were used to extract information such as annual publication rates, corresponding authors’ country, country scientific production (i.e., countries of author affiliations), conceptual structure maps and cumulate occurrence of keywords. A co-word analysis was undertaken using the *bibliometrix* R-package to undertake multiple correspondence analysis (MCA) to examine the conceptual structure of the domain (Aria and Cuccurullo, 2017). MCA is an exploratory multivariate technique for the graphical and numerical analysis of multivariate categorical data (Greenacre and Blasius, 2006). In the co-word analysis undertaken here, the words are plotted on a two-dimensional map.

4.3 Results

4.3.1 Main characteristics of included studies

The total number of articles identified from databases search was 802. Screening the papers based on the exclusion criteria (Table 4.2) resulted in 219 articles, and 177 articles remained after reading the full texts and analysing each article individually. Details of the volume of generated papers and the originating countries of their authors are shown in Figure 4.2.

The number of documents published per year in Figure 4.2(a) indicates that the number of papers has increased continuously since 2010, entering a more rapid growth phase in 2014. This demonstrates that scholars have increasingly studied UGS by using social media data in recent years, or that social media has become more popular. Additionally, Wi-Fi infrastructure may have been improved, with local managers providing Wi-Fi within UGS areas, making it easier to obtain data for research. The increasing number of papers indicates the increasing significance of UGS.

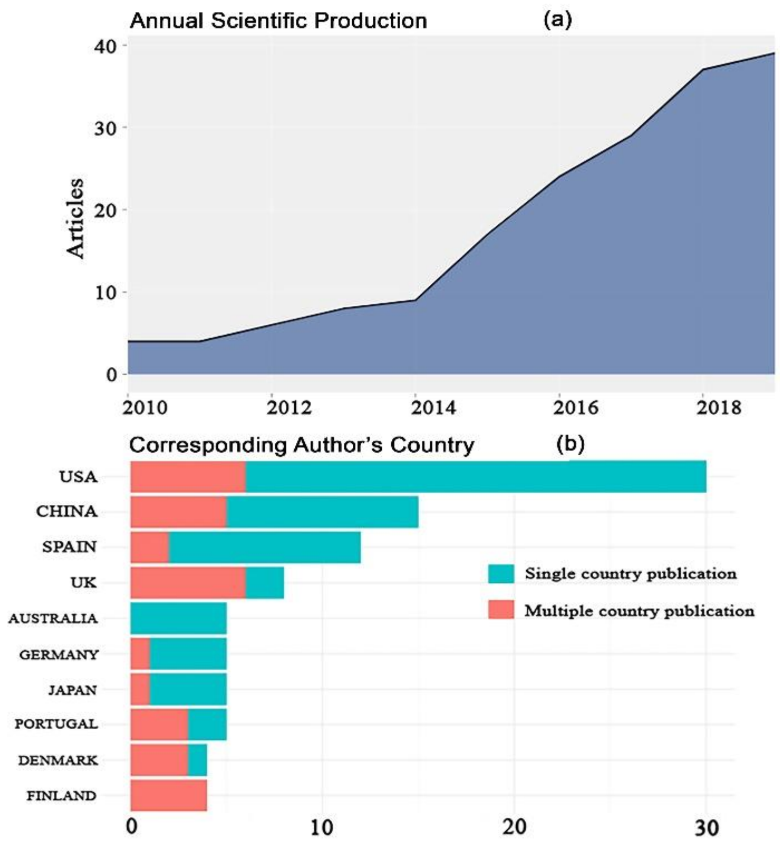

Figure 4. 2 Bibliometric analyses of UGS and social media research, (a) Annual Scientific Production, (b) Corresponding Author's Country

Figure 4.2(b) shows the number of corresponding authors' country and the degree of international collaboration is through the proportions that are associated with single country publications (SCP) and with multiple country publications (MCP). The United States has the largest total number of publications, followed by China, Spain, the United

Kingdom and Australia. Additionally, Finland and the UK have the greatest proportion of MCP, followed by Portugal and Denmark, suggesting that these counties have higher levels of international collaboration than others.

Figure 4.3(a) shows the clustering of the topics identified from the author-specified keywords. This was generated by a Multiple Correspondence Analysis (MCA) of the topics, MCA allows researchers to study the association between two or more nominal categorical data (Abdi et al., 2007), this approach can be used to understand the fields of selected papers from a low-dimensional perspective. Specifically, the nearer the positions of the points the closer the concepts are that they indicate.

Cluster 1 includes words related to urban green space and environment. This shows that the focus of papers was mainly centred on urban areas and green space. In addition, the words related to geographic information systems (GIS) and sentiment analysis were identified as common research methods and analysis tools in this cluster, indicating that these approaches made great contributions in the field. Cluster 2 includes themes related to ecosystem services, tourism, urban planning and behaviour research. Additionally, Twitter, Instagram, Flickr, and OpenStreetMap were also included in this cluster indicating that these social media platforms were selected as the main data sources in this field. In this case Figure 4.2(a) shows that Twitter data are closer to ecosystem services and travel behaviour in this map. This shows that this Twitter was a popular data source in this area of research; Flickr and OpenStreetMap are closer to human mobility and tourism which shows that these sources were more popular in these areas of research in relation to UGS. Social media analysis, urban parks and green space were observed in Cluster 3, indicating that social media can be used as new resources in the analysis of urban parks. Ecosystem system services were found in Cluster 4, indicating the focus on urban parks as the main source of natural landscapes to provide important ecosystem services for urban residents.

This map helps researchers to understand existing research themes in the analysis of UGS by using VGI and social media data, and which data platforms were more popular in which research themes. Figure 4.2(b) shows the cumulate occurrence of the keywords in all 177 articles. The highest number of keywords are social media, followed by Twitter, big data, cultural ecosystem services, Flickr and tourism which indicates that these areas may be important research topics in relation to the studies of VGI data and UGS.

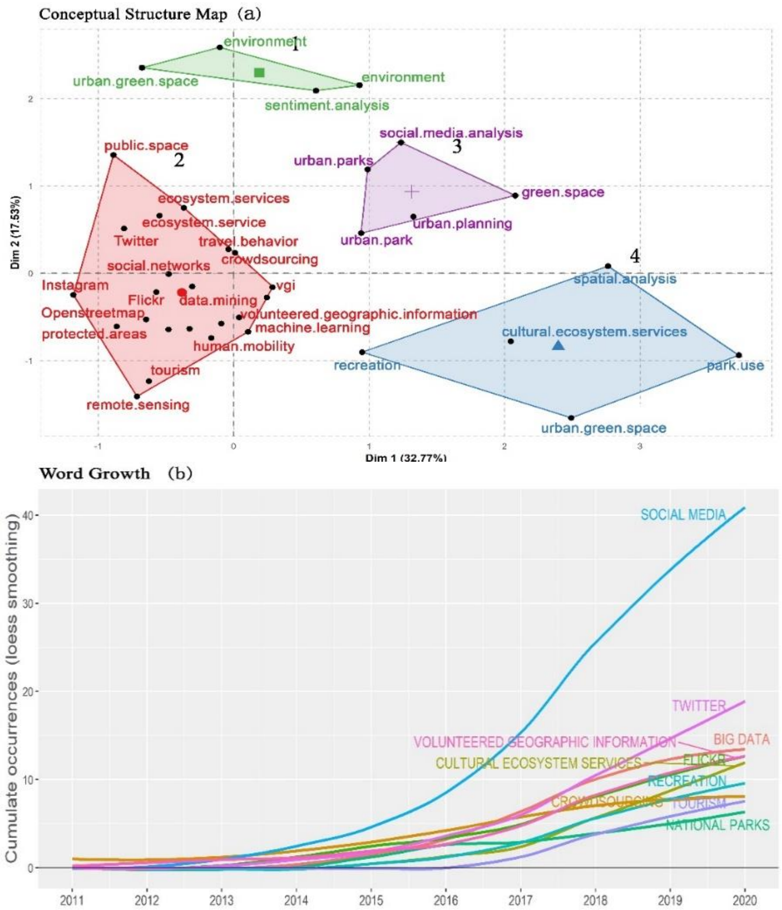

Figure 4. 3 (a) Conceptual Structure Map and (b) Cumulate Occurrence of the Keywords in the UGS and social media literature.

Overall, Figure 4.3 shows that the keywords and abstract terms in the selected articles mainly concentrated on ecosystem services, human behaviour, urban planning and tourism by using various social media data related to urban green space and urban parks. This is not surprise given the search terms of this review, however, the words about physical activities in UGS areas, factors related to urban park use and accessibility of urban green space did not appear in these clusters. This is a potential area for future research as discussed in Section 4.4.

4.3.2 Data Sources in Relation to UGS Analysis

The data sources used in UGS research were summarized from all reviewed articles by scanning the section ***'data resources'*** in each paper. In addition, data acquisition approaches including data collection websites, software and data platform availability were also recorded and summarised in Table 4.3.

The advantages and disadvantages of top 5 popular data platforms are highlighted below. Additionally, in order to understand why certain types of data sources were selected by authors when they studied different themes, the 'introduction' section were summarised to find more detailed descriptions of data sources from the authors' perspective. Figure 4.4 shows the frequency of different data platforms used in the 177 articles over different years. It shows that, overall, social media data including Twitter, Flickr, Instagram and Weibo are becoming increasingly popular in studies relating to UGS, and the data platforms of Twitter and Flickr are the most frequently used as data sources. Twitter is a very popular microblogging service established in 2006. Twitter users 'tweet' about their individual opinions and feelings within a 140-character (now 280) limit (Kwak et al., 2010). Flickr was established in 2004 and is the most popular online photo management and sharing application in the world (Tenkanen et al., 2017). Instagram, established in 2010, is used to share self- and user-generated content (Di Minin et al., 2015). Weibo is a large social network website in China. Weibo users can get up-to-date status information, provide status updates, share views, and communicate with others (Gu et al., 2016).

Twitter was selected as the data platform by 71 articles, accounting for 39 % of all papers, which indicates that this data platform was the most popular in the research works related to UGS, followed by Flickr (40), Instagram (10), Weibo (9) and OpenStreetMap (9). Other, less well known, VGI platforms included MapMyFitness (Norman and Pickering, 2019), Tencent (Chen, Y. et al., 2018), Tuniu (Dai et al., 2019), Wikiloc (Norman and Pickering, 2019) and Wikipedia.

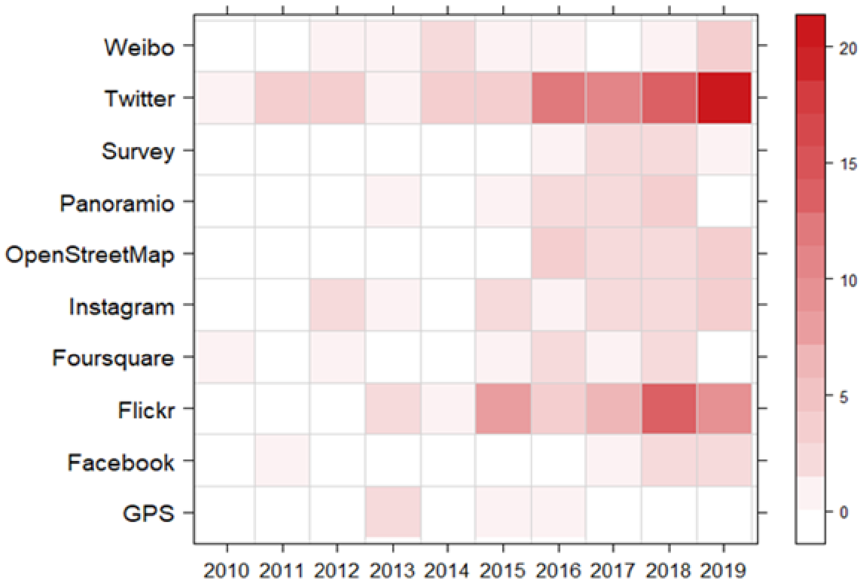

Figure 4. 4 The frequency of occurrence of different data platforms found in the UGS and social media literature.

The social media platforms identified in this review were classified into three categories according to (Senaratne et al., 2017): text-based social media such as Twitter, Weibo;

image-based social media such as Flickr, Instagram; map-based social media such as MapMyFitness, Baidu and Google Maps.

Text-based social media data have been mainly used to investigate park visitation (Hamstead et al., 2018; Li, F. et al., 2020; Sim et al., 2019) , factors affecting park use (Lyu, F. et al., 2019; Norman and Pickering, 2019; Wang, Z. et al., 2018), physical activity and events in park areas (Roberts, H. et al., 2017; Santos et al., 2016; Song, Yimeng et al., 2018), and the emotional response of visitors in park areas (Plunz et al., 2019; Roberts, H. et al., 2019; Sim et al., 2019). The reasons why text-based data were popular in these research topics can be summarised as follows:

1. The data are easy to collect using methods such as public application programming interfaces (APIs), such as Twitter streaming APIs and Weibo APIs (Table 3), and can be downloaded at as frequent a time interval as necessary (Martí et al., 2019; Roberts and Victoria, 2017).
2. There are large numbers of users on these networks generating huge amounts of information (Lyu, F. et al., 2019; Norman and Pickering, 2019; Wang, Z. et al., 2018).
3. The geo-referenced text-based social media data allow researchers to investigate park visitation patterns from a spatial perspective, while achieving greater longitudinal depth (Hamstead et al., 2018; Li, F. et al., 2020; Sim et al., 2019).
4. The time of text-based data (i.e. Tweets) creation can support investigations into the temporal patterns of park visitation (Anna et al., 2018).
5. The content of text data can be used in semantic analysis including sentiment analysis and emotion detection, which can help scholars understand the public perceptions and interest in urban green space areas (Plunz et al., 2019; Roberts, H. et al., 2019).

Image-based social media data (such as Instagram and Flickr) were mainly used in research examining cultural ecosystem services (Guerrero et al., 2016; Sinclair et al., 2018; Gliozzo et al., 2016), park visitation (Hamstead et al., 2018), investigations of factors affecting park use (Dallimer et al., 2014), and physical activities (Heikinheimo et al., 2017) for the following reasons:

1. The photographs that social media users post may reflect their interests, aesthetic values, sentimental attachment and emotional state at a particular time and place (Guerrero et al., 2016; Sinclair et al., 2018).

2. Georeferenced photos allow researchers to detect spatial patterns of park visitation and user behaviour (Hamstead et al., 2018). User profiles help researchers identify where visitors live and their home location (Heikinheimo et al., 2017).
3. Shared pictures provide access to real-time information, allowing researchers to generate temporal patterns of urban green space use (Heikinheimo et al., 2017). Additionally, images are taken and posted throughout the year, enabling longitudinal analysis.
4. These platforms provide free, up-to-date, and high spatial and temporal resolution information sources (Wood et al., 2013; Oteros-Rozas et al., 2018).

There are some limitations associated with social media data that the papers discuss. These include low coverage, data quality, uncertainties, and problems with representativeness and reliability (Johnson, Michelle L et al., 2019; Martí et al., 2019; Plunz et al., 2019). In addition, existing analysis methods for information extraction need to be improved (Steiger et al., 2015). These limitations should not be ignored by researchers. For example, in research examining spatiotemporal park visit patterns using semantic information from Twitter, researchers are often faced with data-specific uncertainties, including identifying the locational information of visitors, which affects the nature of the information extracted (Steiger et al., 2015). In addition, Twitter users only represent a small proportion of the real park visitor population; users are usually younger, wealthier and have more educational qualifications as compared to the general population (Salas-Olmedo and Rojas Quezada, 2017; Blank and Lutz, 2017). This has been an ongoing concern for many of the papers reviewed. Thus, the use of geo-social media data such as georeferenced photo and geo-Tweets should not replace consideration of traditional methods when it comes to the assessment of urban park visitation. However, georeferenced Tweets or photos still have the potential to produce valuable and useful knowledge, particularly in metropolitan areas with a high density of social media users (Plunz et al., 2019).

Research should always consider the validity of social media data before analysing it in order to determine the extent to which the results robustly support management and planning. For example, Lenormand et al. (2014) validated the use of twitter data in Barcelona and Madrid by comparing different data sources including the census, and cell phone data. The results showed that the three data sources provided comparable information for studies of urban human mobility.

Incomplete information such as uncertainty over timestamps and locations can lead to biases in UGS research. For example, the timestamps in Flickr photos can be the time the photo was taken or when it was uploaded, and geo-tagged locations can also be changed by users (Dunkel, 2015). Different types of spatiotemporal analysis (such as seasonal or weekend/weekday comparison) could be affected by the uncertainty of these data (Lee, J.Y. and Tsou, 2018a).

Several researchers combined various datasets in order to overcome the limitations of using a single platform. For instance, some studies (Hamstead et al., 2018; Pickering et al., 2020) used geolocated Twitter and Flickr data to explore park visitors' views and factors affecting urban park visitation. Lyu, F. and Zhang (2019) compared VGI data from Weibo and Baidu to understand the factors affecting urban park use in China. In other research (Levin et al., 2017) two VGI data sources were used, Flickr, and OpenStreetMap (OSM), and then combined with remote sensing data to assess the visitation and perceived importance of UGS. The combination and comparison of different kinds of social media datasets in studies related to UGS allow researchers to generate more comprehensive conclusions about the factors associated with park visitation, UGS physical qualities and events.

However, not all social media data were found to be suitable for the local context. For example, Baidu Map data were found to have more accurate location check-in information than Weibo data (Lyu, F. et al., 2019) in assessing urban parks in Wuhan, but other research was unable to establish whether Baidu Map was better in Beijing (Gu et al., 2016) and Shenzhen (Zhang, S. and Zhou, 2018) as only Weibo data were used to assess the UGS use in these cities. This indicates a potential for bias if studies rely on a single data platform, suggesting the need to consider using a range of social media data from different platforms to enhance the reliability of the research, in other words, future works could focus on the combination of different types of social media data such as text-based data (e.g., Twitter and Weibo) and map-based data (e.g., Baidu map and Open street map) in assess urban park use. Table 4.3 summarises the characteristics of the most popular data platforms in relation to UGS studies.

Table 4. 3 The social media platforms used in UGS analysis.

Data	Platforms				
	Twitter	**Flickr**	**Instagram**	**Weibo**	**Open Street Map**
Data collection website	https://developer.twitter.com (Kovacs-Györi et al., 2018; Johnson, M. L. et al., 2019)	www.flickr.com/api (Dunkel, 2015)	(Vieira et al., 2018) www.instagram.com/developer	(Lyu, F. et al., 2019; Zhang, S. and Zhou, 2018) https://open.weibo.com/development/datacenter	(Dunkel, 2015; Hennig, 2017) http://www.openstreetmap.org
Data type	Text-based VGI	Image-based VGI	Image-based VGI	Text-based VGI	Map-based VGI
Collection methods	Twitter's search API, streaming API, Rest API, research API ; and Twitter's Firehose (Martí et al., 2019; Roberts and Victoria, 2017). Python wrapper. Tweepy (https://www.tweepy.org/) python library (Vieira et al., 2018). Tweet R package (Johnson, Michelle L et al., 2019); TAGS Version 6.0 (Park, S.B. et al., 2018).	Search on the Flickr developer site (Sinclair et al., 2020). The Flickr API (https://www.flickr.com/services/api/) (Donaire et al., 2014). Using standard Hypertext Transfer Protocol (HTTP) methods to retrieve and manipulate data (Clemente et al., 2019).	Using a custom-made tool written for the Python programming language. Using the API of Instagram by (https://www.instagram.com/developer/) (Song, Yang and Zhang, 2020).	The location service dynamic reading interface of the Sina Weibo open platform (https://api.weibo.com/2/place/nearby/photos.json) as the data source (Li, F. et al., 2020). Data collection was facilitated by Weibo application program interfaces (APIs). Through the "to obtain nearby locations" API (Zhang, S. and Zhou, 2018).	QuickOSM (https://plugins.qgis.org/plugins/QuickOSM/) Python module for QGIS was used for collecting data from OSM. The OSM data are freely downloadable from geofabrik website (http://download.geofabrik.de/asia/nepal.html)

Geography	With geo-coordinates	Geotagged posts (including pictures, titles and text)	Geotagged posts (including pictures, titles and text)	With geo-coordinates	Active mapper communities in many locations
Content	User ID, Tweet text, timestamp, geotags and volunteered geolocations	Photo ID and owner ID, title, description, geo-tags, time when a photo was taken and upload time	Photo photo ID, photo title, description, tags, upload time, time when a photo was taken, location, and owner ID	Text and metadata in Weibo with geolocation, and user ID, photographs location, device type .	Open Street Map encodes data in different formats such as points, polylines, and polygons.
Advantages	Free, high spatio-temporal resolution; Lots of Twitter users post messages at various locations, including school, home, restaurants, and touristic sites. Real-time information that potentially reaches a huge audience (Vieira et al., 2018).	Free, spatially and temporally explicit, visitation hotspots. Allows for image analysis and content, User characteristic analysis, actural visitation (Levin et al., 2017).	Online mobile application focused on sharing photographs and providing a platform for social networking (Guerrero et al., 2016).	Weibo users (462 million according to the 2018 Weibo User Development Report) can upload their real-time locations and share their preferences and activities on the Internet. Data from Weibo check-ins can well represent the preferences and activities of people in urban areas (Dunkel, 2015).	A free and up to date map of the world accessible and obtainable for everyone; millions of registered contributors; provides free and flexible contribution mechanisms for data (useful for map provision, routing, planning, geo-visualization, point of interest search). Insight into people's individual perspectives and perceptions (Dunkel, 2015).
Disadvantages	Twitter data have some biases, such as age, gender, and education. Not all the collected Tweets are usable since	Unclear meaning, Confounding factors. Potential sampling and selection biases, noise in the data (Sinclair et al., 2020).	Locational accuracy. The issues of anonymity and privacy arrise. No information was gathered concerning the users, no socio-economic data exist, which makes it difficult to assess	Sina Weibo check-in data has some biases, such as age, gender, a temporal change and social class bias. Weibo users are mainly composed of people between 18 -40 years old, accounting for 89% of the total number of users.	Though OSM has no strict quality control mechanism, studies have indicated that data obtained from OSM are good enough and comparable to authoritative data to some extent (Levin et al., 2017).

	some of them may have been generated by spammer s (Devkota et al., 2019).		representabilit y in detail (Guerrero et al., 2016).		

4.3.3 Research Themes in Relation to UGS Analysis

A set of phrases were manually extracted from keywords, titles and abstracts and then ranked based on their frequency. The first 10 of these were then used to code each paper based on the occurrence or not, as summarised in Figure 4.5.

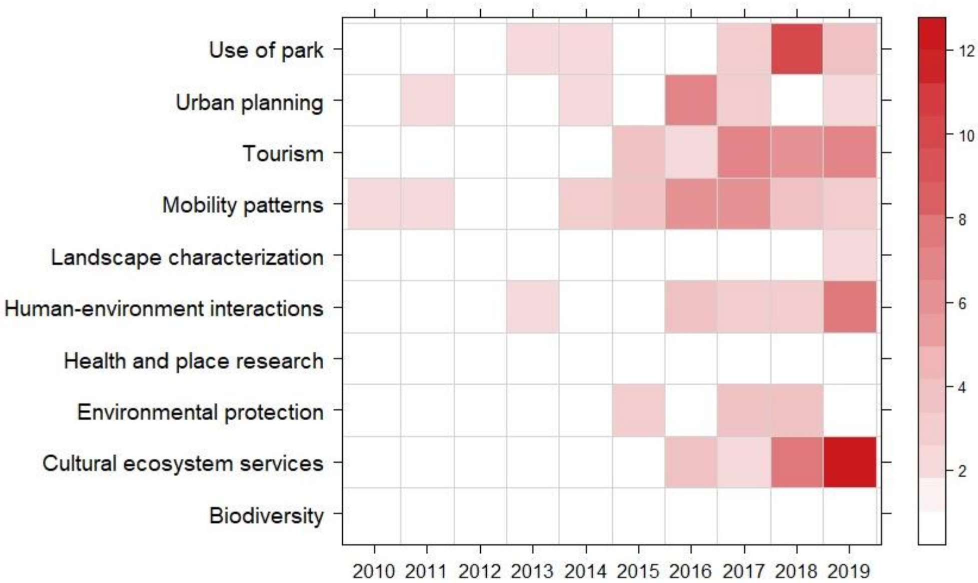

Figure 4. 5 Research topics covered by research using social media and UGS.

The themes of *cultural ecosystem services* and *urban park use* are getting increasing attention from scholars. In detail, 44 papers researched the topic of *culture ecosystem services* provided by UGS, accounting for about 24 % of all papers, making it the most popular topic. This was followed by the theme of *human-environment interactions* (36 papers), with the third most popular topic being *urban tourism* (34 papers). A total of 29 papers considered the theme of *urban park use*, 17 papers studied *environmental protection*, 7 papers focused on *human mobility patterns*, and 5 papers researched biodiversity and *landscape characterization*. In relation to cultural ecosystem services in UGS, various data platforms such as Flickr, Instagram, Twitter, Panoramio (Gliozzo et al., 2016) and Wikilo (Vaz et al., 2019) have been utilized. Amongst these platforms, Flickr was the most commonly used (Clemente et al., 2019; Gosal et al., 2019), whilst research examining the theme of park use, have most commonly used Twitter and Weibo (Lyu, F. et al., 2019; Anna et al., 2018).

4.3.4 Methods used in data analysis

Various data and methods were used in the reviewed articles that relate to UGS studies. These have been divided into three aspects: data pre-processing, spatial and temporal analysis and semantic analysis.

4.3.4.1 Methods used in pre-processing

A key issue is that social media data used by researchers for UGS analysis should be published by human users such as urban dwellers or tourists instead of bots or spammers (Sim et al., 2019). Some have found that advertisers (Devkota et al., 2019) and automated accounts (Plunz et al., 2019) can post a huge number of messages daily or hourly, and even create geolocated messages that are posted in locations a long way from their purported location (> 500 km). Such data should be identified as non-human (Devkota et al., 2019) and removed.

Georeferenced social media data can have high spatial resolution, allowing researchers to observe spatial patterns in the research areas being examined (Hao et al., 2015). Therefore, a second step is often to exclude data lacking relatively high precision location (Park, S.B. et al., 2018) and to exclude geo-located data outside of the study area (Shi et al., 2017; Anna et al., 2018). Gazetteers can also be used to geocode users' locations to latitude/longitude coordinates (Hamstead et al., 2018) and thus allow invalid data to be removed. Donaire et al., (2014) suggested that researchers should take in to account that not all of the users would like to share their locations when posting messages, thus the data used for analysing UGS is a subset of the entire dataset and the users who include spatial information in their messages are not wholly representative of the entire user base.

More broadly, it is estimated that Twitter's Streaming API only released less than 1% of all world-wide generated Tweets (Boyd et al., 2012) and Pew Research Center reported that Twitter users only accounted for about 24% of online adults in 2016 (Greenwood et al., 2016), with users more likely to be younger and wealthier than the general population. However, the total number of social media data are very large, so researchers can still obtain great volumes of georeferenced data and attempt to balance these potential sources of bias (Li, L. et al., 2013).

Individual social media users have different activity characteristics. Individual Twitter user data, for example, typically has a very long tail; a large proportion of Tweets are produced by only few hundred (Li, L. et al., 2013). In order to remove a similar bias in Flickr data, Pickering, et al. (Pickering et al., 2020) suggested capping 10 images per person. In addition to long tail problems, different research aims required specific data sets. For example, Maeda, et al. Maeda et al. (2018) extracted the tourists destinations and generated visitation pattern by using Twitter data and split users into groups of residents and tourists. The sentiment score of geo-Tweets related to UGS in New York were similarly divided into park user and non-park users (Plunz et al., 2019).

4.3.4.2 Methods used in spatial data analysis

Kernel density estimation (KDE) has been frequently used to quantify the spatial distribution of park visitors across a study area (Lee, J.Y. and Tsou, 2018a; Ullah et al., 2020). KDE is a statistical approach used to estimate a smooth and continuous distribution from a limited set of observed points (Maia et al., 2008). It was used to construct density surfaces from point of interest check-ins (Rizwan and Wan, 2018) and Lee and Tsou (2018a) used KDE to analyse geo-tagged Flickr photos, identifying hotspots of tourist behaviours. Han et al. (2015) used KDE to explore spatial activity using Twitter, showing that KDE can be used to study the dynamic evolution of georeferenced data across both time and space. Fundamentally KDE analyses point to the varying distribution of park visitors over fine temporal and spatial scales.

One key variable in the KDE method is the specification of the kernel radius. Adopting different sizes of radius will generate surfaces with different degrees of spatial aggregation or smoothing. Thus, it is important to select a suitable kernel radius when assess the density of park visitors in urban green space areas. For example, Lee and Tsou (2018a) examined two spatial scales of KDE for tourist activity analysis. First, 50km was selected to identify the general regions in the Grand Canyon area and second a 200m kernel, to identify smaller hotspots along roads and trails (with a higher spatial resolution).

In addition to the KDE method, K-means, Mean-Shift and DBSCAN algorithm are commonly used to assess the spatial patterns of tourists (Ghermandi and Sinclair, 2019; Hasnat and Hasan, 2018). In order to measure spatial dependence, Moran's I has been used to measure autocorrelation, allowing researchers to explore the degree to which one object value is similar to other nearby object values (Shi et al., 2017).

4.3.4.3 Methods used in temporal analysis

In terms of temporal analysis, the timestamp of social media contributions have been divided into different temporal categories to trace changes in the number of visitors across the study area (Roberts, H. et al., 2017; Tenkanen et al., 2017). Such studies analysed the temporal patterns from daily to hourly distributions, weekly patterns to distinguish which parks are more popular at the weekends, and seasonal patterns which reflect the effect of climatic factors. Schirpke et al. (2018) and Wakamiya et al. (2011) used the same methods to analyse the temporal patterns of outdoor recreation in the European Alps and their surroundings. Spearman correlation coefficient were used to analyse temporal patterns across data derived from different social media data platforms (Tenkanen et al., 2017).

4.3.4.4 Methods used in semantic analysis

Text mining is very important in social media analysis because it provides the basis for various research objectives including sentiment analysis, emotion detection and topic modelling. Before analysing text data, various preparatory processes must be applied such as tokenization (splitting a sentence into a series of independent words) stemming (removing tenses, capturing singular and plural form of words) and structuring the sentence or text (e.g. “gives”, “gave”, or “given” are all related to “give”). In addition, some users (and researchers) are not fluent in English and effective translation tools such as Google Translate and iTranslate are needed for addressing problems of language confusion when mining text from Tweets (Al-Kodmany, 2019).

Sentiment analysis aims to extract opinions towards a topic or events generally from textual data sources and can be applied after text mining to assess the users' emotion and satisfaction in UGS or urban parks. The approach is to compare the stemmed terms to a sentiment lexicon of some kind. For example, SentiStrength V2.2, an opinion mining tool based on a lexicon of words including positive or negative emotion and scores (e.g. happy: 2, bad: −2), was used to investigate sentiments of texts especially in short texts such as Tweets (Park, S.B. et al., 2018; Gonçalves et al., 2013; Martínez-Cámara et al., 2014). This approach has been proven to achieve a high accuracy in sentiment analysis (Antonakaki et al., 2021). In addition, word polarity analysis can help researchers calculate the probability of the appearance of the word in a given text (Lyu, K. and Kim, 2016), which is a good way for extracting opinions generally from textual data sources (Shi et al., 2017). In the context of UGS, Chapman et al. (2018) used three different approaches to investigate the sentiment of Tweets in relation to UGS. The methods were 1) Manual Annotation referred to a random sample of 1,000 Tweets were annotated by five annotators, this method provides a robust test set which can be used to compared with other methods; 2) Fully Automated Annotation referred to an Affective Norms for English Words (ANEW) resource (Warriner et al., 2013) was used as the basis for emotion annotation instead of manual annotation, and 3) Graph Based Semi-Supervised Learning Annotation where researchers first select a sample of manually annotated Tweets and then used them to train a graph based semi-supervised learning algorithm, which was finally used to annotate the remaining Tweets.

A limitation of the previous study is that each message is assigned one kind of emotion. To overcome this, Park, et al. Park, W. et al. (2018) classified the sentiment scores of Tweets into three categories: positive (scores 1 to 4); neutral (scores of 0); and negative (scores -1 to -4). Other research has used a similar scoring system which allows a larger number of tweets to be classified as ‘neutral’, for example with scores of −2 to 2 (Anna et al., 2018).

4.3.5 Data Quality Issues and Improvement

VGI has proven very successful as a means of obtaining georeferenced information about social media users as frequent a time interval as necessary (Devkota et al., 2019). In addition, these kinds of data can often be freely downloaded via APIs (Table 3) enabling researchers to analyse UGS use at a very low cost. However, VGI has some obvious limitations.

In order to assess the extent to which scholars can rely on Twitter, some researchers have investigated how much information is spam (Donahue et al., 2018). They found that the high volumes of spam made it difficult to generate useful and meaningful information. Hence in order to improve the quality of this type of text-based VGI data, it is important to pre-process the social media data before further analysis (as described in section 4.3.4.1.) to filter out spam (Sim et al., 2019), identifying the data within study areas (Hao et al., 2015), restriction of the number of Tweets from prolific users (Pickering et al., 2020), and the identification of groups of users, such as urban residents and tourists (Plunz et al., 2019).

For image-based VGI data, different types of smart phones and GPS devices may cause various accuracy errors. For example, georeferenced social media data collected from the web application Wikiloc may leads to uncertainty in data quality (Figueroa-Alfaro and Tang, 2017). So, although the photographer may be usually relatively close to the subject of the photo, especially in a UGS and `likely within the geolocation error margin, the geolocations of photographs have found to be influenced by users who prefer to geo-tag the photo with location of the photo subject (e.g. a famous building) rather than the photographer's position (Donaire et al., 2014). Similarly, users who are not familiar with the function of adding geo-locations for photos or lack of enough spatial knowledge sometimes incorrectly geotag their photographs. Study results can also be biased by users post many photos from the same location. This problem should not be ignored and some studies have taken steps to remove this bias (Figueroa-Alfaro and Tang, 2017).

In order to improve the locational quality of image-based VGI data, a series of 200m sided hexagons were set up in which the pictures were aggregated ('binned') and the number of users and photographs were calculated (Lee, Y. et al., 2016). Under this method, the modifiable-areal-unit-problem (MAUP) effect can be minimized (Lee, Y. et al., 2016). Similar studies have also applied this approach to analyse data at the user level (García-Palomares et al., 2015). The number of photos were capped at 10 images per person in order to remove the bias from a few visitors who post lots of images (Pickering et al., 2020). Researchers may also want to consider manual image classification when analysing the content of images. For example, the content of an image was initially interpreted by two people, then a third person cross checked the final interpretation and any discrepancies. (Pickering et al., 2020).

In terms of map-based VGI data, the lack of common standards across platforms and access to accounts for providing and uploading data may further influence the accuracy of data or user attributes (Li, J. et al., 2019). In addition to accuracy, data completeness also exerts obvious influence on providing reliable services (Sun et al., 2017). GPS tracking applications such as Strava, MapMyFitness, and Wikiloc can provide metadata that contain information about physical activities that park users participate in. This allows researchers to detect the mobile patterns of visitors in park areas (Vich et al., 2019). However, GPS tracking data may contain gender bias as men have been found to be more likely to record their activities than women on some applications (Oksanen et al., 2015).

To improve data quality, OSM and authoritative data should be combined to develop an integrated open data source (Hennig, 2017). Levin et al. (2017) presented a semantic analysis to improve data classification, enhancing data quality to overcome cross-cultural and multi-language problems. Some studies have focused on procedures to enhance quality during the acquisition and compilation steps via crowd-sourcing, social, and geographic approaches (Goodchild and Li, 2012).

The evaluation of data validity, accuracy, representativeness, and uncertainty is essential when such data are used to analyse UGS visitation patterns and user behaviours (Levin et al., 2017; Zhang, W. et al., 2015). In order to evaluate and improve the representativeness of different social media data sources, Blank and Lutz Blank and Lutz (2017) evaluated six platforms including Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram in Great Britain. Their results showed that Twitter users tend to be younger and more highly educated. In terms of image-based data, the population representativeness of Flickr was assessed and users represent a specific subsample of visitors to any site with specific motivations to take and share images, hence Flickr represents only a fraction of the actual visitors (Lee, J.Y. and Tsou, 2018b). Twitter data have been widely used in UGS research and some studies (Donahue et al., 2018; Plunz et al., 2019) have suggested that geolocated Twitter data in metropolitan cities can be used as an alternative source of information able to adequately characterize commercial, leisure, and residential areas for urban planners, especially in combination with their geographic location marking and time stamping functions including real-time.

4.4 Discussion

VGI data have been widely used in the research field of UGS analysis. The growing popularity of social networks and social media services have attracted researchers from various disciplines, and this new form of geographic data has been used in a variety of applications. This review has identified the ten most frequent topics from the reviewed articles, with the most common topic related to cultural ecosystem services. This study manually extracted research themes across all selected articles which may be influenced by authors' personal views and knowledge, which was a limitation of this review. Various social media platforms have been used as data resources for different objectives in the reviewed articles. The top 5 popular social media platforms were Twitter, Flickr, Instagram, Weibo and Openstreetmap, with Twitter and Weibo providing text-based data, Flickr and Instagram providing image-based data and Openstreetmap providing map-based data. This review also examined a number of geospatial methods used for data collection and analysis, and highlighted a number of quality issues and suggested methods for improving data quality from the reviewed articles.

4.4.1 Research gaps and opportunities

There are many potential areas for further research that have been highlighted by the process of undertaking this review. These relate to the limitations of social media, as identified in this review, including data acquisition, data representativeness, privacy concerns, data quality, as well as differences across social media platforms. Some of the key research gaps and opportunities in the use of social media data in UGS studies are as follows:

- Using data from multiple sources

Much of the previous research has used only a single data source or platform which may result a biased representation of the target population and fail to capture the important characteristics of that population (Gu et al., 2016; Plunz et al., 2019). Twitter has established a new generation of API (Twitter API 2.0), academic researchers then can collect the full history of public Tweets via Twitter Academic Research API, this provide researchers a window into understanding the use of Twitter and social media (Chen et al., 2021). However, most platforms offer only limited data access to researchers, and the sampling algorithms for platform APIs remain unknown (Toivonen et al., 2019). For example, Wang. et al. (Wang, Z. et al., 2018) used the data that was collected from a social media platform Dazhongdianping (www.dianping.com) which is a website allowing people to provide reviews on local services across China to assess park use in Beijing and recommended that further analysis should be taken using different data. In other studies, Flickr was used as a sole data source (Lee, J.Y. and Tsou, 2018a), however, recent changes to the Flickr API and terms of service have caused difficulties in accessing data. Diffrent platforms can provide data describing different aspects of the same place, whereas using only a single platform may cause biases and uncertainties. Comparisons with different kind of social media platforms and on-site surveys will help improve the generalizability of the studies. An example of an approach that combines multiple sources is that three platforms (Flickr, Panoramio, and Geograph) were used to detect cultural ecosystem services (Gliozzo et al., 2016). Their results show different photo sharing behaviors, with Flickr and Panoramio having almost interchangeable results whereby Flickr placing greater emphasis on human made cultural artifacts. A further extension is possible through recent developments in image analysis, which support the automated classification of photographs into known categories, which could be extended into typical UGS features. Such data would enhance the analysis of social media data, especially in the context of examining the features that are most attractive to UGS users and shared across media platforms.

- The need for combining personal information with data analysis

Information about individual users, including gender, age, occupation, and income, are very meaningful for the study of cultural service perception, park use assessment and UGS planning (Gliozzo et al., 2016). Whilst recognising that some park users may be reluctant to disclose their personal sensitive information, such as income and sexual orientation, such data may allow a more refined analysis or attitudes and perceptions and may provide confounding or modifying factors in an analysis. Analysing this type of user data is an important part of understanding the variations in perceived information. The lack of such data is not conducive to the subdivision of research data, but can be inferred from exploration of user posting histories (Butler et al., 2018). It may be more effective to combine survey data which may cover more comprehensive individual information to supplement the research results. Only two studies (Devkota et al., 2019; Heikinheimo et al., 2017) used both of survey data and social media data in UGS research. In addition, the number of visitors to park areas need to be accurate as much as possible, this data can be used to validate the results from social media data and help researchers to comprehensively understand park use. In order to estimate the actual number of park visitors, counters could be set at some parks, this will give accurate data about the number of people who visit parks. In addition, some municipalities provide free Wi-Fi hubs inside parks and data from these hubs could be used to estimate number of visitors. These types of data can be used as a complement to questionnaires and social network data methods.

- Improving information mining analysis and models

In order to improve the accuracy of language translation there are a number of opportunities for more nuanced analyses of social media such as Twitter. Domain-specific lexicons (Al-Kodmany, 2019) need to be developed specifically for green spaces. In order to generate a more accurate analysis of visitor opinions in social media, future research should consider developing specific, bespoke lexicons for parks, forests, lakes or other related venues as has been done in other domains (Koblet et al., 2020). In addition, there still exists the challenge of analysing and translating polarity related to negative or positive perception in sentence-level sentiment analysis. For data analysis, there are various methods associated with different kinds of social media data used to analyse UGS. Specifically, in terms of text-based data such as Twitter data and Weibo, it is important to process text-based semantics for sentiment and similar analyses. The analysis of geo-tagged social media data requires methods to detect the accurateness of the location information (Dunkel, 2015) and analysis models and workflows need to be further refined. For example, it is difficult to tell whether people mention Bird's Nest and the Water Cube in Olympic parks because they are attractive or simply to use them as a location reference (Wang, Z. et al., 2018). Thus, a stronger unsupervised selection technique is needed to analyse these unlabelled, unstructured and inherently linked datasets online. A further improvement to analyses of social media would be examine the networks suggested by social media posts that are shared, re-posted or reacted. Here classic graph theoretical approaches could be used to infer connections, influencers opinions and spatio-temporal trends in social media data (Comber et al., 2012). This is a hugely under-developed area of research that has yet to gain traction in domain specific analyses of social media such as UGS. Examining such interactions can indicate topics of particular interest and potentially deal with data sparsity issues.

- The representativeness and validation of social media data in UGS research

The representativeness of social media data sources such as Twitter have attracted more and more attention from scholars. For example, British Twitter users tend to be younger, wealthier, and better educated than the general population (Blank and Lutz, 2017). However, when research is limited to urban areas, georeferenced Tweets or photos can produce valuable and useful knowledge due to the high density of social media users (Plunz et al., 2019). It is important that researchers assess the validity of social media data before analysis. For example, Twitter data on park use were validated in Barcelona and Madrid by comparing different data sources including census, and cell phone data

(Lenormand et al., 2014). The results showed that the three data sources provided comparable information in studies of urban human mobility. Twitter data have been widely used in urban green space research and some studies (Donahue et al., 2018) have suggested that geolocated Twitter data in metropolitan cities can be used as an effective tool to characterize commercial, leisure, and residential areas for urban planners. Validation can also be through official data such as contemporary census data or survey data provided by local managers. A further dimension to the issue of representativeness relates to general social media usage. A key area of future work is to examine the context of social media analyses using related data to explore whether the use of social media in relation to UGS is correlated to social media usage generally (for example, ease of access), to local cultural social media usage customs or even to the amount of UGS.

4.4.2 Analysis methods and approaches

Previous studies analysed VGI data from the aspects of spatiotemporal patterns of data points, text mining and semantic analysis. However, the VGI data cleaning and pre-processing plays an important role in whole research works.

Researchers should carefully clean the collected data sets before analyzing them. For example, social media data such as Tweets can be posted by bots or spammers instead of actual Twitter users, this may cause the data bias and over representativeness, and the sentiments of Tweets can also be overestimated by Tweets that were posted by retails, job advertisements and shopping malls. More advanced cleaning methods should be used according to different objectives of research works. For example, some studies (Plunz et al., 2019) focused on the differences between park visitors and non-park users, thus is important to distinguish the users' categories before analyzing the data sets.

As for spatial pattern analysis, this review mainly summarized the KDE as the method which was frequently used in previous studies (Lee, J.Y. and Tsou, 2018a; Ullah et al., 2020; Maia et al., 2008). The key issue in using this method it to determine the kernel radius when assessing the density of data points in study areas. In addition to KDE, K-means, Mean-Shift and DBSCAN algorithm are commonly used to assess the spatial patterns of tourists in some studies (Ghermandi and Sinclair, 2019; Hasnat and Hasan, 2018). The approaches that combing different spatial analysis methods should be therefore developed in future works related to UGS research using VGI data. In temporal analysis, different time scales have been used in previous studies which mainly focused on daily, weekly, monthly visitation patterns. The combination of spatial analysis and temporal analysis could be undertaken in more specific analyses such as at the individual level. For example, a discretized spatial-temporal probabilistic distribution can be used to characterize the Twitter users who posted geo-referenced tweets when visiting UGS areas (Shou et al., 2020). Further, previous studies mainly analysed the UGS visitation to understand the current or past states of UGS use, few studies have paid attention to the prediction of UGS visitation, future research could focus on the prediction of UGS visitation mode especially for holidays such as Christmas and Easter Day.

Text mining is very important in social media analysis because it provides the basis for various research objectives including sentiment analysis, emotion detection and topic modelling. This review summarized the sentiment analysis methods such as SentiStrength V2.2 (Park, S.B. et al., 2018), word polarity (Lyu, K. and Kim, 2016) and Graph Based Semi-Supervised Learning Annotation (Chapman et al., 2018). In the sentiment classification of texts from Tweets for example, it is possible that each Tweet contain more than one kind of emotion or sentiment, thus it is important to determine the overlaps amongst different sentiment categories when classifying the sentiments of Tweets. Topic detection also plays an important role in text mining. However, topics detection from unstructured data such as Tweets is challenging due to the short and unstructured content and dynamic environment. Recently, the methods used to estimate topics from social media platform include Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), Nonnegative Matrix Factorization (NMF) and Latent Dirichlet Allocation (LDA) (Nugroho et al., 2020). The key point of topic modelling for social media data will be combining more text, considering social features and taking the temporal aspect into account as user's environment always changes in real time. In addition, the number of topics and model selection also play an important role in topic modelling. Future research should take care in selecting suitable and appropriately sensitive approaches for detecting topics in different data sources.

4.5 Conclusions

This paper makes a novel contribution by comprehensively reviewing the scientific literature of research using VGI and social media data to understand UGS. Snowballing (Konijnendijk et al., 2013) was used to capture relevant papers that were not part of the original search, but were referenced within the identified papers, and personal knowledge of the literature were used in addition to the systematic search. As such, the literature search is not entirely replicable, which is a limitation. However, it follows well-understood standards for narrative reviews (Wong et al., 2013). The variation in the usage of different data platforms has been described and a number of research areas using these data sources has been discussed, as well as data analysis methods and data quality issues in the context of UGS research. A number of limitations associated with social media data were identified related to their coverage, data quality, representative uncertainties. Researchers using such data should pay particular attention to these especially in the context of spatial or locational research. Social media data can be cross-validated or linked to other data to overcome the limitations of using data from a single platform and combining data sources and types in this way allows some of the limitations to be overcome.

There are a number of opportunities for future research, including the need to evolve methods that have a greater analytical depth beyond sentiment and text mining: in order to increase the depth of information that is extracted from social media data, for example linked to preferences and behaviours. In the specific case of urban green space, future research should focus on factors related to the physical activities in UGS areas, urban park use and accessibility, all which can be captured from social media data. For example, researchers could determine the motivations of contributors to social networks in sharing UGS related text and images, and this has the potential to inform on the specific UGS qualities that are being shared (i.e., park accessibility, design configuration, presence of water, etc.). Automated classification of images posted online also has considerable potential. While some research exists regarding motivations and psychological reasons as to why people share (e.g. a personal cause), further research is needed to determine why a certain UGS feature has been shared, the timing of the shared post, the novelty of the content, etc. In addition, there is a need to assess the usability of social media data analysis in public departments involved in decision making processes around UGS. In terms of data analysis, future research should examine approaches that combine textual, image and map data to extract more representative information for UGS. This would require tools to be developed to do this. Overall, social media data are best used with other data sources to gain a full and dynamic geotags images and text on an urban green space issue, for the benefit of people and living quality.

References

Abdi, H. and Valentin, D. 2007. Multiple correspondence analysis. *Encyclopedia of measurement and statistics*. ***2***(4), pp.651-657.

Al-Kodmany, K. 2019. Improving understanding of city spaces for tourism applications. *Buildings*. ***9***(8), p.187.

Anna, K.G.R., Alina, R., Ronald, K., Bernd, R., Alessandro, C. and Thomas, B. 2018. Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data. *International Journal of Geo-Information*. ***7***(9), p378.

Antonakaki, D., Fragopoulou, P. and Ioannidis, S. 2021. A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. *Expert Systems with Applications*. ***164***, p.114006.

Aria, M. and Cuccurullo, C. 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of informetrics*. ***11***(4), pp.959-975.

Barros, R., Kislansky, P., do Nascimento Salvador, L., Almeida, R., Breyer, M. and Pedraza, L.G. 2015. EDXL-RESCUER ontology: Conceptual Model for Semantic Integration. In: *ISCRAM*.

Basiri, A., Haklay, M., Foody, G. and Mooney, P. 2019. *Crowdsourced geospatial data quality: Challenges and future directions*. Taylor & Francis.

Ben-Harush, O., Carroll, J.A. and Marsh, B. 2012. Using mobile social media and GIS in health and place research. *Continuum*. ***26***(5), pp.715-730.

Blancaflor, E.B., Butalon, J.M.T., Pascual, P.E.S., Yaneza, B.A.U. and Samonte, M.J.C. 2019. Parkpal: a park sharing and crowdsource park monitoring mobile application. In: *Proceedings of the 10th International Conference on E-Education, E-Business, E-Management and E-Learning*, pp.383-388.

Blank, G. and Lutz, C. 2017. Representativeness of social media in great britain: investigating Facebook, Linkedin, Twitter, Pinterest, Google+, and Instagram. *American Behavioral Scientist*. ***61***(7), pp.741-756.

Boyd, D. and Crawford, K. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*. ***15***(5), pp.662-679.

Broadus, R.N. 1987. Toward a definition of “bibliometrics”. *Scientometrics*. ***12***, pp.373-379.

Butler, A., Schafran, A. and Carpenter, G. 2018. What does it mean when people call a place a shithole? Understanding a discourse of denigration in the United Kingdom and the Republic of Ireland. *Transactions of the Institute of British Geographers*. ***43***(3), pp.496-510.

Campbell, L.K., Svendsen, E.S., Sonti, N.F. and Johnson, M.L. 2016. A social assessment of urban parkland: Analyzing park use and meaning to inform management and resilience planning. *Environmental Science & Policy*. ***62***, pp.34-44.

Chapman, L., Resch, B., Sadler, J., Zimmer, S., Roberts, H. and Petutschnig, A. 2018. Investigating the emotional responses of individuals to urban green space using twitter data: A critical comparison of three different methods of sentiment analysis. *Urban Planning*. ***3***(1), pp.21-33.

Chen, E., Deb, A. and Ferrara, E. 2021. # Election2020: the first public Twitter dataset on the 2020 US Presidential election. *Journal of Computational Social Science*. pp.1-18.

Chen, W., Huang, H., Dong, J., Zhang, Y., Tian, Y. and Yang, Z. 2018. Social functional mapping of urban green space using remote sensing and social sensing data. *ISPRS Journal of Photogrammetry and Remote Sensing*. ***146***, pp.436-452.

Chen, Y., Liu, X., Gao, W., Wang, R.Y., Li, Y. and Tu, W. 2018. Emerging social media data on measuring urban park use. *Urban forestry & urban greening*. ***31***, pp.130-141.

Chiesura, A. 2004. The role of urban parks for the sustainable city. *Landscape and urban planning*. ***68***(1), pp.129-138.

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***, pp.59-68.

Cohen, D.A., Lapham, S., Evenson, K.R., Williamson, S., Golinelli, D., Ward, P., Hillier, A. and McKenzie, T.L. 2013. Use of neighbourhood parks: does socio-economic status matter? A four-city study. *Public health*. ***127***(4), pp.325-332.

Cohen, D.A., Marsh, T., Williamson, S., Derose, K.P., Martinez, H., Setodji, C. and McKenzie, T.L.J.P.m. 2010. Parks and physical activity: why are some parks used more than others? ***50***, pp.S9-S12.

Comber, A., Batty, M., Brunsdon, C., Hudson-Smith, A., Neuhaus, F. and Gray, S. 2012. Exploring the geography of communities in social networks. In: *Proceedings of the GIS Research UK 20th Annual Conference*, pp.33-37.

Comber, A., Brunsdon, C. and Green, E. 2008. Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups. *Landscape and urban planning*. ***86***(1), pp.103-114.

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), p.5386.

Daim, T.U., Rueda, G., Martin, H. and Gerdsri, P. 2006. Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological forecasting and social change*. ***73***(8), pp.981-1012.

Dallimer, M., Davies, Z.G., Irvine, K.N., Maltby, L., Warren, P.H., Gaston, K.J. and Armsworth, P.R. 2014. What personal and environmental factors determine frequency of urban greenspace use?. *International Journal of Environmental Research and Public Health*. ***11***(8), pp.7977-7992.

Devkota, B., Miyazaki, H., Witayangkurn, A. and Kim, S.M. 2019. Using volunteered geographic information and nighttime light remote sensing data to identify tourism areas of interest. *Sustainability*. ***11***(17), p.4718.

Di Minin, E., Tenkanen, H. and Toivonen, T. 2015. Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*. ***3***, p.63.

Donahue, M.L., Keeler, B.L., Wood, S.A., Fisher, D.M., Hamstead, Z.A. and McPhearson, T. 2018. Using social media to understand drivers of urban park visitation in the Twin Cities, MN. *Landscape and Urban Planning*. ***175***, pp.1-10.

Donaire, J.A., Camprubí, R. and Galí, N. 2014. Tourist clusters from Flickr travel photography. *Tourism management perspectives.* ***11***, pp.26-33.

Dunkel, A. 2015. Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and urban planning.* ***142***, pp.173-186.

Fan, P., Xu, L., Yue, W. and Chen, J. 2017. Accessibility of public urban green space in an urban periphery: The case of Shanghai. *Landscape and Urban Planning*. ***165***, pp.177-192.

Figueroa-Alfaro, R.W. and Tang, Z. 2017. Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. *Journal of environmental planning and management*. ***60***(2), pp.266-281.

García-Palomares, J.C., Gutiérrez, J. and Mínguez, C. 2015. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*. ***63***, pp.408-417.

Ghermandi, A. and Sinclair, M. 2019. Passive crowdsourcing of social media in environmental research: A systematic map. *Global environmental change*. ***55***, pp.36-47.

Gliozzo, G., Pettorelli, N. and Haklay, M. 2016. Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales, UK. *Ecology and Society.* ***21***(3).

Gonçalves, P., Araújo, M., Benevenuto, F. and Cha, M. 2013. Comparing and combining sentiment analysis methods. In: *Proceedings of the first ACM conference on Online social networks*, pp.27-38.

Goodchild, M.F. and Li, L. 2012. Assuring the quality of volunteered geographic information. *Spatial statistics*. ***1***, pp.110-120.

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***, p.100958.

Greenacre, M. and Blasius, J. 2006. *Multiple correspondence analysis and related methods.* CRC press.

Greenwood, S., Perrin, A. and Duggan, M. 2016. Social Media Update 2016: Pew Research Center: Internet. *Science, and Technology*.

Grose, M.J. 2009. Changing relationships in public open space and private open space in suburbs in south-western Australia. *Landscape and urban planning*. ***92***(1), pp.53-63.

Gu, Z., Zhang, Y., Chen, Y. and Chang, X. 2016. Analysis of attraction features of tourism destinations in a mega-city based on check-in data mining—A case study of ShenZhen, China. *ISPRS International Journal of Geo-Information*. ***5***(11), p.210.

Guerrero, P., Møller, M.S., Olafsson, A.S. and Snizek, B. 2016. Revealing cultural ecosystem services through Instagram images: The potential of social media volunteered geographic information for urban green infrastructure planning and governance. *Urban Planning*. ***1***(2), pp.1-17.

Haaland, C. and van Den Bosch, C.K. 2015. Challenges and strategies for urban green-space planning in cities undergoing densification: A review. *Urban forestry & urban greening*. ***14***(4), pp.760-771.

Hamstead, Z.A., Fisher, D., Ilieva, R.T., Wood, S.A., McPhearson, T. and Kremer, P. 2018. Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*. ***72***, pp.38-50.

Han, S.Y., Tsou, M.H. and Clarke, K.C. 2015. Do global cities enable global views? Using Twitter to quantify the level of geographical awareness of US cities. *PloS one*. ***10***(7), p.e0132464.

Hao, J., Zhu, J. and Zhong, R. 2015. The rise of big data on urban studies and planning practices in China: Review and open research issues. *Journal of Urban Management*. ***4***(2), pp.92-124.

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***, pp.38-54.

Hecht, B. and Stephens, M. 2014. A tale of cities: Urban biases in volunteered geographic information. In: *Proceedings of the International AAAI Conference on Web and Social Media*, ***8***(1), pp.197-205.

Heikinheimo, V., Di Minin, E., Tenkanen, H., Hausmann, A., Erkkonen, J. and Toivonen, T. 2017. User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*. ***6***(3), p.85.

Hennig, S. 2017. OpenStreetMap used in protected area management. The example of the recreational infrastructure in Berchtesgaden National Park. *Journal on Protected Mountain Areas Research and Management.* ***9***, pp.30-41.

Johnson, M.L., Campbell, L.K., Svendsen, E.S. and McMillen, H.L. 2019. Mapping urban park cultural ecosystem services: A comparison of twitter and semi-structured interview methods. *Sustainability*. ***11(21)***, p6137.

Johnson, M.L., Campbell, L.K., Svendsen, E.S. and McMillen, H.L. 2019. Mapping urban park cultural ecosystem services: A comparison of twitter and semi-structured interview methods. *Sustainability*. ***11***(21), p.6137.

Jung, J., Uejio, C.K., Duclos, C. and Jordan, M. 2019. Using web data to improve surveillance for heat sensitive health outcomes. *Environmental Health*. ***18***, pp.1-13.

Kabisch, N., Qureshi, S. and Haase, D. 2015. Human–environment interactions in urban green spaces—A systematic review of contemporary issues and prospects for future research. *Environmental Impact assessment review*. ***50***, pp.25-34.

Kashef, M. 2016. Urban livability across disciplinary and professional boundaries. *Frontiers of Architectural Research*. ***5***(2), pp.239-253.

Kim, D. and Jin, J. 2018. Does happiness data say urban parks are worth it?. *Landscape and Urban Planning*. ***178***, pp.1-11.

Koblet, O. and Purves, R.S. 2020. From online texts to Landscape Character Assessment: Collecting and analysing first-person landscape perception computationally. *Landscape and Urban Planning*. ***197***, p.103757.

Konijnendijk, C.C., Annerstedt, M., Nielsen, A.B. and Maruthaveeran, S., 2013. Benefits of Urban Parks. Copenhagen & Alnarp,(January).

Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A. and Blaschke, T. 2018. Beyond spatial proximity-classifying parks and their visitors in london based on spatiotemporal and sentiment analysis of twitter data. *ISPRS International Journal of Geo-Information.* ***7***(9), p378.

Kwak, H., Lee, C., Park, H. and Moon, S. 2010. What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World wide web*, pp.591-600.

Larson, L.R., Jennings, V. and Cloutier, S.A. 2016. Public parks and wellbeing in urban areas of the United States. *PLoS one*. ***11***(4), p.e0153211.

Lee, J.Y. and Tsou, M.-H. 2018a. Mapping spatiotemporal tourist behaviors and hotspots through location-based photo-sharing service (Flickr) data. In: *LBS 2018: 14th International Conference on Location Based Services*: Springer, pp.315-334.

Lee, J.Y. and Tsou, M.-H. 2018b. Mapping spatiotemporal tourist behaviors and hotspots through location-based photo-sharing service (Flickr) data. In: *Progress in Location Based Services 2018 14*: Springer, pp.315-334.

Lee, Y., Kwon, P., Yu, K. and Park, W. 2016. Method for determining appropriate clustering criteria of location-sensing data. *ISPRS International Journal of Geo-Information*. ***5***(9), p.151.

Lenormand, M., Picornell, M., Cantú-Ros, O.G., Tugores, A., Louail, T., Herranz, R., Barthelemy, M., Frias-Martinez, E. and Ramasco, J.J. 2014. Cross-checking different sources of mobility information. *PloS one*. ***9***(8), p.e105184.

Levin, N., Lechner, A.M. and Brown, G. 2017. An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. *Applied geography*. ***79***, pp.115-126.

Li, F., Li, F., Li, S. and Long, Y. 2020. Deciphering the recreational use of urban parks: Experiments using multi-source big data for all Chinese cities. *Science of the Total Environment*. ***701***, p.134896.

Li, J., Li, J., Yuan, Y. and Li, G. 2019. Spatiotemporal distribution characteristics and mechanism analysis of urban population density: a case of Xi'an, Shaanxi, China. *Cities*. ***86***, pp.62-70.

Li, L., Goodchild, M.F. and Xu, B. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and geographic information science*. ***40***(2), pp.61-77.

Liu, H., Li, F., Xu, L. and Han, B. 2017. The impact of socio-demographic, environmental, and individual factors on urban park visitation in Beijing, China. *Journal of Cleaner Production*. ***163***, pp.S181-S188.

Lyu, F. and Zhang, L. 2019. Using multi-source big data to understand the factors affecting urban park use in Wuhan. *Urban Forestry & Urban Greening*. ***43***, p126367.

Lyu, F. and Zhang, L. 2019. Using multi-source big data to understand the factors affecting urban park use in Wuhan. *Urban Forestry & Urban Greening*. ***43***, p.126367.

Lyu, K. and Kim, H. 2016. Sentiment analysis using word polarity of social media. *Wireless Personal Communications*. ***89***, pp.941-958.

Maeda, T.N., Yoshida, M., Toriumi, F. and Ohashi, H. 2018. Extraction of tourist destinations and comparative analysis of preferences between foreign tourists and domestic tourists on the basis of geotagged social media data. *ISPRS International Journal of Geo-Information*. ***7***(3), p.99.

Maia, M., Almeida, J. and Almeida, V. 2008. Identifying user behavior in online social networks. In: *Proceedings of the 1st workshop on Social network systems*, pp.1-6.

Martí, P., Serrano-Estrada, L. and Nolasco-Cirugeda, A. 2019. Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*. ***74***, pp.161-174.

Martínez-Cámara, E., Martín-Valdivia, M.T., Urena-López, L.A. and Montejo-Ráez, A.R. 2014. Sentiment analysis in Twitter. *Natural language engineering*. ***20***(1), pp.1-28.

Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S. and Danforth, C.M. 2013. The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS one*. ***8***(5), p.e64417.

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G. and Fuller, G. 2019. Sustainable Development Report 2019. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN). *Search in*.

Niemelä, J. 1999. Ecology and urban planning. *Biodiversity & Conservation*. ***8***, pp.119-131.

Nitoslawski, S.A., Galle, N.J., Van Den Bosch, C.K. and Steenberg, J.W. 2019. Smarter ecosystems for smarter cities? A review of trends, technologies, and turning points for smart urban forestry. *Sustainable Cities and Society*. ***51***, p.101770.

Norman, P. and Pickering, C.M. 2019. Factors influencing park popularity for mountain bikers, walkers and runners as indicated by social media route data. *Journal of environmental management*. ***249***, p.109413.

Nugroho, R., Paris, C., Nepal, S., Yang, J. and Zhao, W. 2020. A survey of recent methods on deriving topics from Twitter: algorithm to evaluation. *Knowledge and information systems*. ***62***(7), pp.2485-2519.

Oksanen, J., Bergman, C., Sainio, J. and Westerholm, J. 2015. Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. *Journal of Transport Geography*. ***48***, pp.135-144.

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***, pp.74-86.

Paracchini, M.L., Zulian, G., Kopperoinen, L., Maes, J., Schägner, J.P., Termansen, M., Zandersen, M., Perez-Soba, M., Scholefield, P.A. and Bidoglio, G. 2014. Mapping cultural ecosystem services: A framework to assess the potential for outdoor recreation across the EU. *Ecological indicators*. ***45***, pp.371-385.

Park, S.B., Kim, H.J. and Ok, C.M. 2018. Linking emotion and place on Twitter at Disneyland. *Journal of Travel & Tourism Marketing*. ***35***(5), pp.664-677.

Park, W., You, Y. and Lee, K. 2018. Detecting potential insider threat: Analyzing insiders' sentiment exposed in social media. *Security and Communication Networks*. *2018*.

Petticrew, M. 2001. Systematic reviews from astronomy to zoology: myths and misconceptions. *Bmj*. ***322***(7278), pp.98-101.

Pickering, C., Walden-Schreiner, C., Barros, A. and Rossi, S.D. 2020. Using social media images and text to examine how tourists view and value the highest mountain in Australia. *Journal of Outdoor Recreation and Tourism*. ***29***, p.100252.

Plunz, R.A., Zhou, Y., Vintimilla, M.I.C., Mckeown, K., Yu, T., Uguccioni, L. and Sutto, M.P. 2019. Twitter sentiment in New York City parks as measure of well-being. *Landscape and urban planning*. ***189***, pp.235-246.

Rizwan, M. and Wan, W. 2018. Big data analysis to observe check-in behavior using location-based social media data. *Information*. ***9***(10), p.257.

Roberts, H.V. 2017. Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography*. ***81***, pp.13-20.

Roberts, H., Sadler, J. and Chapman, L. 2017. Using Twitter to investigate seasonal variation in physical activity in urban green space. *Geo: Geography and Environment*. ***4***(2), p.e00041.

Roberts, H., Sadler, J. and Chapman, L. 2019. The value of Twitter data for determining the emotional responses of people to urban green spaces: A case study and critical evaluation. *Urban studies*. ***56***(4), pp.818-835.

Roe, J.J., Thompson, C.W., Aspinall, P.A., Brewer, M.J., Duff, E.I., Miller, D., Mitchell, R. and Clow, A. 2013. Green space and stress: evidence from cortisol measures in deprived urban communities. *International journal of environmental research and public health*. ***10***(9), pp.4086-4103.

Sadhukhan, P. 2017. An IoT-based E-parking system for smart cities. In: *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, dupi, India, 2017, pp.1062-1066.

Salas-Olmedo, M.H. and Rojas Quezada, C. 2017. The use of public spaces in a medium-sized city: from Twitter data to mobility patterns. *Journal of Maps*. ***13***(1), pp.40-45.

Santos, T., Mendes, R.N. and Vasco, A. 2016. Recreational activities in urban parks: Spatial interactions among users. *Journal of outdoor recreation and tourism*. ***15***, pp.1-9.

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***, pp.336-350.

See, L., Estima, J., Pődör, A., Arsanjani, J.J., Bayas, J.C.L. and Vatseva, R. 2017. Sources of VGI for Mapping. *Citizen Sensor.*, p13.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B. and Laakso, M. 2016a. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information.* ***5***(5), p55.

Senaratne, H., Mobasheri, A., Ali, A.L., Capineri, C. and Haklay, M. 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*. ***31***(1), pp.139-167.

Sheng, T., Chen, X.J., Gao, S., Liu, Q.Z., Li, X.F. and Fu, Q.Y., 2018. Pollution characteristics and health risk assessment of VOCs in areas surrounding a petrochemical park in Shanghai. ***39***(11), pp.4901-4908.

Shi, B., Zhao, J. and Chen, P.J. 2017. Exploring urban tourism crowding in Shanghai via crowdsourcing geospatial data. *Current Issues in Tourism*. ***20***(11), pp.1186-1209.

Shou, Z., Cao, Z. and Di, X. 2020. Similarity Analysis of Spatial-Temporal Mobility Patterns for Travel Mode Prediction Using Twitter Data. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*: IEEE, pp.1-6.

Sim, J. and Miller, P. 2019. Understanding an urban park through big data. *International journal of environmental research and public health*. ***16***(20), p.3816.

Sinclair, M., Ghermandi, A. and Sheela, A.M. 2018. A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India. *Science of the total environment*. ***642***, pp.356-365.

Sinclair, M., Mayer, M., Woltering, M. and Ghermandi, A. 2020. Using social media to estimate visitor provenance and patterns of recreation in Germany's national parks. *Journal of Environmental Management*. ***263***, p.110418.

Song, Y., Huang, B., Cai, J. and Chen, B. 2018. Dynamic assessments of population exposure to urban greenspace using multi-source big data. *Science of the Total Environment*. ***634***, pp.1315-1325.

Song, Y. and Zhang, B. 2020. Using social media data in understanding site-scale landscape architecture design: Taking Seattle Freeway Park as an example. *Landscape Research*. ***45***(5), pp.627-648.

Sprake, J. and Rogers, P. 2014. Crowds, citizens and sensors: Process and practice for mobilising learning. *Personal and ubiquitous computing*. ***18***, pp.753-764.

Steiger, E., Resch, B. and Zipf, A. 2016. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science*. ***30***(9), pp.1694-1716.

Stock, K. 2018. Mining location from social media: A systematic review. *Computers, Environment and Urban Systems*. ***71***, pp.209-240.

Sun, Y., Du, Y., Wang, Y. and Zhuang, L. 2017. Examining associations of environmental characteristics with recreational cycling behaviour by street-level Strava data. *International journal of environmental research and public health*. ***14***(6), p.644.

Taylor, M., Wells, G., Howell, G. and Raphael, B. 2012. The role of social media as psychological first aid as a support to community resilience building. *Australian Journal of Emergency Management.* ***27***(1), pp.20-26.

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), p.17615.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H. and Di Minin, E. 2019. Social media data for conservation science: A methodological overview. *Biological Conservation*. ***233***, pp.298-315.

Tsai, W.L., McHale, M.R., Jennings, V., Marquet, O., Hipp, J.A., Leung, Y.F. and Floyd, M.F. 2018. Relationships between characteristics of urban green land cover and mental health in US metropolitan areas. *International journal of environmental research and public health*. ***15***(2), p.340.

Ullah, H., Wan, W., Haidery, S.A., Khan, N.U., Ebrahimpour, Z. and Muzahid, A.A.M. 2020. Spatiotemporal patterns of visitors in urban green parks by mining social media big data based upon WHO reports. *IEEE Access*. ***8***, pp.39197-39211.

Vaz, A.S., Gonçalves, J.F., Pereira, P., Santarém, F., Vicente, J.R. and Honrado, J.P. 2019. Earth observation and social media: Evaluating the spatiotemporal contribution of non-native trees to cultural ecosystem services. *Remote Sensing of Environment*. ***230***, p.111193.

Vich, G., Marquet, O. and Miralles-Guasch, C. 2019. Green streetscape and walking: exploring active mobility patterns in dense and compact cities. *Journal of transport & health*. ***12***, pp.50-59.

Vieira, F.A., Bragagnolo, C., Correia, R.A., Malhado, A.C. and Ladle, R.J. 2018. A salience index for integrating multiple user perspectives in cultural ecosystem service assessments. *Ecosystem Services*. ***32***, pp.182-192.

Wakamiya, S., Lee, R. and Sumiya, K. 2011. Crowd-based urban characterization: extracting crowd behavioral patterns in urban areas from twitter. In: *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks*, pp.77-84.

Wang, A., Zhang, A., Chan, E.H., Shi, W., Zhou, X. and Liu, Z. 2020. A review of human mobility research based on big data and its implication for smart city development. *ISPRS International Journal of Geo-Information*. ***10***(1), p.13.

Wang, Z., Jin, Y., Liu, Y., Li, D. and Zhang, B. 2018. Comparing social media data and survey data in assessing the attractiveness of Beijing Olympic Forest Park. *Sustainability*. ***10***(2), p.382.

Warriner, A.B., Kuperman, V. and Brysbaert, M. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods*. ***45***, pp.1191-1207.

Weiler, A., Grossniklaus, M. and Scholl, M.H., 2016. Situation monitoring of urban areas using social media data streams. *Information Systems*. *57*, pp.129-141.

Wolch, J.R., Byrne, J. and Newell, J.P. 2014. Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and urban planning*. ***125***, pp.234-244.

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J. and Pawson, R. 2013. RAMESES publication standards: Meta-narrative reviews. *Journal of Advanced Nursing*. ***69***(5), pp.987-1004.

Wood, S.A., Guerry, A.D., Silver, J.M. and Lacayo, M. 2013. Using social media to quantify nature-based tourism and recreation. *Scientific reports*. ***3***(1), p.2976.

Zhang, S. and Zhou, W. 2018. Recreational visits to urban parks and factors affecting park visits: Evidence from geotagged social media data. *Landscape and urban planning*. ***180***, pp.27-35.

Zhang, W., Yang, J., Ma, L. and Huang, C. 2015. Factors affecting the use of urban green spaces for physical activities: Views of young urban residents in Beijing. *Urban Forestry & Urban Greening*. ***14***(4), pp.851-857.

Zhe, L., Yong, G., Hung-Suck, P., Huijuan, D., Liang, D. and Tsuyoshi, F. 2016. An emergy-based hybrid method for assessing industrial symbiosis of an industrial park. *Journal of Cleaner Production*. ***114***, pp.132-140.

Chapter 5 aims to examine spatial-temporal changes in UGS use in London associated with COVID-19 related lockdowns by using Twitter. This is one of the first studies to examine social media data over consistent time period before, during and after the COVID-19 lockdown in relation to UGS use. The results show that the land-use type of Public Park and Garden was the most frequently visited type of UGS, the usage of UGS decreased in central London and increased in other areas during lockdown, which may correlated with working from home restrictions, UGS activities were positively associated with Physical activities maybe as a result of allowing people to take a single daily exercise, and finally people spent more time in UGS areas on weekdays than weekends compared to pre-lockdown period. The findings and methods in this chapter can potentially inform policy makers in UGS management and planning, especially in a period of social crisis like the COVID-19 pandemic. It is imperative to acknowledge the presence of certain limitations in this analysis, including, but not limited to, concerns related to data bias and representativeness. For example, demographic information such as age and education background are not available in Twitter metadata, which results in limitations or challenges when attempting to investigate factors associated with UGS usage in relation to these demographic characteristics. Furthermore, this study exclusively investigates the periods preceding, during, and following the initial national lockdown. Nevertheless, it is worth noting that certain restrictions persisted throughout the 2021 period. Subsequent research endeavours could broaden their scope to encompass other lockdown phases and examine periods when policies were eased, eventually leading to the complete removal of lockdown restrictions

Chapter 5 Using social media data to understand the impact of the COVID-19 pandemic on urban green space use

Abstract

Having access to and visiting urban green space (UGS) improves liveability and provides considerable benefits to residents. However, traditional methods of investigating UGS visitation, such as questionnaires and social surveys, are usually time- and resource-intensive, and frequently provide less transferable, site-specific outcomes. This study uses social media data (Twitter) to examine spatio-temporal changes in UGS use in London associated with COVID-19 related lockdowns. It compares georeferenced Tweets posted in a 3 month period from 23 March to 23 June for 3 years covering the first lockdown in the UK in 2020, with Tweets for the same period in 2019 and 2021. The results show that

(1) the land-use type of *Public Park and Garden* was the most frequently visited type of UGS, which may be correlated with these UGS areas remaining opening during the lockdown period; (2) the usage of UGS decreased in central London and increased in other areas during lockdown, which may correlated with working from home restrictions; (3) activities were positively associated with *Physical activities* maybe as a result of allowing people to take a single daily exercise, and (4) people spent more time in UGS areas on weekdays than weekends compared to pre-lockdown. This is the first study to examine social media data over consistent time period before, during and after the lockdown in relation to UGS. The results show that the findings and method can inform policy makers in their management and planning of UGS, especially in a period of social crisis like the COVID-19 pandemic.

Keywords: Urban Green Space, Social Media Data, COVID-19, Lockdown.

5.1 Introduction

Coronavirus Disease 2019 (COVID-19) (Zu et al., 2020) was recognized as a global pandemic by the World Health Organization (WHO) on March 12, 2020 (Dashraath et al., 2020). In response many countries including the United Kingdom adopted a series of public health measures, such as travel restrictions, quarantine, closing non-essential businesses and services, and requiring citizens to stay at home except for essential trips, to mitigate the virus spread (Cameron-Blake et al., 2020). In addition, citizens in the UK had to respect social distancing of 2m, and wear face coverings when entering shops and using public transport (Shoari et al., 2020). In this lockdown, public mental and physical health were affected by these restriction measures (Gao et al., 2020; Erdönmez and Atmiş, 2021). For example, complaints of musculoskeletal and circulatory system disorders increased during the lockdown, and the separation from relatives and friends caused psychological harm (Erdönmez and Atmiş, 2021). Research has showed that that urban parks and green spaces had a significantly positive impact on individuals' mental and physical health in this period (Theodorou et al., 2021), thus parks and green spaces became increasingly important public places for supporting physical and mental wellbeing (Zhu and Xu, 2021; Geng et al., 2021).

This study investigates the impacts of lockdown on urban green space (UGS) use in London by comparing visits to UGS in the before, during and after the first lockdown period in 2020, as captured through social media (Twitter) following Cui et al. (2021).

UGS refers to urban land covered by vegetation (Niemelä, 1999). It plays a critical role in sustaining urban natural environments and the social systems that use these spaces (Kabisch et al., 2015). UGS is one of the key features associated with urban sustainability, as it enhances the quality of life of urban residents (Chiesura, 2004; Houlden et al., 2019). Publicly accessible urban parks are places of solace, recreation, exercise and community enjoyment, and city residents rely on parks and green spaces for physical, mental, and social wellbeing (Houlden et al., 2019).

Twitter is a free social media (microblogging) platform which allows users to post messages of up to 280 characters in length, with additional information such as coordinates and time of posting based on user preference (Steiger et al., 2015). The high spatial and temporal resolution provides researchers with opportunities to analyse the spatio-temporal distributions of Tweets as well as their content. Twitter data has already been used successfully in UGS research recently (Roberts, 2017; Plunz et al., 2019; Cui et al., 2021), but has not yet been used to examine the spatio-temporal changes in UGS visitation, during the COVID-19 pandemic period. This study uses geo-referenced Twitter data extracted for 3 coincident period before, during and after the COVID-19 lockdown to examine changes in the spatio-temporal patterns of UGS visitation.

Twitter data covering 23 March to 23 June were extracted for three years (2019, 2020, 2021) for London and used to capture UGS behaviors and to see how UGS behaviors may have been changed by the lockdown restrictions. This research seeks to answer a number of research questions (RQ):

RQ1: What types of UGS were more frequently visited and how did this change?
RQ2: In which area(s) did UGS visitation increase or decrease?
RQ3: How did UGS activities change, if at all?
RQ4: Did the timing of UGS visitation change?

This paper has been organized as follows: Section 4.2 presents the background and timelines on UK lockdown and restriction measures, as well as related research has explored the impacts of COVID-19 on UGS use. Section 4.3 presents the dataset and the methodology employed. Section 4.4 outlines how UGS visitation patterns changed over the three years. Section 4.5 discusses the results and suggests some areas of further work, before some concluding comments are made (Section 4.6).

5.2 Background

5.2.1 Lockdown rules and restriction measures in the UK

At the outbreak of the COVID-19 pandemic in the UK in early 2020, the government implemented a series of measures in England specifically (Greater London Authority, 2021) to reduce the spread of the disease (Figure 5.1) (the rules in the other Scotland, Wales, and Northern Ireland were different).

On 23 March 2020, the UK government announced national lockdown which stated that all residents must stay at home and work from home, with the exception of shopping for basic necessities. Social events were cancelled and shops selling non-essential goods were closed. Citizens were permitted to take outdoor exercise once per day. Having access to some form of outdoor space such as public parks is vital for people, especially children and those who live in homes without a garden (Poortinga et al., 2021). Fortunately, parks largely remained open, giving people the opportunities to exercise and enjoy fresh air, although gatherings were banned during lockdown. In England citizens were allowed to visit UGS with members of the same households, remaining 2 meters (6ft) away from other people (GOV.UK, 2020b). UGS visitation was also facilitated by the warmer Spring weather in the UK in 2020.

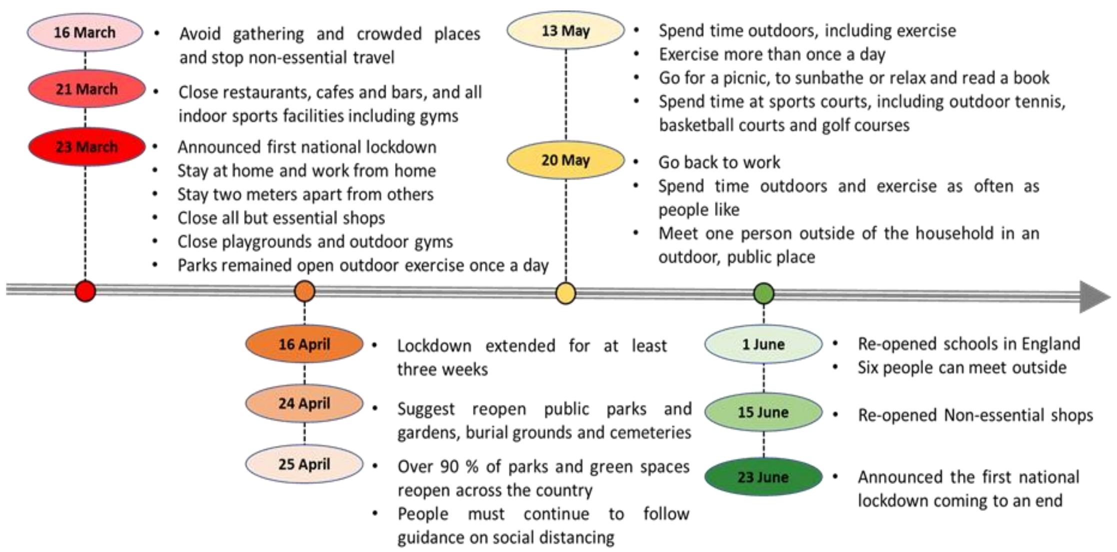

Figure 5. 1 Timeline of UK restriction measures for COVID-19 during 2020.

On 24 April 2020, the government encouraged park managers to keep public parks open to support these activities. Approximately 90% of public parks remained open during the first lockdown in March 2020 (GOV.UK, 2020a) with some restrictions including limiting park opening times, and closing children’s play areas and sports facilities. On 13 May 2020, people in England were allowed to spend more time in park areas to exercise, and to enjoy the outdoors. In addition, sports facilities such as tennis courts, golf courses, and basketball courts were reopened, as sport and exercise provide benefits for people's physical and mental health (Fagerholm et al., 2021), and the government encouraged everyone to stay as active as possible. On 20 May, people were allowed to meet one person from another household. On 23 June, the 2m social distancing restriction was changed to 1m to enable people to meet their family and friends, and the government announced that the first national lockdown was to end in England.

5.2.2 The role of UGS during lockdown

UGS has gained an increasing amount of attention in both academic and policy areas, especially to understand how people visit park areas and their perceptions of UGS (Erdönmez and Atmiş, 2021; Zhu and Xu, 2021). During the COVID-19 pandemic, public mental and physical health were influenced by the restriction measures including working from home, quarantine, and by the constant stream of negative news around the pandemic (such as increasing cases and deaths) (Gao et al., 2020; Erdönmez and Atmiş, 2021). Investigations into park usage during lockdown showed significant changes in visitor behavior patterns, and that urban parks and green spaces had a significantly positive impact on individual mental and physical health in this period (Theodorou et al., 2021). However, the few studies that have analysed changes in the spatial and temporal patterns of urban park visitations during the COVID-19 pandemic period, have offered limited insight. Lopez et al. (2021) focused on changes in park visitor numbers but not changes in spatio-temporal patterns of park visits. Geng et al. (2021) examined changes in people who visited urban parks during the COVID-19 pandemic period but over global, regional, and national scales, rather than local ones. Ugolini et al. (2020) and Erdönmez and Atmiş (2021) explored the impacts of the COVID-19 pandemic on park visits and activities through an online questionnaire, but did not examine the spatio-temporal patterns of park usage during the period. Understanding these patterns may be useful for urban planners and managers especially during a health crisis.

Spatio-temporal analysis of geo-referenced Tweets can provide additional contextual information about the patterns of UGS use and the status of visitors to address RQ (1) 'where', RQ (2) 'what', and RQ (3) 'when'? For example, examining the changes in the spatial patterns of UGS visitation provide information on the actual responses to place-based policy and practice. Changes in the temporal patterns of UGS visitation provide information on how daily routines may have been changed due to restrictions such as working from home and restricting public transport, etc. Geo-referenced Twitter data contains a time-stamp, coordinates, and text information and can be a valuable data source for spatio-temporal analysis (Huang, Y. et al., 2018). Thus, in this study, spatio-temporal analysis of geo-referenced Tweets was used to understand the changes in UGS use before, during and after lockdown.

5.2.3 Social media used to investigate UGS activities

UGS play an essential role in providing physical and social wellbeing for visitors, thus, how to appropriately measure UGS activities is an important research question. For instance, using a questionnaire approach, four types of activities were defined by Sim and Miller (2019): Physical activities, including biking, walking and running; Mental health activities, including relaxing and restoration; Social interaction, including meeting with friends or family, and other activities such as passing the parks. Lesser and Nienhuis (2020) explored how lockdown influenced physical activity during the pandemic but did not examine other types of UGS activities such as social recreational and leisure related ones, despite these representing a large proportion of UGS visits (Sim and Miller, 2019).Social media data have proven to be useful in assessing UGS usage in urban and suburban areas (Roberts, 2017; Cui et al., 2021).

Twitter has become a popular social media platform for many research areas. Tweet data has been used to support land use classifications of urban areas (Mesa-Arango et al., 2016), to investigate emotional responses to different spaces (Zhu and Xu, 2021), to explore how information spreads through urban areas (Huang, Q. and Wong, 2016). In order to categorize Twitter posts into specific kinds of park activity, Salloum et al. (2017) used text mining approaches to extract information from unstructured Tweet. Word frequencies were calculated to identify the park activities and the results were used to conduct sentiment analysis. They found that social media data can capture important activities such as ‘eating’ in parks areas, which could not be distinguished in questionnaire responses. Roberts (2017) extracted information from Twitter data on physical activity in UGS. Tweets were manually screened by authors and word frequencies were calculated for sports-related Tweets, and seven types of physical activity were identified (running, walking, biking, water sports, team sports, and outdoor fitness), demonstrating how Tweets can be used to investigate park physical activities, and to fill the gaps in traditional research methods.

Twitter released a new version of the standard Search Tweets endpoint as part of Twitter Application Programming Interface (API) in 2020. This includes many new features, such as the ability to access the full history of Tweets, allowing researchers to programmatically access public Tweets from the complete archive dating back to March 2006. In this way, the new API provides more complete, and unbiased data than previous Twitter APIs. In the current study, full historical Twitter data were used to explore the impacts of COVID-19 on UGS use, with the aim of assessing the extent to which the government-mandated restriction measures influenced the use of UGS. It explores different types of UGS and people's activities in UGS areas, as well as variations in the temporal and spatial patterns before, during and after the first lockdown period.

5.3 Study area and methods

5.3.1 Twitter data and overview of the analysis

The Twitter datasets used for this study were downloaded via the Twitter academic research API. The API call selected geo-referenced Tweets located in London. They covered a three-month period (23rd March to 23rd Jun) for three consecutive years: 2019, 2020, and 2021. To begin with, the changes in the spatial distributions of geo-located Tweets for three individual years were analysed and the most frequently visited UGSs were identified in each year. Activities in UGS were also identified using text mining (Salloum et al., 2017), and their spatio-temporal changes were investigated. The Twitter data were analyzed over hours and days to identify changes in temporal patterns. A paired sample t-test was used to significant differences between the use of UGS over the three years. All the analyses were undertaken using R software (Ihaka and Gentleman, 1996).

5.3.2 Study area

The study area is London, UK (Figure 5.2), 40% of which is public green space composed of eight main urban parks and about 3000 smaller urban green spaces covering 23,599 ha (GiGL, 2019).

In order to detect whether the COVID-19 lockdown restrictions, including working from home had exerted obvious impacts on UGS visitation, the study area was further separated into London's Central Business District (LCBD) and non-LCBD area (Sulis et al., 2018). The major workplaces, businesses, and tourist attractions are located in LCBD, with many people commuting between LCBD and other areas before lockdown.

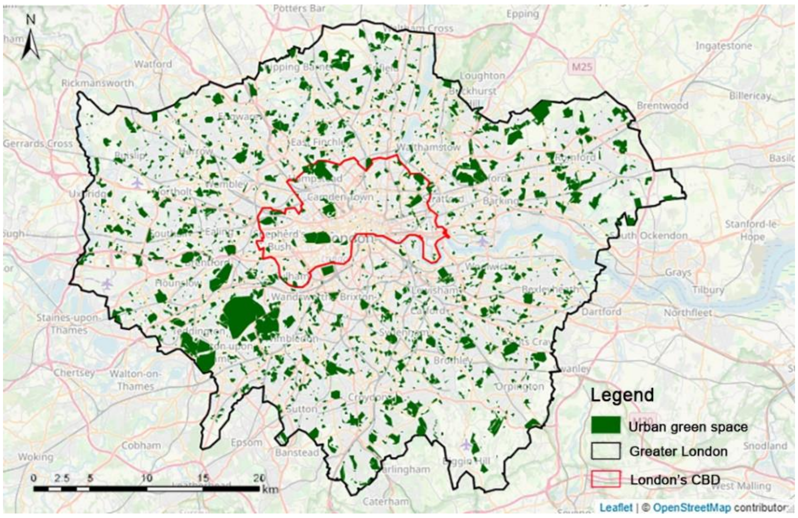

Figure 5.2 The distribution of open green space in London.

5.3.3 Data collection, pre-processing and analysis

Green space in London was derived from the Open Green Space layer from the Ordnance Survey (Ordnance Survey, 2021a). The spatial distribution of green space is shown in Figure 5.2. In order to explore the differences of UGS use among various types of UGS, all open green spaces in study area were classified into 10 functional categories (Ordnance Survey, 2021b), based on the criteria in Table 5.1. The number of Tweets in each type of UGS and their percentages of the total Tweets were calculated.

Table 5.1 Urban green space classification and corresponding descriptions.

	UGS classification	Description
1	*Community Growing Spaces*	"the areas of land where plants such as fruits and vegetables are grown, primarily for the demands of the growers themselves, rather than for commercial activities"
2	*Bowling Green*	"a specially prepared area intended for playing bowls"
3	*Cemetery*	"the areas of land associated with burial areas"
4	*Religious Grounds*	"the areas of land associated with churches and other places of worship"
5	*Golf Course*	"the areas of land that are specially prepared for playing golf".
6	*Other Sports Facility*	"areas of land that were used for sports and not specifically described by other categories".
7	*Play Space*	"a specially prepared area intended for children's play, usually linked to housing areas or parks and containing purpose-built equipment".

8	*Playing Field*	"large, flat areas of grass or specially designed surfaces, generally with marked pitches, used primarily for outdoor sports, i.e. football, rugby, cricket".
9	*Public Park or Garden*	"areas of land designed, constructed, managed and maintained as a public park or garden. These normally have a defined perimeter and free public access".
10	*Tennis Court*	"a specially prepared area intended for playing tennis".

Text mining was used to analyse the Twitter data. It is an analytical method used by researchers to extract meaningful information (He et al., 2013). Text mining processes include text pre-processing, text representation, and information extraction (Hu and Liu, 2012). Text pre-processing includes stop word removal, lower-casing and stemming. Stop words refer to the most common words in data sets, for example, English words such as "a", "the", "is", "are", etc. Stop word removal deletes all meaningless words in textual data. Transforming text to lower case removes all capitalization and stemming identifies the roots of words. Non-text characters are also removed including punctuation, hashtag, URLs, and numbers before conducting text analysis.

In the current study, Twitter data were cleaned in this way to allow topic grouping, sentiment assessment and to identify patterns of user opinions and perceptions. First, duplicated Tweets text from users who posted more than ten times within one day were removed. Tweets posted from bots, fake accounts, and users who posted a same Tweet text more than three times in the data were also removed. Second, only Tweets in English were selected and Tweets with fewer than three words were removed. Third, the Twitter data were cleaned as described above with punctuation, URLs, numbers and stop words removed. The data were converted to lower case and stemmed.

The cumulative effects of the preprocessing steps on the number of Tweets in 2019, 2020 and 2021 are shown in Table 5.2. Note that the geo-referenced locations of individual Tweets were used to examine the spatio-temporal patterns of UGS visitation. User profile information includes a static location, usually their home location, which is difficult to use for analysis of spatial dynamics. This is because a user might post several Tweets in different places which are not differentiated by this attribute. Geo-referenced Tweets are key to this kind of analysis.

Table 5.2 The numbers of Tweets after each step of the data cleaning.

Process	2019	2020	2021
All collected Geo-Tweets	296329	207412	145019
Tweets from bots, fake accounts removed	284229	192535	130155
Identical Tweets posted more than three times removed	283682	192361	129935
Tweets with fewer than three words removed	271752	182666	122938
Geo-Tweets from London Area	257050	170003	112969
Tweets within urban green space (UGS Tweets)	12286	8645	5955
Percentages of Tweets in UGS	4.78%	5.09%	5.27%

In order to investigate the impact of the pandemic on UGS activities, this study identified six main groups of UGS activities based on previous studies using questionnaire, observations and text-mining (Roberts, 2017; Lesser and Nienhuis, 2020; Sim and Miller, 2019; Sim et al., 2020). These were physical activity, viewing art, social interactions, leisure, picnics, and exploring nature. The related keywords for each type of activity are shown in Table 5.3 and Tweets were linked to a specific activity through these keywords. In this way a corpus of Tweets linked to UGS activity for each year was created. All of analyses were undertaken in R (Ihaka and Gentleman, 1996).

Table 5.3 The keywords used to filter the relevant activities.

Activity	Keywords
Art	“art” “museum” “gallery” “photo” “music” “paint” “show” “dance” “sing” “photographer”
Leisure	“view” “see” “watch” “overlook” “enjoy” “look” ‘’listen” “hear”
Nature	“flower” “tree” “bird” “blossom” “sky” “weather” “animals” “zoo” “goose” “swan”
Physical	“walk” “run” “bike” “football” “jog” “swim” “sport” “marathon” “yoga” “skate”
Picnic	“lunch” “snack” “sandwich” “food” “coffee” “yum” “drinking” “beer” “cake” “brunch”
Social	“party” “festival” “friend” “family” “mom” “dad” “wedding” “father” “mother” “meet”

An initial exploratory data analysis was undertaken before investigating UGS activities in more detail to understand user behavior when visiting UGS areas. Differences between years the time periods (2019-2020, 2020-2021, 2019-2021) were explored through paired sample t-tests to investigate whether the impact of COVID-19 restrictions and lockdown rules resulted in differences in changes UGS visitation patterns. The same approach was used to compare differences across UGS types and UGS activities. Finally, the daily and hourly patterns of UGS visitation were investigated to detect whether COVID-19 restrictions influenced the temporal patterns when people visited UGS.

5.4 Results

5.4.1 Changes in UGS visitation during pandemic

Table 5.4 shows the number and percentage of all Tweets in London compared to those in UGS. The total number of Tweets in London continuously decreased from 2019 to 2021, the same trend was found in the number of UGS Tweets. However, the proportion of UGS Tweets slightly increased from 2019 to 2021. Examining the Tweets located in different types of UGS can help to understand UGS use. Table 5.4 shows the changes in the number of Tweets located in each kind of UGS. As larger UGSs are likely to contain more Tweets than smaller ones, Tweet density (Tweets per hectare) is included. Overall, 80% of all Tweets were posted in *Public Park or Garden*, *Playing Fields* in 2020 and a distinct trend of increased visitation in park types related to physical activities was found during lockdown. *Public Park or Garden* was the most frequently visited type among all UGS. *Playing Fields* showed an increasing trend from 2019 to 2021, indicating increased and sustained demands for this type of urban green space over the lockdown periods. The percentages of Tweets in *Community Growing Spaces*, *Bowling Green* and *Tennis Court* were less than 1.00% in all years. However, downward trends were found in *Cemetery*, *Other Sports Facilities*, *Play Space*, *Religious Grounds* and *Tennis Court* from 2019 to 2020, with *Other sports facilities*, *Religious Grounds* and *Tennis Court* increasing in 2021.

However, the results in this table indicates that simply using the number of Tweets in an area is not an appropriate measure to estimate the visitation of UGS and that the number of Tweets by area is universal. By measuring Tweet densities, the results suggests that the frequency of UGS visitation are affected by factors including but not limited to UGS size and the types of UGS.

Table 5.4 Tweets number and Tweets density in different types of UGS.

	UGS class	Area (ha)	Number of Tweets 2019		2020		2021		Number of Tweets / ha 2019	2020	2021
1	Community Growing Spaces	966	60	0.46%	53	0.59%	22	0.36%	0.06	0.05	0.02

2	Bowling Green	95	93	0.71%	79	0.88%	54	0.88%	0.98	0.84	0.57
3	Cemetery	1251	276	2.11%	148	1.65%	96	1.56%	0.22	0.12	0.08
4	Golf Course	4516	277	2.12%	195	2.17%	293	4.77%	0.06	0.04	0.06
5	Other Sports Facility	2752	806	6.17%	500	5.57%	605	9.84%	0.29	0.18	0.22
6	Play Space	289	191	1.46%	103	1.15%	29	0.47%	0.66	0.36	0.10
7	Playing Field	3263	840	6.43%	682	7.60%	523	8.51%	0.26	0.21	0.16
8	Public Park Or Garden	11776	9853	75.45%	7046	78.48%	4355	70.82%	0.84	0.60	0.37
9	Religious Grounds	283	546	4.18%	131	1.46%	140	2.28%	1.93	0.46	0.49
10	Tennis Court	190	117	0.90%	41	0.46%	32	0.52%	0.62	0.22	0.17

Tweet densities show that *Bowling Greens* had the highest Tweet density: 0.98 in 2019, 0.84 in 2020, and 0.57 in 2021, but with the lowest number of Tweets (93) and the smallest size (95ha). *Public Park and Garden*, the UGS with the largest areas, had lower densities of 0.84 (2019), 0.60 (2020), and 0.37 (2021). The lowest density was found in *Golf Course:* 0.06 in 2019 and 0.04 in 2020 and 0.06 which was the second-lowest in 2021, but with second largest area (4516ha).

Figure 5.3 shows the spatial distributions of UGS visitation in 2019, 2020, and 2021. In 2019, high levels of UGS visitation were mainly concentrated in LCBD and south central London, with lower visitation levels located in the periphery. However, this status appeared to change slightly during lockdown, with UGS visitation levels appearing to disperse to some degree especially around the LCBD in 2020 and 2021. However, it is still unclear to what extent UGS visitation levels have changed over these three years. This limitation arises from our representation of visitation levels as the percentage of tweets within each park in comparison to the total number of UGS-related tweets for that year. This approach does not capture the absolute changes in UGS visitation levels across the three years, and this limitation has also been identified in subsequent figures, such as Figure 5.4. Future research requires a deeper exploration of how UGS visitation levels have changed in different regions of London.

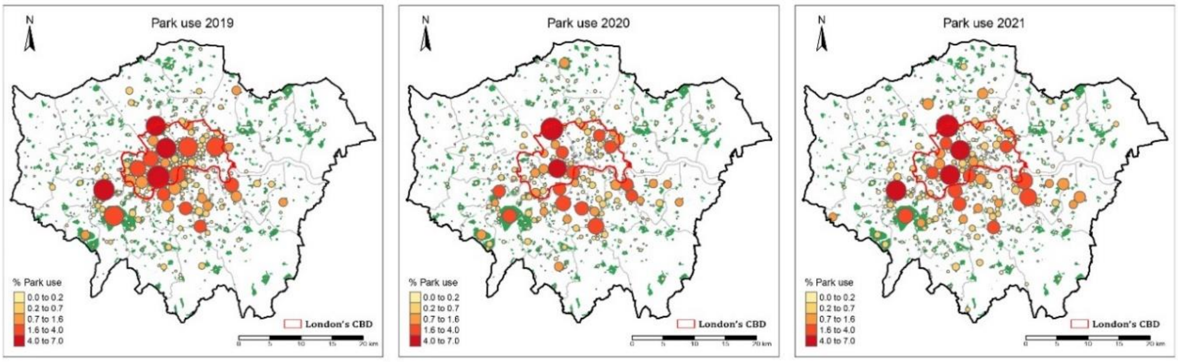

Figure 5.3 Spatial distributions of UGS use in 2019, 2020 and 2021, as a percentage of all Tweets.

Table 5.5 The changes in UGS visitation across the three years.

	Paired Years	Mean (London)	Sig (p)	Mean (LCBD)	Sig (p)	Mean (Non-LCBD)	Sig (p)
Pair 1	From 2019 to 2020	- 0.32	0.023 *	-0.86	0.002 **	+0.04	0.736 ns
Pair 2	From 2020 to 2021	+ 0.29	0.008 **	+0.23	0.074 ns	+0.23	0.046 *
Pair 3	From 2019 to 2021	- 0.03	0.837 ns	-0.63	0.019 *	+0.26	0.010 *

Significant differences between the every two groups were identified by the paired samples test and marked by * at $p < 0.05$, ** at $p < 0.01$, and ns for no significance.

To determine whether the observed changes in UGS use levels were statistically significant, a paired sample *t*-test was conducted to investigate the changes in UGS visitation levels among the three time periods. Table 5.5 shows the differences between each time period in all London, LCBD and Non-LCBD. Overall, in the whole study area (London), significant differences were found in *Pair 1* (from 2019 to 2020) and *Pair 2* (2020 to 2021), and no significant difference was observed in *Pair* 3 (2019 to 2021). The change in the mean value of UGS use levels in *Pair 1* (-0.32, $p < 0.05$) shows a significant decrease in 2020 compared to 2019. In contrast, *Pair 2* (-0.29, $p < 0.01$) shows a significant increase in 2021 compared to 2020, while in *Pair 3* no significant difference was observed between 2019 and 2021, indicating that the UGS visitation levels might not have been significantly different before and after the lockdown periods. Similar trends were found in the LCBD. Significant differences were found in *Pair 1* and *Pair 3* with changes in UGS visitation levels in LCBD significantly decreased from 2019 to 2020 (-0.86, $p < 0.01$) and to 2021 (-0.63, $p < 0.05$). This indicates that the restriction measures such as working from home, closure of shops, etc. may have significantly influenced the levels of UGS visitation in the LCBD, even though there was an increase from 2020 to 2021. Opposite trends were found in the non-LCBD. Increasing trends were found across all 3 periods with significant differences in *Pair 2* (0.23, $p < 0.05$) and *Pair 3* (0.26, $p < 0.05$), indicating that UGS visitation levels might continuously increase from 2019 to

2020 and to 2021 in the non-LCBD. The results in Table 5.5 show that the UGS visitation levels were likely to be changed from 2019 to 2020, and to 2021.

It should be noted that the results presented in this table exhibit limitations, despite the indications that UGS visitation levels may have fluctuated across the three years within and outside the LCBD areas, as well as throughout the entirety of London. For example, our analysis solely focused on UGS-related Tweets, without taking into account the overall volume of Tweets, including those outside of UGS areas. Consequently, the findings presented in this table only illuminate potential trends in UGS visitation levels. Future studies should aim to incorporate more specific factors when investigating the use of UGS.

To further explore the changes in spatial patterns of UGS visitation, the spatial distributions of UGS visitation were investigated. These are shown in Figure 5.4 Figs 5.4a, 5.4b, and 5.4c show the spatial distributions of *increasing* trends in the levels of UGS visitation between each pair of years, and *decreasing* trends are shown in Figures 5.4d, 5.4e and 5.4f.

From 2019 to 2020, Figure 5.4a suggests an increasing trend in UGS visitation outside of LCBD; in the southern part of London and along the northern LCBD boundary. Figure 5.4d shows that UGS use *within* the LCBD broadly decreased, suggesting that UGS use moved away from the centre of the city during the first lockdown in 2020. From 2020 to 2021, Figure 5.4b suggests an increase in UGS use within the LCBD, indicating that UGS visitation levels may have started to return to pre-pandemic levels. This is supported by a decreasing trend (Figure 5.4e) that dominates the changes in UGS visitation outside of LCBD. Finally, Figures 5.4c, and 5.4f compare 2019 to 2021 and show that although increases in UGS use (Figure 4.4c) were widely distributed within and without the LCBD, there was a concentrated decline within the LCBD (Figure 5.4f). This suggests that overall UGS visitation levels may have increased throughout London, but the central areas were more likely to see a decrease.

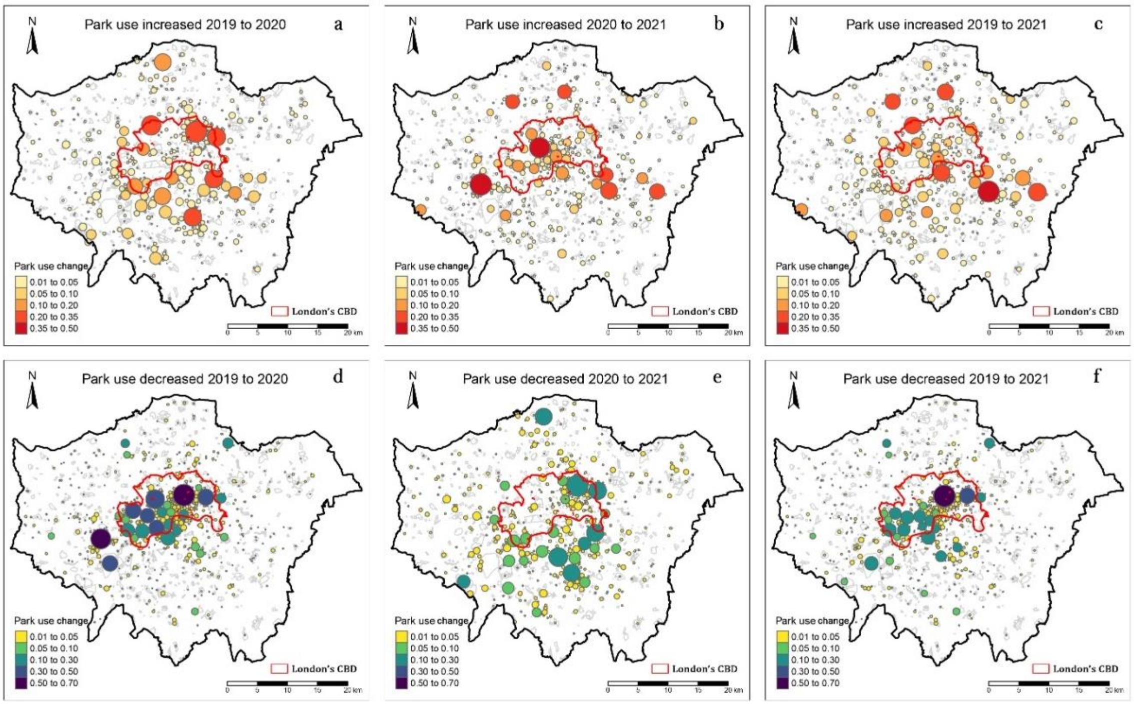

Figure 5.4 The spatial changes of UGS use from 2019 to 2020 and 2020 to 2021 in London.

5.4.2 Changes in the temporal patterns of UGS visitation

To investigate the temporal distribution of UGS users, the Tweets were grouped to daily and weekly patterns (Figure 5.5). There are obvious differences in the 3 periods. Specifically, the number of UGS Tweets on weekends was much higher than weekdays in the pre-pandemic period (2019). However, there is no significant difference in the number of Tweets on weekdays and weekends in 2020, indicating that UGS areas were more likely to experience increased visitation during the pandemic period compared to the pre-pandemic period. It should be noted that this figure solely calculated the number of UGS-related tweets in daily and weekly patterns over the three years, without taking into account the volume of whole Tweets. Therefore, further analysis is needed to substantiate the conclusion that UGS visitation levels might have increased on weekdays compared to the UGS visitation levels in 2019.

In 2021, the number of UGS Tweets on Sunday was higher than other days. When the hourly patterns are examined, there are very few Tweets between 22:00 (10pm) and 05:00 (5am), as might be expected across the 3 periods. What is evident is the increased morning and afternoon usage in 2020, especially on weekdays as people either undertook their permitted daily exercise or later in the year used the UGS to socialise, when the weather was generally warm and dry during the spring of 2020 in the UK (time and date, 2021). This may show the result of restrictions on working from home, giving people more flexibility in arranging their daily lives. In 2021, with the beginnings of a return to normal life, the daily UGS visitation pattern was similar to that of 2019, indicating that the UGS in London became important recreational destinations for citizens especially during the pandemic period and people prefer to visit parks and green spaces in the afternoon and evening.

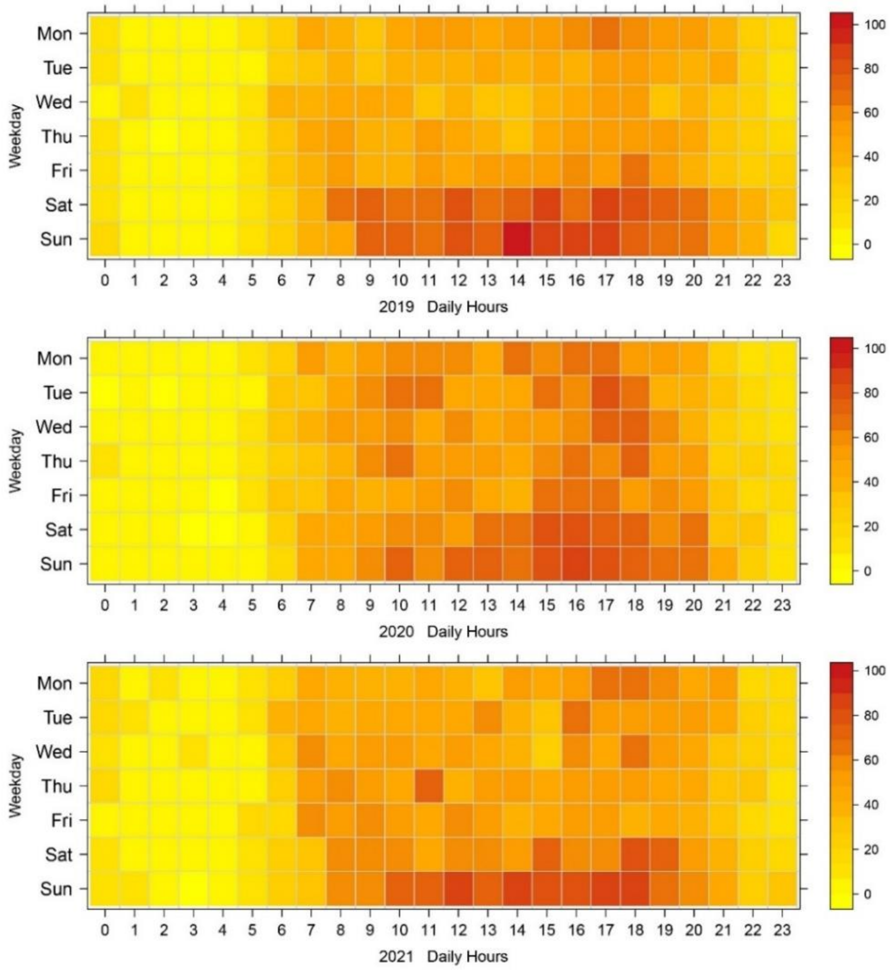

Figure 5.5 Daily pattern and hourly pattern of counts of UGS Tweets in 2019, 2020 and 2021.

5.4.3 Changes in UGS activities during lockdown

Table 5.6 shows the percentages of Tweets in relation to each kind of UGS activity relative to all Tweets in each year. Overall, the total percentages of UGS activity Tweets increased from 80.07% to 84.26% and 89.48% in 2019, 2020 and 2021, respectively, indicating that UGS activities became more frequent during and after the lockdown period.

Specifically, the percentages of *Art Activity* Tweets were the highest among all types of UGS activity across three years, indicating that users were mainly taking photos, visiting galleries, museums, or painting when they visited UGS. The trajectory of *Art Activity*

Tweets (decreased from 2019 to 2020, increased in 2021) may be as a result of restriction measures, for example, outdoor music events were cancelled (Onderdijk et al., 2021) and people were only able to visit UGS once a day during the lockdown period.

The second most participated UGS activity was *Social Activity* in 2019 (13.08%), indicating that people may have met their friends or family members when visiting UGS pre lockdown. This showed an evident decline from 2019 to 2020 and 2021 (9.28% and 6.97%, respectively), indicating that the restriction rules such as social distancing during lockdown may have exerted impact on *Social Activity* in UGS areas.

Leisure Activity accounted for 8.11% in 2019, then increased slightly to 8.95% in 2020 and increased more greatly to 16.59% in 2021, indicating that leisure activities were becoming more frequent after the lockdown. The percentages of *Nature activity* increased from 2019 to 2021, suggesting that users may be more interested in observing nature during and after the lockdown.

Physical Activity was the third most frequent UGS activity in 2019 (12.75%) and increased during the lockdown period. It accounted for 18.81% in 2020, indicating that users preferred running, walking, and other physical exercise during lockdown over other activities. The percentage of *Physical Activity* decreased to 10.16% in 2021, indicating that *Physical Activity* may have played a very important role in people's daily life especially during a pandemic period.

The percentages of *Picnic Activity* was the lowest across all study periods, indicating that a small part of UGS users were having picnic when they visited UGS before, during and after the first lockdown period.

Table 5.6 The percentages of activity related Tweets to all Tweets in each year.

Activity	2019	2020	2021
Art	30.49%	28.66%	38.38%
Leisure	8.11%	8.95%	16.59%
Nature	9.65%	12.16%	13.27%
Physical	12.75%	18.81%	10.16%
Picnic	5.99%	6.40%	4.11%
Social	13.08%	9.28%	6.97%
Total	80.07%	84.26%	89.48%

5.5 Discussion

5.5.1 Discussion of results

UGS is widely recognized as an essential component of urban areas, as it provides considerable public health and environmental benefits, especially during difficult times such as the COVID-19 lockdown periods (Ugolini et al., 2020). To our knowledge, *full* historical Twitter data sets were used for the first time in this study to examine the changes in UGS visitation before, during and after the first lockdown in London. This study explored the potential impacts of COVID-19 on the usage of UGS using Twitter data, and the results demonstrate that this data source could be a reliable proxy for assessing UGS visitation. The findings of this analysis indicate the changes in types of UGS used, the types of activities undertaken and their spatio-temporal patterns in London. This study addressed the following research questions:

RQ1. What types of UGS were more frequently visited and how did this change?

Public Park or Garden and *Playing Field* were more frequently visited, indicating that the demand for visiting these types of UGS increased during the pandemic period. Other studies have also claimed that the importance of UGS significantly increased during the pandemic (da Schio et al., 2021). For example, Ugolini et al. (2020) found that a large proportion of urban residents missed UGS and would readily travel long distances to access UGS during the pandemic period for basic needs such as fresh air and scenery. However, restriction measures on closing *Cemetery, Other Sports Facilities, Play Space, Religious Grounds,* and *Tennis Court* and social distancing may have resulted in their usage during lockdown in 2020. The closure of these types of UGS may have resulted in people changing their UGS destinations and activities, which may increase the possibility of physical and psychological disorders of users (Erdönmez and Atmiş, 2021). In 2021, the percentages of Tweets in *Other Sports Facilities, Religious Grounds,* and *Tennis Court* increased, indicating that reopening these UGSs led to a higher frequency of UGS visitation in the period of the post lockdown. Other studies have found fewer psychological disorders in areas where the use of UGS was not prohibited (Freeman and Eykelbosh, 2020), indicating the potential importance of UGS remaining accessible during a health crisis like the COVID-19 pandemic. Future planning could be encouraged to create more spaces for individual activities such as enjoying nature and walking or running freely in UGS (Erdönmez and Atmiş, 2021).

Considering the Tweet number and Tweet density in Table 4 indicates that simply using the number of Tweets in an area is not an appropriate measure to estimate the visitation of UGS and that the number of Tweets by area is universal. By measuring Tweet densities this study found that the frequency of UGS visitation are affected by factors including but not limited to UGS size and the types of UGS. This requires further exploration.

RQ2. In which area(s) did UGS visitation increase or decrease?

Opposite trends were found in the UGS visitation levels in LCBD and non-LCBD areas during the lockdown period. The visitation levels in LCBD decreased compared with pre-lockdown. The lockdown rules resulted in reduced UGS visitation levels within the LCBD area, but the opposite trend was found in non-LCBD areas. The explanation for this was likely to be the restriction of working from home and the closure of all non-essential shops, resulting a decline in the number of workers and customers in LCBD. In a related study using surveys to investigate the changes in spatio-temporal patterns of UGS use over lockdown, Korpilo et al. (2021) in Helsinki, Finland found that usage and recreational activity in the central part of Helsinki decreased during the pandemic. In a study in Oslo, Norway (Venter et al., 2020) found that physical activities such as cycling and walking decreased in the city centre while they increased in the urban periphery. In this study, an increasing trend was found in the UGS visitation in non-LCBD areas. It is likely that people who work from home were be able to spend more time in local parks and greenspaces (Ipsen et al., 2021). Further studies could be conducted to investigate the UGS visitation of other districts outside of LCBD and non-LCBD areas.

RQ3. How did UGS activities change, if at all?

This study showed that *Art Activity* accounted for the largest proportion of all activities across the three years, even though this decreased during lockdown, indicating that people were taking photos when they visited UGS (this study covered the Spring season). *Social Activity* decreased during lockdown, indicating that the restriction measures may have resulted in decrease *Social Activity* because people were practicing social-distancing when they visited UGS. These activities are non-essential which are especially important in times of crisis (Ugolini et al., 2020), and demonstrates that UGSs are fundamentally different from other kinds of urban public spaces such as pubs, theatres, and non-essential stores. *Physical Activity* became more frequent during lockdown, as more UGS people took exercise during their visits. Other studies using survey methods to explore the impacts of COVID-19 have found that people who more frequently participated in physical exercise during lockdown had better physical and mental health than others (Geng et al., 2021; Lesser and Nienhuis, 2020), and that they found new experiences in natural settings (Korpilo et al., 2021).

RQ4. Did the timing of UGS visitation change?

Different temporal patterns of UGS visitation were found during the lockdown period compared with before and after the first lockdown periods. Specifically, UGS visitation increased in the daytime during the lockdown compared with the before and after the lockdown periods and the frequency of UGS visits on weekdays increased. These changes are likely to be due to different working patterns: in lockdown people worked at home, they had more autonomy and flexibility over their time, they did not have to commute to work and were able to choose when they visited local UGSs in the daytime. This is supported by other research using surveys which found that respondents preferred to visit UGS closer to their house during the lockdown (Ugolini et al., 2020) and that UGS visitation frequencies were greater in daylight hours than in pre-lockdown times (Venter et al., 2020). This may also be linked the increased ability to spend more time with family members, to support each other mentally and physically within the family group through doing some physical or leisure activities together, facilitated by this enhanced flexibility (Pierce et al., 2020).

5.5.2 Discussion of methods

This is the first study to examine the impact of COVID-19 related lockdowns on behaviours UGS over before, during and after the first lockdown in 2020. Previous studies have selected specific time intervals within a single year to make comparisons between lockdown and pre-lockdown periods, which may introduce problems caused by seasonal weather and climate patterns. For example, Lesser and Nienhuis (2020) selected time intervals from January to February in 2020 as the 'baseline' to detect the influence of COVID-19 which outbreaks from March to summer; da Schio et al. (2021) implemented a survey from April 25th and July 10th to investigate the changes in UGS visitation before, during, and after the 2020 COVID-19 lockdown in Belgium. In the study presented here, the same periods (spring season from March to June in 2019, 2020, and 2021) were selected across the three years, to reduce the impacts of seasonal UGS visitation variations.

This study has several limitations. First, although Twitter data provides much useful information for understanding spatio-temporal changes in UGS use, there is potential for several inherent biases in these datasets (Roberts, 2017; Steiger et al., 2015). For example, information such as age and education background is not available through Twitter metadata (Plunz et al., 2019). This means that the sample of the population used in this analysis is unknown. A further improvement to analyses of social media data users would be determine the demographics of people who visit urban green space areas. For example, it may be possible to infer user age and educational background by examining their profiles. Second, the study considered only periods before, during and after the first lockdown and some restriction measures were still in place during the 2021 period. This study could be extended to include other lockdown periods and to cover times when policies were relaxed, and lockdown restrictions were fully removed. Third, UGS visitation patterns will not only be affected by restriction policies during lockdown, but also by the characteristics of the UGS such as accessibility, size, quality and facilities (Cohen et al., 2010). A further extension would be to examine UGS visitation in relation to the specific characteristics of each. Fourth, people's opinions about UGS were not only directly captured.

Twitter data provide a proxy for UGS visitation, attitudes and preferences and Twitter users represent a small proportion of the actual UGS visitors, although the data reflects the increase in people visiting UGS during lockdown as found in other studies (Fagerholm et al., 2021; Ugolini et al., 2020). However, as yet no study has examined social media data over three-time periods to explore the spatio-temporal patterns of UGS use, and this study is one of the first to confirm behavioural shifts during the lockdown period.

There are a number of other methodological suggestions. The potential effects of geo-location errors in Twitter data could be overcome by using a buffer around UGS boundaries (Plunz et al., 2019) and more data could be collected over other periods to explore the impact of the full lockdown release on UGS visitation and the permanency of the changes observed in this study. Future work could also on a deeper understanding of preferences and attitudes around UGS by examining social media content in greater detail, thereby expanding the analysis of the impact of the COVID-19 pandemic on UGS-related perceptions relative to park use, and other factors related to public space.

5.6 Conclusion

This study examined the impact of the COVID-19 pandemic on UGS visitation in London, UK. This analysis is novel due to its use of the full Twitter archive and is the first to examine a consistent time period before, during and after the lockdown. It shows how geo-referenced Tweets can be used to investigate the spatio-temporal patterns, including those related to UGS visitation.

The results identified clear differences in UGS visitation related to types of UGS and to UGS activity, with high levels of visitation in *Public Park and Garden* over whole study area, as public parks were one of the few places remaining open during lockdown. This suggests that the enhancement of public parks and gardens provision could be considered in future urban planning, but this does not mean other types of UGS are not important. An increasing trend of physical activity was found suggesting that the demands for sport facilities during a time of pandemic increased, and that people preferred to green space settings to do this. More sport facilities could be included within large parks along with flower and wildlife gardens and greater flexibility in pandemic closure policies could be considered. The provision of a variety of UGS services should be re-considered in response to the change in behaviors and demands that have arisen from the COVID-19 pandemic, as identified in this study and others. Decreasing trend in UGS visitation were found in the London central business district (LCBD), balanced by increases outside of this area, as result of people working from home. They visited UGS at different times of the day on weekdays during the pandemic due to spending more time with their family and being able to be more flexible in their UGS visitation. This suggests that small, highly localised pocket parks and gardens could enhance people’s lives and should be considered by policy to support dynamic and flexible UGS access as people work at home more (Venter et al., 2020).

This study demonstrates the value and potential of analysis of social media data to understand UGS usage and dynamics and to inform UGS managers and policy-makers, both in normal times and during periods of lockdown. As the potential for situations increases with rapid changes in climate, socio-economic disruptions, expanding global populations etc., such an understanding of how people react could be crucial for planning responses to future crises.

References

Cameron-Blake, E., Tatlow, H., Wood, A., Hale, T., Kira, B., Petherick, A. and Phillips, T. 2020. Variation in the response to COVID-19 across the four nations of the United Kingdom. *Blavatnik School of Government Working Paper.*

Chiesura, A. 2004. The role of urban parks for the sustainable city. *Landscape and urban planning.* ***68***(1), pp.129-138.

Cohen, D.A., Marsh, T., Williamson, S., Derose, K.P., Martinez, H., Setodji, C. and McKenzie, T.L. 2010. Parks and physical activity: why are some parks used more than others? *Preventive medicine.* ***50***, pp.S9-S12.

Cui, N., Malleson, N., Houlden, V. and Comber, A. 2021. Using VGI and Social Media Data to Understand Urban Green Space: A Narrative Literature Review. *ISPRS International Journal of Geo-Information.* ***10***(7), p425.

da Schio, N., Phillips, A., Fransen, K., Wolff, M., Haase, D., Ostoić, S.K., Živojinović, I., Vuletić, D., Derks, J. and Davies, C. 2021. The impact of the COVID-19 pandemic on the use of and attitudes towards urban forests and green spaces: Exploring the instigators of change in Belgium. *Urban Forestry & Urban Greening.* ***65***, p127305.

Dashraath, P., Wong, J.L.J., Lim, M.X.K., Lim, L.M., Li, S., Biswas, A., Choolani, M., Mattar, C. and Su, L.L. 2020. Coronavirus disease 2019 (COVID-19) pandemic and pregnancy. *American journal of obstetrics gynecology.* ***222***(6), pp.521-531.

Erdönmez, C. and Atmiş, E. 2021. The impact of the Covid-19 pandemic on green space use in Turkey: Is closing green spaces for use a solution? *Urban Forestry & Urban Greening.* ***64***, p127295.

Fagerholm, N., Eilola, S. and Arki, V. 2021. Outdoor recreation and nature's contribution to well-being in a pandemic situation-Case Turku, Finland. *Urban Forestry & Urban Greening.* ***64***, p127257.

Freeman, S. and Eykelbosh, A. 2020. COVID-19 and outdoor safety: Considerations for use of outdoor recreational spaces. *National Collaborating Centre for Environmental Health.* ***829***, pp.1-15

Gao, J., Zheng, P., Jia, Y., Chen, H., Mao, Y., Chen, S., Wang, Y., Fu, H. and Dai, J. 2020. Mental health problems and social media exposure during COVID-19 outbreak. *Plos one.* ***15***(4), pe0231924.

Geng, D., Innes, J., Wu, W. and Wang, G. 2021. Impacts of COVID-19 pandemic on urban park visitation: a global analysis. *Journal of forestry research.* ***32***(2), pp.553-567.

GiGL. 2019. *Key London Figuress – Greenspace Information for Greater London.* [Online]. [Accessed Oct 10]. Available from: https://www.gigl.org.uk/keyfigures/

GOV.UK. 2020a. *Communities Secretary welcomes response to his call for parks to open.* [Online]. Available from: https://www.gov.uk/government/news/communities-secretary-welcomes-response-to-his-call-for-parks-to-open

GOV.UK. 2020b. *PM address to the nation on coronavirus: 23 March 2020.* [Online]. [Accessed July 5]. Available from: https://archive.vn/VnrbU

Greater London Authority. 2021. *COVID-19 Restrictions Timeseries.* [Online]. Available from: https://data.london.gov.uk/dataset/covid-19-restrictions-timeseries

He, W., Zha, S. and Li, L. 2013. Social media competitive analysis and text mining: A case study in the pizza industry. *International journal of information management.* ***33***(3), pp.464-472.

Houlden, V., de Albuquerque, J.P., Weich, S. and Jarvis, S. 2019. A spatial analysis of proximate greenspace and mental wellbeing in London. *Applied Geography.* ***109***, p102036.

Hu, X. and Liu, H. 2012. Text analytics in social media. *Aggarwal, C., Zhai, C. (eds) Mining Text Data.* Springer, Boston, MA, pp.385-414.

Huang, Q. and Wong, D.W.S. 2016. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science.* ***30***(9), pp.1873-1898.

Huang, Y., Li, Y. and Shan, J. 2018. Spatial-temporal event detection from geo-tagged tweets. *ISPRS International Journal of Geo-Information.* ***7***(4), p150.

Ihaka, R. and Gentleman, R. 1996. R: a language for data analysis and graphics. *Journal of computational and graphical statistics.* ***5***(3), pp.299-314.

Ipsen, C., van Veldhoven, M., Kirchner, K. and Hansen, J.P. 2021. Six key advantages and disadvantages of working from home in Europe during COVID-19. *International Journal of Environmental Research and Public Health.* ***18***(4), p1826.

Kabisch, N., Qureshi, S. and Haase, D. 2015. Human–environment interactions in urban green spaces—A systematic review of contemporary issues and prospects for future research. *Environmental Impact Assessment Review.* ***50***, pp.25-34.

Korpilo, S., Kajosaari, A., Rinne, T., Hasanzadeh, K., Raymond, C.M. and Kyttä, M. 2021. Coping with crisis: green space use in Helsinki before and during COVID-19. *Frontiers in Sustainable Cities. 3*, p713977.

Lesser, I.A. and Nienhuis, C.P. 2020. The impact of COVID-19 on physical activity behavior and well-being of Canadians. *International journal of environmental research and public health.* ***17***(11), p3899.

Lopez, B., Kennedy, C., Field, C. and McPhearson, T. 2021. Who benefits from urban green spaces during times of crisis? Perception and use of urban green spaces in New York City during the COVID-19 pandemic. *Urban forestry & urban greening.* ***65***, p127354.

Mesa-Arango, R., Zhan, X., Ukkusuri, S.V. and Mitra, A. 2016. Direct transportation economic impacts of highway networks disruptions using public data from the United States. *Journal of Transportation Safety & Security*. ***8***(1), pp.36-55.

Niemelä, J. 1999. Ecology and urban planning. *Biodiversity & Conservation.* ***8***(1), pp.119-131.

Onderdijk, K.E., Acar, F. and Van Dyck, E. 2021. Impact of lockdown measures on joint music making: playing online and physically together. *Frontiers in psychology*. ***12***, p1364.

Ordnance Survey. 2021a. *OS Open Greenspace.* [Online]. [Accessed 23.02]. Available from: https://www.ordnancesurvey.co.uk/business-government/products/open-map-greenspace

Ordnance Survey. 2021b. *OS Open Greenspace support: Technical information.* [Online]. [Accessed 26.02]. Available from: https://www.ordnancesurvey.co.uk/business-government/tools-support/open-map-greenspace-support#technicalInformation

Pierce, M., Hope, H., Ford, T., Hatch, S., Hotopf, M., John, A., Kontopantelis, E., Webb, R., Wessely, S. and McManus, S. 2020. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. *The Lancet Psychiatry.* ***7***(10), pp.883-892.

Plunz, R.A., Zhou, Y., Vintimilla, M.I.C., Mckeown, K., Yu, T., Uguccioni, L. and Sutto, M.P. 2019. Twitter sentiment in New York City parks as measure of well-being. *Landscape and Urban Planning.* ***189***, pp.235-246.

Poortinga, W., Bird, N., Hallingberg, B., Phillips, R. and Williams, D. 2021. The role of perceived public and private green space in subjective health and wellbeing during and after the first peak of the COVID-19 outbreak. *Landscape and Urban Planning.* ***211***, p104092.

Roberts, H.V. 2017. Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography.* ***81***, pp.13-20.

Salloum, S.A., Al-Emran, M., Monem, A.A. and Shaalan, K. 2017. A survey of text mining in social media: facebook and twitter perspectives. *Advances in Science, Technology and Engineering Systems Journal.* ***2***(1), pp.127-133.

Shoari, N., Ezzati, M., Baumgartner, J., Malacarne, D. and Fecht, D. 2020. Accessibility and allocation of public parks and gardens in England and Wales: A COVID-19 social distancing perspective. *PloS one.* ***15***(10), pe0241102.

Sim, J. and Miller, P. 2019. Understanding an urban park through big data. *International journal of environmental research and public health.* ***16***(20), p3816.

Sim, J., Miller, P. and Swarup, S. 2020. Tweeting the High Line Life: A Social Media Lens on Urban Green Spaces. *Sustainability.* ***12***(21), p8895.

Steiger, E., Westerholt, R., Resch, B. and Zipf, A. 2015. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems.* ***54***, pp.255-265.

Sulis, P., Manley, E., Zhong, C. and Batty, M. 2018. Using mobility data as proxy for measuring urban vitality. *Journal of Spatial Information Science.* ***16***, pp.137-162.

Theodorou, A., Panno, A., Carrus, G., Carbone, G.A., Massullo, C. and Imperatori, C. 2021. Stay home, stay safe, stay green: The role of gardening activities on mental health during the Covid-19 home confinement. *Urban Forestry & Urban Greening.* ***61***, p127091.

time and date. 2021. *High & Low Weather Summary for Past Weather.* [Online]. [Accessed August 12]. Available from: https://www.timeanddate.com/

Ugolini, F., Massetti, L., Calaza-Martínez, P., Cariñanos, P., Dobbs, C., Ostoić, S.K., Marin, A.M., Pearlmutter, D., Saaroni, H. and Šaulienė, I. 2020. Effects of the COVID-19 pandemic on the use and perceptions of urban green space: An international exploratory study. *Urban forestry & urban greening.* ***56***, p126888.

Venter, Z.S., Barton, D.N., Gundersen, V., Figari, H. and Nowell, M. 2020. Urban nature in a time of crisis: recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. *Environmental research letters.* ***15***(10), p104075.

Zhu, J. and Xu, C. 2021. Sina microblog sentiment in Beijing city parks as measure of demand for urban green space during the COVID-19. *Urban Forestry & Urban Greening.* ***58***, p126913.

Zu, Z.Y., Jiang, M.D., Xu, P.P., Chen, W., Ni, Q.Q., Lu, G.M. and Zhang, L.J. 2020. Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology.* ***296***(2), pp.E15-E25.

Chapter 6 aims to use structural topic model (STM) to detect the dynamics of UGS-related topics pre-, during- and after the COVID-19 outbreaks in London. The results show that there were seven main topics categories under discussion in UGS during the COVID-19 pandemic period. Specifically, an increase in Tweets proportions was found for the topics Nature engagement and Dog walking, indicating that the popularity of these activities increased during the COVID-19 pandemic period. The topic Social events showed a decline in topic proportion, which might be related to restriction measures such as social distancing and mask wearing. This study not only identified the main types of topics in UGS during the COVID-19 pandemic period but also revealed the dynamics of spatial-temporal patterns of the topics over three years, which provide suggestions and guidance for future urban policies, especially in times of crisis. A number of limitations should be acknowledged in this analysis. First, only the COVID-19 period was used as the covariates in STM model. Future studies can incorporate the UGS user characteristics, such as gender, age, and expertise into the model to investigate how different user groups interact with UGSs and how their experiences are related to different topics. By addressing these limitations, future studies can provide a more comprehensive understanding of the dynamics of UGS-related topics and how they are influenced by various factors, especially during the time crisis.

Chapter 6 Urban green space topics based on structural topic modelling during the COVID-19 pandemic

Abstract

Popular social media platforms such as Twitter can provide information about important events (e.g., activities, accidents, crises), thereby providing new knowledge about human dynamics in urban green space (UGS) use. Although recent developments in approaches such as text-based word frequency analysis and spatial analysis offer new perspectives on UGS, they are often stationary and non-continuous in nature. This limits their ability to capture the complexity and diversity of UGS use, especially during the time of crises such as COVID-19 pandemic. This study uses structural topic model (STM) to detect the dynamics of UGS-related topics pre-, during- and after the COVID-19 outbreaks in the UK. Geo-referenced Tweets were used to generate the topics, the content covariate of the COVID-19 was added into this step which enable researchers to deep understand how UGS topics were influenced by the pandemics. The spatial variation of probabilities for all topics were investigated to understand the dynamics of topical spatial patterns across

the whole of London. The results suggested that there were seven main topics categories under discussion in UGS during the COVID-19 pandemic period. Specifically, an increase in Tweets proportions was found for the topics *Nature engagement* and *Dog walking*, indicating that the popularity of these activities increased during the COVID-19 pandemic period. The topic *Social events* showed a decline in topic proportion, which might be related to restriction measures such as social distancing and mask wearing. This study not only identified the main types of topics in UGS during the COVID-19 pandemic period but also revealed the dynamics of spatial-temporal patterns of the topics over three years, which provide suggestions and guidance for future urban policies, especially in times of crisis.

6.1 Introduction

The use of UGS changed during the COVID-19 pandemics. Research has shown that the spatial-temporal characteristics of visits, visitor activities engagement, and attitudes towards UGS all changed (Cui et al., 2022; Da Schio et al., 2021; Erdönmez and Atmiş, 2021; Geng et al., 2021). Geng et al. (2021) analysed the impacts of COVID-19 and government restriction measures on park visitation at global, regional and national levels. They found that the demand from residents for UGS has increased since the outbreak began, and parks could be utilized during pandemics to increase the physical and mental health and social well-being of individuals. Cui et al. (2022) examined the impacts of COVID-19 lockdown restrictions on UGS visitation patterns in London. They found that the usage of UGS decreased in central London and increased in around city centre areas during lockdown, and UGS activities were positively associated with physical activities during the COVID-19 lockdown period. Da Schio et al. (2021) investigated the relationship between demographic characteristics in Belgium and actual green space use during the COVID-19 pandemic. Their findings show that citizens with higher education levels expressed an increase use of UGS during the pandemic. However, the dynamics of topics in relation to UGS during the COVID-19 period have not been studied and could provide useful insights to support future UGS planning and management.

Topic modelling is gaining increasing attention from scholars. It allows hidden information to be extracted from large volumes of textual data. Latent Dirichlet Allocation (LDA) is a commonly used topic modelling approach. It is a powerful method which can detect topics from textual data such as articles, newspapers, and social media posts such as Tweets (Jelodar et al., 2019; Blei, D.M. et al., 2003; Lansley and Longley, 2016). LDA is used to determine the probability of a given document (such as a Tweets) being a member of a given topic through a "bag-of-words" interpretation of its contents (Blei, D.M. et al., 2003). However, the themes in a corpus (i.e. a collection) of a document may evolve over time. As yet little research gas examined the temporal evolution of topics and their dynamics. Tracking such temporal changes can be reached by using structural topic model (STM), introduced by (Roberts, M.E. et al., 2014b), this model can capture the evolution of topics in a sequentially organized corpus of documents, thus this model evaluates the changes of topic proportions and word probability in a corpus of documents along a time series text data.

Various data sources have been used to explore UGS usage. Traditional methods, including questionnaires and on-line surveys, have been used to explore the impact of COVID-19 pandemic on UGS visitation (Ugolini et al., 2020; Erdönmez and Atmiş, 2021; Sim and Miller, 2019). However, these approaches have limitations related to low response numbers and a lack of spatial-temporal information and granularity (Cui et al., 2021). Twitter is a free social media (microblogging) platform which allows users to post messages of up to 280 characters in length, with additional information such as coordinates and time of posting based on user preference (Steiger et al., 2015). The high spatial and temporal resolution provides researchers with opportunities to analyse the spatial-temporal distributions of Tweets as well as their content. Recently, Twitter have also been used to investigate the usage of UGS. Roberts, H.V. (2017) used Twitter data to detect UGS activities by manually classifying Tweets and then determining activity-related Tweets. This was time-consuming and inefficient, even though this study demonstrated how Twitter data could be used in UGS use investigation. Cui et al. (2022) and Niță et al. (2021) used Twitter and Instagram data respectively to reveal how spatial-temporal UGS visitation patterns changed during the COVID-19 pandemic period, by using frequency counting and classical statistic methods such as paired sample t-tests. Other types of social media data such as Google's Community Mobility Reports have also been used to investigate the changes in urban park visitation during the COVID-19 pandemic period (Geng et al., 2021).

Previous studies have demonstrated that Twitter data could be effectively used to detect changes in UGS visitation pattern. However, these types of data have not yet been used to examine the spatial-temporal changes in topics expressed by UGS visitors, particularly during the COVID-19 pandemic period. This limits the ability of these datasets to capture the complexity and diversity of the impacts of the COVID-19 on UGS use. Recently, the spatial variation of STM has been used in previous studies in relation to other research fields. For example, Sachdeva et al. (2017) analysed the geographic variation of STM results by extracting geographic information from geo-referenced Tweets, to investigate the spatial-temporal patterns of air quality impacts after a wildfire event in in northern California. A more effective method for automatically detecting and monitoring events in UGS through the analysis of geo-referenced-Tweets can help scholars to understand what UGS visitor discuss and how these topics change over time and space. This in turn can inform policy and UGS management and help them better understand how UGS is used and public perceptions (Yao and Wang, 2020; Fu et al., 2018). This study uses geo-referenced Twitter data to extract topics from 3 coincident periods – before, during and after the COVID-19 pandemic – to examine changes in the topics of UGS Tweets.

The work addresses the following research questions: What topics and attitudes were expressed through Tweets before, during and after the COVID-19 pandemic? And how did the observed topics change over space and time? To answer these, structural topic modelling (STM) (Roberts, M.E. et al., 2014b; Roberts, M.E. et al., 2014a) is used to create and analyse the time evolution of topics in a large collection of UGS-related Tweets. The results can reveal trends in the frequency with which topics appear over time as well as relationships between covariates and topic prevalence or word use within a topic. We also investigate the changes in spatial distribution of each topic to see whether pandemic-related policies influenced the spatial variation of the identified topics.

The rest of this study is organised as follows: In the Background section 2, the literatures related to topic modelling and structural topic modelling (STM) are reviewed, and for investigating topic spatial changes over three years. The methodology section 3, describes the collection and prepossessing Twitter datasets, identifying topics by using STM, and investigating spatial patterns of topics. The results section 4 describes how the number of topics were determined, followed by a brief explanation of the meaning and underlying themes of each topic. Then the analysis of the evolution of each topic over time is described analysed by comparing time periods before, during and after the COVID-19 pandemic. Finally, the spatial-temporal changes of topics were investigated to understand the spatial variation of topics, before discussing the results of STM and limitations in relation to social media data. The Conclusion section 5 states the contribution of this research to a broader understanding of both the topics surrounding the COVID-19 pandemic and their evolution over space and time, especially in relation to UGS use.

6.2 Research background

Recently, dynamic topic modelling (DTM) has been used in detecting topics in geographical spaces. For example, Bogdanowicz and Guan (2022) used DTM to detect topics in relation to the COVID-19 pandemic and analysed the changes of notable topics over time. Twelve most popular topics were identified by using the Sequential Latent Dirichlet Allocation model, which allow researchers to investigate the growth and changes of topics over time. Du et al. (2020) improved LDA model to generate and track evolution trend of the topics, the results showed that the improved LDA model has a lower perplexity and higher coverage rate than LDA under the same conditions. However, these models have not yet been developed to detect spatial changes in topics. This section aims to summarize and compare the models that have been used to investigate the spatial-temporal changes of topics, and then discuss the research gap in topic modelling, particularly in UGS use.

6.2.1 Dynamic topic modelling using structural topic modelling

Building upon earlier approaches to topic modelling such as latent Dirichlet allocation (LDA) (Blei, D.M. et al., 2003) and correlated topic models (Blei, D. and Lafferty, 2006). Roberts, M.E. et al. (2014b) developed structural topic modelling (STM) that allows researchers to make greater use of observed structural variables that characterize documents. STM directly estimates the impacts of metadata (covariates such as posts dates, spatial distribution) on topic prevalence and allows analysts to model the trends and relationships between topics and documents (Kuhn, 2018). STM allows researchers to examine the prevalence of particular topics in a body of text and examine how it varies based on other factors of interest along the timeline. Based on the research questions as mentioned above, STM is therefore a better fit for the current research objective, as it allowed examination of how the content of users' Tweets may vary by time and geographic area (Sachdeva et al., 2017).

STM has been widely used in previous studies, for example, Chen et al. (2020) used STM to investigate the evaluation of topics in computers and education academic publications over 40 years. Sachdeva et al. (2017) used STM to detect the topics of Tweets in order to investigate the air quality impacts of wildfire events in the United States. The results suggested that social media can be a valuable tool for providing insight into the socio-psychological dimensions of fire and smoke and their impact on people residing in affected areas. These studies demonstrated that STM modelling enabled the creation of tools for investigating the topic prevalence and relevant trends in the research. However, they failed to detect the spatial patterns of topics even though the geo-referenced Tweets were utilised. More details on LDA and STM models are provided in later sections of this study.

6.2.2 Spatial-temporal trajectories of topics

The coordinates information and the time stamps in Tweets metadata allow researchers to identify the spatial patterns of different topics within a certain area, and further to track the changes of these topics over time – i.e. the spatial-temporal patterns of topics. Previous studies have used geo-referenced Tweets to identify the topics in geographical spaces. For example, geo-tagged Tweets from Inner London were used to detect the topics and their geographical location (Lansley and Longley, 2016). The results revealed the relationships between human behaviours and surrounding characteristics, topics and attitudes expressed through Tweets are found to vary substantially across Inner London. This study demonstrated that topic modelling could be usefully applied to short text data over urban areas but did not investigate topics dynamics or change over time or space aspects. Fu et al. (2018) used a LDA based model to detect the spatial-temporal patterns of different human activity topics within cities. Their study demonstrated that Twitter posts could be used to extract activities by applying LDA based topic modelling, but without any consideration of topic dynamic over time and space.

6.3 Methodology

6.3.1 Data collection and pre-processing

The Twitter datasets used for this study were downloaded via the Twitter academic research application programming interface (API). The API call selected geo-referenced Tweets located in London. They covered a three-month period (23rd March to 23rd Jun) for three consecutive years: 2019, 2020, and 2021. A bounding box was used on the area of Greater London to select Tweets. In addition, green space profile in Greater London was superimposed to extract Tweets that located within urban green space area, which was derived from the Open Greenspace layer from the Ordnance Survey (Ordnance Survey, 2021). The spatial distributions of Greater London and UGS are shown in Figure 6.1.

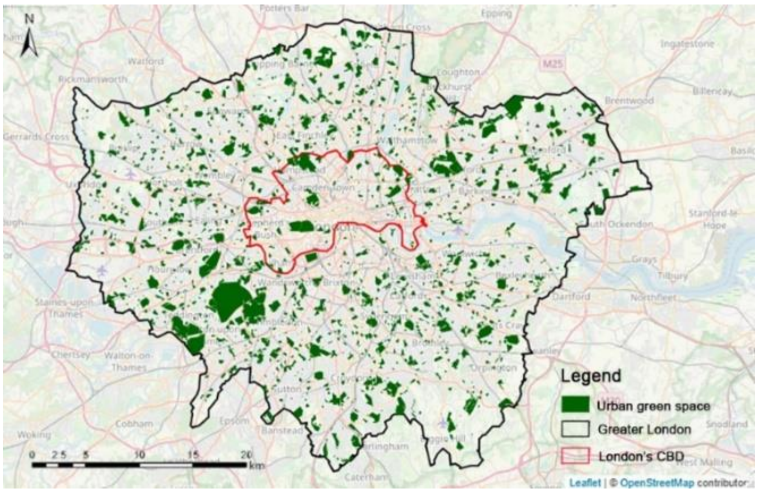

Figure 6.1 The distribution of open green space in Greater London (Cui et al., 2022).

Text mining was used to analyse the Twitter data. It is an analytical method used by researchers to extract meaningful information and knowledge from unstructured text data (He et al., 2013). The role of text mining has increased in the field of social media data analysis (Salloum et al., 2017). Text mining processes include text pre-processing, text representation, and information extraction (Hu and Liu, 2012). Text pre-processing includes stop word removal, lower-casing and stemming. Stop words refer to the most common words in data sets, for example, English words such as “a”, “the”, “is”, “are”, etc. Stop word removal deletes all meaningless words in textual data. Transforming text to lower case removes all capitalization and stemming identifies the roots of words. Non-text characters are also removed including punctuation, hashtag, URLs, and numbers before conducting text analysis.

In the current study, Tweets were cleaned in this way to allow topic grouping and sentiment assessment, to identify patterns of user opinions and perceptions. First, duplicated Tweets were removed, and Tweets posted from bots, fake accounts, and users who posted a same Tweet text more than three times in the data were removed. Second, only Tweets in English were selected and Tweets with fewer than three words were removed. Third, the Twitter data were cleaned as described above with punctuation, URLs, numbers and stop words removed. The data were converted to lower case and stemmed. The cumulative effects of the preprocessing steps on the number of Tweets can be find in (Cui et al., 2022). All the analyses were undertaken using R software (Ihaka and Gentleman, 1996).

6.3.2 Structural topic modelling (STM)

STM is a type of probabilistic topic modelling that extends the LDA framework (a common form of topic modelling). Topic modelling considers a single document as a combination of topics, and each topic as a distribution over words. This section briefly introduces the LDA model before discussing STM.

LDA is a generative probabilistic model which considers each document *d* as being generated from 1 to *D* based on word frequency from a set of documents as follows (Blei, D.M. et al., 2003)(Figure 6.2):

1. *N* represents the number of words which is a random variable drawn from a Poisson distribution, θ represents the topic proportion which is a random variable drawn from a Dirichlet(α) distribution.
2. For each word in the document (words being indexed by n), the topic of the word is *Z*, is a random variable drawn from a multinomial (θ) distribution.
3. For each topic (topics being indexed by *k*), the model parameter β represents word proportions within the topic.
4. The word themselves, *W* terms, are random variables drawn from another multinomial distribution defining $p(W_n|Z_n,\beta)$ terms.

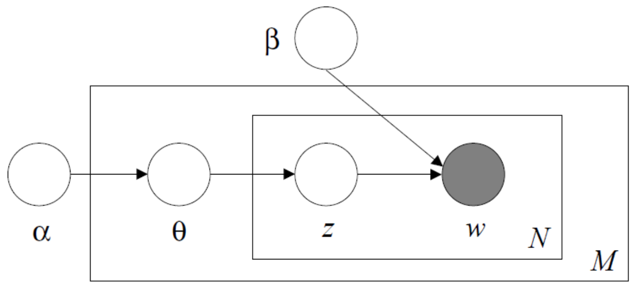

Figure 6.2 Graphical model representation of LDA. The boxes are "plates" representing replicates.

The outer plate represents documents, while the inner plate represents the repeated choice of topics and words within a document. (Blei, D.M. et al., 2003)

Both of LDA and STM are generative approaches, built on the assumptions that documents are comprised of a distribution of topics and that topics are made up of a semantically coherent distribution of words. LDA assumes that a parameter θ_d for each document *d* was assumed to be a random variable drawn from a Dirichlet(α) distribution, in addition, this distribution was common across all documents. In the progress of STM, a parameter θ_d for each document *d* is a random variable drawn from a Log-normal distribution that is based on document-level data (Roberts, M.E. et al., 2014b).

LDA also assumes that there were β terms, model parameters that represented word proportions within a topic. These also were common across the corpus. In STM, a multinomial logit model is used for word distributions where a word's prevalence is based on topic, document covariate data, and topic-covariate interactions (Roberts, M.E. et al., 2014b). Thus, STM allows researchers to calculate the correlations among topics. Topic prevalence within documents and word distribution within topics is defined by covariate data including metadata. In a word, STM performs two major tasks toward estimating the distributions of the document-topic θ_d and the topic-term β_k, representing topic proportion and word distribution within topics, respectively. A variational expectation maximization (VEM) method was applied to the estimation of parameters (Roberts, M.E. et al., 2016). Further technical details on structural topic modelling are provided in (Roberts, M.E. et al., 2014a) and (Roberts, M.E. et al., 2014b). In this analysis, the *stm* R package (Roberts, M.E. et al., 2014a) was used to conduct the STM topics.

An important point for STM analysis is the determination of the number of topics for social media data, topic modelling needs to specify the number of topics before generating the topics. Although there are numerous methods available to calculate the number of topics, there is currently no scientific consensus for determining the optimal number of topics within a defined model (Grimmer and Stewart, 2013; Rodriguez and Storer, 2020; Kuhn, 2018). In this study, the search *k* algorithm in the *stm* package was used to determine the optimal number of topics (Mimno and Lee, 2014). Diagnostic testing was conducted to examine the goodness of fit for a topic model with varying number of topics from 2 to 10, in 1 topic increments.

6.3.3 Dynamics in spatial patterns of topics

After computing the probability values for all Tweets, individual Tweets were then assigned to a particular topic according to the highest probability value among all seven probability values. Thus, each Tweet was assigned to its most probable topic. Then the geographical variation of each topic across three years over London was investigated by using spatial interpolation method. Spatial interpolation is defined as predicting the values of a primary variable at points within the same region of sampled locations (Gómez-Losada et al., 2019). There are various spatial interpolation methods (Li and Heap, 2011) of which inverse distance weighting (IDW) is one of the most popular or frequently used method, which has been developed for and applied in many disciplines (Li and Heap, 2014), such as environmental sciences analysis including air pollution and water quality, epidemiology, and agriculture analysis (Gu et al., 2021; Fatima et al., 2022; Pereira et al., 2022). However, previous studies rarely employed IDW to assess the spatial patterns of topics, particularly in UGSs. In the current study, IDW approach was used to explore the spatial variations of topics across the whole of London over all time periods.

IDW is a kind of deterministic interpolation method that creates a continuous surface of values based on point data, where the values at any given location are determined by the weighted average of nearby points. The weight assigned to each point is inversely proportional to the distance from that point to the location being estimated. The formula for IDW is as follows:

$$Z = \sum_{i=1}^{n} \frac{1}{(d_1)^p} Z_i \div \sum_{i=1}^{n} \frac{1}{(d_1)^p}$$

where Z_i is the given height value of the reference point; p is the index for determining the weight of the reference point (most often p = 2); n is the number of reference points located in the nearest neighbourhood; $(d_i)^p$ is the distance between points with known and unknown height values (Tomczak, 1998; Abdulmanov et al., 2021). The computational implementation of the IDW was performed using the *gstat* package (Gräler et al., 2016) from R software.

6.4 Results

6.4.1 STM results

Diagnostic testing was conducted to examine the goodness of fit for a topic model with varying number of topics from 2 to 10, in 1 topic increments. The relative goodness of fit for each number of topics were shown in Figure 6.3. The Held-out likelihood refers to the log probability of topics in the test set correctly replicating topics in the training set (Wallach et al., 2009). Similar to cross-validation, when some of the data is removed from estimation and then later used for validation, the held-out likelihood helps the user assess the model’s prediction performance. The lower bound refers to the lower bound of the marginal log likelihood. Residuals refers to the difference between expected (training set) and predicted (test set) topic predictions. Semantic coherence refers to the co-occurrence of words in each topic.

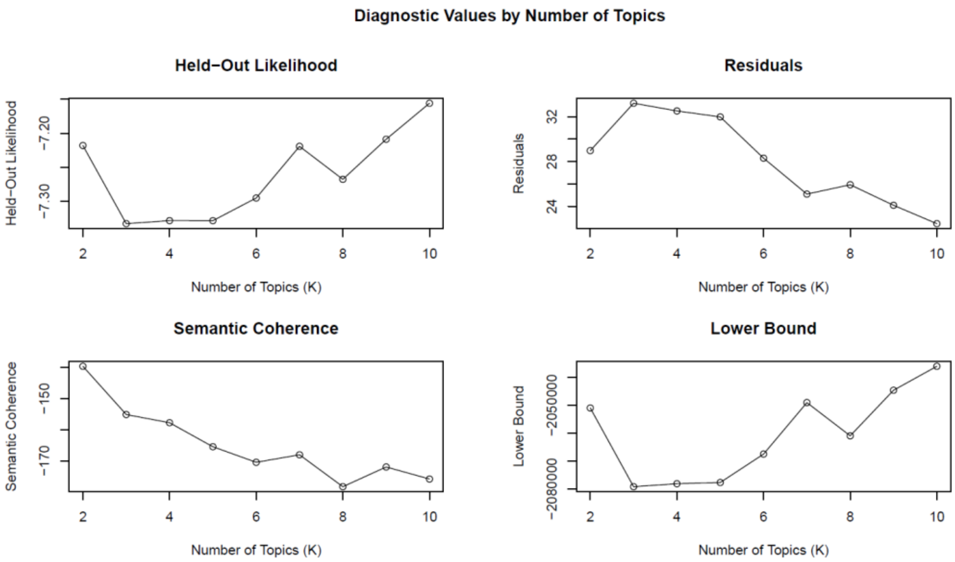

Figure 6.3 This figure illustrates the relative goodness of fit for each number of topics. Optimal results would demonstrate relatively low residuals, high semantic coherence, a maximized lower bound, and a high held-out likelihood.

In this figure, 7 topics generated relatively high Held-out likelihood and high Lower Bound, low residuals. As a result, 7 was selected as optimal number of topics for this dataset. This is fewer than some other studies – for example Sachdeva et al. (2017) determined that 20 topics was optimal – but is advantageous here because it makes it easier to label and adequately distinguish the different topics.

Figure 6.4 displays the proportions of each topic to all topics over all periods, along with the corresponding name of each topic. The name of each topic was manually labelled by the authors based on the highest probability topic words (Table 6.1). The most popular topic among all topics was Topic 6, named *Nature engagement*, followed by *Social events*, *Crowd events*, *Art and exhibition*, *Sport events*, *Drinking and leisure*, and *Dog walking.* In the current study we focus on three topics as these are the most strongly related to the impact of the COVID-19 pandemic on UGS use. The most popular topics, *Nature engagement* (Topic 6) and *Social events* (Topic 5), were selected as an illustration of the types of insight that might be gleaned from the method. In addition, the topic *Dog walking* (Topic 4), with the smallest proportion, was also selected to investigate the impact of specific policies related to park use (such as dog walking) (Jackson, 2020).

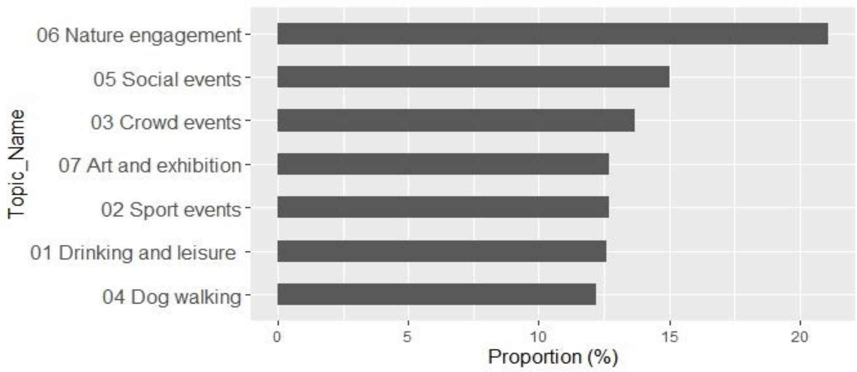

Figure 6.4 The proportions of all identified topics.

Table 6.1 Topic words with the highest probability.

Topic name	Topic words with the highest probability
1. Drinking and leisure	Lockdown, good, make, night, well, train, home, fit, exercise, last
2. Sport and music events	day, run, happy, music, weekend, festival, house, start, artist, marathon
3. Crowd events	year, easter, see today, first, point, people, amazing, club, friend
4. Dog walking	walk, love, dog, common, even, like, morn, tree, enjoy, cockerspaniel
5. Social events	get, one, back, look, go, time, thank, week, wedding, cake, food
6. Nature engagement	park, garden, beauty, nature, spring, flower, photo, photography, life, blossom
7. Art and exhibition	new, work, design, museum, open, made, show, think, exhibition, draw

6.4.2 The evolution of topics over time

Figure 6.5 shows the trends of topic proportions of the topics *Dog walking* (Topic 4), *Social events* (Topic 5), *and Nature engagement* (Topic 6). The proportions of Topic 4 show an increasing trend from 2019 to 2020, indicating that the *Dog walking* related activities were becoming popular during the COVID-19 pandemic period, which might be the results of policies such as people being allowed to walk dogs during the lockdown period. The proportions of this topic slightly decreased from 2020 to 2021, but it remained higher than in 2019 which suggests that *Dog walking* remained a popular activity. In addition, visitors took their dog outside both in the morning and evening according to the key words of this topic (see Table 6.1). The proportions of *Social events* (Topic 5) declined from 2019 to 2020 and 2021, indicating that *Social events* accounts for decreasing proportions among all types of topics. This may be the result of restriction measures such as forbidden social events including weddings, celebrating parties, and group activities, according to the key words of this topic. The proportions of Topic 6 increased from 2019 to 2020 and 2021, indicating that *Nature engagement* related activities were becoming popular year by year, especially after the first peak of the

COVID-19 pandemic. This may be the results of restriction measures such as closing non-essential shops and cancelling all public events, but encouraging residents to visit parks.

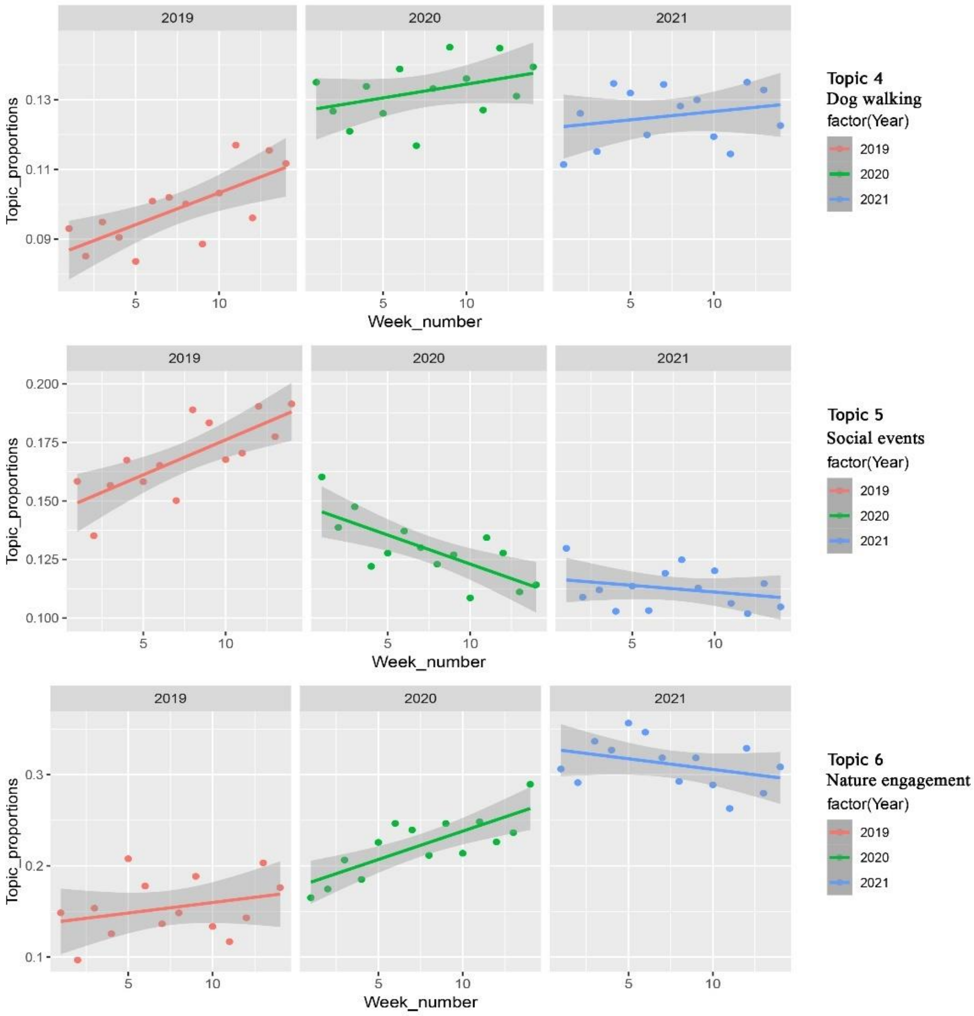

Figure 6.5 Trends of the topic proportions.

In an STM analysis, a content covariate is a variable that is used to explain the variation in the prevalence of topics across documents. In this case, whether the Tweets were posted before or after the COVID-19 outbreaks was used as a content covariate to examine how the COVID-19 pandemics have affected the topics discussed in the analysed texts. Figure 6.6 is a graphical display of topical perspectives with content covariate of COVID-19. The figure shows the distribution of topic words across documents and how the notable topics correlated to the COVID-19 covariate. Overall, vocabulary differences by rating were plotted for Topics 4, 5 and 6, which display the obvious differences in topic words from 2019 to COVID-19 pandemics, the vertical position of the words is distributed simply to aid understanding (the vertical axis has no meaning other than to prevent words from overlapping), while the horizontal position represents the tendency of the activity to occur during a particular period (before or during the COVID-19 pandemics).

Furthermore, the size of the words reflects the degree of correlation between the activity and the corresponding period, with larger sizes indicating stronger correlations.

Specifically observing Topic 4 (*Dog walking*), before the COVID-19 period, topics frequently expressed words about enjoyment such as ‘love’ and ‘like’. However, during COVID-19 pandemics, words of ‘walk’, ‘dog’, ‘morning’ and ‘daily’ were frequently mentioned, indicating that UGS visitors tended to take dog-related activities during the pandemic periods, compared with before the COVID-19 period.

Topic 5 (Social events) mentioned daily routine activities such as ‘Sunday’, ‘week’, ‘get’ and ‘come’ before the COVID-19 outbreaks, whereas during the COVID-19 pandemics, positive and encouraging words such as ‘live’, ‘hill’, ‘thank’, ‘need’, ‘food’ and ‘back’ were more commonly used, indicating that UGS users emphasised contents that might help to fight against the pandemics, and encourage people to save lives and stay healthy.

Topic 6 (*Nature engagement*) expressed words of ‘garden’, ‘spring’, ‘sunshine’ and ‘beautiful’ both before and during the COVID-19 period. However, during the COVID-19 pandemics, this topic emphasised the expression such as 'park', 'flower', 'blossom', 'nature' and 'wildlife', suggesting that that UGS visitors were more likely to spend time observing nature-related objects such as flowers and wildlife during the pandemic, rather than simply enjoying the sunshine in gardens.

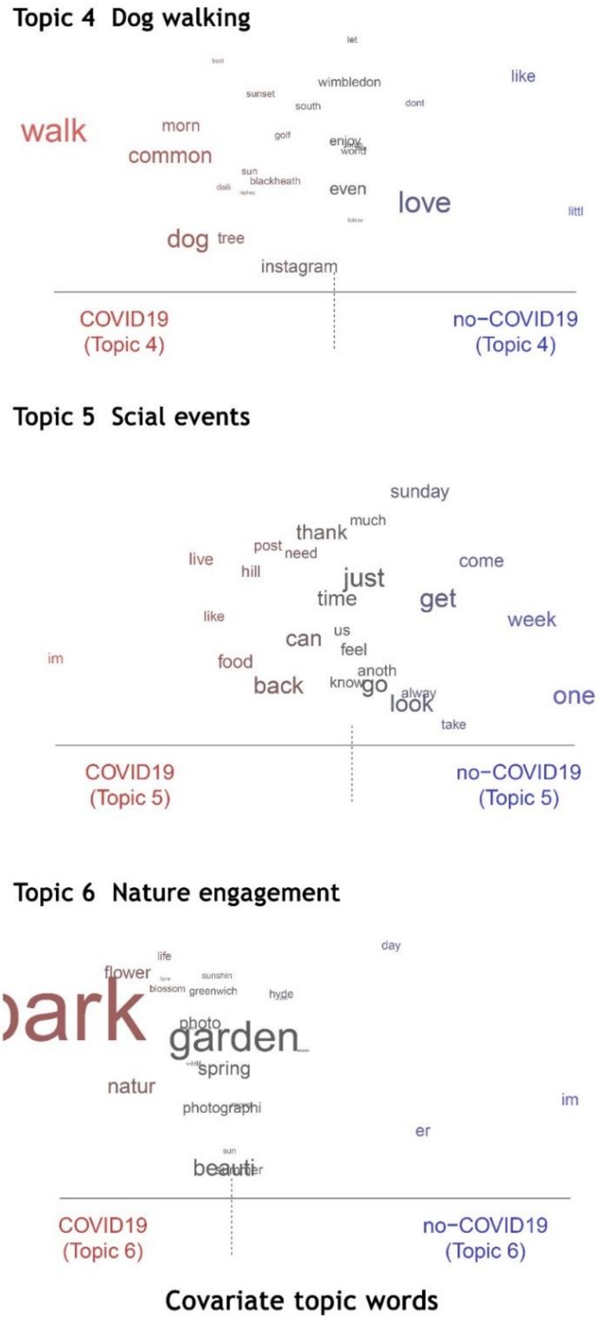

Figure 6.6 Graphical display of topical perspectives with content covariate of COVID-19.

6.4.3 Dynamics in spatial patterns of topics

Figure 6.7 shows the spatial variations of topics *Dog walking, Social events, and Nature engagement,* with darker colours indicating a higher hotspot that a place corresponds to the topic. Overall, the three topics show different trends in spatial patterns across three years. Topic 4 displays an initially increasing trend from 2019 to 2020, followed by a decrease to 2021. Topic 5 shows a decreasing trend from 2019 to 2020, continuing through 2021. Topic 6 shows an increasing trend across all study periods.

Specifically, in 2019, Topic 4 (*Dog walking)* shows a relatively lower probability across the whole study area, indicating that the related activities were less popular compared to other types of UGS activities, even though some hotspots located in the southwest part of London were found. However, the probability increased from 2019 to 2020 across the whole of London, which may the results of lockdown policy such as dog walkers being allowed to go outside every day, which benefited both humans and pets, and people might spend more time in UGSs with their dogs. In 2021, the probability remains higher than that in 2019, indicating that this type of activities remains attractive compared to the pre-COVID 19 period.

The topic *Social events* show a decreasing trend from 2019 to 2020, with decreasing probabilities across the study area, except for one hotspot found in southeast part of London. This might be related to the lockdown restriction measures such as 'stay home and stay healthy', social distancing, and wearing masks, which cut off the connections between people, thus there was an obviously decreasing trends from 2019 to 2020. In 2021, most parts of the study area show a decline in probability, with several hotspots found on the periphery.

The spatial patterns of topic *Nature engagement* dynamic changed from 2019 to 2020 and 2021. The areas with the topic probability were smoothly distributed across London in 2019, then many hotspots came out in centre part of London in 2020, with some hotspots located on the boundary of London, indicating that the restriction measures such as encouraging people to visit parks every day made people more likely to take part in this type of activity. In 2021, the hotspots extended to almost the whole city. The areas with higher probability continuously increased in 2021, indicating this type of UGS activity were popular across the London, which may benefit to visitor's mental and physical wellbeing (Vujcic et al., 2019).

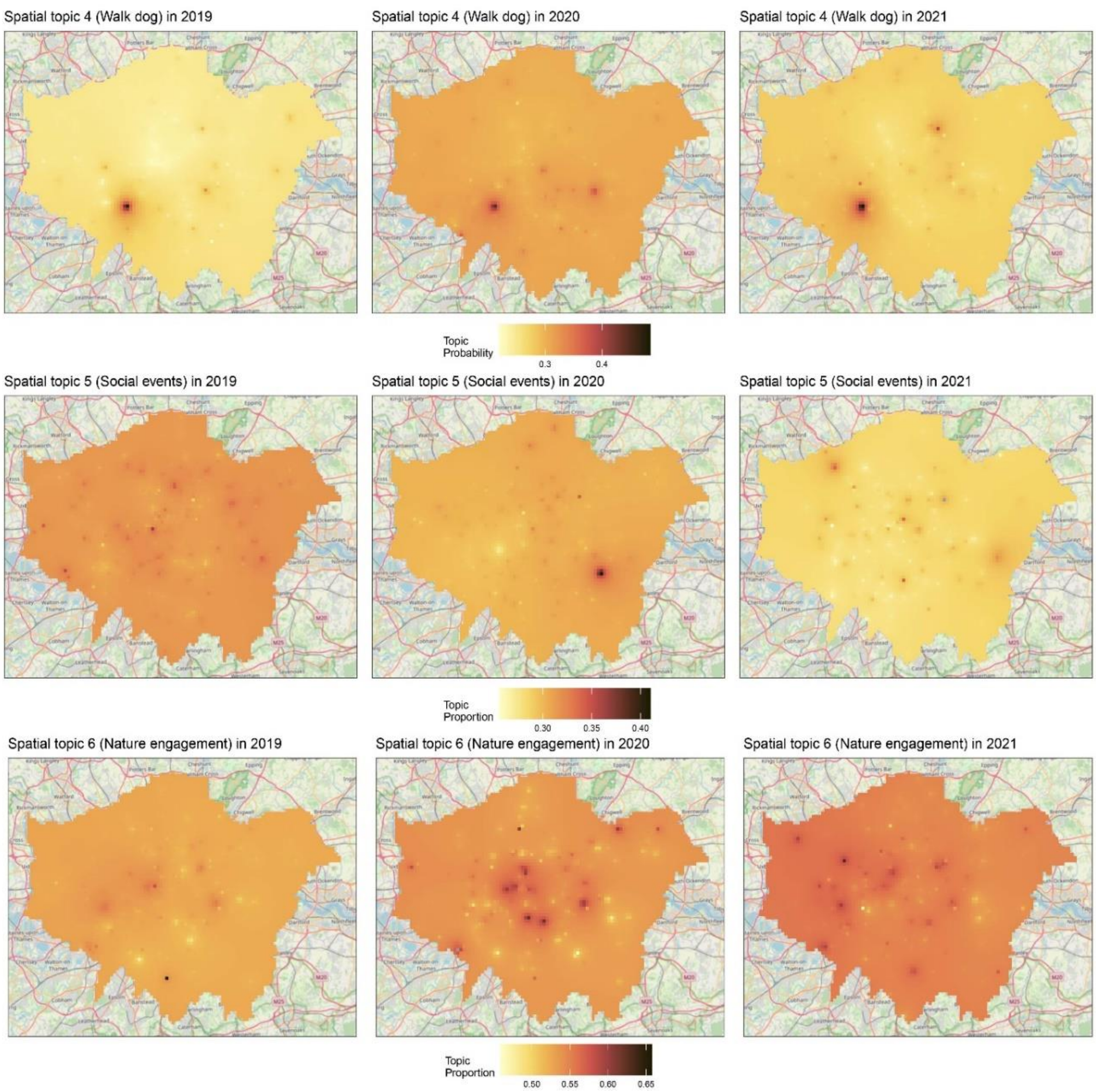

Figure 6.7 Spatial-temporal patterns of the topics *Dog walking*, *Social events*, and *Nature observation.*

6.5 Discussion

This study investigated the impact of the COVID-19 pandemic on UGS-related topics from the spatial and temporal perspectives. In summary, this study utilized STM and IDW to detect dynamic changes in spatial-temporal patterns of UGS topics in London before-, during-, and after the COVID-19 pandemic period. This section discussed the research results and the applications of STM and IDW in relation to UGS use.

6.5.1 What topics and attitudes expressed through Tweets during the COVID-19 pandemic?

Seven UGS topics were identified in this study, of which *Nature engagement* and *Social events* accounted the top two proportions, followed by *Social events*, *Crowd events*, *Art* *and exhibition*, *Sport events*, *Drinking and leisure*, and *Dog walking.* Previous studies focused on physical activities and their benefits to human wellbeing in relation to UGS (Knobel et al., 2021; Schipperijn et al., 2013). However, the research results in the current study suggested that other types of activities also played an essential role especially during a time a crisis such as COVID-19 pandemic. For example, *Dog walking* related activities showed a relative increasing trend among all topics, indicating that the dog ownership may benefit human mental wellbeing during the COVID-19 pandemics (Owczarczak-Garstecka et al., 2021). This finding is consistent with previous study, which suggested that dogs can provide psychological and social support to their owners, as well as motivate them to engage in physical activity (Christian et al., 2018). Future study could focus on details of the UGS activity types, combining with the socio-demographics of local people, characteristics of their homes, and accessibility of UGSs surrounding their household.

6.5.2 How did the notable topics change over space and time?

This study focused on investigating topics *Nature engagement*, *Social events*, and *Dog walking* to detect the impact of COVID-19 pandemics on topic evolution and topic probability over space and time. The results suggested that the topic of *Nature engagement* displayed an increasing trend in topic proportion, indicating that activities related to 'flower', 'blossom', 'nature' and 'wildlife' became increasingly popular during COVID-19 pandemics. Additionally, the spatial patterns of this topic displayed an increasing trend across the whole study area, with some hotspots found during the pandemics. Previous studies that used Twitter points as data source to account for the frequency of UGS visits (Cui et al., 2022) also revealed diverse trends in the UGS visitation during the lockdown period. However, this study treated the whole Twitter points as one level, which could not capture the spatial variations in specific UGS topics and related activities. The topic *Social events* displayed a decreasing trend in both topic proportion and spatial patterns, suggesting that restriction measures such as social distancing, staying home and wearing masks may have resulted in decrease social activity when they visited UGS. The increasing trend in *Dog walking* suggested that walking-dog related activities may be a kind of popular UGS activity that is relatively unaffected by COVID-19 pandemic restrictions. This result is consistent with previous studies that suggest dog walking can be a reliable and consistent form of physical activity, even during periods of social distancing and quarantine (Carr et al., 2021).

6.5.3 Approaches for tracking dynamics in topics

Previous studies used key-words analysis (Sim and Miller, 2019) and manually methods (Roberts, H.V., 2017) to detect UGS related activities through Tweet texts, the mentioned approaches were time and energy consuming, and may lack the ability to identify and capture the details from datasets. The main existing event-detection methods are based on LDA and have had satisfactory performances on a wide range of applications (Yao and Wang, 2020; Fu et al., 2018). However, LDA is a useful method for topic detection but not for topic dynamics over time and space. The current study employed STM to detect the topics, which allowed researchers to examine how the covariate variable (whether before or during COVID-19 pandemics) influenced the UGS topics during this time. By utilizing STM, the study was able to capture more specific variations in UGS topics and related activities.

In terms of investigating the spatial changes of topics, this study used IDW as an interpolation method to estimate the spatial variation of topic probability across the whole London, which can reveal a relative real spatial patterns of topics. The study was able to identify the areas where certain topics were more prevalent, and how these patterns changed across the city. This information can be used to gain insights into the underlying factors that contribute to the spatial distribution of topics and how they relate to urban green space types and attribute, UGS accessibilities, and user characteristics using UGS. Overall, using IDW to investigate the spatial changes of topics can provide a more precise and informative analysis compared to other interpolation methods, particularly when considering the topic probability.

This study remains some limitations. First, this study only used data of Twitter in Greater London, future research can collect data of different regions from other platforms to explore the generalizability of this approach. Second, the COVID-19 period was used as the covariates in STM model. Future studies can incorporate the UGS user characteristics, such as gender, age, and expertise into the model to investigate how different user groups interact with UGSs and how their experiences are related to different topics. By addressing these limitations, future studies can provide a more comprehensive understanding of the dynamics of UGS-related topics and how they are influenced by various factors, especially during the time crisis.

6.6 Conclusion

This study provided a framework for identifying, analysing, and visualising topics from social media data by combining text mining techniques, STM and IDE approaches. Specifically, this study initially collected Twitter data from the academic research API which provides full historical datasets with high resolution of space and time information. This datasets enabled the researchers to track the dynamics in spatial and temporal patterns related to UGS use across three years pre-, during-, and after the COVID-19 outbreaks. Then STM was employed to identify the topics in UGSs and analyse changes in topic proportions across the study periods. The covariate of COVID-19 period was used to understand how the users visited UGSs before and after the COVID-19 outbreaks. Finally, spatial interpolation method (IDW) was used to explore the dynamics of topical spatial patterns across the whole of London, by mapping the spatial variations of the probability values of each Tweet's correspondence to the respective topic, which provided insights into the distribution of UGS-related topics across the city. Overall, this study demonstrated how combining text mining techniques, STM and IDE approaches can enable researchers to gain a more comprehensive understanding of social media data related to UGSs and how they are impacted by factors such as COVID-19 outbreaks.

This approach identified seven types of topics, of which *Nature engagement* was the most popular, followed by *Social events*, *Crowd events, Art and exhibition, Sport events, Drinking and leisure*, and *Dog walking.* From these, *Dog walking*, *Social events*, and *Nature engagement* were selected as a case study to evaluate the impacts of the COVID-19 pandemic restriction measures and related policies on UGS use. The results show that *Dog walking* and *Nature engagement* were becoming popular during the lockdown period, while *Social events* were less popular, potentially due to restriction measures such as stay home orders and social distancing. Spatial patterns of topic probability also show spatial various trends over three years. It is notable that nature related activities were increasingly popular across the whole study area. This study provides an effective method in exploring the dynamic changes in spatial-temporal patterns of UGS topics, which is helpful in UGS planning and management, especially in times of crisis such as COVID-19 outbreak.

References

Abdulmanov, R., Miftakhov, I., Ishbulatov, M., Galeev, E. and Shafeeva, E. 2021. Comparison of the effectiveness of GIS-based interpolation methods for estimating the spatial distribution of agrochemical soil properties. *Environmental Technology & Innovation*. ***24***, p101970.

Blei, D. and Lafferty, J. 2006. Correlated topic models. *Advances in neural information processing systems*. ***18***, p147.

Blei, D.M., Ng, A.Y. and Jordan, M. 2003. Latent dirichlet allocation. *Journal of machine Learning research*. ***3***(01), pp.993-1022.

Bogdanowicz, A. and Guan, C. 2022. Dynamic topic modeling of twitter data during the COVID-19 pandemic. *PloS one*. ***17***(5), pe0268669.

Carr, D., Friedmann, E., Gee, N.R., Gilchrist, C., Sachs-Ericsson, N. and Koodaly, L. 2021. Dog walking and the social impact of the COVID-19 pandemic on loneliness in older adults. *Animals*. ***11***(7), p1852.

Chen, X., Zou, D., Cheng, G. and Xie, H. 2020. Detecting latent topics and trends in educational technologies over four decades using structural topic modeling: A retrospective of all volumes of Computers & Education. *Computers & Education*. ***151***, p103855.

Christian, H., Bauman, A., Epping, J.N., Levine, G.N., McCormack, G., Rhodes, R.E., Richards, E., Rock, M. and Westgarth, C. 2018. Encouraging dog walking for health promotion and disease prevention. *American journal of lifestyle medicine*. ***12***(3), pp.233-243.

Cui, N., Malleson, N., Houlden, V. and Comber, A. 2021. Using VGI and Social Media Data to Understand Urban Green Space: A Narrative Literature Review. *ISPRS International Journal of Geo-Information*. ***10***(7), p425.

Cui, N., Malleson, N., Houlden, V. and Comber, A. 2022. Using social media data to understand the impact of the COVID-19 pandemic on urban green space use. *Urban Forestry and Urban Greening*. ***74***, p127677.

Da Schio, N., Phillips, A., Fransen, K., Wolff, M., Haase, D., Ostoić, S.K., Živojinović, I., Vuletić, D., Derks, J. and Davies, C. 2021. The impact of the COVID-19 pandemic on the use of and attitudes towards urban forests and green spaces: Exploring the instigators of change in Belgium. *Urban Forestry & Urban Greening*. ***65***, p127305.

Du, Y., Yi, Y., Li, X., Chen, X., Fan, Y. and Su, F. 2020. Extracting and tracking hot topics of micro-blogs based on improved Latent Dirichlet Allocation. *Engineering Applications of Artificial Intelligence*. ***87***, p103279.

Erdönmez, C. and Atmiş, E. 2021. The impact of the Covid-19 pandemic on green space use in Turkey: Is closing green spaces for use a solution? *Urban Forestry & Urban Greening.* ***64***, p127295.

Fatima, M., Butt, I. and Arshad, S. 2022. Geospatial clustering and hot spot detection of malaria incidence in Bahawalpur district of Pakistan. *GeoJournal.* ***87***(6), pp.4791-4806.

Fu, C., McKenzie, G., Frias-Martinez, V. and Stewart, K. 2018. Identifying spatiotemporal urban activities through linguistic signatures. *Computers, Environment and Urban Systems.* ***72***, pp.25-37.

Geng, D., Innes, J., Wu, W. and Wang, G. 2021. Impacts of COVID-19 pandemic on urban park visitation: a global analysis. *Journal of forestry research.* ***32***(2), pp.553-567.

Gómez-Losada, Á., Santos, F.M., Gibert, K. and Pires, J.C. 2019. A data science approach for spatiotemporal modelling of low and resident air pollution in Madrid (Spain): Implications for epidemiological studies. *Computers, Environment and Urban Systems.* ***75***, pp.1-11.

Gräler, B., Pebesma, E.J. and Heuvelink, G.B. 2016. Spatio-temporal interpolation using gstat. *The Royal Journal.* ***8***(1), p204.

Grimmer, J. and Stewart, B.M. 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis.* ***21***(3), pp.267-297.

Gu, K., Zhou, Y., Sun, H., Dong, F. and Zhao, L. 2021. Spatial distribution and determinants of PM 2.5 in China’s cities: Fresh evidence from IDW and GWR. *Environmental monitoring and assessment.* ***193***, pp.1-22.

He, W., Zha, S. and Li, L. 2013. Social media competitive analysis and text mining: A case study in the pizza industry. *International journal of information management.* ***33***(3), pp.464-472.

Hu, X. and Liu, H. 2012. Text analytics in social media. *Aggarwal, C., Zhai, C. (eds) Mining Text Data.* Springer, Boston, MA, pp.385-414.

Ihaka, R. and Gentleman, R. 1996. R: a language for data analysis and graphics. *Journal of computational and graphical statistics.* ***5***(3), pp.299-314.

Jackson, D. 2020. *Can I walk my dog during the Corona lockdown?* [Online]. Available from: https://www.allaboutdogfood.co.uk/articles/can-i-walk-my-dog-during-the-corona-lockdown

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y. and Zhao, L. 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools Applications.* ***78***(11), pp.15169-15211.

Knobel, P., Maneja, R., Bartoll, X., Alonso, L., Bauwelinck, M., Valentin, A., Zijlema, W., Borrell, C., Nieuwenhuijsen, M. and Dadvand, P. 2021. Quality of urban green spaces influences residents' use of these spaces, physical activity, and overweight/obesity. *Environmental Pollution.* ***271***, p116393.

Kuhn, K.D. 2018. Using structural topic modeling to identify latent topics and trends in aviation incident reports. *Transportation Research Part C: Emerging Technologies.* ***87***, pp.105-122.

Lansley, G. and Longley, P. 2016. The geography of Twitter topics in London. *Computers, Environment and Urban Systems.* ***58***, pp.85-96.

Li, J. and Heap, A.D. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics.* ***6***(3-4), pp.228-241.

Li, J. and Heap, A.D. 2014. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software.* ***53***, pp.173-189.

Mimno, D. and Lee, M. 2014. Low-dimensional embeddings for interpretable anchor-based topic inference. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp.1319-1328.

Niță, M.R., Arsene, M., Barbu, G., Cus, A.G., Ene, M., Serban, R.M., Stama, C.M. and Stoia, L.N. 2021. Using Social Media Data to Evaluate Urban Parks Use during the COVID-19 Pandemic. *International Journal of Environmental Research Public Health.* ***18***(20), p10860.

Ordnance Survey. 2021. *OS Open Greenspace.* [Online]. [Accessed 23.02]. Available from: https://www.ordnancesurvey.co.uk/business-government/products/open-map-greenspace

Owczarczak-Garstecka, S.C., Graham, T.M., Archer, D.C. and Westgarth, C. 2021. Dog walking before and during the COVID-19 pandemic lockdown: experiences of UK dog owners. *International Journal of Environmental Research and Public Health.* ***18***(12), p6315.

Pereira, L.C., dos Santos, G.R., Marques, E.A.G., Pires, J.D. and Renó, R. 2022. Construction of multidimensional geomechanical models with IDW and using R language. *Journal of South American Earth Sciences.* ***116***, p103775.

Roberts, H.V. 2017. Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography*. ***81***, pp.13-20.

Roberts, M.E., Stewart, B.M. and Airoldi, E.M. 2016. A model of text for experimentation in the social sciences. *Journal of the American Statistical Association*. ***111***(515), pp.988-1003.

Roberts, M.E., Stewart, B.M. and Tingley, D. 2014a. Stm: An R package for structural topic models. *Journal of Statistical Software*. ***91***, pp.1-40.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G. 2014b. Structural topic models for open-ended survey responses. *American journal of political science*. ***58***(4), pp.1064-1082.

Rodriguez, M.Y. and Storer, H. 2020. A computational social science perspective on qualitative data exploration: Using topic models for the descriptive analysis of social media data. *Journal of Technology in Human Services*. ***38***(1), pp.54-86.

Sachdeva, S., McCaffrey, S. and Locke, D. 2017. Social media approaches to modeling wildfire smoke dispersion: Spatiotemporal and social scientific investigations. *Information, Communication & Society*. ***20***(8), pp.1146-1161.

Salloum, S.A., Al-Emran, M., Monem, A.A. and Shaalan, K. 2017. A survey of text mining in social media: facebook and twitter perspectives. *Advances in Science, Technology and Engineering Systems Journal*. ***2***(1), pp.127-133.

Schipperijn, J., Bentsen, P., Troelsen, J., Toftager, M. and Stigsdotter, U.K. 2013. Associations between physical activity and characteristics of urban green space. *Urban forestry & urban greening*. ***12***(1), pp.109-116.

Sim, J. and Miller, P. 2019. Understanding an urban park through big data. *International journal of environmental research and public health*. ***16***(20), p3816.

Steiger, E., Westerholt, R., Resch, B. and Zipf, A. 2015. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*. ***54***, pp.255-265.

Tomczak, M. 1998. Spatial interpolation and its uncertainty using automated anisotropic inverse distance weighting (IDW)-cross-validation/jackknife approach. *Journal of Geographic Information and Decision Analysis*. ***2***(2), pp.18-30.

Ugolini, F., Massetti, L., Calaza-Martínez, P., Cariñanos, P., Dobbs, C., Ostoić, S.K., Marin, A.M., Pearlmutter, D., Saaroni, H. and Šaulienė, I. 2020. Effects of the COVID-19 pandemic on the use and perceptions of urban green space: An international exploratory study. *Urban forestry & urban greening*. ***56***, p126888.

Vujcic, M., Tomicevic-Dubljevic, J., Zivojinovic, I. and Toskovic, O. 2019. Connection between urban green areas and visitors' physical and mental well-being. *Urban forestry & urban greening*. ***40***, pp.299-307.

Wallach, H.M., Murray, I., Salakhutdinov, R. and Mimno, D. 2009. Evaluation methods for topic models. In: *Proceedings of the 26th annual international conference on machine learning*, pp.1105-1112.

Yao, F. and Wang, Y. 2020. Tracking urban geo-topics based on dynamic topic model. *Computers, Environment and Urban Systems*. ***79***, p101419.

Chapter 7 Discussion and Conclusions

7.1 Introduction

The work presented in this thesis is novel and the research findings have the potential to inform urban green space planning and management, especially in a period of social crisis like the COVID-19 pandemic. There are no other published examples of synthesising existing knowledge on the use of social media data in the research field of UGS research, a number of future research directions associated with data availability, quality, and analysis methods have been identified by conducting a review with meta-analysis in this thesis. Furthermore, the thesis is one of the first studies to investigate the impact of the COVID-19 pandemic on UGS use from spatial-temporal perspectives, and to explore topics in relation to UGS by using social media data analysis. This work is also one of the first studies to compare georeferenced Tweets posted in a 3 month period from 23 March to 23 June for 3 years covering the first lockdown in the UK in 2020, with analysis for the same period in 2019 and 2021. Additionally, the work is one of the first studies to examine the usability of Twitter full historical datasets especially in the field of UGS research, the datasets were collected from academic research API that can provide more precise, unbiased, and high representative datasets (Pfeffer et al., 2022; Barrie and Ho, 2021).

This chapter will begin with a summary of the thesis based on each chapter in Section 6.2. The main research findings will be discussed, followed by a conclusion that emphasizes the extent to which the research objectives, initially outlined in Chapter 1, have been achieved in Section 6.3. Additionally, the limitations of the analysis will be included in Section 6.4. Finally, the chapter will end with recommendations for future work and concluding remarks.

7.2 Thesis summary

As stated in Chapter 1, the overall aim of this thesis will be achieved through a series of analysis in three chapters (Chapters 3, 4 and 5), with each chapter comprising several specific research objectives (Table 1). This section provides a summary of each chapter.

7.2.1 Summary of Chapter 1: Introduction

Chapter 1 introduced the research background and overall aim. The overall aim of this work is to evaluate the contributions of UGS to human wellbeing during a time of crisis, by investigating the characteristics and spatial-temporal patterns of UGS use across three periods: pre-, during- and after the COVID-19 pandemic. A number of specific research objectives were listed in order to achieve the overall research aim, thesis structure were also introduced in this chapter. This chapter highlighted the importance of UGS as it provides considerable various benefits to residents and plays an essential role in maintaining and enhancing sustainable city.

7.2.2 Summary of Chapter 2 and 3: Research contexts

Chapter 2 provided research contexts of the relevant studies on UGS and its visitation, social media and applications of the relevant datasets, as well as study areas and datasets used in the thesis. The works presented in Chapter 2 have demonstrated that it is extremely important to investigate the contributions of UGS to human wellbeing, especially during a time of crisis such as COVID-19 pandemic. Additionally, Chapter 2 highlighted the contributions of volunteered geographic information (VGI) and social media data to UGS research. The results suggested that the VGI and social media data can provide valuable information for UGS research compared to traditional analysis methods such as observation and questionnaires. After the above early observations, the remainder of Chapter 2 reviewed the social media analysis with the aim of identifying the methodological challenges of using social media data for UGS research from the perspectives of spatial-temporal analysis and semantic information extraction methods such as topic modelling.

7.2.3 Summary of Chapter 4: Using VGI and social media data to understand urban green space: A narrative literature review

Chapter 3 provided an in-depth and comprehensive literature review on the use of VGI and social media data in research examining UGS. A number of 177 international research articles were reviewed in this analysis, which covered major social media platforms around the world. It examined methods for data collection and data analysis, then explored advantages and disadvantages associated with different social media datasets. The research findings suggested that scholars have increasingly studied UGS by using social media data in recent years. Several limitations including issues of data acquisition, representativeness, and data quality concerning the use of social media have been identified, thereby providing insights for informing future research efforts in this domain. The work in Chapter 3 also suggested that the bibliometric analysis methods can provide an objective way to measure the literature on UGS analysis in relation to social media.

7.2.4 Summary of Chapter 5: Using social media data to understand the impact of the COVID-19 pandemic on urban green space use

Chapter 4 used social media data (Twitter) to examine spatial-temporal changes in UGS use in London associated with the COVID-19 pandemic. It compared georeferenced Tweets posted in a 3 month period from 23 March to 23 June for 3 years covering the first lockdown in the UK in 2020, with Tweets for the same period in 2019 and 2021. The results suggested that the land-use type of *Public Park and Garden* was the most frequently visited type of UGS during the first COVID-19 lockdown period. The results also demonstrated the spatial variations of UGS Tweet points within and out of the LCBD boundaries. Additionally, physical activities suggested an increasing trend which maybe as a result of allowing people to take daily exercises during the lockdown period. Finally, the timestamps of the datasets were used to explore the temporal patterns of UGS use, the results suggested that people spent more time in UGS areas on weekdays than weekends compared to pre-lockdown. The findings and methods in this analysis can potentially inform UGS planning and management, especially in a period of social crisis like the COVID-19 pandemic.

7.2.5 Summary of Chapter 6: Urban green space topics based on structural topic modelling during the COVID-19 pandemic

Chapter 5 explored the UGS-related topics under discussed by UGS users. This study uses structural topic model (STM) to detect the dynamics of UGS-related topics pre-, during- and after the COVID-19 outbreaks in the UK. Geo-referenced Tweets were used to generate the topics, the content covariate of the COVID-19 was added into this step which enable researchers to deep understand how UGS topics were influenced by the pandemics. The results suggested that there were seven main topics categories under discussion in UGS. Specifically, an increase in Tweets proportions was found for the topics *Nature engagement* and *Dog walking*, indicating that the popularity of these activities increased during the COVID-19 pandemic period. The topic *Social events* showed a decline in topic proportion, which might be related to restriction measures such as social distancing and mask wearing. This work demonstrated how geo-referenced Tweets can be used to reflect the impact of COVID-19 on UGS topics and relevant activities. The work suggested that using the STM can be an appropriate approach for examining the dynamic effects of the COVID-19 pandemic on the utilization of UGS.

7.3 Discussions and contributions of research findings

This section discusses the research findings in response to the overall aims and specific research objectives proposed in Chapter 1. The discussions are divided into three main parts in response to the specific work in chapters 4, 5 and 6. This section also summarises the potential contributions of research findings in a broader context of previous research works.

7.3.1 Research findings of Chapter 4

In the research field of UGS analysis, VGI and social media data can provide massive information with high spatial and temporal resolution, which enable researchers to understand the UGS in various perspectives, such as ecosystem services (Liu and Russo, 2021; Sinclair et al., 2018; Dai et al., 2022), tourism analysis (Gu et al., 2016), sentiments and perceptions (Heather E. Wright Wendel, 2012; Zhu and Xu, 2021), and public health (Hu and Sinnott, 2019). This work makes a novel contribution as this is one of the first studies to synthesis existing knowledge in this specific research field. The main research findings and potential contributions are as follows:

- The primary research objective of Chapter 4 is to **obtain a comprehensive understanding of the use of VGI and social media data in the context of UGS visitation,** with the aim of summarising the current state-of-the-art in this research field. The results showed that VGI and social media data have been widely and frequently used in the field of UGS related research. This indicates that social media data can contribute to UGS research, capitalizing on its strengths such as high spatial and temporal resolution, real-time and free available datasets, and large volume of data. However, it is important to acknowledge that VGI and social media data cannot replace traditional methods such as surveys, observations, and questionnaires in the field of UGS research (Lansley and Longley, 2016). Researches using APIs should always include a critical assessment of the sample and the in-built limitations to generalization when reporting findings.
- The second research objective is to **summarise the research topics in the selected studies**. This has been achieved by using multiple correspondence analysis (MCA) and manually summarising the research themes. The results identified four groups of research fields. Two groups of them associated with UGS, urban parks, natural environment, and social media. The third group includes themes related to ecosystem services, tourism, urban planning and behaviour

research. The final group was related with cultural ecosystem services. The results have the potential to identify new research themes for analysing UGS and extracting meaningful insights that can inform urban policy and practice. For example, future analysis could focus on investigating *cultural ecosystem services* and *urban park use* by using social media data. However, this suggestion was based on the analysis of MCA and manually annotation, there are lots of other algorithms such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) that can be used to detect article themes (Nugroho et al., 2020), which can possibly generate different results. This could be further explored in next steps of analysis.

- The third objective of **understanding data availability, quality and analysis methods of previous studies** has been achieved by summarising the popular social networks in this research field. The results show that Twitter was the most frequently used in UGS analysis, followed by Flickr, Instagram, Weibo and OpenStreetMap. This chapter contributed to literature on UGS analysis by providing researchers information about the ways in which various characteristics of social networks can be utilized in different research themes in relation to UGS. However, it should be noted that different types of VGI and social media data might exert various influences on research. For example, text-based VGI data can be used to detect topics related to UGS and to analyse the sentiment and perceptions concerning UGS (Plunz et al., 2019; Roberts, H. et al., 2019); image-based VGI data can be used to examine cultural ecosystem services (Guerrero et al., 2016; Sinclair et al., 2018; Gliozzo et al., 2016); and Map-based VGI data can be employed for the investigation of human mobility patterns and trajectory analysis (Dunkel, 2015; Hennig, 2017). It is important to choose appropriate datasets within varying contexts and research objectives. This principle potentially extends beyond UGS research and can be applied to a broad spectrum of studies, including urban mobility analysis, public health, and disaster detection.
- Chapter 4 also summarised a series of methods that were used in data analysis, including data pre-processing, spatial and temporal analysis, and semantic analysis including sentiment analysis and topic modelling. The work makes a contribution to literature by providing an overview for each step of analysis, which can potentially guide future data analysis, particularly in UGS research field.

- This study also makes a contribution to the field of social media analysis in UGS by identifying a range of ways to enhance the quality of datasets and improve the precision of research results. Specifically, the study highlights the potential benefits of using datasets from multiple social networks, integrating personal information with data analysis, and selecting appropriate analysis models for specific research questions. These findings can inform the development of more effective and efficient social media analysis tools and techniques that account for the complex dynamics of social media use in urban contexts. By incorporating these recommendations into research methodologies, scholars and practitioners can develop more sophisticated and nuanced approaches to social media analysis in broader urban analysis, with the potential to generate more accurate and insightful findings that can inform policy and practice.

7.3.2 Research findings of Chapter 5

The overall aim of Chapter 5 is to use social media data to investigate the impact of the COVID-19 pandemic on UGS visitation. This aim has been achieved through a comprehensive analysis of several specific research objectives as stated in Chapter 1. In this thesis, Twitter academic research API were used to collect unbiased datasets, which can provide a more comprehensive results than the previous APIs (Pfeffer et al., 2022; Jing-Huei et al., 2022). The main research findings in response to each research objective and potential contributions are summarised as follows:

- Chapter 5 selected Twitter as data source based on the results of Chapter 4. **The work successful designed a strategy for data collection, data pre-processing, storing and managing Tweets in field of UGS research**. This strategy can potentially serve as a means for researchers interested in using social media data for urban analysis. The work makes a novel contribution by demonstrating that the full historical Twitter datasets could be a reliable proxy for assessing UGS visitation. However, it is necessary to adjust the process when dealing with different research purposes. For example, the key words during data collection could be added or adjusted when focusing on specific research questions such as the assessment of UGS physical activities. The key words could be ‘running’, ‘cycling’, ‘walking’ and related words. Additionally, during the data pre-processing, this analysis has demonstrated how to deal with fake accounts and bots, but this might result in a loose of useful information. For example, this analysis removed the Tweets that were posted by only one user for more than ten times within one day. This could potentially result in the exclusion of true human users who tend to generate a substantial volume of Tweets during their visits to UGSs. Such exclusions may impose limitations on data analysis.

- The work presented in Chapter 5 is one of the first studies **to examine the impact of the COVID-19 pandemic on UGS use before-, during- and after the first lockdown period in 2020**. Previous studies have selected specific time intervals within a single year to make comparisons between lockdown and pre-lockdown periods, which may have problems caused by seasonal weather and climate patterns. This work contributed to existing literature by selecting a consistent study period over three years, which mitigates the potential impact of climate and weather changes on the research findings. Thus the work has the potential to provide a more robust and reliable analysis and inform future urban planning and management. However, there is a limitation in the datasets, as only the spring season was considered for the analysis. Future research could expand the analysis to encompass the entire three-year period, yielding a more comprehensive understanding of how UGS usage evolved before, during, and after the COVID-19 pandemic.
- The results in Chapter 5 highlighted the importance of *Public Park or Garden* and *Playing Field* in providing spaces for UGS visitation during the COVID-19 lockdown period. The results shed light on the preferences and behaviours of urban residents towards parks, gardens, and other types of green spaces. However, these results were based on the closure or restrictions of other types of UGSs and public spaces, including sports courts and facilities, private gardens, as well as public spaces like museums and art galleries. Therefore, the findings could vary if the circumstances changed during the COVID-19 pandemic. Future analyses could be expanded to investigate differences under varying conditions, and the utilization of agent-based modelling related methods may contribute to such related analyses.
- Additionally, the results suggested that *Physical Activity* became more frequent during the COVID-19 pandemic period, as more UGS users took exercise during their visits (Geng et al., 2021; Lesser and Nienhuis, 2020). This highlighted the importance of the engagement of physical activity, and suggested that future UGS planning could consider to provide more spaces for outdoors physical activities in

response the times of health crisis. However, the results were also limited by the absence of demographic information specific to the study area, which are subject to privacy policies and the choices made by social media users. Therefore, UGS planning and management should take into account the characteristics of different age groups, their health statuses, and preferences. To achieve this, it is essential to integrate datasets from various sources, including city hall records, local hospitals, and gym centers, among others with the approval of ethics review.

- The work also makes a contribution by revealing the impact of restriction measures of working from home, closing non-essential shops on different trends in UGS visitation levels within and out of the LCBD boundaries, and the changes in temporal patterns of UGS use as well. However, the current study solely selected UGS Tweets when detecting spatial and temporal variations in UGS use, without considering potential disparities in spatial and temporal patterns between UGS and non-UGS Tweets. Future analyses should incorporate this aspect to offer more robust recommendations and insights for UGS planning and management.

7.3.3 Research findings of Chapter 6

The use of UGS has changed during the COVID-19 pandemics. Previous studies have shown that the spatial-temporal characteristics of visits, activities engagement, and attitudes towards UGS all changed (Cui et al., 2022; Da Schio et al., 2021; Erdönmez and Atmiş, 2021; Geng et al., 2021). Although recent developments in approaches such as text-based word frequency analysis and spatial analysis offer new perspectives on UGS, they are often stationary and non-continuous in nature. This limits their ability to capture the complexity and diversity of UGS use, especially during the time of crises such as COVID-19 pandemic. Structural topic model (STM) (Roberts et al., 2014) can help researchers to track such temporal changes. This means that STM can capture the evolution of topics in a sequentially organized corpus of documents. Furthermore, it can assess changes in topic proportions and word probabilities within a corpus of documents over a time series of textual data.

This study used geo-referenced Twitter data to extract topics from three coincident periods – before, during and after the COVID-19 pandemic – to examine changes in the topics of UGS Tweets. The discussion of research findings and related contributions of are summarised as follows:

- The first specific research objective in Chapter 6 was to use **STM to process the textual datasets that were collected from Twitter academic research API**.

Previous studies usually used LDA to detect the UGS topics and activities (Yao and Wang, 2020; Fu et al., 2018), the method cannot capture the dynamics of topics over space and time. The work presented in this chapter demonstrated the usability of STM on investigating the dynamics of topics discussed under the UGS users. Nevertheless, alternative machine learning algorithms, including Naive Bayes, Support Vector Machines (SVM), and Word Embeddings, could also be applied to text classification. In the context of this specific analysis, the effectiveness of these algorithms remains uncertain. Therefore, it is advisable to assess the performance of these techniques in UGS topic detection. This evaluation would provide valuable insights into their suitability for the task, which would be our next research direction.

- **The second research objective was to explore the UGS topics expressed by Twitter users. Seven topic categories were identified for the whole three years.** Three important topics *Nature engagement, Social events,* and *Dog walking* were further analysed to explore the impact of the COVID-19 on UGS-related topics. However, other detected topics, including *Drinking and leisure, Sports events, and Crowd events* are still worth exploring. The results of the analysis of these topics have been published as an article (Cui et al., 2023). Integrating all the information from the analysed topics would enhance the comprehensiveness and reliability of the recommendations and suggestions for UGS planning and management. Additionally, it should be noted that STM needed to determine the number topics before generating the topics, which is a limitation in the analysis. There is currently no scientific consensus for determining the optimal number of topics within a defined model (Grimmer and Stewart, 2013; Rodriguez and Storer, 2020). Thus future analysis could focus on testing the goodness of fit for a topic model with varying number of topics. Some previous studies manually determine the number of topics within a specific research context (Kuhn, 2018), which could provide a reasonable number of topics.
- This work also makes a contribution by adding the covariate variable (whether before or during COVID-19 pandemics) into the model to capture the changes of the topic prevalence for each of the topics. This has the potential to reflect the impact of the COVID-19 pandemic on UGS related topics. The results provided potential insights of how UGS users response to a time of crisis such as COVID-19. However, this analysis exclusively chose COVID-19 as a covariate,

potentially overlooking the influence of other factors such as age and gender groups, education levels, and health status. The model could be enhanced to encompass multiple covariates when dealing with a number of variables in the analysis. This gap provides researchers with new directions for future analyses.

- **The third research objective was to investigate the spatial patterns of the identified topics**. The results suggested an increasing trend for the topic of *Nature engagement* across the whole study area, with some hotspots found during the pandemics. The topic *Social events* displayed a decreasing trend in both topic proportion and spatial patterns, suggesting that restriction measures such as social distancing, staying home and wearing masks may have resulted in decrease social activity when they visited UGS. The increasing trend in *Dog walking* suggested that walking-dog related activities may be a kind of popular UGS activity that is relatively unaffected by COVID-19 pandemic restrictions.

Overall, the thesis has made three important contributions. First, the thesis has critically summarised the literature on using VGI and social media data to understand the UGS. The research results in chapter 4 have provided an overview of the research topics using VGI data sources in relation to UGS, and characterised the research approaches based on data pre-processing, data quality assessment and enhancement, as well as data analysis and modelling. This has the potential to provide guidance to future researchers regarding the utilization of social media data in their analyses, especially in the field of UGS research. Second, the thesis contributes to social media data analytics. The analyses in chapter 5 have demonstrated that the geo-information and timestamps of social media data have the potential to reveal the spatial and temporal patterns of UGS visitation. This provides researchers and urban planners with the information for UGS planning, especially in a period of social crisis like the COVID-19 pandemic. The third contribution of the thesis is the application of topic modelling in the field UGS research associated with the COVID-19 pandemic. The analyses in chapter 6 have provided evidence to suggesting that the textual content of social media data such as Tweets have the potential to reflect the dynamics of topics discussed before, during, and after the COVID-19 outbreaks. This analysis can be utilised as a means of exploring the impact of broader social crises, including but not limited to the COVID-19 pandemic, on the interactions between humans and the environment.

7.4 Limitations of the study

It is important to acknowledge the limitations of this thesis. The limitations are highlighted in this section, and they should be take into account and addressed when applying the research outcomes and conducting UGS analysis using social media data.

7.4.1 Limitations of literature review (Chapter 4) and bibliometric analysis

The work presented in chapter 4 reviewed the literatures on the UGS studies associated with social media data. While bibliometric analysis can be a useful tool for examining the characteristics of the selected studies, there are also some limitations to this approach. The limitations of the analysis in Chapter 4 include:

- Only English research articles were selected which may not capture the related knowledge presented in other languages such as Chinese, Spanish, French and so on. This may limit the comprehensiveness and generalizability of the study's findings. Future researches should consider selecting a broader range of languages to chieve a relative global review of the UGS researches associated with social media data. Additionally, future literature review could consider the research output published in books, chapters, and patents for example.
- This review selected three scientific databases Web of Science, Scopus and IEEE as major data sources without considering other data sources. This may result in an incomplete representation of research output in the field of UGS research. Additionally, it should be noted that this review only obtained data sources that were freely available for download, which may lead to bias and potentially limit the comprehensiveness of research output. Future research could consider selecting the research outcomes that that requires payment, however, this is often limited by researchers' funding and support.
- The search strategies in this review excluded the articles that concerned with disaster events and emergency situations within a certain period. However, during times of crisis, the strategies should be improved by considering the COVID-19 outbreaks in order to comprehensively synthesise the knowledges about researches on UGS in relation to social media data. Additionally, it is important to update search terms with a timeline to account for the constantly evolving nature of social media networks.
- The metadata analyses were conducted by using *bibliometrix* R-package (Aria and Cuccurullo, 2017), there are a number of limitations of using this tool. First, the

metadata of literatures can vary which means that inaccurate or incomplete metadata can influence the quality of the research results. For example, the lack of information about the of publication years and authors' affiliation might affect the results of Annual Scientific Production and Corresponding Author's Country analyses. Secondly, the package provides predefined functions for bibliometric analysis, which may limit the ability to customize the analysis to specific research needs. The results could be influenced by the predefined parameters. Thirdly, different researchers might interpret the results based on their individual research experiences and backgrounds, potentially influencing the reliability of the results. Therefore, the interpretation of results may require domain-specific knowledge.

7.4.2 Limitations in understanding UGS use during times of crisis (Chapter 5 and 6)

The research presented in this thesis focused on detecting the UGS use from perspectives of spatial-temporal patterns of Tweet points and potential related topics. While the research results have reflected the impact of COVID-19 on UGS use over space and time, there are some limitations as follows:

- It should be noted that this study only selected spring season when the first lockdown was announced, and selected only three years 2019, 2020, and 2021 as study periods, which may result in less comprehensive outcomes for detecting the effects of COVID-19 pandemic on the other time period such as summer, autumn and winter. Future research should consider the whole of year to generate a relative complete researches on UGS over a year. Additionally, the year of 2022, and 2023 could be selected for investigating the long tale effect of the COVID-19 on UGS use.
- Only London was selected in this thesis, which may limit the comprehensiveness of the findings. Future study should consider including more urban areas around the world to provide a global view of the impact of the global pandemics on UGS use, which help to understand UGS users' responses to global pandemics. Additionally, lacking consideration in Twitter user's socioeconomic and demographic characteristics may lead to the risk of making an inaccurate inference. It is a paradox that despite the emerging social media data at finer spatial-temporal granularity, it is still difficult to get access to the individual users' personal information. Future analysis of social media data may require various sources, including city hall records, local hospitals, and gym centers, along with the approval of ethics review.

- This study initially investigated the occurrence of Tweet points across different types of UGS in London. This enable us to discover the changes of UGS use in different types of UGS, which could potential provide policy makers and UGS planners with the understanding of UGS visitation during times of crisis such as COVID-19 pandemic. However, some further analysis should be conducted in future study. For example, the spatial associations between Tweet points and UGS types, the sentiment analysis of UGS users over three study periods, and the accessibility, size, quality and facilities of UGS could also affect the UGS visitation (Guo et al., 2019) during the special period as many of UGS have been closed around the world (Geng et al., 2021). Further extension would be to examine UGS visitation in relation to the specific characteristics of each.
- Structural topic modelling (STM) has some limitations that should be acknowledged. One limitation is that STM relies on pre-specified parameters, such as the number of topics, which can impact the results and limit their generalizability. Although STM has been successfully used for detecting topics for Twitter datasets, the model may not be well-suited for analysing short texts, as it requires a minimum number of words to identify meaningful topics. In this thesis, the Tweets with the number of words less than three were removed, which could potentially reduce the influences of the results. Finally, STM is a relatively new method and its performance and validity compared to other topic modelling approaches is still an area of ongoing research.

7.4.3 Limitations of datasets and representativeness

This study selected Twitter as data sources which have been proved to be an useful proxy for investigating human behaviours. However, it should be noted that the datasets only represented a part of urban residents, which may lead to biases and less representativeness. Additionally, only Tweets with coordinates information were selected as input to data analysis due to the purpose of the study is to investigate the spatial patterns of UGS use, which may also lead to a less estimation of the UGS use. Finally, only Twitter was selected as data sources, future analysis should consider selecting multiple data sources, for example, including map- and image- based VGI and social media data.

Additionally, Twitter fake accounts and bots, the accuracy of GPS location, the availability of Wi-Fi or cellular data, and users' preferences (i.e. some of UGS users are more likely to post Tweets when back to home) can also affect the data quality, which may lead to biases and over- or underrepresentation. There are a number of methodological suggestions. The potential effects of geo-location errors in Twitter data could be overcome by using a buffer around UGS boundaries and more data could be collected over other periods to explore the impact of the full lockdown release on UGS visitation and the permanency of the changes observed in this study. Future work could also focus on a deeper understanding of preferences and attitudes around UGS by examining social media content in greater detail, thereby expanding the analysis of the impact of the COVID-19 pandemic on UGS-related perceptions relative to park use, and other factors related to public space.

7.5 Conclusion and outlook

In this thesis, the usage of UGS (urban green space) were analysed to reveal how UGS users response to a global pandemic such as COVID-19 outbreaks. London was the first national park city around the world and one of the most diverse and international city, making it as an ideal location for studying the challenges and opportunities associated with UGS planning and management particularly during times of crisis such as COVID-19 pandemic. Twitter was selected as data sources as its availability, high space and time resolution, and Twitter datasets that can be collected via new unbiased academic research API. Through evaluating spatial-temporal patterns of Tweet points across the study areas, the research fill the gap on the interactions between human (i.e. visits to UGS and outdoor activities) and natural environments particularly during the COVID-19 pandemic. Additionally, the thesis explored the topics discussed under UGS users during the COVID-19 pandemic, the dynamics of topic prevalence and spatial patterns of identified topics were investigated. The results have the potential to provide urban planners and UGS managers with the map of specific topics and concerns by people during the times of crisis. Before the data analysis, the thesis conducted a comprehensive review for understanding the knowledge about UGS researches associated with social media. The results of this review provided a comprehensive overview of the current state of the art of social media analysis technologies, and highlight a number of challenges and opportunities for future research in the field of UGS researches.

A number of planning and management recommendations and strategies are proposed from the research results in this thesis. Specifically, the differences in UGS visitation related to types of UGS suggested that the enhancement of *Public parks and Gardens* provision and more sport facilities could be included within large parks along with flower and wildlife gardens in response to future crisis. Additionally, the analysis found that small, highly localised pocket parks and gardens could enhance people's lives and should be considered by policy to support dynamic and flexible UGS access as people spend more time at home during the times of crisis. Finally, the work identified seven types of topics, of which *Nature engagement* was the most popular, followed by *Social events, Crowd events, Art and exhibition, Sport events, Drinking and leisure*, and *Dog walking.* The dynamics of these topics and related activities potentially reflected the impact of restriction measures such as stay home orders and social distancing on UGS use. Overall, the thesis provided a comprehensive understanding of dynamic changes of UGS use during a time of crisis such as COVID-19. With the situations increases with rapid changes in climate, socio-economic disruptions, expanding global populations etc., such an understanding of how people react could be crucial for planning responses to future crises.

To conclude, the successful application in urban green space planning and management, and policy-making relies on incorporating the knowledge of the interactions between human and natural environments. The analysis of social media data and the identification of key topics may support evidence-based decision-making in the planning and management of UGS, especially during times of crisis. The research outcomes can possibly improve the understanding of people's UGS behaviours and promote sustainable UGS development. This are likely to have potential to provide guidance in both normal and special times. The findings of this study may also have broader implications for understanding the role of social media in shaping public opinion and behaviour, and may inform the development of UGS policy making and urban planning.

References

Barrie, C. and Ho, J.C.T. 2021. academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software.* ***6***(62), p3272.

Cui, N., Malleson, N., Houlden, V. and Comber, A. 2022. Using social media data to understand the impact of the COVID-19 pandemic on urban green space use. *Urban Forestry and Urban Greening.* ***74***, p127677.

Cui, N., Malleson, N., Houlden, V. and Comber, A. 2023. The impact of the COVID-19 pandemic on the dynamics of topics in urban green space. *AGILE: GIScience Series*. ***4***, p.22.

Da Schio, N., Phillips, A., Fransen, K., Wolff, M., Haase, D., Ostoić, S.K., Živojinović, I., Vuletić, D., Derks, J. and Davies, C. 2021. The impact of the COVID-19 pandemic on the use of and attitudes towards urban forests and green spaces: Exploring the instigators of change in Belgium. *Urban Forestry & Urban Greening.* ***65***, p127305.

Dai, P., Zhang, S., Gong, Y., Zhou, Y. and Hou, H. 2022. A crowd-sourced valuation of recreational ecosystem services using mobile signal data applied to a restored wetland in China. *Ecological Economics.* ***192***, p107249.

Erdönmez, C. and Atmiş, E. 2021. The impact of the Covid-19 pandemic on green space use in Turkey: Is closing green spaces for use a solution? *Urban Forestry & Urban Greening.* ***64,*** p127295.

Fu, C., McKenzie, G., Frias-Martinez, V. and Stewart, K. 2018. Identifying spatiotemporal urban activities through linguistic signatures. *Computers, Environment and Urban Systems.* ***72***, pp.25-37.

Geng, D., Innes, J., Wu, W. and Wang, G. 2021. Impacts of COVID-19 pandemic on urban park visitation: a global analysis. *Journal of forestry research.* ***32***(2), pp.553-567.

Gu, Z., Zhang, Y., Chen, Y. and Chang, X. 2016. Analysis of attraction features of tourism destinations in a mega-city based on check-in data mining—A case study of ShenZhen, China. *ISPRS International Journal of Geo-Information.* ***5***(11), p210.

Guo, S., Yang, G., Pei, T., Ma, T., Song, C., Shu, H., Du, Y. and Zhou, C. 2019. Analysis of factors affecting urban park service area in Beijing: Perspectives from multi-source geographic data. *Landscape and Urban Planning.* ***181***, pp.103-117.

Heather E. Wright Wendel, R.K.Z., James R. Mihelcic. 2012. Accessibility and usability: Green space preferences, perceptions, and barriers in a rapidly urbanizing city in Latin America. *Landscape and Urban Planning.* ***107***, pp.272-282.

Hu, Y. and Sinnott, R.O. 2019. Big Data Analytics Exploration of Green Space and Mental Health in Melbourne. In: *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, Larnaca, Cyprus, 2019, pp. 648-657.

Jing-Huei, H., Myron F., F., Laura G., T. and J., A.H. 2022. Exploring public values through Twitter data associated with urban parks pre- and post- COVID-19. *Landscape and Urban Planning.* ***227,*** p104517.

Lesser, I.A. and Nienhuis, C.P. 2020. The impact of COVID-19 on physical activity behavior and well-being of Canadians. *International journal of environmental research and public health. **17***(11), p3899.

Liu, O.Y. and Russo, A. 2021. Assessing the contribution of urban green spaces in green infrastructure strategy planning for urban ecosystem conditions and services. *Sustainable Cities and Society. **68***, p102772.

Pfeffer, J., Mooseder, A., Hammer, L., Stritzel, O. and Garcia, D. 2022. This Sample seems to be good enough! Assessing Coverage and Temporal Reliability of Twitter's Academic API. *arXiv preprint arXiv.**2204***, 02290.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G. 2014. Structural topic models for open-ended survey responses. *American journal of political science. **58***(4), pp.1064-1082.

Sinclair, M., Ghermandi, A. and Sheela, A.M. 2018. A crowdsourced valuation of recreational ecosystem services using social media data: An application to a tropical wetland in India. *Science of the total environment. **642***, pp.356-365.

Yao, F. and Wang, Y. 2020. Tracking urban geo-topics based on dynamic topic model. *Computers, Environment and Urban Systems. **79***, p101419.

Zhu, J. and Xu, C. 2021. Sina microblog sentiment in Beijing city parks as measure of demand for urban green space during the COVID-19. *Urban Forestry & Urban Greening. **58***, p126913.

List of Abbreviations

API	Application Programming Interfaces
DW	Inverse Distance Weight
GPS	Global Position System
KDE	Kernel Density Estimation
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
MCA	Multiple Correspondence Analysis
MCP	Multiple Country Publications
ML	Machine Learning
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
NMF	Non-Negative Matrix Factorization
NMF	Nonnegative Matrix Factorization
OS	Ordnance Survey
PLSA	Probabilistic Latent Semantic Analysis
PPGIS	Participation Geographic Information Systems
SCP	Single Country Publications
STM	Structural Topic Modelling
TF-IDF	Term Frequency-Inverse Document Frequency
UGS	Urban Green Space
VGI	Volunteered Geographic Information