Article:

Cui, N., Malleson, N. orcid.org/0000-0002-6977-0615, Houlden, V. orcid.org/0000-0003-2300-2976 et al. (2 more authors) (2025) Using Twitter to understand spatial-temporal changes in urban green space topics based on structural topic modelling. Cities, 157. 105601. ISSN 0264-2751

https://doi.org/10.1016/j.cities.2024.105601

Using Twitter to understand spatial-temporal changes in urban green space topics based on structural topic modelling

Cui N [a,b], Malleson N [a], Houlden V [a], Yan Y [b] and Comber A [a]

[a] School of Geography, University of Leeds, UK

[b] Department of Geography, National University of Singapore, Singapore

Abstract: Social media data offers urban planners insights into human activities in urban green spaces (UGSs). While recent methods like text-based word frequency analysis provide new perspectives on UGS, they are often lack stationary and non-continuous in nature. This limits their ability to capture the complexity and diversity of UGS use. This study conducts a structural topic model (STM) analysis of geo-referenced Tweets posted in London to investigate the dynamics of UGS-related topics before-, during- and after the COVID-19 outbreaks. Additionally, an approach of inverse distance weighting (IDW) was used to investigate the spatial patterns of topics probabilities. The results found that there were seven main topics categories expressed in UGS over study periods. Specifically, the increasing trends in topics proportions were found for the topics *Nature engagement* and *Dog walking*, indicating that these activities became increasingly popular during the pandemic. However, the topic *Social events* showed a decline in topic proportion, which might be the results of restriction measures such as practicing social distance. This study further discussed the potential factors that affecting the dynamics of these topics in spatial and temporal patterns. The results can potentially support future UGS planning and management especially during a time of crisis.

Keywords: Structural topic modelling; social media data; Twitter, urban green space; COVID-19; spatial temporal analysis

1. Introduction

The restriction measures of the COVID-19 pandemics have changed human behaviour patterns when they visited UGS. Researches shown that the spatial-temporal characteristics of visits, visitor activities engagement, and attitudes toward UGS have all changed (Cui et al., 2022; Geng et al., 2021; Marchi et al., 2022; Sikorska et al., 2023; Taczanowska et al., 2024). For example, Geng et al. (2021) analysed visitation patterns of UGS at global, regional, and national scales. They found that during the COVID-19 outbreaks, the demand from residents for UGS increased particularly for engaging in physical activities, enjoying nature settings, and improving human well-being. Cui et al. (2022) explored the associations between different types of UGS and corresponding visitation patterns across the COVID-19 pandemic in London, the results showed that UGS equipped with physical facilities was more attractive to visitors than other types of UGS. Their findings suggested that citizens with higher education levels visited UGS more frequently during this period. Grzyb et al. (2021) used Instagram data to investigate the recreational use of UGS in Poland before and during the COVID-19 outbreak, the findings suggested that UGS users' recreational activities became more focused towards wilder green spaces during the pandemic. As yet, the dynamics of peoples' activities, as captured through social media, in relation to UGS during the COVID-19 period have not been sufficient studied, the results could potentially provide valuable insights to inform future UGS planning and management.

Twitter dataset has been widely used to investigate UGS visitation. Twitter users can post messages at any time and from any location, with the option to include geographic information in their posts (Lansley & Longley, 2016; Müller et al., 2023). Thus geo-referenced Tweets can provide researchers with high resolution spatial-temporal information and valuable contents, enabling them to accurately evaluate UGS visitation patterns. For example, Roberts (2017) used Twitter data to detect UGS activities-related events by manually classifying Tweets . While this study demonstrated how Twitter data could be used in UGS use investigation, the methods were time-consuming and inefficient. Cui et al. (2022) used Twitter data 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 datasets such as Google's Community Mobility Reports (Geng et al., 2021) and Instagram data (Grzyb et al., 2021) have also been used to investigate the changes in urban park visitation during the COVID-19 pandemic period.

Topic modelling is gaining an increasing attention from scholars as this approach can extract hidden information from large volumes of textual data such as articles, newspapers, and social media posts such as Tweets (Blei et al., 2003; Lansley & Longley, 2016). Latent Dirichlet Allocation (LDA) is a commonly used approach for topic modelling by determining the probability of a given document (such as a Tweet) being a member of a given topic through a "bag-of-words" interpretation of its contents (Blei et al., 2003). However, the themes in a corpus (i.e. a collection of documents) may evolve over time. As yet little research has examined the temporal evolution of topics and their dynamics. Tracking such temporal changes can be done using STM, introduced by (Roberts et al., 2019). STM can capture the evolutions of topic proportions and word probabilities in a set of time-series documents (e.g., Tweets with timestamps), thus this model evaluates the changes of topic proportions and word probability in a corpus of documents along a time series of text data. STM has been wildly used in various research fields. For example, Ding et al. (2023) explored how Airbnb users' preferences, as expressed in online reviews, differ across listings with varying levels of shared space and price ranges.

This study conducts a STM analysis of geo-referenced Tweets to extract topics from 3 coincident periods – before-, during- and after the COVID-19 pandemic – to examine changes in the topics over three comparable time periods. The work addresses the following research questions: i) What topics and attitudes were expressed through Tweets before-, during- and after the COVID-19 pandemic? ii) How did the observed topics change over space and time? To answer these, a covariate indicating the COVID-19 stage was integrated in STM to explore how UGS topics were influenced by the pandemic. Additionally, an approach of inverse distance weighting (IDW) was used to investigate the spatial patterns of topics probabilities 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. Changes in spatial distribution of each topic were also investigated to determine whether pandemic-related policies influenced the spatial variation of topics. The results in turn can inform policy and UGS management and help them better understand how UGS is used and public perceptions of it.

The rest of this study is organised as follows: In the Research background section 2, the literatures on UGS in city areas are reviewed, followed by a review of literatures related to topic modelling and STM. The methodology section 3 outlines the process of collecting and prepossessing Twitter dataset, identifying topics by using STM, and investigating spatial patterns of these topics. The Results section 4 describes the evolution of topics by comparing time periods before-, during- and after the COVID-19 pandemic, and explores the spatial-temporal changes of topics. The section 5 discusses the research results and assesses how these results answer the research questions. The Conclusion section 6 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.

2. Research background

The global urban population is expected to rise significantly, from 50% in 2010 to nearly 70% by 2050 (Mahtta et al., 2022), leading to an increasing demand for UGS, particularly in large metropolitan areas. UGS plays a critical role in promoting the physical and mental well-being of city residents and offers important social benefits (Houlden et al., 2019; Taczanowska et al., 2024). Consequently, the effective planning and management of UGS are essential to meet the growing needs of urban inhabitants, in particular during a health emergency (Grzyb et al., 2021; Marchi et al., 2022).

During the COVID-19 pandemic in the UK, the government implemented a series of limitation measures in England to reduce the virus's spread. On March 23 in 2020, the first nationwide lockdown was announced, requiring all residents to stay home except for essential needs. On June 23 2020 the end of the initial lockdown in England was announced with a 1-meter social distancing rule (Ghosh et al., 2020). During this lockdown period, city residents were permitted to visit UGS for physical and social activities (Cui et al., 2022; Owczarczak-Garstecka et al., 2021).

2.1 UGS research

UGS has been widely analysed and diverse benefits of UGS to humans have been confirmed by previous studies. These have investigated UGS visitation patterns (Guan & Zhou, 2024), the social activities that occur in UGS areas (Wilson et al., 2024), and the benefits of UGS to human wellbeing (Houlden et al., 2019). These studies showed that parks and green spaces provide key roles in physical, mental, and social well-being of city residents. The methods used in many previous studies included questionnaires, on-line surveys, and onsite-observations. These are usually time- and resource intensive, and providing less transferable generalizable outcomes (Müller et al., 2023). In addition, these approaches frequently suffer from low response rates and a lack of spatial-temporal information (Cui et al., 2021; Marchi et al., 2022).

Recently, social media data has been increasingly used to investigate the UGS to take advantage of real-time information, large numbers of users on social networks, and information with high spatial and temporal resolutions (Cui et al., 2022; Grzyb et al., 2021). An increasing number of studies have highlighted the importance of using social media data to analyse UGS especially when people are restricted to participate in public and social events in order to reduce the spread of the virus (Grzyb et al., 2021; Lopez et al., 2021; Marchi et al., 2022; Müller et al., 2023).

2.2 Structural topic modelling for dynamics in topics

Topic modelling has been used to investigate the dynamics of topics over study periods. For example, Bogdanowicz and Guan (2022) used dynamic topic modelling to detect topics in relation to the COVID-19 pandemic and analysed their temporal changes. They identified twelve most popular topics by using sequential LDA model, allowing the growth and changes of topics over time to be investigated. Du et al. (2020) enhanced the LDA model to generate and track topic evolution trends. The results showed that the improved LDA model had lower perplexity and higher coverage rate compared to the traditional LDA under the same conditions. However, these studies have not sufficiently investigated the spatial changes in topics.

STM was developed by (Roberts et al., 2019) based on earlier approaches such as LDA (Blei et al., 2003) and correlated topic models (Blei & Lafferty, 2006). STM incorporates covariates (i.e., post-dates, coordinates information) during the topic identification process, this enables researchers to explore spatial and temporal dynamics within the identified topics. Additionally, STM integrates structural variables among documents, which allows researchers to investigate the topic prevalence and their relationships with documents (Kuhn, 2018). This functionality allows for the exploration of how specific topics vary concerning other factors over time, providing insights into the prevalence of particular topics and their fluctuation in association with other variables.

STM has been widely used in previous studies. For example, Chen et al. (2020) used STM to investigate the evolutions of topics in computers and education academic publications over 40 years. All the previous results suggested that STM can not only identify the topics expressed on social media platforms, journals and websites, but also can capture the dynamics of topic prevalence and related trends (Ding et al., 2023). However, these studies failed to detect spatial patterns of topics even though the geo-referenced Tweets were utilised. STM has the features to integrate time and space covariates into the processing, which can potentially reflect the impacts of diverse events on the identified topics from dimensions of space, time and contents. The current study will discuss more details on LDA and STM models in later sections.

2.3 Spatial-temporal trajectories of topics

The coordinates information in Tweets metadata allow researchers to investigate the spatial patterns of identified topics within a certain area, the time stamps in Tweets metadata enable researchers to further investigate the spatial changes of topics over time. Previous studies have used geo-referenced Tweets to identify the topics in geographical spaces. For example, geo-referenced Tweets were used to detect the topics and their geographical locations within London (Lansley & Longley, 2016). The results have successfully revealed the relationships between human behaviours and surrounding characteristics. 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 how Twitter posts could be used to investigate location-based activities by applying LDA topic modelling, but without considering the dynamics of topics over time and space. The current study used STM to analyse how the UGS-related topics changed over space and time, which can potentially reveal the impacts of the COVID-19 pandemics on UGS-related events and activities, providing insights into future UGS planning and management, especially during a time of crisis or emergency events.

3. Methodology

3.1 Study area

This study selected Greater London, the United Kingdom, as the study area (Figure 1). Greater London contains 33 local districts with an estimated population of 8,866,180 in 2022 (Office for National Statistics, 2022). The region covers a land area of 1,569 km^2 , with 40% consisting of approximately 3,000 parks. London has eight Royal Parks including Hyde Park, Green Park, and Richmond Park. These iconic green spaces and landscapes are enjoyed by millions every year. In July 2019, London was declared the world's first National Park City with plans to make over 50 percent of the city green by 2050. During the COVID-19 pandemic, UGS in Greater London played a crucial role in providing essential spaces for physical and mental health benefits (Cui et al., 2022; Lee et al., 2023), as well as facilitating social interaction while adhering to social distancing guidelines. Therefore, this study selected Greater London to explore the topics expressed by UGS users during the pandemic and how these topics evolved over time.

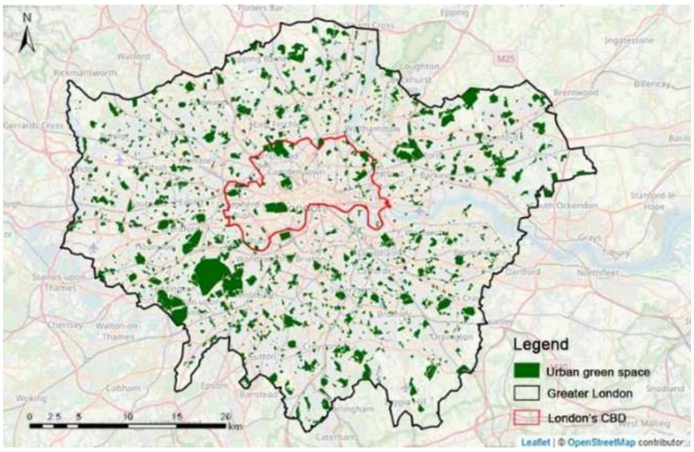

Figure 1. The spatial distribution of open green space in Greater London (Cui et al., 2022)

3.2 Data collection and pre-processing

This study utilized Twitter academic research application programming interface (API) to collect Twitter dataset, this API provides many new features such as the ability to access the full historical Tweets dating back to March 2006, allowing researchers to obtain more complete and unbiased data than previous Twitter APIs. This API provides a feature that has been used to exclude Tweets that were posted through the advertisers or at business.x.com, this feature can enhance the data quality especially for the current study. The dataset covered a three-month duration spanning from March 23rd to June 23rd for three consecutive years: 2019, 2020, and 2021. This period was chosen to capture the first lockdown period in the UK in 2020 and the increase spring related outdoor activity.

The aim of the research is to explore the topics and sentiment expressed by UGS visitors when they visited UGSs. Therefore we extract all Tweets in UGS areas regardless of the message content, a user's message does not need to explicitly discuss UGS for inclusion in the collected data. Geo-referenced Tweets in London were extracted using a bounding box covering the Greater London area. The study utilized the Open Greenspace layer sourced from the Ordnance Survey (Ordnance Survey, 2021) to overlay and filter Tweets situated within UGSs in Greater London, as depicted in Figure 1 showcasing the spatial distribution of these areas.Text mining is the process of extracting meaningful information and deriving insightful knowledge from unstructured textual data, such as Tweets (Müller et al., 2023). There are three main steps in the process of text mining, including dataset preprocessing, text representation and information extraction (Hu & Liu, 2012). In detail, dataset preprocessing comprises data cleaning and text preprocessing. First of all, only English Tweets were selected. Secondly, users who posted identical Tweets more than three times were recognised as potential fake accounts (Lansley & Longley, 2016). All the Tweets posted by such accounts were removed from the dataset. Finally, the remaining Tweets were further cleaned by removing punctuations, numbers, URLs and English stop words. In the process of text-preprocessing, all words were stemmed and converted to lower case. All the dataset preprocessing and analysis in this study were undertaken using R software (Ihaka & Gentleman, 1996). The total number of original Tweets were 296,329, 207,412 and 145,019 in 2019, 2020 and 2021, respectively. After data cleaning, the final number of Tweets being analysed were 12,286, 8,645 and 5,955 in the three years respectively.

3.3 Structural topic modelling (STM)

Both of LDA and STM are generative probabilistic approaches of a corpus, they can be used for the purposes of identifying topics from unstructured textural data such as Tweets. STM was developed based on LDA by Roberts et al. (2019). This section introduces LDA before discussing STM. LDA assumes that, for a given corpus containing a set of Tweets, each Tweet consists of multiple topics, where each topic can be represented by a distribution of words. These words are sorted by their probabilities of occurrence within that topic (Blei et al., 2003). Based on this assumption, each Tweet in a corpus is generated by following steps (Figure 2).

1. For each of the topics, choose a multinomial distribution from a Dirichlet distribution with parameter β. This process is depicted in the inner box in Figure 2.
2. For each of the Tweets, choose a multinomial distribution θ_d from a Dirichlet distribution with parameter α. This process is depicted in the outer box in Figure 2.
3. For each of the words w in a Tweet d,
   (a) select a topic z from Multinomial distribution (θ_d).
   (b) select a word w from $p(w_n|z_n, \beta)$, a multinomial probability conditioned on the topic z_n.

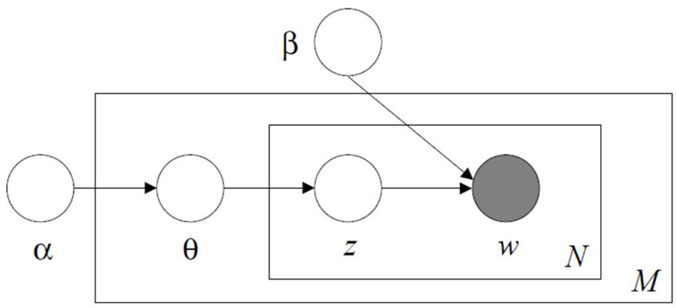

Figure 2. Graphical model representation of LDA (Blei et al., 2003)

In above processes, variable θ_d estimates the probability of a topic z occurring in a Tweet d, parameter β estimates the probability of a word w occurring in a topic z. The two types of probabilities are calculated by using the LDA algorithm below:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n, \beta) \quad (1)$$

Where the parameters α and β are corpus-level parameters, assumed to be selected during corpus generation. The variable θ are document-level parameters, and the z and w are word-level variables which are selected once for each word in each Tweet. In LDA, the topic distributions in all Tweets share a common Dirichlet prior α and the word distributions of topics share a common Dirichlet prior β as well.

STM extends this basic model, in STM, θ_d is a random variable drawn from a Log-normal distribution that is based on document-level data (Roberts et al., 2014). This enable researchers to explore the evolution or dynamic changes in the detected topics by integrating covariates such as time and other structural variables among Tweets. In this study, the research time periods were divided as weekly patterns and the '*weeknumber'* was used as time covariate in STM models. For example, week 1 in 2020 refers to from March 23rd to 29th, and week 2 refers to from March 30th to April 5th. There were 15 weeks for each year. Weekly topic proportions were calculated to capture the temporal patterns of topics to reveal the impact of COVID-19 on topics in relation to UGS use.

Additionally, 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 et al., 2014). Topic prevalence within Tweets and word distribution within topics is defined by covariate data including metadata. STM uses a variational expectation maximization (VEM) method to estimate the parameters. Further technical details on STM are provided in (Roberts et al., 2014). In this analysis, the *stm* R package (Roberts et al., 2019) was used to generate 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 (Kuhn, 2018; Roberts et al., 2019). In this study, the search k algorithm in the *stm* package was used to determine the optimal number of topics (Roberts et al., 2019; Roberts et al., 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.

The analyses of the STM results included the interpretation of topic contents and analyses of topic proportions. The interpretations were based on the highest probability words and frequency–exclusivity (FREX) words (Roberts et al., 2019). The FREX metric evaluates terms and allows researchers to identify the distinguishing words of each topic. Here these were used to label the topics and two example Tweets were selected for each topic to illustrate each label. The analyses of topic proportions and their trends were conducted based on the θ matrixes generated by STM (Chen et al., 2020; Roberts et al., 2019). In this study, for Tweet d, the probability of d being assigned to topic z was θ_{dz}; for topic z, the proportion of this topic within the corpus was denoted as $\theta_{1z}+\theta_{2z}+\theta_{3z}+\ldots+\theta_{dz}$. The weekly topic proportions were finally calculated by integrating time covariate in STM models (Roberts et al., 2019). To examine whether the topic proportions increased or decreased over the three-year period, STM employed a function that used linear regression to estimate the relationship between time and topic proportions. The results allows us to compare the changes of topic proportions across the three years and can also clearly show how the topic proportions changed within each year, this method has been used by previous studies in relation to STM topic models (Chen et al., 2020).

3.4 Dynamics in spatial and temporal patterns of topics

The probabilities of each topic occurring within each Tweet were calculated by using STM algorithm in the above section, then each of the Tweet was assigned to the highest probable topic. For example, there are 5 topics in a corpus containing a set of Tweets, the probability values of the $Topic_{1...5}$ to Tweet *d* are 15%, 30%, 40%, 5%, and 10%, respectively, thus Tweet *d* should be assigned to topic 3. In order to investigate the spatial patterns of the identified topics, the geographical variations of each topic over the three years across London were investigated by using spatial interpolation method.

Spatial interpolation refers to the estimation of the values of a main variable at locations situated within the region based on the data collected from specific sampled points within that same area (Gómez-Losada et al., 2019). There are various spatial interpolation methods (Li & Heap, 2014) of which inverse distance weighting (IDW) is one of the most popular or frequently used method, which has been applied in many disciplines (Li & Heap, 2014), such as assessing air pollution and water quality (Gu et al., 2021; Khouni et al., 2021). However, previous studies rarely employed IDW to investigate the spatial patterns of UGS-topics. 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:

$$Z = \sum_{i=1}^{n} \frac{1}{(d_1)^p} Z_i \div \sum_{i=1}^{n} \frac{1}{(d_1)^p} \quad (2)$$

where Z_i refers to the height value of a reference point, p is used as an index to calculate the weight of the reference point, with p set to 2; n refers to the total number of reference points located in the nearest neighbourhood; $(d_i)^p$ denotes the distance between points with known and unknown height values (Abdulmanov et al., 2021). The computational implementation of the IDW was performed using the *gstat* package (Gräler et al., 2016) from R software.

4. Results

4.1 Topics within UGSs

The results of diagnostic tests show the relative goodness of fit for each number of topics in Figure 3. The Held-out likelihood denotes the logarithmic probability that the topics present in the test set accurately reproduce the topics identified in the training set (Wallach et al., 2009). Similar to cross-validation, this method involves excluding a portion of the original dataset during estimation, which is later utilized for validation. This approach allows researchers to assess the model's predictive capacity. The lower bound indicates the minimum value of the marginal log likelihood. Residuals represent the variance between anticipated (training set) and projected (test set) topic predictions. Semantic coherence relates to the co-occurrence of words together within individual topics. In this figure, relatively high Held-out likelihood and high Lower Bound with low residuals were found when 7 was topic number. Therefore 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.

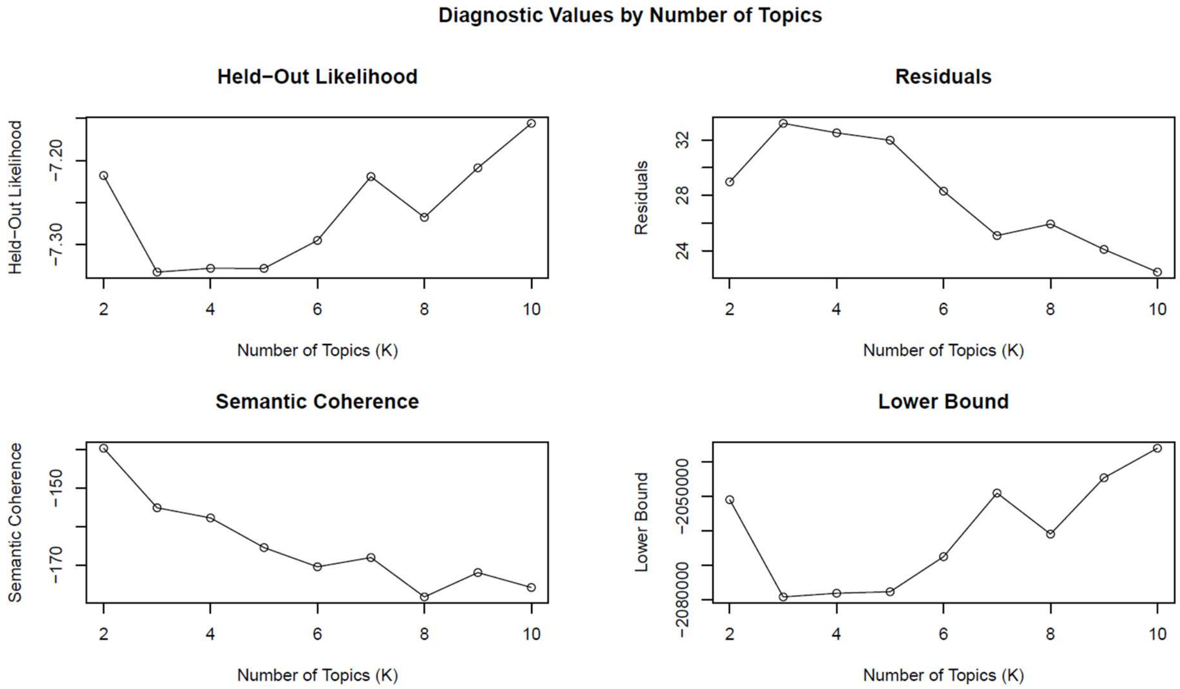

Figure 3. The relative goodness of fit for each topic number.

The total number of Tweets under each of the topics were 3,636, 4,109, 3,922, 4,396, 3,776, 5,688, and 3,894 for topics 1 to 7, respectively. Topic 6 had the highest number of Tweets, suggesting that this type of topic and related activities were more frequently mentioned or participated in. Table 1 shows the labels assigned to the 7 topics that were identified with the top 10 highest probability words and FREX words. For example, Topic 1 was labelled *Lockdown and exercise as* the co-occurrence words in the two types of words (highest probability and FREX) within this topic include "lockdown", "stayhome", "train", "fit" and 'exercise', all these words were lockdown- or exercise- related. Next, the sample Tweets were identified to verify the representativeness of each label. For example, there are two example Tweets that are related to this topic "*People take in their daily exercise at Greenwich Park, South East London during the national lockdown due to the Coronavirus outbreak.*" and "*stay at home and take daily exercise #coronadays #exerciseonceaday #socialdistancing*". Following this approach, all the labels were determined and used to name the topics.

Table 1. Topic labels and top 10 most probable and FREX words

Topic name	Topic words	
1. Lockdown and exercise	Most Probable	lockdown, good, make, night, well, train, home, fit, exercise, last
	FREX	fit, lockdown, covid, stayhome, body, train, exercise, social, drink, night
2. Sport and music	Most Probable	day, run, happy, music, weekend, festival, house, start, artist, marathon
	FREX	run, hackney, sport, festival, marathon, marsh, mile, birthday, bio, bank
3. Crowd events	Most Probable	year, easter, see, today, first, point, people, amazing, club, friend
	FREX	point, people, year, queen, east, Olympic, club, march, first, win

4. Dog walking	Most Probable	walk, love, dog, common, evening, like, morning, tree, enjoy, cockerspaniel
	FREX	dog, walk, cockerspaniel, Wimbledon, spaniel, se, village, labrador, instagram, golf
5. Social events	Most Probable	get, one, back, look, go, time, thank, week, wedding, cake, food
	FREX	food, wed, class, help, post, know, bexley, get, create, tri
6. Nature engagement	Most Probable	park, garden, beauty, nature, spring, flower, photo, photography, life, blossom
	FREX	graden, nature, park, zoo, swan, er, chihuli, flower, wildlife, hyde
7. Art and exhibition	Most Probable	new, work, design, museum, open, made, show, think, exhibition, draw
	FREX	design, exhibit, tate, crystal, museum, room, palace, draw, kubrick, bit

Topic 2 was characterized by key words including ‘run’, ‘marathon’, ‘sport’, ‘music’, and ‘festival’; all words largely related to sports- and music- related activities. Hence Topic 2 was labelled *Sport and music*. Topic 3 was labelled *Crowd events* with topic words of ‘Easter’, ‘today’, ‘people’, and ‘club’. These words are associated with gatherings, celebrations and festivals, indicating that people were likely participating in crowd events when they posted Tweets. Topic 4 was labelled *Dog walking* as it is characterized by words related to walking-dogs and some specific dog breeds e.g. ‘walk’, ‘dog’, ‘spaniel’, and ‘labrador’, suggesting that people frequently shared dog-related content via Twitter when they visited UGS with their dogs. Topic 5 was labelled *Social events* with topic words such as ‘thank’, ‘wedding’, ‘cake’, and ‘food’. These terms imply that the topic is likely associated with social gatherings involving friends or family. Topic 6 was labelled *Nature engagement* with topic words of ‘nature’, ‘flower’, ‘blossom’, ‘zoo’, and ‘wildlife’. These nature-related terms indicate that UGS visitors were likely observing natural settings, such as plants, flowers, and animals, during their visits to UGS. Topic 7 was labelled *Art and exhibition* with topic words of ‘design’, ‘museum’, ‘exhibition’, and ‘show’. These terms suggest that some users were likely engaging in art-related activities during their visits to UGS. Table 2 provides two example Tweets for each topic, which can be used to evaluate the representativeness of the assigned topic labels.

Table 2. Example Tweets under each of the topics

Topic name	Example Tweet 1	Example Tweet 2
1. Lockdown and exercise	*“People take in their daily exercise at Greenwich Park, South East London during the national lockdown due to the Coronavirus outbreak”*	*“stay at home and take daily exercise. #exerciseonceaday #socialdistancing”.*
2. Sport and music	*“Parkrun and a bit of old railway (sort of) ! 18:11 for my best performance by some way in 3 years”*	*“The feel is good after a 5k runs”*
3. Crowd events	*“What a fun night! #muse #simulationtheory #london #uk #amazing #fan #show #stadium”*	*“#Put it To The People #peoples vote #fightbrexit half way up Park Lane”*
4. Dog walking	*“Arwen love playing fetch #doggydaycare #doggiedaycare”*	*“I like taking pictures with him. #dog #family #love #mansbestfriend”*
5. Social events	*“There are times that you need good friends and there are times that friend need you.”*	*“Wedding guest with an Apple iPad. #mayfair #hydepark”*
6. Nature engagement	*“Lovely #spring morning view from the window this morning”*	*“Beautiful nature: #kewgardens #trees”*
7. Art and exhibition	*“My favourite pieces from our trip to @kewgardens. What incredible works of art. ”*	*“I'm at Imperial War Museum”*

Figure 4 shows the 7 identified topics with their relative proportions to the overall distribution of topics. The figure shows that the topics are ranked in a descending order, with the highest associated words within each topic, and the proportion of each topic to all topics. The most popular topic among all topics was Topic 6, named *Nature engagement,* followed by *Social events*, *Crowd events, Art and exhibition, Sport and music events, Lockdown and exercise*, and *Dog walking.*

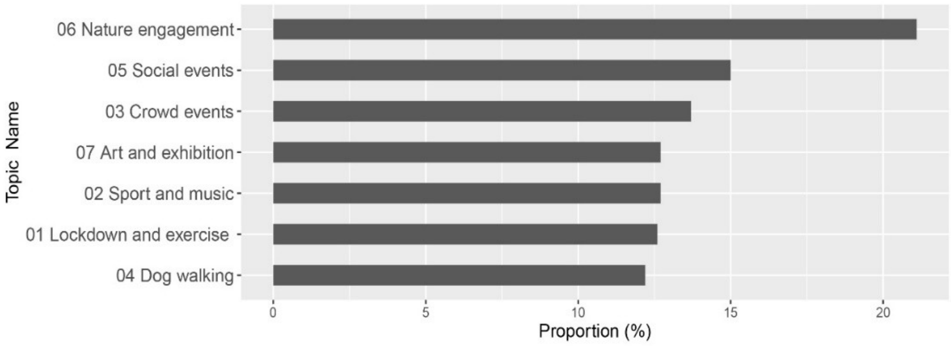

Figure 4. The proportions of all identified topics

4.2 The evolution of topics over time

Among all identified topics, the topics 4, 5, and 6 were selected for the purpose of understanding how the pandemic influenced the UGS users' topics when they visited UGS areas. Specifically, topic 4 was because *Dog-walking* and related activities became one of the most important UGS activities during the COVID-19 pandemics (Owczarczak-Garstecka et al., 2021). A number of studies have highlighted that dog-walking activities provided essential mental and physical health benefits for both owners and their pets, particularly during periods of restrictions (Hoy et al., 2024; Oliva & Johnston, 2021; Owczarczak-Garstecka et al., 2021). Topics 5 and 6 were chosen as they represented the two largest proportions among all topics in this study.

Figure 5 shows the topics' proportions and their trends over the study periods. In detail, the proportions of *Dog walking* (Topic 4) show an increasing trend from 2019 to 2020. This means 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, UGS users took their dog outside both in the morning and evening according to the key words of this topic (see Table 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 *Nature engagement* (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.

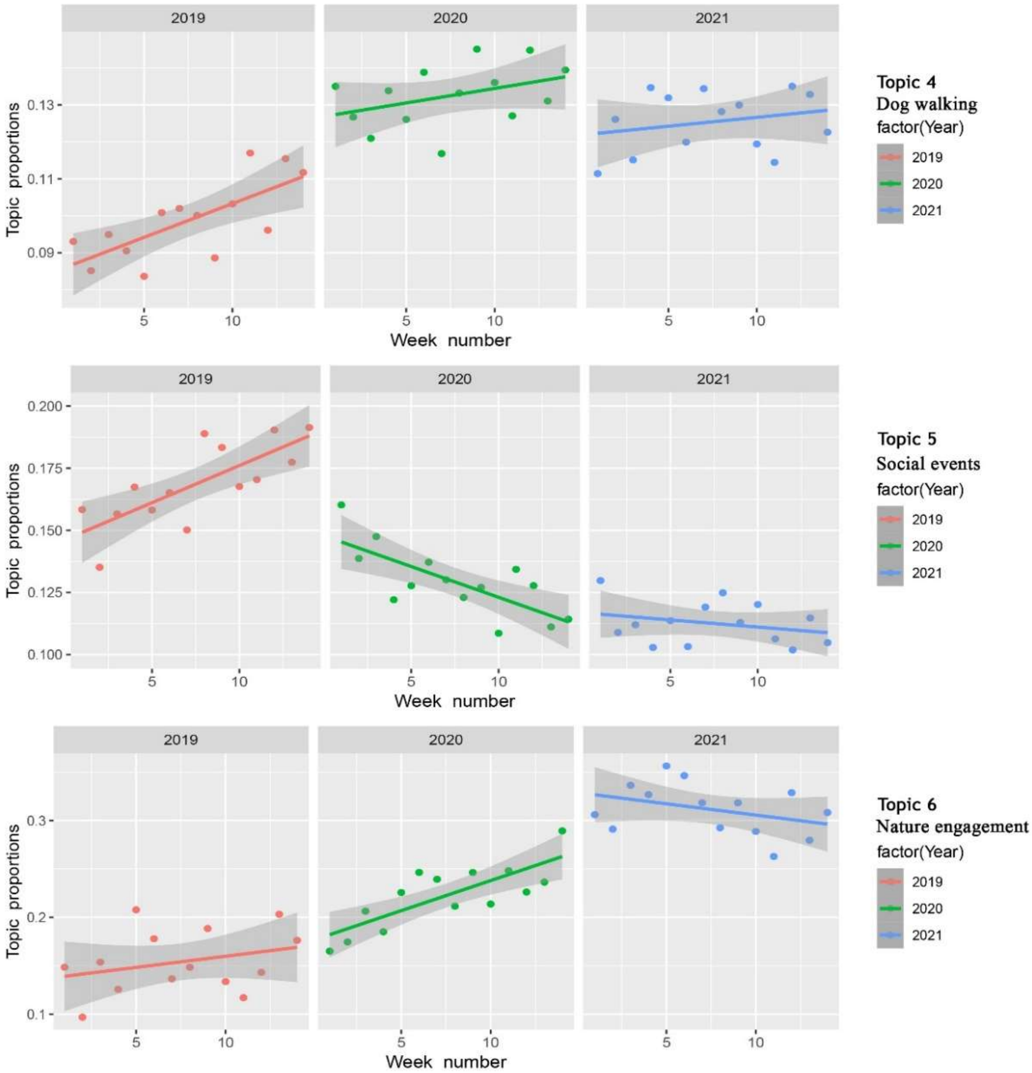

Figure 5. Trends of the topic proportions

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 pandemic has affected the topics discussed in the analysed texts. Figure 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 the COVID-19 pandemic. 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 indicates the likelihood of the activity occurring in a specific timeframe (prior to or during the COVID-19 pandemic). Additionally, the word sizes convey the extent of correlation between the activity and the corresponding period, where larger sizes signify stronger correlations.

Specifically consider Topic 4 (*Dog walking*). Before the COVID-19 period, this topic frequently expressed words about enjoyment such as ‘love’ and ‘like’. For example, the Tweets “*Arwen love playing fetch #doggydaycare #doggiedaycare*” and “*I like taking pictures with* *him. #dog #family #love #mansbestfriend*" expressed more love- and like-related words. However, during COVID-19 pandemic, words of 'walk', 'dog', 'morning' and 'daily' were frequently mentioned. For example, Tweets "*Today's daily walk: So much quieter this morning, lovely weather, room to walk, run and play with the dog! #MorningWalk #DogWalk*" and "*Beautiful Easter Friday morning dog walk*" expressed walking dog-related activities. This indicates that UGS visitors tended to undertake more specifically dog-walking related activities during the pandemic periods, compared to those before.

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 pandemic, 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 pandemic, 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.

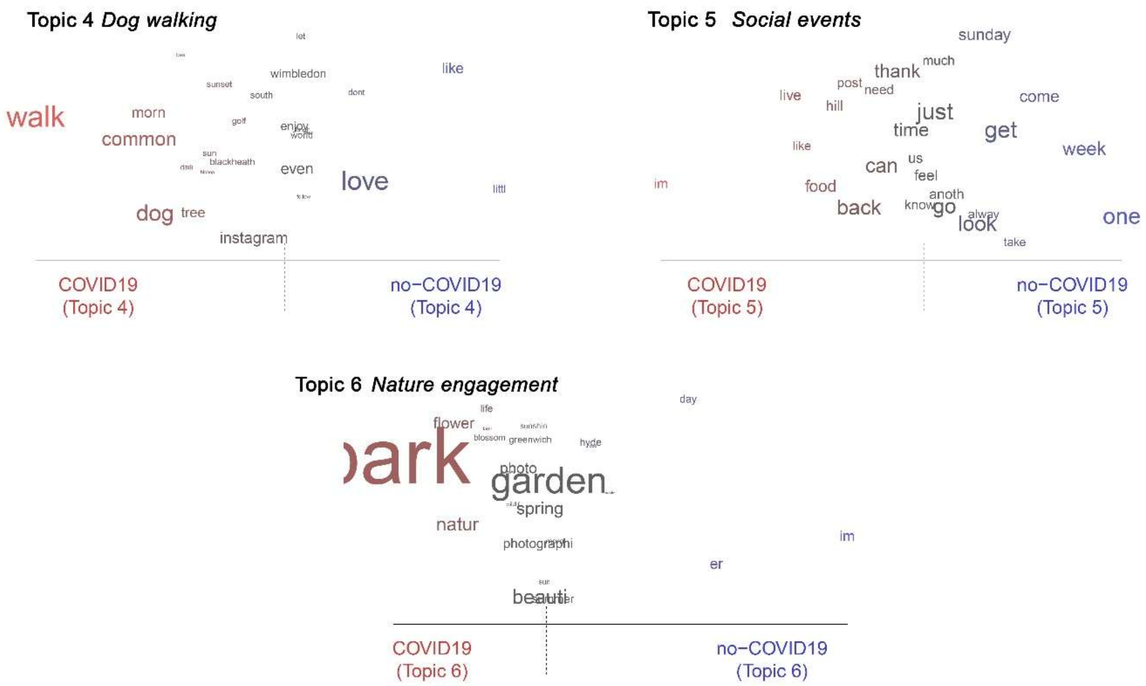

Figure 6. Graphical display of topical perspectives with content covariate of COVID-19

4.3 Dynamics in spatial patterns of topics

Figure 7 shows the spatial variations of topics *Dog walking, Social events, and Nature engagement*, with darker colours indicating a higher estimated value of that topic at that location. 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 relatively lower topic probability values across the whole study area, indicating that the related activities were less frequently compared to other types of UGS activities, even though some hotspots located in the southwest part of London were found. However, the probability values 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 (Owczarczak-Garstecka et al., 2021). In 2021, the probability remains higher than that in 2019, indicating that this type of activities remains attractive compared to the before-COVID 19 period.

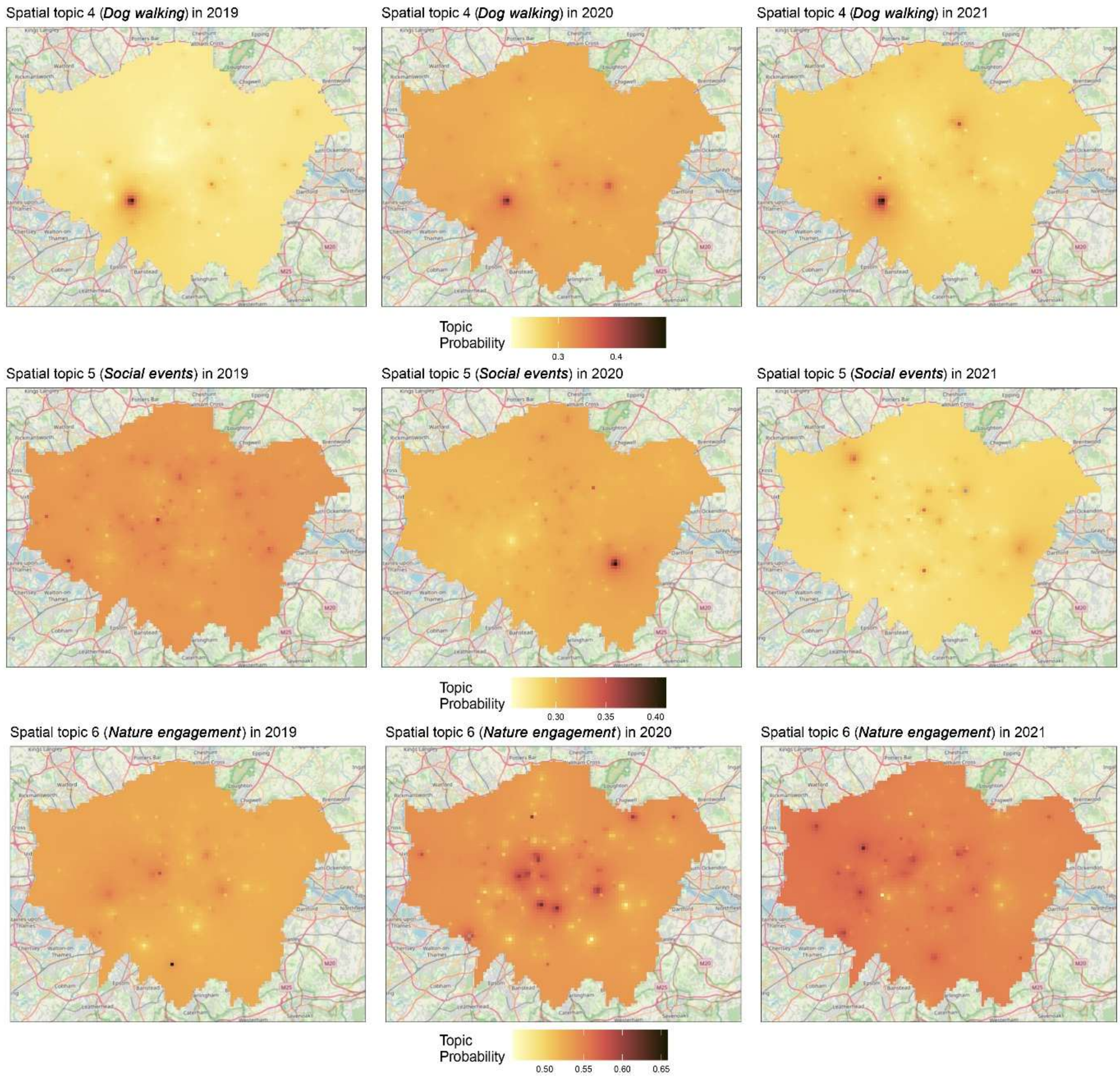

Figure 7 Spatial-temporal patterns of the topics *Dog walking, Social events,* and *Nature engagement*

The topic *Social events* shows 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 (Houlden et al., 2019).

5. Discussion

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

5.1 What topics have been expressed through Tweets that were posted within UGSs.

Seven UGS-related topics were identified in this study, of which *Nature engagement* and *Social events* accounted the top two proportions, followed by *Crowd events, Art and exhibition, Sport events, Lockdown and exercise*, and *Dog walking.* The seven types of topics suggested that UGS users engaged with various UGS and related activities when they visited UGSs. It is not surprise that nature and physical related topics and activities were identified in UGSs. Previous studies highlighted the importance of physical activities within UGSs and their benefits to human wellbeing (Cui et al., 2022; Lopez et al., 2021). Unlike these studies, the findings of this study showed that UGS visitors also engaged with a range of different activities such as *Crowd events* and *Art exhibitions* which are not ostensibly directly related with UGS, reflecting the role of UGS in supporting diverse arts and cultural events (Van der Hoeven & Hitters, 2019). The selection and analysis of all Tweets within UGS areas, without filtering for keywords, enabled the STM to generate a more comprehensive set of topics and suggest potential activities and related facilities that UGS planners could focus on improving.

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

This study selected topics *Dog walking, Social events,* and *Nature engagement* to investigate the topics' evolution over space and time. 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. Owczarczak-Garstecka et al. (2021) found that people spent more time on dog walking activities during the pandemic, and overall duration of dog-walking activities did not significantly change, particularly in households with multi-person households. The benefits of dog-walking to reduce stress and have been highlighted by a number of studies. For example, Hoy et al. (2024) suggested that dogs can provide psychological and social support to their owners, as well as motivate them to engage in physical activity. Oliva and Johnston (2021) suggested that dog walking can be a reliable and consistent form of physical activity, even during periods of social distancing and quarantine.

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 decreasing social activities when they visited UGS. In addition, these restriction measures may cause a series of health problems such as increasing the risk of depression especially for vulnerable communities such as elderly population and people living in compact neighbourhoods (Marchi et al., 2022).

The topic *Nature engagement* displayed an increasing trend in topic proportion, indicating that the activities related to observing 'flower', 'blossom', 'nature' and 'wildlife' seemed became increasingly popular during COVID-19 pandemic. Previous studies have highlighted the numerous benefits of nature observation and related activities, as these activities help people alleviate self-isolation issues such as depression and anxiety (Houlden et al., 2019; Marchi et al., 2022). A number of studies also found that people living in urban areas increased UGS-related activities during the pandemic (Geng et al., 2021; Marchi et al., 2022). Similarly, the spatial patterns of this topic displayed an increasing trend across London, suggesting that pandemic-related restrictions impacted not only urban centers but also peripheral urban areas.

5.3 Approaches for tracking dynamics in topics

Previous studies used key-words analysis (Sim & Miller, 2019) and manually methods (Roberts, 2017) to detect UGS-related activities through Tweet texts. These approaches are time and energy consuming, and may lack the ability to systematically identify and capture details from dataset. The prevailing event-detection techniques based-on LDA have demonstrated satisfactory performance across various applications (Fu et al., 2018). Nonetheless, while LDA serves as a valuable method for topic detection, it lacks the capability to address the dynamics of topics concerning time and space. The current study employed STM to detect the topics, which allowed researchers to examine how the covariate variable (whether or not during COVID-19 pandemic) influenced the UGS topics over time. By utilizing STM, the study was able to capture more specific variations in UGS topics and related activities.

Previous studies using social media data to track UGS visits revealed diverse trends during lockdown but did not capture local spatial variations in specific UGS topics and activities (Geng et al., 2021; Lopez et al., 2021). To investigate the spatial changes of topics, this study used IDW as an interpolation method to estimate the topic probability across the London study area, to reveal the relative spatial patterns of topics incidence. 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 spatially informative analysis compared to other methods, particularly when considering the topic probability.

5.4 Limitations and future research

This study has several limitations which could be considered in a future analysis. Regarding study periods, this study mainly focused on the first national lockdown period in the UK, future analysis could cover longer study periods such as the whole three years from 2019 to 2021, thereby expanding the analyses of the UGS-related topics and their dynamics over time. Additionally, previous studies suggested that Twitter may represent only a partial topics of general public (Marchi et al., 2022). Future research could broaden the data collection scope by encompassing additional regions and alternative social media platforms to examine the applicability of this approach in a more comprehensive manner. Only geo-referenced Tweets were selected for the purpose of exploring spatial patterns of topics but non-geo-referenced Tweets could also be included by linking to users' home locations.

The current study focused on the temporal changes in activities through topic patterns rather than examining the specific number of users during each time period. Future research could explore the number of users and their tweets under each topic across three years to investigate the impact of the pandemic on visitor numbers and how individuals responded to the pandemic. While the COVID-19 period was utilized as a covariate in the STM model, upcoming studies could link to the socio-demographics of local people, the characteristics of their communities, and measures such as household accessibility. Finally, more complex analytical methods for examining topic prevalence could be applied in future studies. This would facilitate an exploration of how distinct user groups engage with UGS and how their experiences correlate with different topics. By addressing these limitations, future studies could provide a more comprehensive understanding of the dynamics of UGS-related topics and how they are influenced by various factors, especially during a crisis.

6. Conclusion

This study demonstrated how the combining text mining techniques, STM and IDW approaches, can be used to gain a comprehensive understanding of social media posts located within UGSs. Additionally, it investigated how the detected topics were impacted by social crises, such as COVID-19 outbreaks. This study provides a framework for identifying, analysing, and visualising topics expressed by social media users when they visited UGS. The analysis identified seven types of topics, of which *Nature engagement* was the most popular, followed by *Social events*, *Crowed events, Art and exhibition, Sport events, Lockdown and exercise*, and *Dog walking.*

During the COVID-19 lockdown period, the increasing trends of topic *Nature engagement* highlighted the importance of UGS to urban citizens, especially during a health emergency. UGS planners and managers are advised not only to protect existing UGS areas but also to enhance their quality of UGS by providing more user-friendly and sustainable facilities. For example, UGS managers could consider incorporating diverse vegetation types, such as gardens and various plant species, within UGS areas, rather than providing solely on trees and grasslands. Additionally, the accessibility of UGS is also crucial to urban residents and close-to-home UGSs can effectively and timely provide local communities with accessible spaces and recreational areas especially during lockdown periods (De Luca et al., 2021). To ensure the appropriate use of UGS during a health emergency, urban administrators and UGS managers could assess UGS usage or monitor visitor numbers, rather than having to close urban parks and open green spaces entirely. The increasing trend in *Dog-walking* also highlights the need for designing UGS that provide more dog-friendly places and facilities. Finally, these findings and insights could inform future urban environmental planning, rather than solely addressing lockdown-related issues.

The spatial patterns of the topics' probabilities also revealed various changing trends over three years. These findings not only help identify areas where topics have increased in probability but also enable local administrators to implement targeted measures for specific districts or even individual communities. For example, the hotspot areas of *Dog-walking* and related activities across London help UGS managers to effectively provide dog-friendly facilities and spaces, especially during a time of movement restrictions. Previous studies in other metropolitan areas have also conducted big-data analyses to provide targeted-insights into UGS planning and management for specific regions or demographic groups. For example, Guan and Zhou (2024) examined access and inequalities among residents across Tokyo's 23 special wards and various visitor groups. Their findings suggested that future urban parks could focus on developing cycling paths to promote sustainable transport and help reduce carbon footprints. Wen et al. (2020) and Kim et al. (2023) examined spatial inequalities in access to UGS in Germany and the U.S., respectively, with a particular focus on elderly populations. Their findings suggested that enhancing mobility for the elderly could enable them to access high-quality green spaces located father away. Overall, by integrating demographic characteristics of local communities or neighbourhoods, urban planner and policy makers will be able to make more appropriate limitation rules and restrictions. Our study demonstrated an effective approach to examine the evolving spatial-temporal patterns of UGS topics. These insights are valuable for UGS planning and management, particularly during crises like the COVID-19 outbreak.

Acknowledgements

This study was supported by China Scholarship Council-University of Leeds Scholarship (201906390033) and the Ministry of Education, Singapore, under the Academic Research Fund Tier 1 (FY2022-FRC2-009). The authors thank the editor and the anonymous reviewers for their insightful comments and suggestions that have helped improve the quality of the study.

References

Abdulmanov, R., Miftakhov, I., Ishbulatov, M., Galeev, E., & 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*, 101970. https://doi.org/10.1016/j.eti.2021.101970

Blei, D., & Lafferty, J. (2006). Correlated topic models. *Advances in neural information processing systems*, *18*, 147.

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

Bogdanowicz, A., & Guan, C. (2022). Dynamic topic modeling of Twitter data during the COVID-19 pandemic. *PloS one*, *17*(5), e0268669. https://doi.org/10.1371/journal.pone.0268669

Chen, X., Zou, D., Cheng, G., & 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*, 103855. https://doi.org/10.1016/j.compedu.2020.103855

Cui, N., Malleson, N., Houlden, V., & 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), 425. https://doi.org/10.3390/ijgi10070425

Cui, N., Malleson, N., Houlden, V., & 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*, 127677. https://doi.org/10.1016/j.ufug.2022.127677

De Luca, C., Libetta, A., Conticelli, E., & Tondelli, S. (2021). Accessibility to and availability of urban green spaces (UGS) to support health and wellbeing during the COVID-19 pandemic—the case of Bologna. *Sustainability*, *13*(19), 11054. https://doi.org/10.3390/su131911054

Ding, K., Choo, W. C., Ng, K. Y., & Zhang, Q. (2023). Exploring changes in guest preferences for Airbnb accommodation with different levels of sharing and prices: Using structural topic model. *Frontiers in psychology*, *14*, 1120845. https://doi.org/10.3389/fpsyg.2023.1120845

Du, Y., Yi, Y., Li, X., Chen, X., Fan, Y., & Su, F. (2020). Extracting and tracking hot topics of micro-blogs based on improved Latent Dirichlet Allocation. *Engineering Applications of Artificial Intelligence*, *87*, 103279. https://doi.org/10.1016/j.engappai.2019.103279

Fu, C., McKenzie, G., Frias-Martinez, V., & Stewart, K. (2018). Identifying spatiotemporal urban activities through linguistic signatures. *Computers, Environment and Urban Systems*, *72*, 25-37. https://doi.org/10.1016/j.compenvurbsys.2018.07.003

Geng, D., Innes, J., Wu, W., & Wang, G. (2021). Impacts of COVID-19 pandemic on urban park visitation: a global analysis. *Journal of forestry research*, *32*, 553-567. https://doi.org/10.1007/s11676-020-01249-w

Ghosh, A., Nundy, S., Ghosh, S., & Mallick, T. K. (2020). Study of COVID-19 pandemic in London (UK) from urban context. *Cities*, *106*, 102928. https://doi.org/10.1016/j.cities.2020.102928

Gómez-Losada, Á., Santos, F. M., Gibert, K., & 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*, 1-11. https://doi.org/10.1016/j.compenvurbsys.2018.12.005

Gräler, B., Pebesma, E. J., & Heuvelink, G. B. (2016). Spatio-temporal interpolation using gstat. *the R Journal*, *8*(1), 204-218. https://doi.org/10.32614/RJ-2016-014

Grzyb, T., Kulczyk, S., Derek, M., & Woźniak, E. (2021). Using social media to assess recreation across urban green spaces in times of abrupt change. *Ecosystem Services*, *49*, 101297. https://doi.org/10.1016/j.ecoser.2021.101297

Gu, K., Zhou, Y., Sun, H., Dong, F., & 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*, 1-22. https://doi.org/10.1007/s10661-020-08749-6

Guan, C., & Zhou, Y. (2024). Exploring environmental equity and visitation disparities in peri-urban parks: A mobile phone data-driven analysis in Tokyo. *Landscape and urban planning*, *248*, 105104. https://doi.org/10.1016/j.landurbplan.2024.105104

Houlden, V., de Albuquerque, J. P., Weich, S., & Jarvis, S. (2019). A spatial analysis of proximate greenspace and mental wellbeing in London. *Applied Geography*, *109*, 102036. https://doi.org/10.1016/j.apgeog.2019.102036

Hoy, L. S., Stangl, B., & Morgan, N. (2024). Leisure with dogs in the UK: the importance of shared outdoor leisure spaces highlighted by the COVID-19 pandemic. *Leisure/Loisir*, 1-23. https://doi.org/10.1080/14927713.2024.2308919

Hu, X., & Liu, H. (2012). Text analytics in social media. *Mining text data*, 385-414. https://doi.org/10.1007/978-1-4614-3223-4_12

Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, *5*(3), 299-314. https://doi.org/10.1080/10618600.1996.10474713

Khouni, I., Louhichi, G., & Ghrabi, A. (2021). Use of GIS based Inverse Distance Weighted interpolation to assess surface water quality: Case of Wadi El Bey, Tunisia. *Environmental Technology & Innovation*, *24*, 101892. https://doi.org/10.1016/j.eti.2021.101892

Kim, Y., Corley, E. A., Won, Y., & Kim, J. (2023). Green space access and visitation disparities in the phoenix metropolitan area. *Landscape and urban planning*, *237*, 104805. https://doi.org/10.1016/j.landurbplan.2023.104805

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*, 105-122. https://doi.org/10.1016/j.trc.2017.12.018

Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, *58*, 85-96. https://doi.org/10.1016/j.compenvurbsys.2016.04.002

Lee, K. O., Mai, K. M., & Park, S. (2023). Green space accessibility helps buffer declined mental health during the COVID-19 pandemic: evidence from big data in the United Kingdom. *Nature Mental Health*, *1*(2), 124-134. https://doi.org/10.1038/s44220-023-00018-y

Li, J., & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, *53*, 173-189. https://doi.org/10.1016/j.envsoft.2013.12.008

Lopez, B., Kennedy, C., Field, C., & 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*, 127354. https://doi.org/10.1016/j.ufug.2021.127354

Mahtta, R., Fragkias, M., Güneralp, B., Mahendra, A., Reba, M., Wentz, E. A., & Seto, K. C. (2022). Urban land expansion: the role of population and economic growth for 300+ cities. *Npj Urban Sustainability*, *2*(1), 5. https://doi.org/10.1038/s42949-022-00048-y

Marchi, V., Speak, A., Ugolini, F., Sanesi, G., Carrus, G., & Salbitano, F. (2022). Attitudes towards urban green during the COVID-19 pandemic via Twitter. *Cities*, *126*, 103707. https://doi.org/10.1016/j.cities.2022.103707

Müller, M., Salathé, M., & Kummervold, P. E. (2023). Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. *Frontiers in artificial intelligence*, *6*, 1023281. https://doi.org/10.3389/frai.2023.1023281

Office for National Statistics. (2022). *Mid-Year Population Estimates, UK, June 2022*. https://www.ons.gov.uk/

Oliva, J. L., & Johnston, K. L. (2021). Puppy love in the time of Corona: Dog ownership protects against loneliness for those living alone during the COVID-19 lockdown. *International Journal of Social Psychiatry*, *67*(3), 232-242. https://doi.org/10.1177/0020764020944195

Ordnance Survey. (2021). *OS Open Greenspace*. https://www.ordnancesurvey.co.uk/products/os-open-greenspace

Owczarczak-Garstecka, S. C., Graham, T. M., Archer, D. C., & 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), 6315. https://doi.org/10.3390/ijerph18126315

Roberts, H. V. (2017). Using Twitter data in urban green space research: A case study and critical evaluation. *Applied Geography*, *81*, 13-20. https://doi.org/10.1016/j.apgeog.2017.02.008

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of statistical software*, *91*, 1-40. https://doi.org/10.18637/jss.v091.i02

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American journal of political science*, *58*(4), 1064-1082. https://doi.org/10.1111/ajps.12103

Sachdeva, S., McCaffrey, S., & Locke, D. (2017). Social media approaches to modeling wildfire smoke dispersion: Spatiotemporal and social scientific investigations. *Information, Communication & Society*, *20*(8), 1146-1161. https://doi.org/10.1080/1369118X.2016.1218528

Sikorska, D., Wojnowska-Heciak, M., Heciak, J., Bukowska, J., Łaszkiewicz, E., Hopkins, R. J., & Sikorski, P. (2023). Rethinking urban green spaces for urban resilience. Do green spaces need adaptation to meet public post-covid expectations? *Urban Forestry & Urban Greening*, *80*, 127838. https://doi.org/10.1016/j.ufug.2023.127838

Sim, J., & Miller, P. (2019). Understanding an urban park through big data. *International Journal of Environmental Research and Public Health*, *16*(20), 3816. https://doi.org/10.3390/ijerph16203816

Taczanowska, K., Tansil, D., Wilfer, J., & Jiricka-Pürrer, A. (2024). The impact of age on people's use and perception of urban green spaces and their effect on personal health and wellbeing during the COVID-19 pandemic—A case study of the metropolitan area of Vienna, Austria. *Cities*, *147*, 104798. https://doi.org/10.1016/j.cities.2024.104798

Van der Hoeven, A., & Hitters, E. (2019). The social and cultural values of live music: Sustaining urban live music ecologies. *Cities*, *90*, 263-271. https://doi.org/10.1016/j.cities.2019.02.015

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. *Proceedings of the 26th annual international conference on machine learning*, 1105-1112. https://doi.org/10.1145/1553374.1553515

Wen, C., Albert, C., & Von Haaren, C. (2020). Equality in access to urban green spaces: A case study in Hannover, Germany, with a focus on the elderly population. *Urban Forestry & Urban Greening*, *55*, 126820. https://doi.org/10.1016/j.ufug.2020.126820

Wilson, B., Neale, C., & Roe, J. (2024). Urban green space access, social cohesion, and mental health outcomes before and during Covid-19. *Cities*, *152*, 105173. https://doi.org/10.1016/j.cities.2024.105173