



The impact of the COVID-19 pandemic on the dynamics of topics in urban green space

Nan Cui¹, Nick Malleson¹, Vikki Houlden¹ and Alexis Comber¹

¹ School of Geography, University of Leeds, Leeds, United Kingdom

Correspondence: Nan Cui (gync@leeds.ac.uk)

Abstract

Urban residents' daily lives have been impacted by the COVID-19 pandemic in various aspects such as social, leisure, and physical activities. Fortunately, urban green spaces (UGSs) have become a main outdoor destination, due to the policies encouraging people to visit UGS and keeping them open. This study aimed to comprehensively investigate the impact of the COVID-19 pandemic on topics discussed on social media by UGS visitors over space and time. Data was collected from geo-referenced Tweets across London in spring 2019, 2020, and 2021. Structural Topic Modelling (STM) was used to identify UGS topics and describe the dynamics of topic proportions. The inverse distance weighted (IDW) interpolation method was used to explore spatial distributions of all topics. The study identified seven main types of UGS topics over all study periods, with topics such as *Lockdown and exercise* and *Social and friends* showing a decreasing trend in topic proportions, indicating that visitors' outdoor activities were restricted. The study not only identifies the main types of topics in UGS during the COVID-19 pandemic period but also reflects people's attitudes and perceptions towards restriction measures, which can provide guidance for future urban policies, especially during crises.

Keywords. Urban green space, Topic detection, Spatial-temporal analysis, COVID-19

1 Introduction

Urban green space (UGS) plays an essential role in supporting people's daily lives (Houlden et al., 2019). For example, UGS provides people with outdoor activity areas such as running tracks, tennis courts, and rowing places; and entertainment activity areas such as flower

gardens, grasslands, and picnic areas. During the COVID-19 pandemic, a series of public health measures were implemented to mitigate the spread of the virus, such as travel restrictions, quarantine, closing non-essential business services, and requiring citizens to stay at home (Cameron-Blake et al., 2020). UGS therefore became a major outdoor destination during the COVID-19 pandemic period (Wang and Li, 2022). In this context, questions about what types of topics people discussed and how these topics changed over space and time became increasingly popular research issues (Geng et al., 2021).

Traditional methods such as questionnaires and surveys can provide valuable information about the perceptions, attitudes, and behaviors of people towards UGS. However, these methods often face challenges related to their time and energy consumption (Cui et al., 2021). Popular social media platforms such as Twitter provide new data sources on important events, providing rich knowledge about urban systems and human dynamics (Ilieva and McPhearson, 2018). In this study, Twitter was selected as a data source due to its popularity and availability (Cui et al., 2021), and all Tweets used included geo-reference positions within UGS in Greater London.

STM was used to investigate the topics in UGS over the three years. STM is a type of topic modelling that allows researchers to analyse how the prevalence and distribution of topics in the corpus vary across different groups or time periods (Roberts et al., 2014a). This can be used to detect the dynamics of the topics over time and potentially reflect the impact of the COVID-19 pandemic on the interactions between humans and UGS.

This study aims to detect the impact of the COVID-19 pandemic on topics UGS. This can effectively track important information during accident events like the pandemic and help with the planning and management of

UGS. The analysis consists of three steps: data collection and data pre-processing, dynamic topics generation, and investigation of spatial-temporal patterns of topics. In detail, the first step was geo-referenced Twitter data collection, Tweets cleaning and data pre-processing. The second step refers to using STM to identify UGS topics and detect the trends in topic proportion and topic label words. The third step focuses on monitoring the changes in UGS topics in semantic, spatial, and temporal dimensions pre-, during- and after the COVID-19 pandemic.

2 Data collection and analysis

This section give the description of datasets and analysis tools. Twitter was selected as a data source and R was employed to conducted the data analysis.

2.1 Data collection and data pre-processing

The study used Twitter academic research API to download geo-referenced Tweets from London, covering a three-month (23rd March to 23rd June) period for 2019, 2020, and 2021. The Tweets were cleaned by selecting only English Tweets, removing Tweets with fewer than three words and stop-words, and eliminating those from bots, fake accounts (classifying users who post more than 10 UGS Tweets as fake accounts), and users who posted the same Tweet more than three times. Third, punctuation, URLs, and numbers were removed. Finally all words were converted to lower case (See Cui et al. (2022) for further details). In addition, urban green space layers (Open greenspace) were collected from Ordnance Survey (<https://www.ordnancesurvey.co.uk/>). Only the tweets geo-referenced in UGS areas were selected in this analysis.

R was used as a main software in data collection and pre-processing, topic detection, and spatial-temporal analysis (Ihaka and Gentleman, 1996). In detail, the *tm* package was used to conduct text mining works, such as importing data, cleaning text, creating corpus, and managing data (Feinerer, 2013). After data preprocessing, the topics were

identified by using *stm* package (Roberts et al., 2014a), which enables researchers to conduct a comprehensive investigation of text data (Tweets in the current study); the *stm* package also can explore the evolution of topics over time. In terms of spatial-temporal analysis, the *gstst* package (Pebesma, 2004) was used to estimate the probabilities of each Tweet belonging to each topic from a spatial perspective. Finally, packages such as *ggplot2* (Wickham et al., 2016) and *tmap* (Tennekes, 2018) were used for data visualization.

2.2 Dynamic topics generation

This study used Structural Topic modelling (STM) to detect the topics within collected Tweets. Topic modelling considers a single document as a combination of topics, and each topic as a distribution over words. STM performs two major tasks toward estimating the distributions of the document-topic θ_i and the Topic-term β_k , representing topic proportion and word distribution within topics, respectively. Further technical details on STM are provided in (Roberts et al., 2014a) and (Roberts et al., 2014b).

An important consideration for STM analysis is the a priori determination of the number of topics. The held-out likelihood estimation was used to estimate the probability of words appearing within a document when those words have been removed from the document in the estimation step (Wallach et al., 2009). The goodness-of-fit of STM models with varying numbers of topics were evaluated from 2 to 10, in 1 topic increments. Then the name of each topic was manually labelled according to the highest probability topic words.

2.3 Spatial-temporal analysis

IDW was used to detect the spatial patterns of the topics. This method is a kind of deterministic interpolation method that creates a continuous surface of values based on point data (Li and Heap, 2014). The probability values of each topic across the study area were calculated, this can potentially reflect the spatial variations of the identified topics. All Tweets were grouped by week which allows researchers to track the dynamics of topic

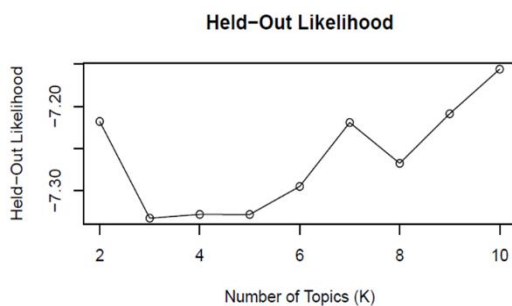


Figure 1. Figure 1 The number of topics.

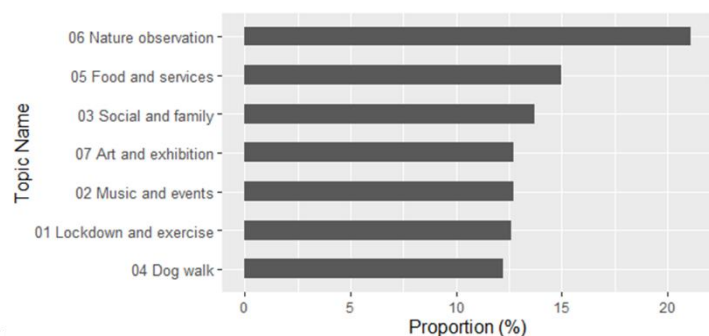


Figure 2 . Topic name and proportion.

prevalence on a weekly basis, this can potentially provide information for detecting how people response to UGS policies during the COVID-19 pandemic period.

3 Results

3.1 Topics in urban green space

The optimal number of topics for the STM analysis was 7 (Fig. 1). Figure 2 shows the proportions of each topic to all topics over all periods, and the corresponding name of each topic. Topic 6 with the name of *Nature observation* was the most popular topic among all topics, following by *Food and service*, *Social and family*, *Art and exhibition*, *Music and events*, *Lockdown and exercise*, and *Dog walking*.

To detect the impact of COVID-19 pandemics on UGS topics over space and time, the study focused on three representative topics as a case study. In detail, the topic 01 *Lockdown and exercise* was selected as this topic mainly described sports- and exercise-related activities before and after the COVID-19 outbreaks, which can potentially reveal how outdoor physical activities changed during the COVID-19 pandemic. The restriction measures such as stay at home, keeping fit and exercising can exert impacts on people’s park life. In addition, the restriction measures such as cancelling all public events including music festival and sport events also may also have a profound influence on people’s life, thus in the current study, topics 02 (*Music and events*) and 03 (*Social and family*) were selected for the further exploration. Table 1 show the top highest probability related words to each topic, which helps us to understand the meaning of each topic and explore related activities. Topic 1 (*Lockdown and exercise*) was represented by words *lockdown, night, home, train, fit and exercise*, indicating that this topic was correlated to physical health activities during the lockdown period. Topic 2 (*Music and events*) was represented by words *day, run, happy, music, artist and marathon*, indicating that this topic was correlated to public events in UGSs such as music festival and marathon. Topic 3 (*Social and family*) was represented by words *easter, people, club, and friend*, which indicates

this this topic was more related to social activities such as meeting friend and go club.

It should be noted that this study did not consider other topics such as the top frequent topic 6 *Nature observation*, this is because people can experience UGS in various ways such as observing animals including birds, swans and geese, or seeing plants such as flowers and trees. The authors aim to conduct further analysis on this specific topic by implementing a more detailed classification system in future analysis, which will potentially provide more precise information about the impact of COVID-19 on UGS.

3.2 Dynamics in topic proportions

To explore how the popularities of the topics 01, 02 and 03 changed pre-, during- and after the COVID-19 pandemic, the same 3 months of datasets from 23rd March to 23rd June in 2019, 2020 and 2021 were extracted, which can minimize the influence of seasonal and climate factors on the results to the greatest extent. Besides, all Tweets were grouped by week unit to show the weekly patterns of topic proportions over time periods.

Figure 3 shows the trends of topic proportions in the three topics, and corresponding covariate topic words as well. In proportions part, various trends in all three topics reflected t cancelling cancelling he impacts of covid-19 pandemic on UGS visitors’ daily lives. In covariate word part, the study period was separated as two parts: no COVID-19 period

(2019) and COVID-19 period (2020 and 2021). Take Figure 3 (a) as an example, the vertical position of the words is randomly distributed, while the horizontal position represents the tendency of the activity to occur during a particular period (before or during the COVID-19 pandemic). Furthermore, the size of the words reflects the degree of correlation between the activity and the corresponding period, with larger sizes indicating stronger correlations. This description pertains to the visualization or design of information graphics. This enables us to deeply explore the temporal dynamics of specific activities within each topic, thereby providing deeper insights into the impact of the COVID-19 pandemic on UGS visitation patterns.

Table 1. Topic words with the highest probability.

Topic name	Topic words with the highest probability
Topic 1 Lockdown and exercise	Lockdown, good, make, night, well, train, home, fit, exercise, last
Topic 2 Music and events	day, run, happy, music, weekend, festival, house, start, artist, marathon
Topic 3 Social and family	year, easter, see today, first, point, people, amazing, club, friend

Figure 3(a) shows that the topic 1 accounted for approximately 12% of all topics in 2019, but significantly decreased during the pandemic periods, stabilizing at around 6.2% of all topics in 2021. The obvious decline in the proportion during the lockdown period indicated that the lockdown restriction measures such as stay home, keeping social distance, closing pubs have largely limited the people's outdoor activities. The corresponding covariate words of this topic show how the key words of this topic changed before and during the COVID-19 pandemics. The activities such as 'night', 'Saturday' were frequently happened pre COVID-19 pandemic period, however, activities such as 'fit', 'exercise', and 'social distancing' show a high correlation with COVID-19 pandemic periods.

Figure 3(b) shows that the topic 2 accounted for roughly 14% of all topics in 2019, slightly higher than that of topic

were more likely organised before COVID-19 outbreaks, and the 'run' activity was found to be popular in both periods. However, it was observed that activities related to 'day' life, 'happy birthday' and 'sports' were more strongly associated with the COVID-19 pandemic.

Figure 3(c) showed that topic 3 had a significant increase in topic proportions in 2019, possibly due to increasingly warm weather. However, the proportions did not return to high levels later, only slightly increasing to around 9% during the lockdown period in 2020, followed by a recovery trend in 2021. The COVID-19 pandemics may have influenced the topic's temporal patterns due to restriction measures such as social distancing and reduced social activities in UGSs. Covariate words associated with this topic included 'people', 'team', 'vote', 'queen', 'tonight', and 'party' before COVID-19. "Friend" was a common word in this *social and friends* topic.

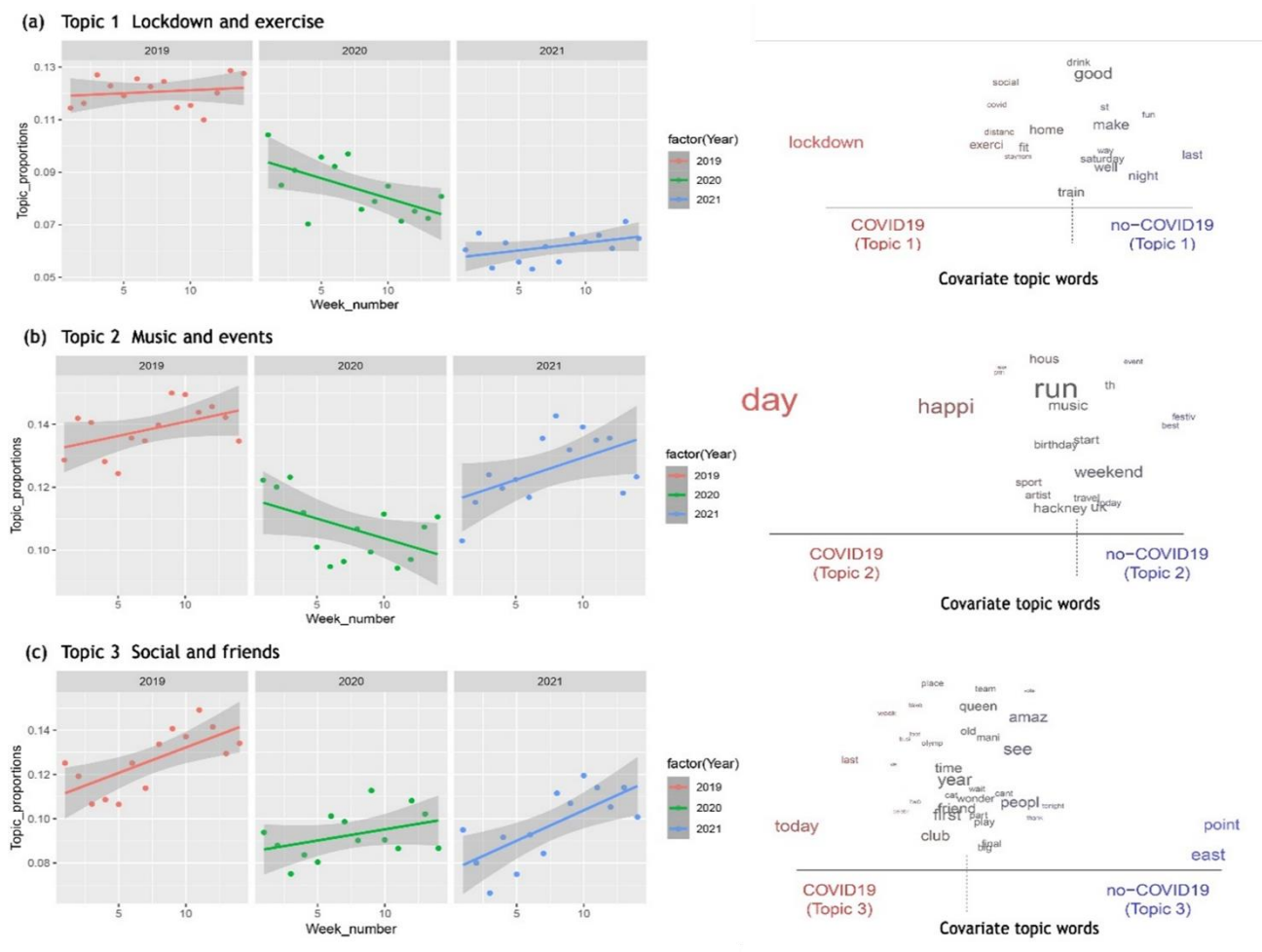


Figure 3. Dynamics in topic proportions and corresponding covariate topic words.

1. During the COVID-19 pandemic, this proportion experienced an initial decrease in 2020 and subsequently climbed in 2021. This may due to the COVID-19 restriction measures such as cancelling public events within UGS. The corresponding covariate words of this topic show that 'weekend' activities, and 'festival' events

3.3 Dynamics in spatial patterns of topics

Figure 4 shows dynamic changes in spatial patterns of topics. Specifically, topic 1 shows a different trend from 2019 to 2020. The areas with high probability mainly distributed around the boundary of London in 2019, with some hotspots in the city centre. While in 2020, the high probability areas were mainly distributed in the south part of London, and mid-high probability areas were distributed across the whole study area, indicating that the lockdown related topics and exercise activities were discussed or participated around the large part of London especially in the south part. However, the high probability areas obviously disappeared in London, with only two main hotspots were found in the very north and south part

In 2021, the spatial distribution of this topic showed a similar pattern with that in 2019, indicating that people were gradually back to normal life after the lockdown period. The topic 3 suggests that the areas with high probability mainly concentrated in northeast part of London, with a big hotspot found near the city centre, then the areas with high and higher probability in 2020 distributed in both northeast and southwest part of London, finally the obvious declines were found in 2021.

4. Discussion and conclusion

This study explored the impact the COVID-19 pandemic on UGS topics over space and time perspective by using STM and IDW approaches. Twitter has been selected as

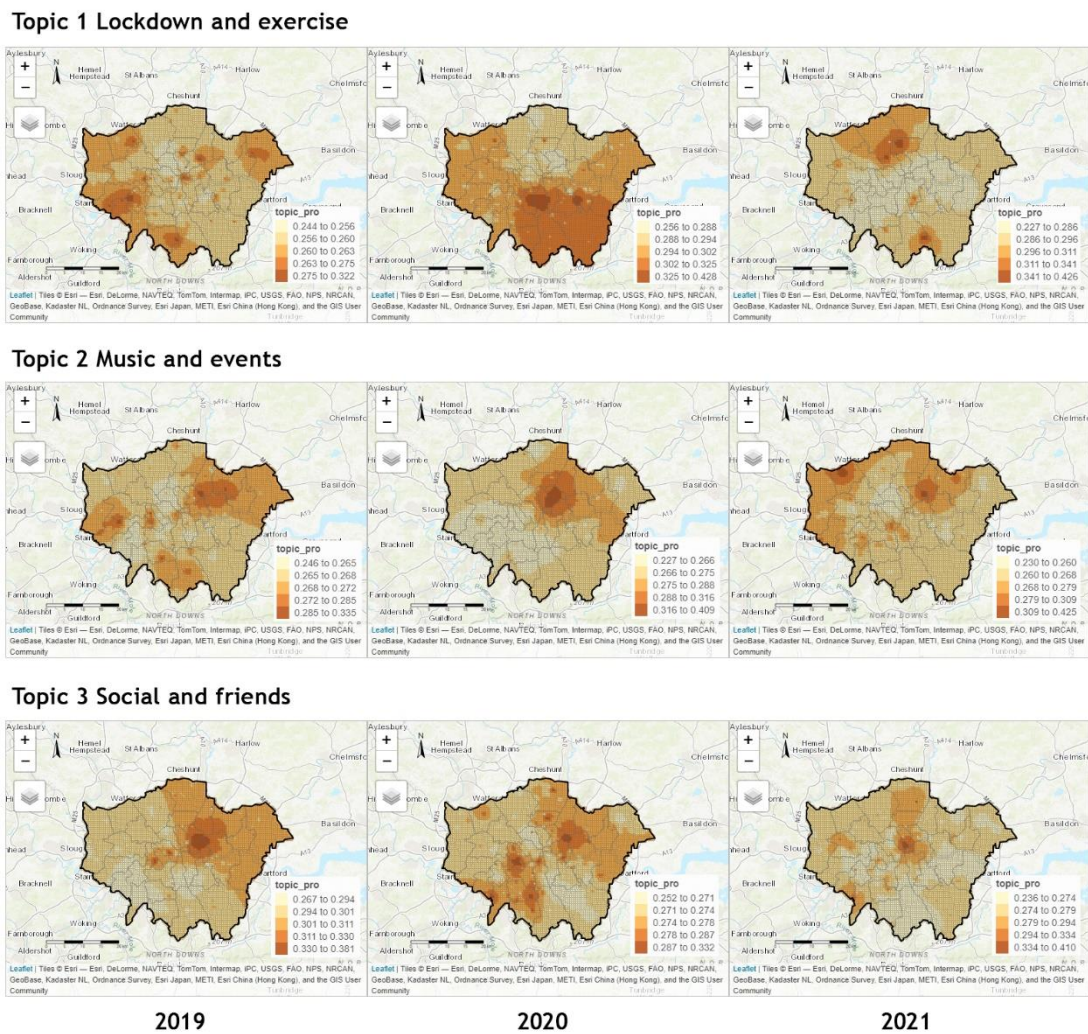


Figure 4. Spatial patterns of topics.

of London. In term of the spatial patterns of topic 2, lots of places showed a high probability of correlation with this topic, then the areas with high probability decreased during the lockdown period, and there is only one hotspot with higher probability, indicating that the restriction measures have influenced this topic and related activities.

data source for detecting UGS use. Twitter has been used by previous study (Roberts, 2017) to investigate human use and interaction with UGS. However, this study manually classified the Tweets, which was time and energy consuming. The current study used STM to identify the UGS topics and related activities, which can

efficiently reveal the impact of COVID-19 on UGS topics.

Seven types of topics were identified in this study, of which *Nature observation* accounted for the largest proportions, followed by *Food and service*, *Social and family*, *Art and exhibition*, *Music and events*, *Lockdown and exercise*, and *Dog walking*. Topics *Lockdown and exercise*, *Music and events*, and *Social and friends* 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 showed that all three topics declined during the COVID-19 lockdown period, while topic *Music and events* gradually back to normal status till the end of pandemics. The decreasing trends of the topics were potentially due to restriction measures such as stay home orders and social distancing. Previous studies confirmed the similar influence of the impact of the COVID-19 on UGS use (Ugolini, 2021). Spatial patterns of topic probability also show spatial various trends over three years. 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.

It is important to acknowledge that this study has limitations in terms of data sources and analyse methods. For example, this study only selected Twitter as the data source, which may have led to a loss of information about users who do not use Twitter. Future analyses should consider utilizing multiple social media networks to capture a more diverse and representative sample. Second, 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.

Overall, the findings of this study have the potential to contribute to a better understanding of the ways in which the COVID-19 pandemic has affected UGS activities and topics, which could potentially inform UGS policy and management for addressing the associated social crisis.

References

- Cameron-Blake, E., Tatlow, H., Wood, A., Hale, T., Kira, B., Petherick, A., and Phillips, T.: Variation in the response to COVID-19 across the four nations of the United Kingdom, Blavatnik School of Government Working Paper, <https://www.bsg.ox.ac.uk/research/publications/variation-on-response-covid-19-across-four-nations-united-kingdom>, 2020.
- Cui, N., Malleson, N., Houlden, V., and Comber, A.: 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>, 2021.
- Cui, N., Malleson, N., Houlden, V., and Comber, A.: Using social media data to understand the impact of the COVID-19 pandemic on urban green space use, *Urban Forestry and Urban Greening*, 74, 127677, <https://doi.org/10.1016/j.ufug.2022.127677>, 2022.
- Feinerer, I.: Introduction to the tm Package Text mining in R, Accessible En Ligne, <http://Cran.r-Project.Org/Web/Packages/Tm/Vignettes/Tm>. Pdf, 2013.
- Geng, D., Innes, J., Wu, W., and Wang, G.: Impacts of COVID-19 pandemic on urban park visitation: a global analysis, *Journal of forestry research*, 32(2), 553-567, <https://doi.org/10.1007/s11676-020-01249-w>, 2021.
- Houlden, V., de Albuquerque, J. P., Weich, S., and Jarvis, S.: A spatial analysis of proximate greenspace and mental wellbeing in London. *Applied Geography*, 109, 102036, <https://doi.org/10.1016/j.apgeog.2019.102036>, 2019.
- Ihaka, R., and Gentleman, R.: R: a language for data analysis and graphics, *Journal of computational and graphical statistics*, 5(3), 299-314, doi:10.1080/10618600.1996.10474713, 1996.
- Ilieva, R. T., and McPhearson, T.: Social-media data for urban sustainability, *Nature Sustainability*, 1(10), 553-565, <https://doi.org/10.1038/s41893-018-0153-6>, 2018.
- Li, J., and Heap, A. D.: Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173-189. <https://doi.org/https://doi.org/10.1016/j.envsoft.2013.12.008>, 2014
- Pebesma, E. J.: Multivariable geostatistics in S: the gstat package, *Computers & Geosciences*, 30(7), 683-691, <https://doi.org/10.1016/j.cageo.2004.03.012>, 2004.
- Roberts, H. V.: 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>, 2017.
- Roberts, M. E., Stewart, B. M., and Tingley, D.: Stm: An R package for structural topic models, *Journal of Statistical Software*, 91, 1-40, <https://doi.org/10.18637/jss.v091.i02>, 2014.
- Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B. and Rand, D.G.: Structural topic models for open-ended survey

responses. *American journal of political science*, 58(4), 1064-1082, <https://doi.org/10.1111/ajps.12103>, 2014.

Tennekes, M.: tmap: Thematic Maps in R. *Journal of Statistical Software*, 84, 1-39, 2018

Ugolini, F., Massetti, L., Pearlmutter, D., and Sanesi, G.: Usage of urban green space and related feelings of deprivation during the COVID-19 lockdown: Lessons learned from an Italian case study, *Land Use Policy*, 105, 105437, <https://doi.org/10.1016/j.landusepol.2021.105437>, 2021.

Wallach, H. M., Murray, I., Salakhutdinov, R., and Mimno, D.: Evaluation methods for topic models. Paper presented at the proceedings of the 26th annual international conference on machine learning, <https://doi.org/10.1145/1553374.1553515>, last access: 14 January 2023, 2009.

Wang, S. and Li, A.: Impacts of COVID-19 Lockdown on use and perception of urban green spaces and demographic group differences. *Land* 2022, 11, 2005. <https://doi.org/10.3390/land11112005>.

Wickham, H., Chang, W., and Wickham, M. H.: Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics, Version, 2(1), 1-189, 2016.