

Topics and sentiments on Twitter during lockdown in London

Nan Cui^{*1}, Nick Malleon^{†1} and Alexis Comber^{‡1}

¹School of Geography, University of Leeds, Leeds LS2 9JT, UK

February 10, 2021

Summary

The aim of this study is to investigate the topics that people discussed in Greater London during the Coronavirus disease 2019 (COVID-19) lockdown period. The Latent Dirichlet Allocation (LDA) method was used to investigate the topics discussed by people, then the sentiments and perceptions of social media users were evaluated by using *Syuzhet* R package, and the spatial patterns of urban green space visitation were also studied. LDA analysis identified eight topic groups, among which the word “lockdown” was frequently mentioned. Further, the sentiment analysis showed that users posted Tweets that were more negative when they referred to ‘lockdown’ during this period.

KEYWORDS: COVID-19, Social Media Data, Latent Dirichlet Allocation, Sentiment Analysis

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 (Cucinotta and Vanelli, 2020). The United Kingdom adopted a series of public health measures such as travel restrictions, quarantine, closing non-essential business services, and requiring citizens to stay at home to mitigate the virus spread. In addition, the UK government announced a second national lockdown for England from 5 November 2020 to 2 December 2020. In this context, questions about the topics that people discussed, their attitudes during lockdown and how they use urban green space became increasingly popular research topics (Geng et al., 2020).

Most approaches for investigating topics and sentiments use survey questionnaires or observations. Some limitations of surveys have been identified including low response rates which may result in biases (Evenson et al., 2013). In comparison to traditional methods, micro-blog data such as Twitter has been widely used for urban research (Longley et al., 2015). Twitter also has drawbacks, including data quality and representativeness, but it still offers some potential advantages over traditional surveys.

In this study, Twitter data was used to explore the impacts of COVID-19 on urban citizens life, drawing on methods including topic analysis, text mining and spatiotemporal analysis. To begin with, Latent Dirichlet Allocation (LDA) (Blei et al., 2003) modelling was used to investigate probable topics from Tweets during lockdown period by using the *topicmodels* R package. After this, sentiment analysis was performed using the *Syuzhet* R package (Jockers, 2017) which has been used to assess the sentiment in previous studies (Dubey, 2020; Pokharel, 2020). Finally Kernel Density Estimation (KDE) (Ullah et al., 2019) was used to quantify the spatial patterns of park visitation during the lockdown period.

* gync@leeds.ac.uk

† N.S.Malleon@leeds.ac.uk@leeds.ac.uk

‡ A.Comber@leeds.ac.uk

2. Data collection and analysis

Geo-referenced Tweets were downloaded via Twitter’s streaming Application Programming Interface (API) between 6th November 2020 and 2th December 2020. The downloaded Twitter data needed to be thoroughly cleaned. First, only English Tweets were selected, Tweets with fewer than three words were removed, and stopwords in each of the Tweets were also removed. Second, the Tweets posted from bots, fake accounts and users who posted a same tweet more than three times were removed. Third, punctuation, URLs, numbers, and words were removed, finally all words were converted to lower case. In order to detect the sentiment of users who mentioned lockdown in their tweets, the keyword “lockdown” was used to filter the relevant tweets, all the collected tweets were classified into lockdown-related tweets and unrelated tweets. Table 1 shows the number of Tweets after each preprocessed steps.

Table 1. The numbers of Tweets after each step of the data cleaning.

Process	Count
Collected Tweets from the Greater London Area	65398
Tweets from bots, fake accounts removed	37245
Non English Tweet removed	32896
The identical Tweets posted more than three times	18965
Tweets with fewer than three words removed (total tweets for the study)	15131
Lockdown-related Tweets	870
Lockdown-unrelated Tweets	14261
Tweets within urban green space	1198

The study area was Greater London, England, which covers an area of 1594 km² (Figure 1) and contains many parks and open spaces. London became the world’s first National Park City in 2019 (<https://www.nationalparkcity.london/>). Some 40% of London’s area is public green space, including 3,000 parks and totalling 23599 ha; derived from the Ordnance Survey Open Green Space layer.

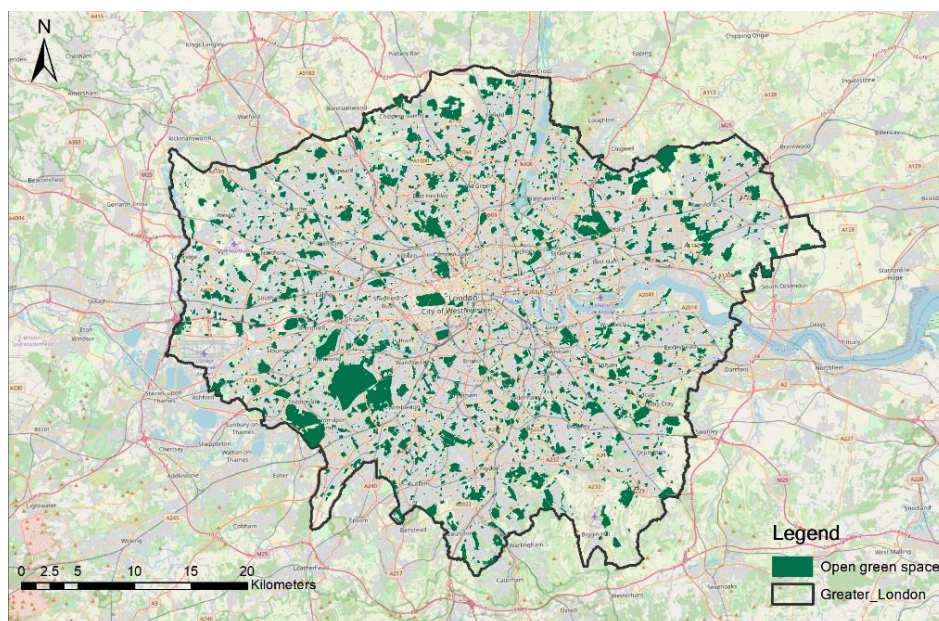


Figure 1 Greater London and urban green space

3. Results

3.1 Topic modelling

To begin with the goodness-of-fit of LDA models with varying numbers of topics was evaluated and the perplexity value showed that the optimal number of topics for the LDA analysis was 8. Figure 2 shows the top 10 popular words in each group. Twitter users frequently mentioned “get”, “like” and “look” in topic 1; “lockdown”, “sunday” and “winter” were frequent words in topic 2. It is not surprising that people often mentioned lockdown, “love”, “garden”, “video” and “photo” were mentioned in topic 3, which showed that art and photography were the popular topics; “happy (or happy birthday)”, “weekend”, “good” in topic 4 indicated that Twitter users tend to post tweets when celebrating birthday or during weekends; In topic 5, “London”, “park”, “walk”, “beautiful” and “autumn” were frequently mentioned; “today” “drink” were popular words in topic 6, indicating that food and drink were main content in this topic; “tonight”, “home” and “green ” were often mentioned in topic 7, indicating that Twitter user tend to post tweets related family activities. Topic 8 mentioned “Christmas”, “week” and “Friday”. Future work will explore the topics in more detail and capture a larger, non-lockdown data set in order to reveal the impacts of COVID-19 on residents’ life by re-running the analysis in normal times.

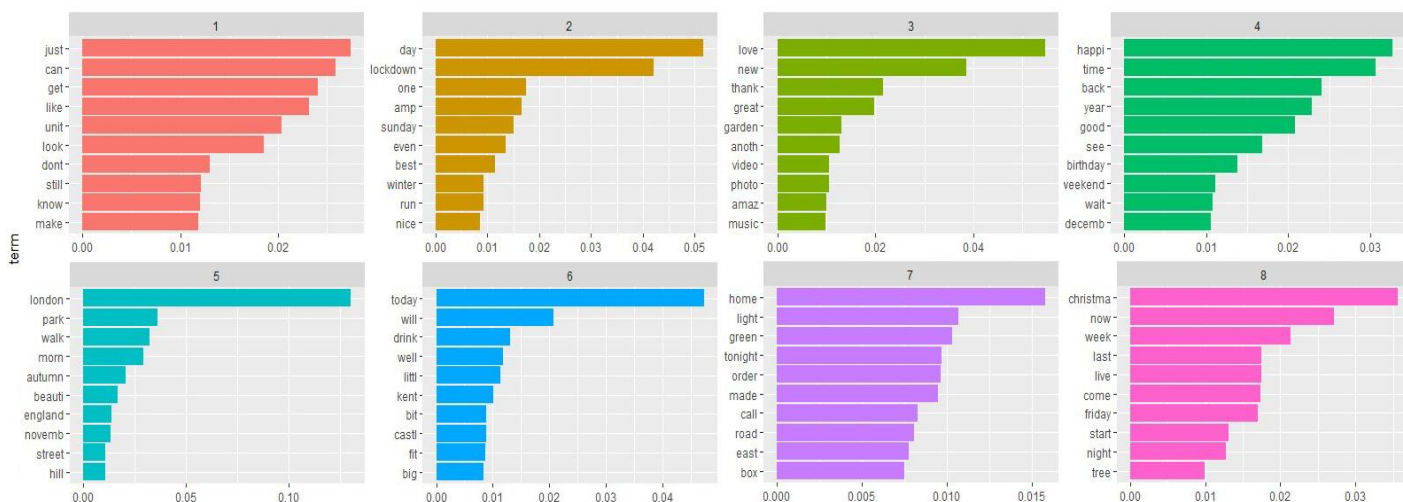


Figure 2 Terms in each topic

3.2 Sentiment analysis

The sentiment of the lockdown-related Tweets and unrelated Tweets were assessed by using *Syuzhet* R package and Figure 3 shows the average sentiment scores per day. Overall, the mean daily sentiment scores of lockdown-related Tweets is 0.129, and the mean sentiment of unrelated Tweets is 0.155. In detail, the average sentiment of lockdown-related Tweets largely fluctuated between 0.055 on 11 November and 0.230 on 26 November. This indicated that users expressed different sentiment when posted Tweets including word “lockdown”, for example, Tweet “*lockdown birthday big thank making birthday special even lockdown*” expressed high sentiment and “*second lockdown will take heavy toll mental health charities warn*” showed a negative attitude. However, the mean daily sentiment of unrelated tweets remained at a stable level, and the highest daily average sentiment of unrelated Tweets is 0.184 on 02 December, on this day the second national lockdown ended.

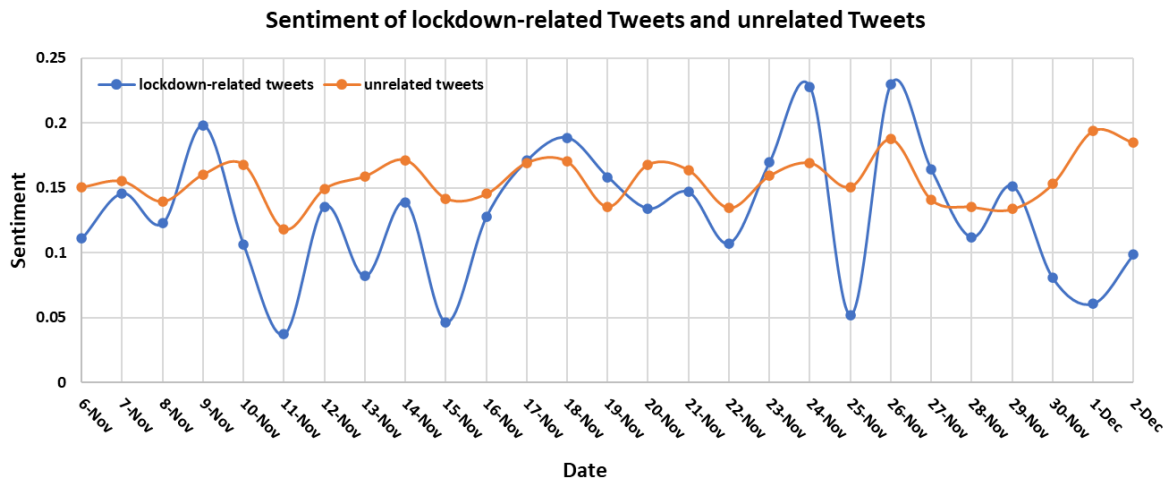


Figure 3 Average sentiment scores of Lockdown-related Tweets and unrelated Tweets per day

The independent T-test was used to detect the mean values of sentiment scores for lockdown-related Tweets and unrelated Tweets (Table 2). There was a difference in the sentiments for unrelated Tweets ($M = 0.156$, $SD = 0.288$) and lockdown-related Tweets ($M = 0.130$, $SD = 0.239$). The results of independent sample t-tests shows that on average, the sentiments of lockdown-related Tweets were more negative than unrelated Tweets, indicating that Twitter users tended to post about lockdown more negatively than other activities.

Table 2. Differences between lockdown-related tweets and unrelated Tweets during lockdown

Tweets	M	SD	Sig(<i>p</i>)
unrelated Tweets	0.156	0.288	
Lockdown Tweets	0.130	0.239	0.002

3.3 Spatial pattern of park Tweets

Tweets posted within urban green space (UGS) were selected to understand the role of UGS during lockdown period. KDE analysis was used to detect the spatial patterns of park visitation during lockdown period (Figure 4).

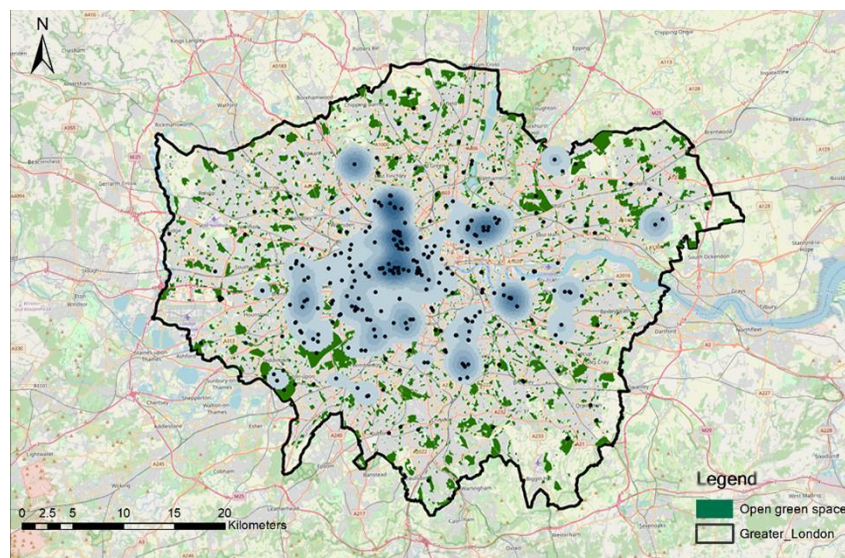


Figure 4. Park Tweets points density during lockdown period

Overall, it can be observed that the areas shaded in navy blue have a higher density of park visitors and higher concentration of social media use. It is no surprising that the green areas in the map, in the city centre, have large clusters of twitter points, followed by blue and light blue areas scattered in other parts of the study area. These areas were recognized as the hot spots of park users, indicating that these urban green spaces are more popular than others during this period, and potentially reflect higher UGS demand by local residents. In order to detect the impact of COVID-19 on UGS visitation, Tweets data in normal time will be studied.

4. Conclusion

This study has demonstrated that the Twitter data can be used to investigate the social media users' topics and sentiment during lockdown in Greater London. The LDA analysis identified eight different topics for all Tweets, the findings reveals that the words about "lockdown", "photo", "park", "garden", "home", "day", "Christmas" and "today" were frequently mentioned. The sentiment analysis showed that users posted Tweets that were more negative when they referred to 'lockdown', indicating that the lockdown rules had broken peoples' normal life which made people upset and people needed to look after their mental health. Additionally, this study analysed the spatial patterns of UGS visitation, the findings identified the hotspots of park user across the Greater London, which showed the high urban green space demand during this period. Future work will focus on the comparison of sentiment between tweets collected during lockdown and normal times. In addition, the changes in spatial patterns of UGS visitation will be investigated.

References

- Blei, D.M., Ng, A.Y. and Jordan, M.I.J.t.J.o.m.L.r. 2003. Latent dirichlet allocation. **3**, pp.993-1022.
- Cucinotta, D. and Vanelli, M.J.A.B.M.A.P. 2020. WHO declares COVID-19 a pandemic. **91**(1), p157.
- Dubey, A.D.J.A.a.S. 2020. Twitter sentiment analysis during COVID19 outbreak.
- Evenson, K.R., Wen, F., Hillier, A., Cohen, D.A.J.M., sports, s.i. and exercise. 2013. Assessing the contribution of parks to physical activity using GPS and accelerometry. **45**(10), p1981.
- Geng, D.C., Innes, J., Wu, W. and Wang, G.J.J.o.f.r. 2020. Impacts of COVID-19 pandemic on urban park visitation: a global analysis. pp.1-15.
- Jockers, M.J.T.C.R.A.N. 2017. Introduction to the Syuzhet package. **13**.
- Longley, P.A., Adnan, M., Lansley, G.J.E. and A, P. 2015. The geotemporal demographics of Twitter usage. **47**(2), pp.465-484.
- Pokharel, B.P.J.A.a.S. 2020. Twitter sentiment analysis during covid-19 outbreak in nepal.
- Ullah, Wan, Haidery, Khan and Geo-Information, L.J.I.J.o. 2019. Analyzing the Spatiotemporal Patterns in Green Spaces for Urban Studies Using Location-Based Social Media Data. **8**(11), p506.
- Zu, Z.Y., Jiang, M.D., Xu, P.P., Chen, W., Ni, Q.Q., Lu, G.M. and Zhang, L.J.J.R. 2020. Coronavirus disease 2019 (COVID-19): a perspective from China. **296**(2), pp.E15-E25.

Biographies

Nan Cui is a second year PhD student at the School of Geography, University of Leeds. His interests are in urban green space use, social media data and big data analysis.

Professor Lex Comber is a leading international researcher in many areas of spatial science and geocomputation, with publications in accessibility, facility location optimisation, graph and network theory, spatial data uncertainty, citizen science, land use / land cover and remote sensing.

Professor Malleson is a Professor of Spatial Science and a member of the Centre for Spatial Analysis and Policy (CSAP). His primary research interest is in developing spatial computer models of social phenomena with a particular focus on crime simulation and models of urban areas.