Where is Tweeting What about London?: Investigating Discontent in Left Behind Places

Rachael Sanderson*[1], Rachel Franklin†[2], Danny MacKinnon‡[2] and Joe Matthews§[3]

[1]Geospatial Systems CDT, School of Engineering, Newcastle University
[2]Centre for Urban and Regional Development Studies (CURDS), Newcastle University
[3] School of Mathematics, Statistics and Physics, Newcastle University

GISRUK 2023

Summary

The "levelling up" agenda set out by the UK government was designed to tackle entrenched regional patterns of social inequality, which have been associated with "left behind places". These places are united through discontent resulting from being 'left behind' in regional development, experiencing stagnation whilst other, often metropolitan, areas experience rapid growth. Alternative datasets, including social media data, have been identified as possible ways to evaluate this discontent. This study will use a specific case study, perspectives of London from outside London, to evaluate how this dataset can be used to understand the dynamics of levelling up.

KEYWORDS: Left Behind Places, Sentiment Analysis, Location-based social media, Twitter, Volunteered Geographic Information

1. Introduction

A decade of political discontent in the UK has drawn attention to left behind places. These lagging areas of under-development have been contrasted to thriving cities, with London being highlighted as the centre of economic power. This paper uses social media to explore the geographies of this sentiment, analysing mentions of London outside of London. This methodology offers a GIScience perspective to further understanding of social dynamics of the levelling up agenda.

2. Literature

The UK is one of the most spatially unequal developed economies. London was identified as the richest region in Europe, despite six of the ten poorest areas being in the UK (UK2070, 2020, p. 20). This inequality contributes to discontent, with places being 'left behind' – a category associated with disadvantage (OCSI, 2019) and being framed negatively compared to dynamic cities (Rodríguez-Pose, 2018),. This divide is experienced internationally, with the favouritism of elite areas argued to contribute to this political discontent, and dissatisfaction targeted towards the capital city observed in survey data (Rickardsson et al., 2021).

This paper will undertake similar analyses using social media data. As a platform for voicing grievances, Twitter has detailed data used to measure quality of life (Zivanovic et al., 2020) and sense of place (Butler, Alice et al., 2018), with tweets classified by sentiment, which can reflect perspectives of a location (Kovacs-Györi et al., 2018). London has been a case study to analyse tweet density (Ballatore

* r.sanderson4@newcastle.ac.uk
† rachel.franklin@newcastle.ac.uk
‡ danny.mackinnon@newcastle.ac.uk
§ joe.matthews@newcastle.ac.uk

and De Sabbata, 2018) or topics within the city (Lansley and Longley, 2016). This study offers an alternative contribution by mapping perspectives of London outside the city. It will investigate links between discontent towards the capital and being left behind, supporting the levelling up agenda by exploring the intricacies of associated social dynamics.

3. Methodology

The data was collected using Twitter API, with the *academictwittR* package in RStudio. 1429996 tweets were collected across 223853 accounts. Filtering for tweets mentioning London, all geo-located tweets were collected for the time-period 1st November 2019 – 31st October 2022. To avoid false hot spots, tweets were included with geo-tag or place-name data accurate to city level or better, then aggregated to Local Authority 2022 (LAD) polygons. Only tweets from outside of the region of London were analysed. LADs were classified as left behind based on Martin et al (2021). Bots were removed in PostgresSQL, removing overly active accounts based on established criteria (Arthur and Williams, 2019; Ostermann, 2021), leaving 1331864 tweets for analysis. Sentiment analysis was undertaken using the VADER lexicon computed using the *vader* package in RStudio (Hu et al., 2020), which analyses complete tweets including punctuation, to provide a compound score between -1 and 1. A score higher than 0.05 was classed as positive, and lower than 0.05 as negative. To normalise the number of positive or negative tweets by tweet population, the odds ratio (OR) is calculated as (Equation 1, Shelton, Taylor et al., 2014):

$$OR_{\text{lower}} = e^{ln(OR_i) - 3.29x\sqrt{\frac{1}{P_i}+\frac{1}{P}+\frac{1}{R_i}+\frac{1}{R}}} \quad (1)$$

P relates to the number of positive or negative tweets, *R* the total number tweets, and *i* indicates the LAD being measured. An output higher than 1 implies more positive or negative tweets than expected, whilst under 1 means vice versa. Further language analysis used tweets with three or more words, and removed hyperlinks, non-Latin symbols, punctuation, and other notation, as well as meaningless stop words. Linear regression assessed the association between the sentiment OR for an LAD and socio-economic variables associated with left behind classifications. All variables were normalised, except the dummy left behind classification variable.

4. Analysis

4.1 Where is talking about London?

The LADs with more tweets per resident mostly neighbour London (**Figure 1**). Other areas of high concentration of London-related tweets are towards the North-West. Lower concentrations are in Wales, potentially due to language variation filtering non-English tweets. The relationship between distance from London and tweets per 1000 residents (**Figure 2**) reveals a weak correlation of further LADs containing less tweets per resident. Anomalies include Manchester and Norwich. There is no significant difference in the number of London-related tweets per 1000 residents in left behind LADs compared to those not left behind, to 99% significance, shown by the spread of left behind LADs in **Figure 2**. This means that whilst the digital footprint of tweets mentioning London is not spatially even nationally, it is comparable between LADs that are left behind and those not left behind.

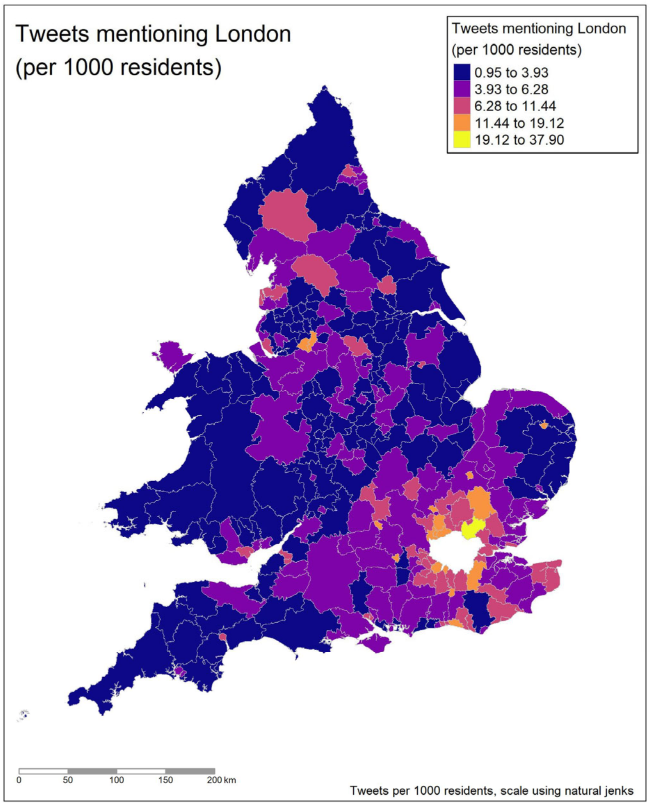

Figure 2 Map of London-related tweets per 1000 residents, aggregated by LAD

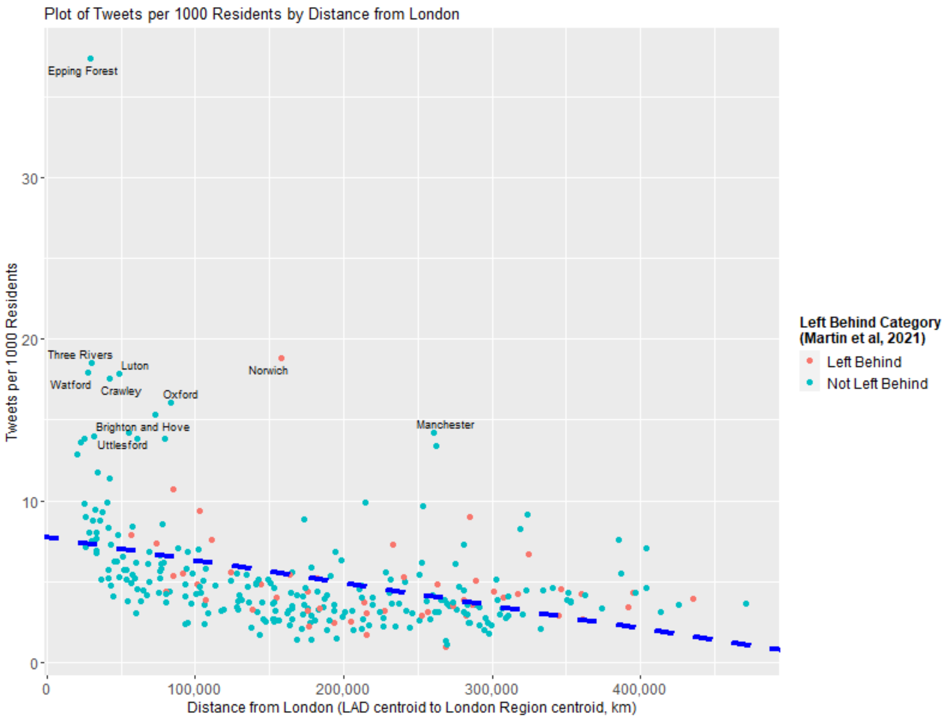

Figure 2: Scatter plot comparing quantity of London-related tweets to distance from London

4.2 What is being said?

After classification, 48% of tweets were positive, and 28% were negative. The wordcloud shows how London is represented differently in positive and negative tweets (**Figure 3**). The dataset is affected by the timescale of data collection, for example the prominence of 'Covid' frequently links to complaints about new variants in the Covid-19 pandemic and the UK's 2020 policy of a tiered lockdown system.

Using OR, the geographies of positive and negative tweets are mapped in **Figure 4.** There are more areas with a higher concentration of negative tweets than positive tweets; this is potentially because of the higher occurrence of positive tweets, causing them to be more evenly spread.

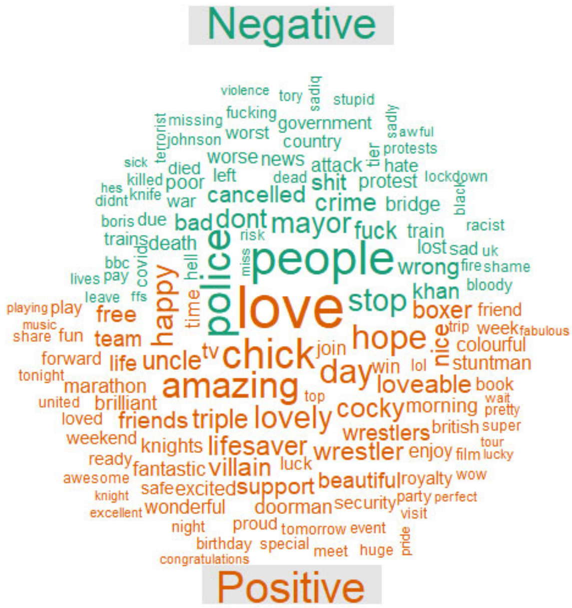

Figure 3 Words included in London-related tweets, grouped by sentiment

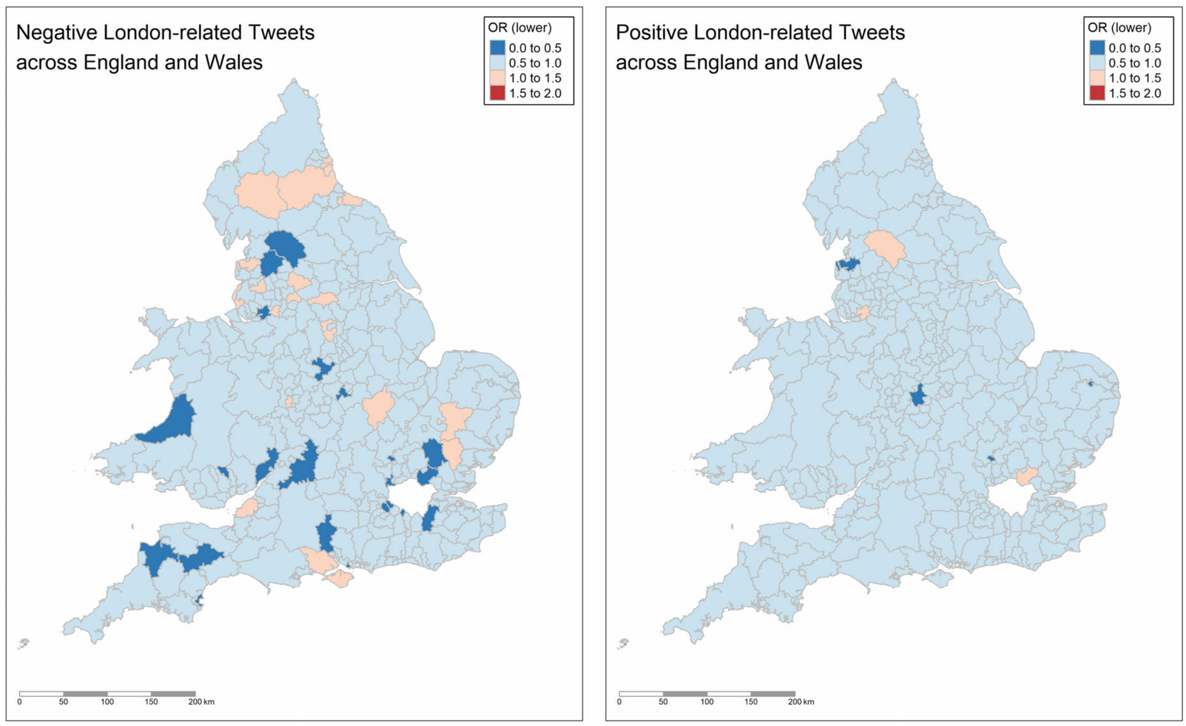

Figure 4 Odds ratio for positive and negative London-related tweets

4.3 What does this suggest about left behind places?

The ORs for left behind and not left behind LADs were compared in **Figure 5**. There is no difference in the OR for the quantity of positive tweets between these classifications – being left behind does not mean more positive tweets. However, left behind LADs are significantly more likely to post with negative London-related tweets with 99% confidence level.

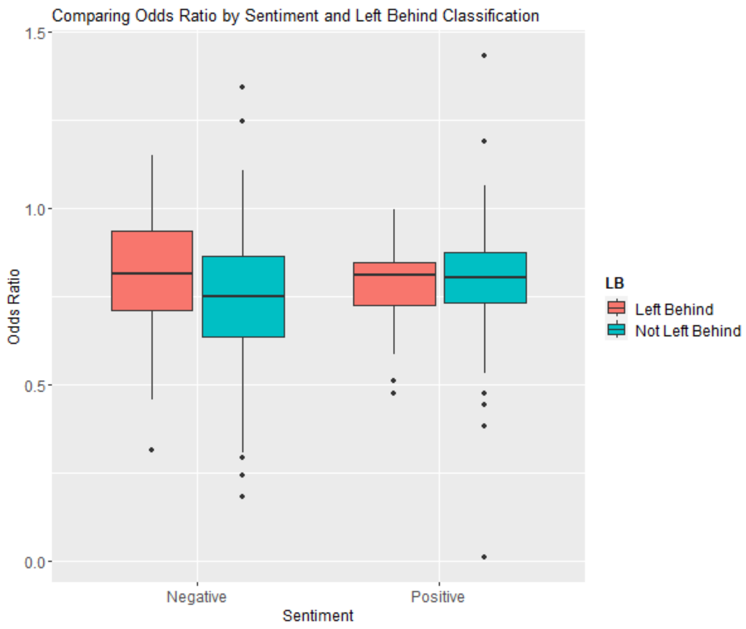

Figure 5 Boxplot comparing OR for positive and negative tweets by left behind classification

Linear regression assessed if this variation is influenced by the characteristics of the LAD (**Table 1**). We can observe the characteristics of places more likely to tweet, including places closer to London, and places with higher qualifications. This indicates bias towards particular users across the dataset, reflecting *who* is talking about London. For positive tweets, the most significant variables were 4G connection quality and qualification level. For negative tweets, 4G connection quality was also significant, implying this influences the likelihood of a tweet showing strong sentiment. Distance from London was highly significant for negative tweets, suggesting increasing distance from London correlates with increased negative discussion. The R^2 value is less than 0.3 for both sentiment models, so there is scope to develop these models further. Overall, it is observed that left behind LADs display a higher occurrence of negative London-related tweets, coinciding with LADs further from London.

5. Conclusions

This work has introduced potential analysis to understand social relations between places, as understood through a digital footprint. Future work will consider what this can inform us about the future of levelling up, as well as reflecting how classifications can be refined with alternative methods, such as machine learning approaches. Linear regression could be more spatially aware, for example using Geographically Weighted Regression.

6. Acknowledgements

The authors would like to thank EPSRC for their funding, received through Geospatial Systems CDT (Ref: EP/S023577/1). Data was collected through the Twitter API, with supporting data through ONS and Ofcom.

Table 1 Linear Regression Results

	Tweets per 1000 Residents	Positive	Negative
Intercept	0.2809***	0.1345*	-0.0287
	(0.0573)	(0.0643)	(0.0628)
Distance to London	-0.2658***	-0.1031	0.3921***
	(0.0561)	(0.0629)	(0.0615)
GDP change (£, 2010 - 2020)	-0.3261+	-0.3442+	-0.2734
	(0.1841)	(0.2064)	(0.2018)
Age 16-29 change (%, 2010 - 2020)	0.0007	0.0333	-0.0663
	(0.0599)	(0.0672)	(0.0657)
Age 65 or over change (%, 2010 - 2020)	0.3432***	-0.0395	-0.1020
	(0.0783)	(0.0877)	(0.0858)
Population density (per sqkm)	0.3869**	-0.0534	-0.1025
	(0.1426)	(0.1599)	(0.1563)
Population Change (persons, 2010 - 2020)	0.2832**	-0.1872+	-0.2107*
	(0.0896)	(0.1005)	(0.0982)
Qualification Level 4 or above (%)	0.3714***	0.3046***	-0.0913
	(0.0732)	(0.0821)	(0.0802)
Data Availability in 4G	-0.0889	0.2492***	0.3742***
	(0.0543)	(0.0609)	(0.0595)
Left Behind Classification	0.0095	0.0427	0.0143
	(0.0519)	(0.0582)	(0.0569)
R^2	0.3783	0.2388	0.2725
Adj. R^2	0.3589	0.2150	0.2497
Num. obs.	298	298	298

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; +$p < 0.1$

References

Arthur, R., Williams, H.T.P., 2019. Scaling laws in geo-located Twitter data. PLoS ONE 14.

Ballatore, A., De Sabbata, S., 2018. Charting the Geographies of Crowdsourced Information in Greater London, in: Geospatial Technologies for All. Presented at the The Annual International Conference on Geographic Information Science.

Butler, Alice, Schafran, Alex, Carpenter, Georgina, 2018. What does it mean when people call a place a shithole? Understanding a discourse of denigration in the United Kingdom and the Republic of Ireland. Transactions of the Institute of British Geographers 43, 496–510.

Hu, T., She, B., Duan, L., Yue, H., Clunis, J., 2020. A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets. IEEE Access 8.

Kovacs-Györi, A., Ristea, A., Kolcsar, R., Resch, B., Crivellari, A., Blaschke, T., 2018. Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data. ISPRS International Journal of Geo-Information 7, 378. https://doi.org/10.3390/ijgi7090378

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

Martin, R., Gardner, B., Pike, A., Sunley, Pe., Tyler, P., 2021. Levelling Up Left Behind Places: The Scale and Nature of the Economic and Policy Challenge. Taylor and Francis, Oxon.

OCSI, 2019. Left Behind? Understanding communities on the edge.

Ostermann, F.O., 2021. Linking Geosocial Sensing with the Socio-Demographic Fabric of Smart Cities.

International Journal of Geo-Information 10.
Rickardsson, J., Mellander, C., Bjerke, L., 2021. The Stockholm Syndrome: the view of the capital by the “Places Left Behind.” Cambridge Journal of Regions, Economy and Society 14, 601–617.
Rodríguez-Pose, A., 2018. The revenge of the places that don’t matter (and what to do about it). Cambridge Journal of Regions, Economy and Society. https://doi.org/10.1093/cjres/rsx024
Shelton, Taylor, Poorthuis, Ate, Graham, Mark, Zook, Matthew, 2014. Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of “big data.” Geoforum 52, 167–179.
UK2070, 2020. Make No Little Plans: acting at a scale for a fairer and stronger future UK2070.
Zivanovic, S., Martinez, J., Verplanke, J., 2020. Capturing and mapping quality of life using Twitter data. GeoJournal 85. https://doi.org/10.1007/s10708-018-9960-6(0123456789().,-volV)(0123456789().,-volV)

Biographies

Rachael Sanderson is a PhD Student in the Geospatial Systems CDT at Newcastle University, having previously studied Human Geography. Her research interests focus on the defining of left behind places, applying on novel datasets and reflecting on the politics behind different classifications.

Rachel Franklin is Professor of Geographical Analysis in the Centre for Urban and Regional Development Studies (CURDS) and the School of Geography, Politics and Sociology at Newcastle University. Her research interests include population decline and shrinking cities, spatial inequality, and demographic change.

Danny MacKinnon is a Professor of Regional Development and Governance and Director of Centre for Urban and Regional Development Studies (CURDS) and the School of Geography, Politics and Sociology at Newcastle University. His research is concerned with the institutions and politics of local and regional development.

Joe Matthews is a lecturer in statistics, interested in real world applications of statistical techniques in the Bayesian paradigm. His research interests encompass virtually any area of applied statistics, but historically focused on the areas of environmental extremes and road safety analysis, and more recently geospatial analysis and population dynamics.