Using Twitter to track internal migration in the UK before and during the COVID-19 pandemic

Yikang Wang[1], Chen Zhong[1]*, Carmen Cabrera-Arnau[1], Qili Gao[1]
[1]Centre for Advanced Spatial Analysis, University College London, London, UK
*Corresponding author, e-mail: chen.zhong@ucl.ac.uk

Keywords: Human Mobility; Migration; Geocoding; Twitter; COVID-19;

Background: The COVID-19 pandemic has greatly impacted people's location choices [1]. National and regional lockdowns, economic depression, and working from home policies have significantly decreased short-term human mobility such as daily trips and tourism, and have also affected people's migration choices – namely, residents have been moving out of large cities due to the pandemic [2]. The accurate and timely measurement of changes in internal migration during a pandemic is essential for quantitatively assessing the impact and effectiveness of restrictive policies and to help develop recovery strategies for the post-pandemic phase.

Objective: We aim to propose a privacy-preserving framework for monitoring migration at national and local authority scales based on geocoding and estimating home locations using Twitter data.

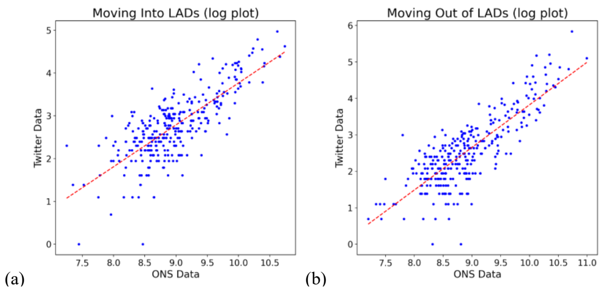
Data: This study was based on an anonymized dataset of 182 million tweets sent from the UK from the beginning of 2019 to the end of 2022, collected using the Twitter Developer API. UK geography and census data released by the Office for National Statistics (ONS) [3] and COVID-19 data [4] were also used.

Methodology: The process of tweet place geocoding consisted of four steps. (1) To avoid ambiguity, we removed from the dataset those tweets whose place attributes were of a scale larger than the LAD scale. (2) We constructed a lookup table matching lower-level geographic hierarchies (such as wards) to LADs. (3) We used Bing Maps' Geocoding API and tried to match the returned administrative divisions or address with our LAD list and lookup table. (4) For the remaining unmatched places, the intersection index of its bounding box according to Bing Maps and each LAD was calculated as $I(Place, LAD) = \frac{Area\ of\ intersection\ with\ LAD}{Area\ of\ the\ place\ bounding\ box}$. If I(Place, LAD) > 65%, the place was considered to belong to that LAD.

We estimated the home location of each user as the LAD associated with the highest number of their posted tweets [5]. When a user stayed in two different locations during two non-overlapping periods, we considered that a migration or a change of regular activity area, rather than an occasional trip.

In order to evaluate the migration patterns and heterogeneity before and during the COVID-19 pandemic, we defined the following indicators: (1) **Migration rate:** the number of migrations divided by the total number of Twitter users whose home locations were successfully detected in both months; (2) **Migration index of moving in/out of large cities:** the number of migrations in/out of LADs belonging to large cities divided by the total number of migrations. LADs were attributed to large cities on the basis of population density; (3) **Net migration rate:** the net number of migrations (inflows minus outflows) for a LAD divided by its census population; (4) **Migration share proportion:** the number of migrations moving from a selected city to each LAD, divided by the total number of migrations from that city; (5) **Recovery index:** the total number of migrations (inflows plus outflows) for a LAD in 2021 divided by the corresponding number in 2020.

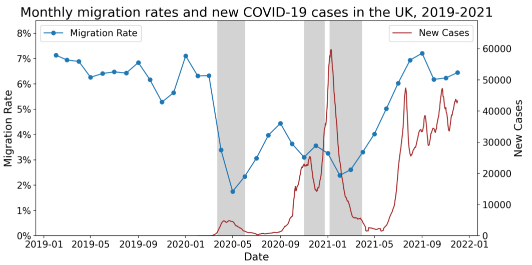
Result and Conclusion: Out of the total collected tweets, 161 million (89%) had valid place attributes at or below the LAD scale, in which 160 million were successfully geocoded at the LAD scale. We validated the yearly matrix by comparing it with annual internal migration data from ONS in England and Wales in the year ending June 2020 (Figure 1). Our results are consistent with official migration flow data, showing that despite the biases present in Twitter data, it is highly valuable for trend analysis.



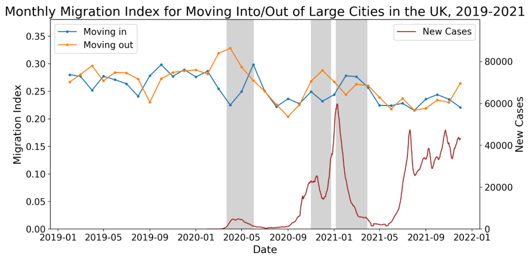
(a) (b)

Figure 1. Comparison plot of LAD migration numbers as determined from Twitter data (localized Twitter users > 50 in England and Wales in the year ending June 2020) with corresponding ONS internal migration data: (a) moving into LADs, (b) moving out of LADs.

As ONS only provides an annual migration matrix, one of the advantages of the Twitter method of migration analysis is that Twitter data can provide more frequently updated matrices that allow us to monitor the market and detect any abnormal patterns and behaviours. Figure 2 illustrates migration rates and migration indexes of moving into and out of large cities for every pair of consecutive months in the time period spanning from 2019 to 2021. Our findings point to complex social processes unfolding differently over space and time, likely driven by variations in policy adherence, vaccine relaxation, and regional interventions. In particular, the pandemic and lockdown policies significantly reduced the rate of migration, but mass vaccination soon mitigated the effect.



(a)



(b)

Figure 2. (a)Monthly migration rates, (b) Monthly migration index for moving into and out of large cities and new COVID-19 cases in the UK, 2019-2021. The periods during which most parts of the UK went into lockdown were marked as grey.

To explore the population losses of large cities, we examined the relationship between the net migration rate and (log) population density in 2019, 2020 and 2021. In 2019 and 2021, the R square values were less than 0.022, indicating no occurrence of city shrinking or urban depopulation. Meanwhile, the 2020 data yielded a substantially higher R square value and a regression coefficient of -6.46, reflecting population loss for large cities in that year. While such population loss was suspended in 2021, most of the lost population has not returned to the cities so far.

Restrictive policies have had uneven impacts on those cities of England and Wales that are top-ten in terms of population. Figure 3 illustrates the home locations of people who moved out of those cities during the first wave (February to May 2020) and the second wave (October 2020 to January 2021) of city-exiting migration. The majority of those who moved out of the cities choose to go to the nearby countryside.

The recovery index reflects how quickly each LAD recovered in 2021 relative to their status in 2020 in terms of migrations. Of the 370 LADs analysed, 311 (84%) had a recovery index of more than 1, meaning most parts of the UK were experiencing more internal migrations and hence recovering. Additionally, when the index was plotted against population density (Figure 4), the slope of the trendline was greater than 0, meaning that LADs with higher population density recovered faster.

Code Availability

The code that support the findings of this study are available: https://github.com/ykangw/Twitter-Internal-Migration.

Funding

This work was supported by the ESRC under JPI Urban Europe/NSFC (No. ES/T000287/1), the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (No. 949670).

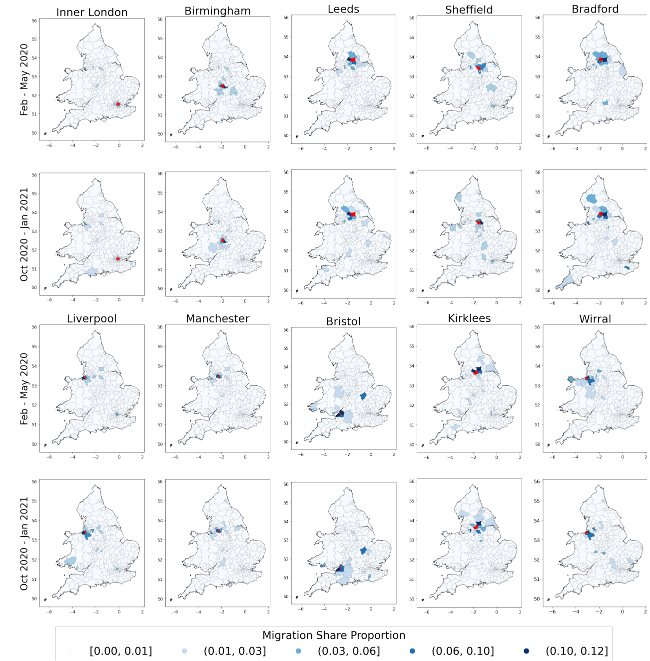


Figure 3. Destination places for those who moved out of the top 10 largest cities in England and Wales (defined in terms of population) during the first and second city-exiting waves. The selected city is marked in red. Other locations are coloured in blue shades according to their migration share proportions from the selected city.

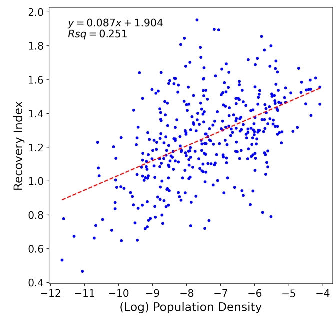


Figure 4. Plot of the recovery index against (log) population density.

References

[1] M. Batty, 'The Coronavirus crisis: What will the post-pandemic city look like?', *Environment and Planning B: Urban Analytics and City Science*, vol. 47, no. 4, pp. 547–552, May 2020, doi: 10.1177/2399808320926912.

[2] P. H. Haslag and D. Weagley, 'From L.A. to Boise: How Migration Has Changed During the COVID-19 Pandemic', *SSRN Journal*, 2021, doi: 10.2139/ssrn.3808326.

[3] Office for National Statistics, 'Population estimates for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics', Jun. 25, 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest#movement-of-people-within-the-uk (accessed Mar. 17, 2022).

[4] 'Coronavirus (COVID-19) in the UK', Apr. 08, 2022. https://coronavirus.data.gov.uk (accessed Apr. 08, 2022).

[5] M. C. González, C. A. Hidalgo, and A.-L. Barabási, 'Understanding individual human mobility patterns', *Nature*, vol. 453, no. 7196, pp. 779–782, Jun. 2008, doi: 10.1038/nature06958.