GI_Forum 2018, Issue 1
Page: 316 - 335
Full Paper
Corresponding Author:
ltclasper@aol.com
DOI: 10.1553/giscience2018_01_s316

Exploring Vernacular Perceptions of Spatial Entities: Using Twitter Data and R for Delimiting Vague, Dinformal Neighbourhoods in Inner London, UK

Luke Thomas Clasper
UNIGIS, United Kingdom

Abstract

The informal and unofficial nature of how citizens discuss and conceive geographical entities such as neighbourhoods has traditionally been difficult to capture. Ambient Geographic Information (AGI) from social media services offers researchers an opportunity to collect large amounts of geo-referenced information concerning vernacular geography. Twitter data was harvested and analysed in R statistical software in order demonstrate whether using geodata from social media is a feasible method for spatially defining vague, vernacular neighbourhoods in Inner London, UK. The results suggest that social media data can be a valuable source for capturing vernacular geography from which vernacular neighbourhoods could be delimited. The study also revealed factors which may have contributed to vernacular neighbourhood demarcation. Twitter data was seen to both mirror the physical form of the underlying topography and reflect the social character of the city's land use. This work builds upon previous attempts to investigate vernacular geography which used more traditional methods, such as sketch maps and interviews. It also examines how manual qualitative coding can improve data quality and demonstrates how R statistical software can be used to capture, analyse and present geospatial data.

Keywords:

vernacular geography, ambient geographic information, R project, qualitative GIS, urban morphology

1 Introduction

The relationship between urban form and function (Batty, 2013; Batty & Longley, 1994), as well as how social activity interacts with and replicates the physical structure of cities (Tonkiss, 2013) have long been theorized. However, the city's inhabitants themselves can also offer us a strong insight into geographic form and the way that space is used, perceived mentally and referred to. This bottom-up, unofficial discourse concerning geographic place is fundamental to our understanding of cities (Lynch, 1960). The question of how to reproduce these colloquial, vague notions of space in rigid computer representations is also central to the advancement of GIS software and GIScience (Goodchild, 2011).

This study is concerned with the individual's awareness of fuzzy, abstract geographic areas in relation to their own location. These areas could be informal or defined by an administrative boundary. Official legislative bodies have long attempted to demarcate geographic areas for administrative purposes (Fletcher, 1844). However, in the minds and conversations of citizens, these imposed, precise boundary lines are often much vaguer, guided by landmarks, street networks, architecture, land use, transport hubs (Huck et al., 2014) and less visible factors, such as demographics, class, politics and socio-economics (Galster, 2001).
Vernacular geography has important applications. The emergency services find these types of colloquial indications of place invaluable when locating reported incidents. There are commercial applications for deliveries and in-vehicle navigation, and government uses for the allocation of services and collection of census data. Neighbourhoods also form geographical entities that people and communities can relate to and feel associated with (Brindley et al., 2014).

Capturing this type of qualitative, casual, ambiguous information has proved challenging (Montello et al., 2003). However, due to the proliferation of GPS-enabled devices, social media posts are often geo-tagged, which leads to the unconscious generation of massive amounts of geodata. In turn, this enables spatial analysis (Sui & Goodchild, 2011) and spatial modelling (Lovelace et al., 2014) of AGI. Thus, the acquisition of AGI can help us study how large numbers of people use vernacular language to describe where they are (Hollenstein & Purves, 2010), and the corpus obtained is certainly far more substantial than it would be possible to collect using traditional techniques.
In this paper, I present how AGI from the social networking service Twitter can be harvested, processed and analysed in R in order to capture vernacular indications of which neighbourhoods individuals think they are located in.

AGI is used to describe passively volunteered data where the volunteers of the information are often the focus of the study (See et al., 2016). The qualitative nature of social-media AGI is evident in its content. For this reason, manual qualitative coding techniques are employed to test for dataset veracity, in line with Cope (2003), who describes the coding of qualitative textual data as a way of interpreting and filtering data in order to classify it into themes.
Qualitative GIS (or mixed-method GIS) is the integration of qualitative data with the quantitative analysis capabilities of GIS (Cope & Elwood, 2009). Elwood & Cope (2009) describe Qualitative GIS as an extension of GIS which includes non-numerical data, the mixing of methodologies, technologies and data, citizen participation and social practices.
After this quality testing, the point-based dataset is validated against government administrative polygons and place-name seeds. The delimitation of precise neighbourhood boundaries is then attempted, along with an investigation into the epicentres and demarcation factors affecting these vernacular neighbourhoods.

2 Related Work

Lynch (1960) set the foundations for understanding mental images held by individuals about their environment, describing how cities are comprised of imagined elements and views, which form continuous patchworks of distinct regions. The five elements that Lynch adopted were paths (movement channels), edges (boundaries), landmarks (familiar buildings etc.), districts (areas with a common recognizable character or architecture) and nodes (places of navigational decision).

Vernacular geography has traditionally been captured from participants who are aware that they are involved in a study, mainly by asking them to draw sketch maps (Coulton et al., 2012; Doran & Young, 2013; Stanton Fraser et al., 2013), by conducting interviews and questionnaires (Vallée et al., 2015), or through a combination of these techniques (Raanan & Shoval, 2014).

There is a growing body of work in which vernacular geography has been captured from participants who are unaware that they are involved in a study. The techniques used have involved the use of web-scraping (Liesch et al., 2015; Brindley et al., 2014; Twaroch et al., 2008) and geo-tagged Flickr photographs (Feick & Robertson, 2015; Hu et al., 2015).

Twitter data has been used in various GIScience research fields (Zimmer & Proferes, 2014), as well as in other analytical social sciences (Batrinca & Treleaven, 2015). The most notable studies include ones on the effects of geographical distance on social networks (Stephens & Poorthuis, 2015); analysis of visitor flows to attractions (Lovelace et al., 2014); exploration of urban social-spatial inequalities (Shelton et al., 2015), and cartographic display (Field & O'Brien, 2010). There are also a profusion of studies based on Twitter data that analyse the spatial distribution of phenomena. These include: investigation of the tempo-spatiality of earthquake activity (Crooks et al., 2013); mapping the course of hurricanes using Tweets (Shelton et al., 2014); accessing Tweets to determine location, frequency and time of forest fires (De Longueville et al., 2009); monitoring of infectious disease outbreaks (Padmanabhan et al., 2014); predicting spatiality and severity of traffic congestion (Lécué et al., 2014), and mapping Tweet topics (Lansley & Longley, 2016).

The study of how invisible social factors affect and mirror the physical fabric of cities has successfully been explored using AGI. Batty et al. (2013) saw street networks and population densities replicated virtually by Tweets, and Ferrari et al. (2011) extracted urban mobility flows from AGI. These studies show us how AGI links the virtual world with the underlying physical urban structure by revealing virtual traces of processes and activities (Steiger et al., 2015).

3 Data and Methods

Data

Twitter (2017) states that it has 313 million active monthly users; Lansley & Longley (2016) calculate that this results in 500 million Tweets daily. For this study, Twitter data was collected for selected neighbourhoods within Inner London. The dataset consists of individual Tweets, each containing numerous fields of information, two of which are the longitude and latitude from which the Tweet was sent, which are used to form a point geometry object. Other fields include the status-update text of the Tweet, a temporal field with Tweet creation date and time, the source of the Tweet (e.g. from a linked social media site such as Instagram), the screen name of the user, and a unique identifier for the Tweet. In fact, Twitter data fulfils all the characteristics that geographical data should adhere to, as proposed by Worboys (1994).

Software

The open-source statistical software environment R and the R language were used for all data collection and analysis (https://www.r-project.org). R offers research reproducibility and self-documentation thanks to its command line format. It is also efficient at analysing large datasets and repeating tasks, and it can draw down base maps through internet calls. R's capabilities are enhanced by thousands of user-created packages which provide functions and code libraries for statistics, visualization, data handling and data collection.

Data collection from the Twitter API

The Twitter API is accessed from R using the twitteR library, and authentication codes are requested from Twitter (an API Key, an API token and an API secret). Tweets are filtered based on the searchTwitter() function's geocode argument and a query for keywords and hashtags which reference an Inner London neighbourhood (e.g. Soho, #Soho). The geocode argument specifies a geographic location (a latitude/longitude) and a search radius, which both remain constant. The geographic location chosen was Charing Cross (traditionally thought of as the centre of London), and the search radius was set at 5 miles, which is the extent of the study area (and coincides with the current Congestion Charge Zone https://tfl.gov.uk/modes/driving/congestion-charge/congestion-charge-zone).

The keywords and hashtags are changed each time the query is run depending on the neighbourhood that is being researched. The query is run multiple times for each neighbourhood, at different times of the week and day. Tweet retrievals from the Twitter API are limited to 1,500 each time the query is run. Longley *et a*l (2015) suggest that this is roughly 1% of a random selection of Tweets. However, Lansley & Longley (2016) suggest that this small percentage may still allow access to over 90% of all geo-tagged Tweets. 28 neighbourhoods were studied, based on place-name seeds from OpenStreetMap and Ordnance Survey to give an even spread throughout the study area. Table 1 shows the neighbourhoods selected and whether there is currently an eponymous official administrative area.

Table 1: Neighbourhoods selected from OpenStreetMap and Ordnance Survey data for keywords and hashtags

Neighbourhood	Official Administrative Boundary?
Aldgate	Aldgate Ward
Barbican	No
Bishopsgate	Bishopsgate Ward
Blackfriars	No
Bloomsbury	Bloomsbury Ward
Brick Lane	No
Clerkenwell	Clerkenwell Ward
Covent Garden	Holborn and Covent Garden Ward
Elephant and Castle	No
Euston	No
Farringdon	The Ward of Farringdon Within, The Ward of Farringdon Without
Fitzrovia	No
Holborn	Holborn and Covent Garden Ward
Hoxton	Hoxton Ward
Kings Cross	Kings Cross Ward
Lambeth	Lambeth London Borough, Lambeth and Southwark Greater London Assembly Constituency
Leicester Square	No
Marylebone	Marylebone High Street Ward
Mayfair	No
Paddington	No
Shoreditch	No
Soho	No
Southbank	No
Southwark	Southwark London Borough, Lambeth and Southwark Greater London Assembly Constituency
Spitalfields	Spitalfields and Banglatown Ward
Strand	No
Vauxhall	No
Waterloo	No

Qualitative thematic coding

Lovelace et al. (2016) concede that social media data suffers from a lack of veracity. To improve data quality, a methodology of unautomated quantitative coding was employed. A manual scrutinizing of geo-tagged social-media data for locational errors was implemented by Hollenstein & Purves (2010). A qualitative examination of Tweets for topic-related errors was considered by Albuquerque et al. (2015). In this study, a combination of these two approaches was used to produce a derived, quality controlled, dataset. Tweets are first visualized geographically and assessed for outlying Tweets in unexpected locations. The text of all Tweets is then examined manually for off-topic subject matter that indicates the user is not located within the neighbourhood about which they are Tweeting. Finally, Tweets are filtered by assigning each with a textual code depending on its content, as implemented by Jung (2015).

The qualitative coding is designed to find and categorize Tweets that may be sent from outside a neighbourhood, e.g. Tweets sent while travelling to or from a neighbourhood (Tweet example: 'Made it as far as Covent Garden en-route to Soho, gotta experience the gay night life here in London…'). Coding will also find a neighbourhood keyword used in the wrong context, e.g. a Tweet about a person or entity with the same name as a neighbourhood (Tweet example: 'Paul Strand 1890–1976 arguably one of the greatest documentary photographers of 20th century'), or one about an event that took place in a neighbourhood or will take place in the future (Tweet example: 'Still feeling stuffed after yesterday's meal at @BodeansBBQ in soho. huge massive portions, can't wait to go again.'). A Tweet for which the Tweet subject matches the Tweet located is classed as a Well Located Tweet (Tweet example: 'I'm at Gail's Artisan Bakery in Soho'). Table 2 shows the coding categories and gives a description of the criteria for determining how Tweets fit into each category.

Table 2: Coding categories and criteria for how Tweets are attributed to them

Category	Description
Travelling to or from neighbourhood.	The Tweeter is travelling to or from the neighbourhood that they mention; use of phrases like 'en-route to', 'bound', 'cycling to', 'on my way to' etc.
Tweeting from another location about a neighbourhood.	The Tweet is sent after being in a neighbourhood, past tense is used, or Tweet is about an event happening in another neighbourhood.
Tweeting from a venue named after a neighbourhood.	A venue named after a neighbourhood but not located in that neighbourhood, e.g. The Hoxton Hotel, Holborn
Well Located Tweet	Tweet where the subject matches the location.
GPS Positional Error	Many Tweets at exactly the same latitude and longitude from many different neighbourhood keyword searches.
Tweet about person or entity named after a neighbourhood	Tweet about a person or entity that has the same name as a neighbourhood.
Truncated Coordinates	Cannot locate Tweet accurately.
Uncertain Outlier	Anomalies, possibly due to network coverage.

Point clustering and polygon delimitation

A combination of methods is used to investigate the Tweet point clusters, both before and after qualitative coding. The mean centres of the Tweet point patterns are calculated and the dispersions of neighbourhood Tweets around official place-name seeds are recorded. Standard Distance Deviation (SDD) of Tweet dispersals is calculated to find the central tendency of the points to validate the dataset and assess the qualitative coding process. Buliung & Remmel (2008) used SDD to measure the spread of a set of points. Standard Deviational Ellipses (SDE) measure directional dispersal and are used to look for any directional factors affecting the point clusters. This checks whether any underlying topographical factors are affecting dispersal (Ayhan & Cubukcu, 2010). Kernel Density Estimation (KDE) is employed to determine neighbourhood point concentrations. These are compared geographically to official administrative boundaries (where they exist) to again validate the dataset. 2D contours are used to research the basis of neighbourhoods, identifying centres of neighbourhoods in order to draw conclusions about the reasons for vernacular neighbourhood demarcations. Finally, hexagonal binning and convex hulls are built to spatially determine neighbourhood extents and create discrete vernacular neighbourhood polygons.

The techniques described above provide point clusters of Tweets concerning where people believe they are located. These techniques then provide a quality-controlled dataset of Tweets from which vernacular neighbourhood extents can be analysed and delimited.

4 Results

Results for Tweet collection

Over a period of two months, 31,692 Tweets were collected, sent by 14,832 individual Twitter users. Figure 1 shows the uneven distribution of Tweets collected between the neighbourhoods. Shoreditch and Soho have by far the greatest number of Tweets, followed by Covent Garden and Mayfair.

When the Tweets are viewed spatially (Figure 2), denser point clusters of Tweets can be seen to the west and east of the study area. As well as highlighting the areas of high Tweet intensity, the 2D density estimation contours (Figure 3) highlight the areas of sparse Tweet coverage. These can be seen around the City of London, Westminster, Hyde Park, Regent’s Park, Green Park, and large swathes to the north and south of the study area.

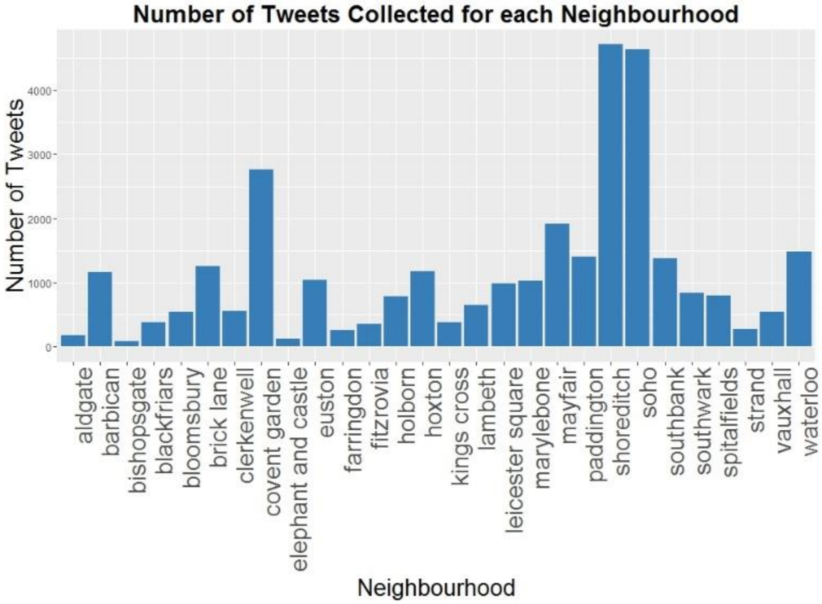

Figure 1: Distribution between neighbourhoods of Tweets collected

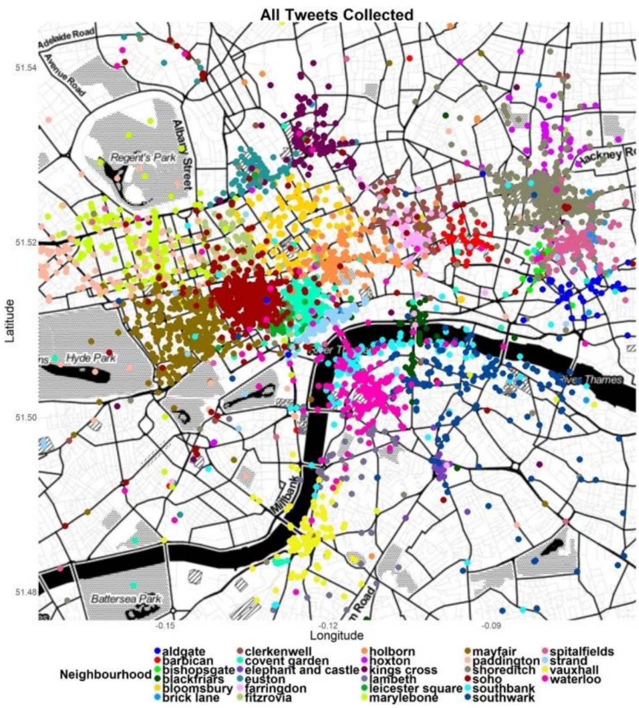

Figure 2: All Tweets collected for each neighbourhood. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

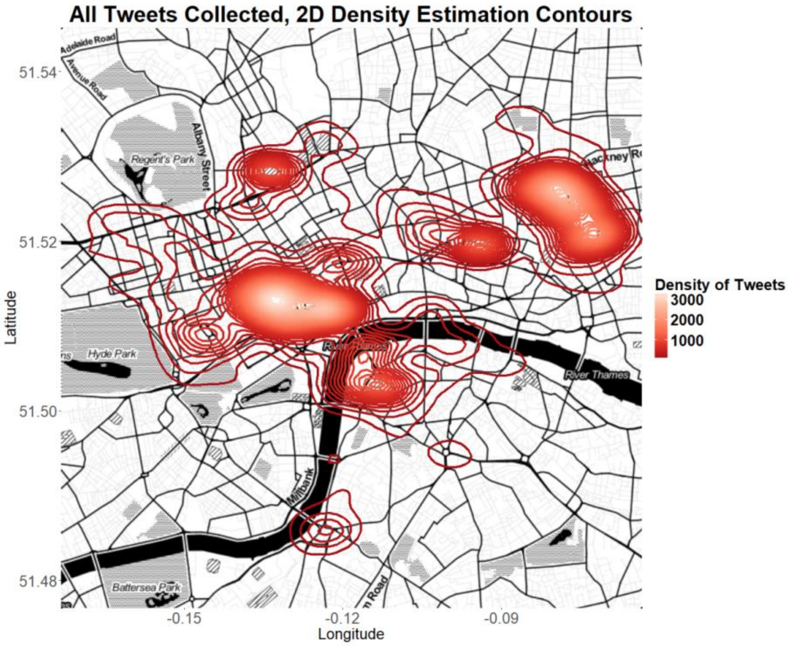

Figure 3: 2D density estimation contours for all Tweets collected. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

Qualitative coding results

The results of the qualitative coding exercise are presented in Table 3. What is clear is that the Well Located Tweet category includes by far the greatest number of Tweets, with around 95%.

Table 3: Results of the qualitative coding

Qualitative Coding Category	Number of Tweets
GPS positional error	393
Travelling to or from neighbourhood	69
Truncated coordinates	49
Tweet about person or entity with the same name as a neighbourhood	108
Tweeting from a venue named after a neighbourhood	286
Tweeting from other location about a neighbourhood	430
Uncertain outlier	47
Well Located Tweet	30,210

As an example of the manual qualitative coding exercise, a cartographic output of the results for the neighbourhood of Shoreditch is presented in Figure 4. This shows the distributions of each coding category and indicates that Shoreditch forms a large region to the east of the study area.

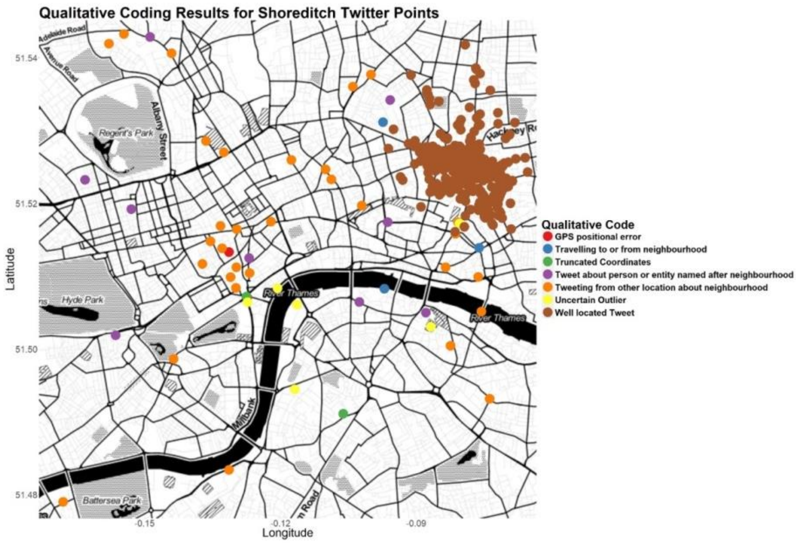

Figure 4: Qualitative coding results for Shoreditch. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

Cluster analysis results

The standard distance between official place-name seeds and mean centres of clusters decreased after qualitative coding. The SDD results saw the dispersal of Tweets around the means decrease for all neighbourhoods after qualitative coding. They also gave an indication of how dispersed or compact a neighbourhood is. In all cases, the KDEs of the neighbourhood Tweets were within or near their official boundaries. Figure 5 shows the KDE for Marylebone. The results of the SDE analysis demonstrated the directional tendencies of the neighbourhood Tweets. Figure 6 illustrates this for the Brick Lane neighbourhood, showing a linear directional trend along the eponymous thoroughfare.

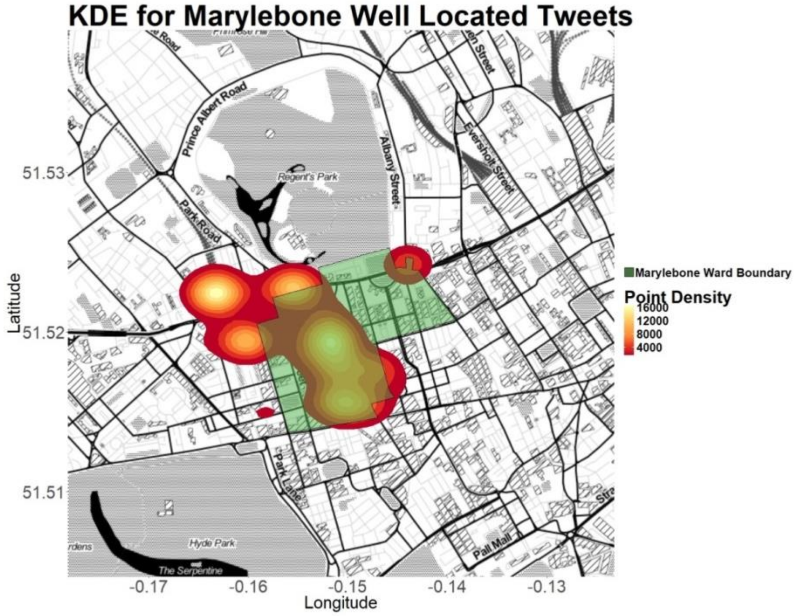

Figure 5: KDE for Marylebone, along with the Marylebone Ward official boundary. Contains OS OpenData © Crown Copyright/database right 2018. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

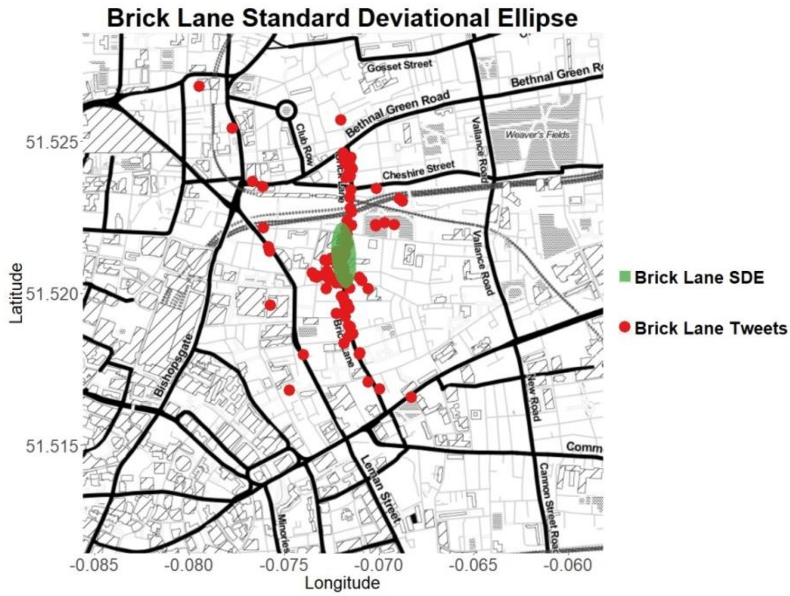

Figure 6: SDE result for the Brick Lane neighbourhood. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

4.1 Results for delimiting vernacular neighbourhoods and centres

Hexagonal binning was applied to the dataset to explore any preliminary vernacular neighbourhood formation (Figure 7). Following this, based on Tweets, vernacular neighbourhood polygons with convex hulls were delimitated. The results for the neighbourhoods in West London are presented in Figure 8. The epicentres of neighbourhoods were examined with 2D density contours (Figure 9).

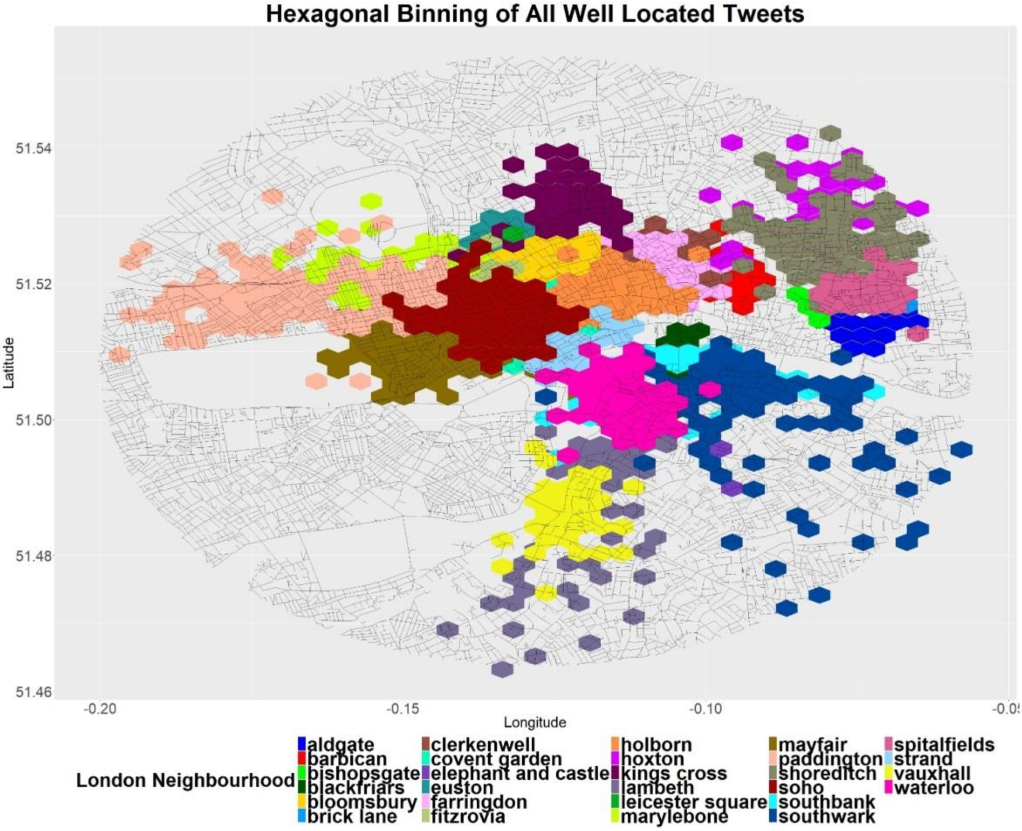

Figure 7: Hexagonal binning results for all Well Located Tweets. Contains OS OpenData © Crown Copyright/database right 2018

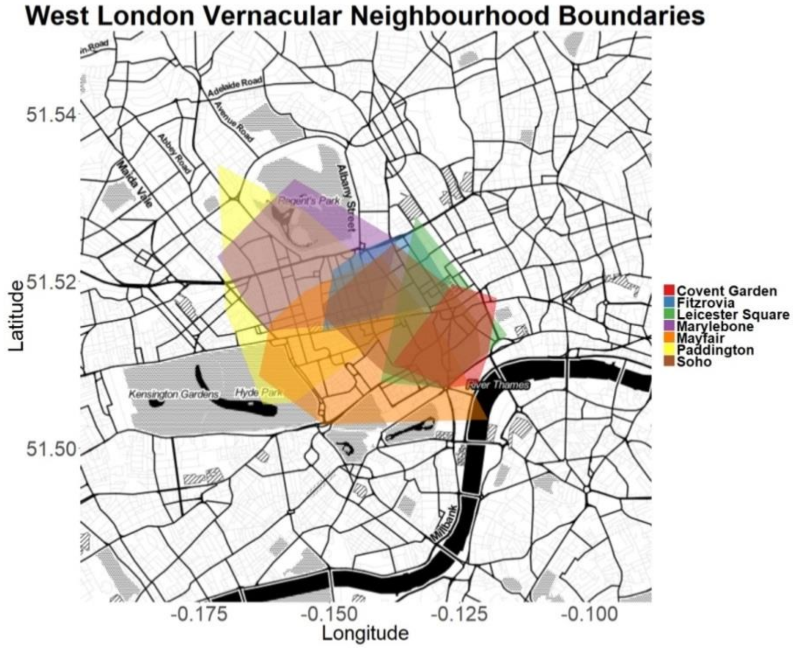

Figure 8: West London vernacular neighbourhood boundaries. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

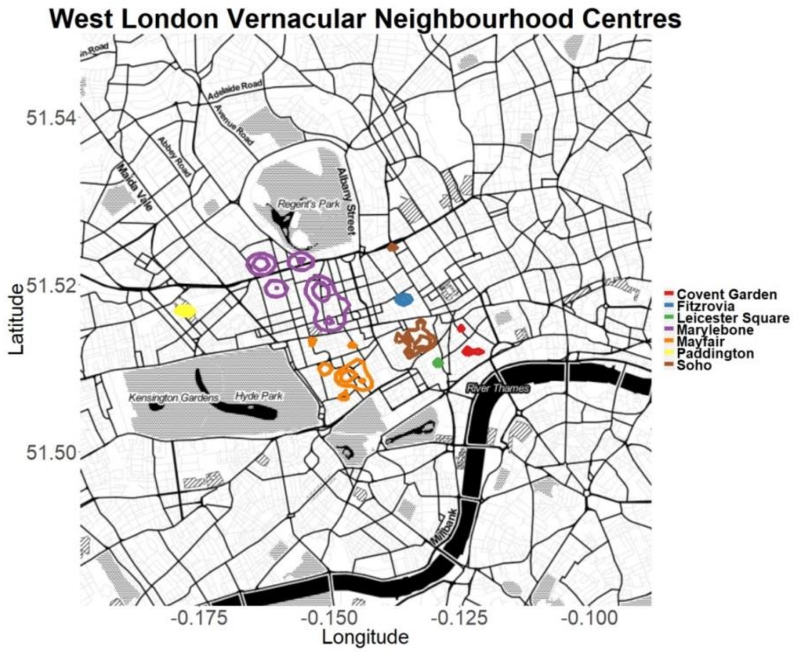

Figure 9: West London neighbourhood centres. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

5 Interpretation of Results

A point-based dataset of Tweets concerning vernacular perceptions of place was compiled. The accuracy and overall certainty of the dataset were then enhanced by qualitative coding.
The delimitation of discrete vernacular neighbourhood polygons from the fuzzy Tweet dataset proved effective. The overall visualization of neighbourhood polygons (Figure 8) provides a very overlapping picture, reflecting the underlying differences between individuals' spatial perceptions. There were also positive correlations between vernacular neighbourhood polygons and the official administrative boundaries, as highlighted in Figure 5. Figure 10 demonstrates the overlap between the Bishopsgate vernacular neighbourhood polygon and the official Bishopsgate Ward polygon. To provide a quantitative measure of overlap between the vernacular neighbourhood polygons and the official boundary polygons, the percentage of the vernacular polygon which intersects with the official boundary polygon (where this exist) was calculated. The results are illustrated in Figure 11.

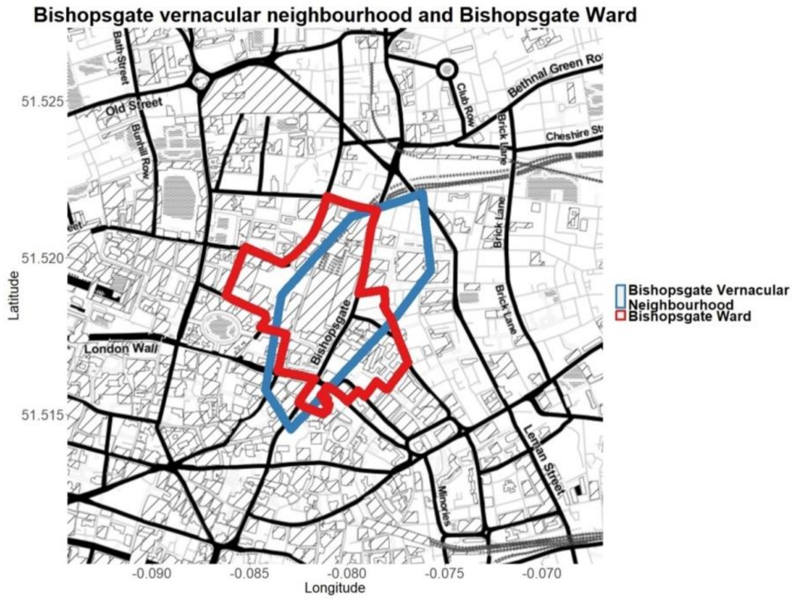

Figure 10: Bishopsgate vernacular neighbourhood and Bishopsgate Ward. Contains OS OpenData © Crown Copyright/database right 2018. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

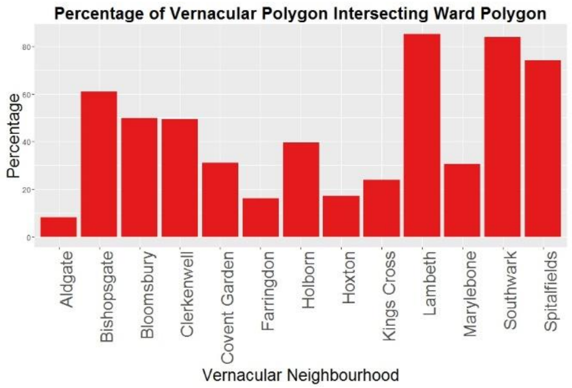

Figure 11: Percentage of vernacular polygon intersecting with its eponymous ward polygon

There are few Tweets in the City of London, in parks and in the large swathes of residential land to the north and south of the study area. Here we see open spaces, rivers, roads and zones of lower social activity effectively acting as edges (Lynch, 1960). Conversely, areas of social functionality exhibit higher social media activity and mirror the underlying topography and density of London. An example of an edge delimiting a vernacular neighbourhood is presented in Figure 12, where the River Thames acts as a perimeter, or a physical boundary, to the Southbank vernacular neighbourhood. Landmarks and transport hubs are also seen to form the basis of neighbourhoods (illustrated in Figure 13 with Waterloo), as are linear features such as roads (Figure 14).

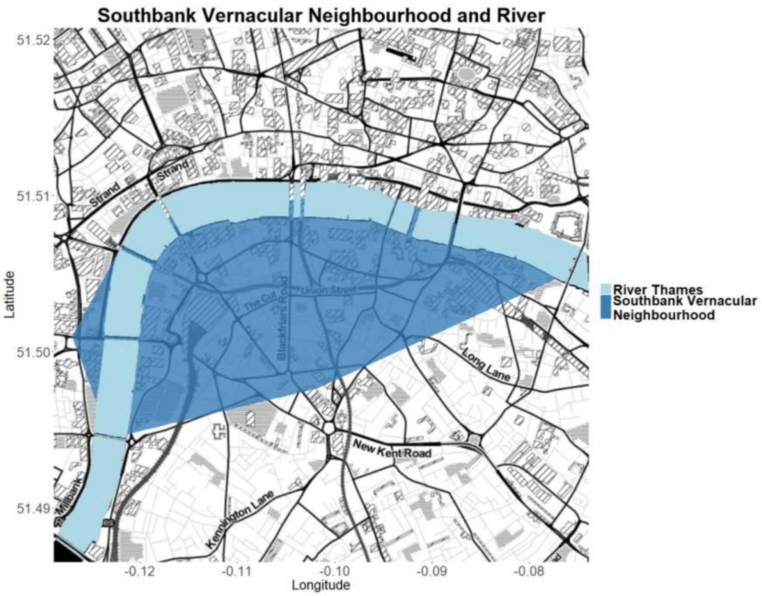

Figure 12: River Thames and Southbank vernacular neighbourhood. Contains OS OpenData © Crown Copyright/database right 2018. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

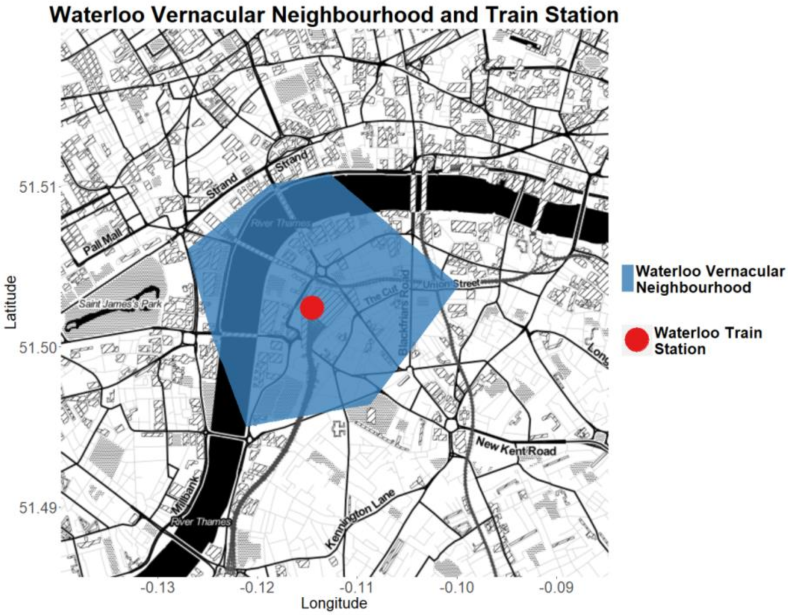

Figure 13: Waterloo vernacular neighbourhood. Contains OS OpenData © Crown Copyright/database right 2018. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

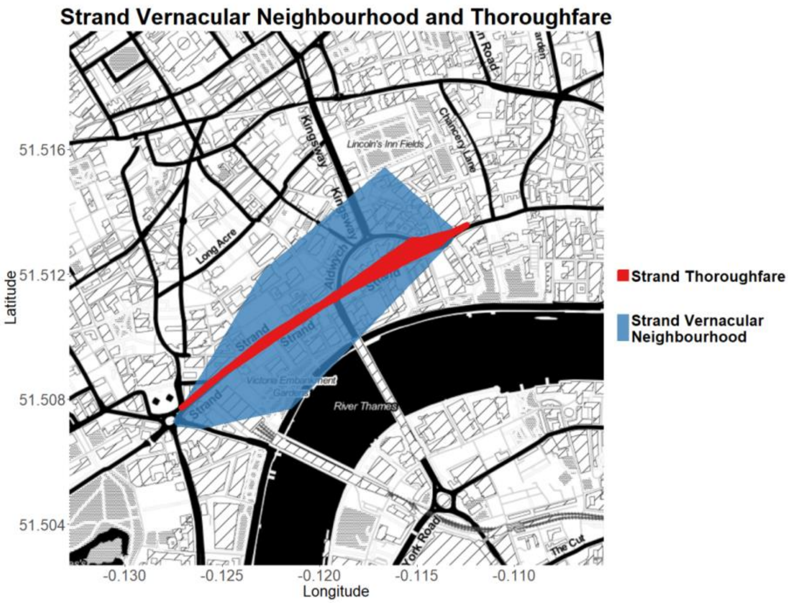

Figure 14: The Strand, a famous London thoroughfare, forming the basis of its vernacular neighbourhood. Contains OS OpenData © Crown Copyright/database right 2018. Map tiles by Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

Twitter allows users to Tweet from other platforms, which links data between websites. Within the Source field of each Tweet's metadata is the name of the platform used to post a Tweet. When plotted (Figure 15), it becomes evident that a very high proportion of the Tweets collected were also cross-posted on Instagram and Foursquare, demonstrating the interconnected nature of social media platforms.

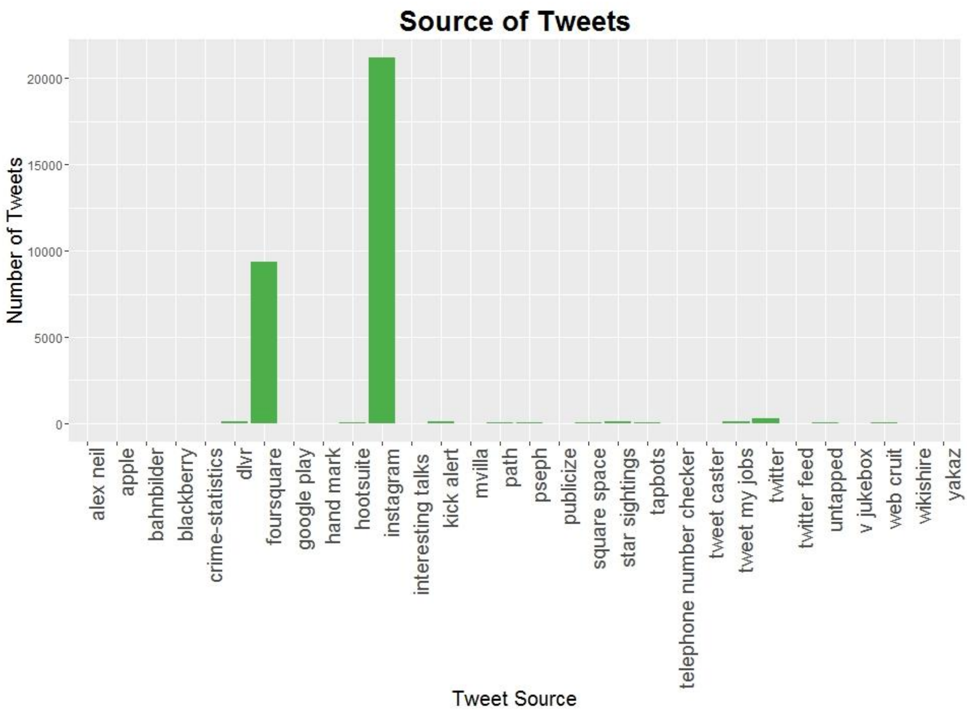

Figure 15: The source platforms of the Tweets collected

6 Conclusions and Further Work

31,692 Tweets were collected; this is considerably more responses than it would be possible to collect using traditional questionnaire or sketch map techniques. Slightly concerning is that these Tweets were sent by only 14,832 Twitter users, which means only 14,832 individual perceptions from which to study vernacular geography. This phenomenon of a few highly active users dominating Twitter output, and affecting research, was also observed by Shelton et al. (2015).

The subjective and time-consuming nature of the manual qualitative coding process could be a limitation if the study were to be expanded. To negate these concerns, a process of machine learning using training datasets could be devised. However, Hahmann et al. (2014) found in the context of their study that human Tweet classification proved to be more accurate than automated text detection techniques.

This research has demonstrated the feasibility of capturing vernacular geography from Twitter. As Twitter is a global social networking service, the study could readily be applied to other cities in the UK and worldwide. To allow the collection of sufficient data, the city would need to have a large population of potential Twitter users and a wide variety of neighbourhoods, both formal and informal. Paris, France would be a perfect city to explore vernacular geography in next as it fulfils both criteria. It would also be interesting to test data quality further by comparing the vernacular neighbourhood polygons from this study with polygons derived from other AGI sources such as Flickr.

References

Albuquerque, J, P. de., Herfort, B., Brenning, A. & Zipf, A. (2015). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, 29 (4): 667-689.

Ayhan, I. & Cubukcu, M. (2010). Explaining historical urban development using the locations of mosques: A GIS/spatial statistics-based approach. *Applied Geography* 30:229-238.

Batrinca, B. & Treleaven. (2015) Social media analytics: a survey of techniques, tools and platforms. *AI & Society* 30: 89-116.

Batty, M. (2013). *The New Science of Cities*. Cambridge (Massachusetts): The MIT Press.

Batty, M., Gray, S., Hudson-Smith, A., Milton, R., O'Brien, O. & Roumpani, F. (2013). Visualising spatial and social media. *CASA working papers series, paper 190*: London: University College London.

Batty, M. & Longley, P. (1994). *Fractal Cities: A Geometry of Form and Function.* San Diego, CA and London: Academic Press.

Brindley, P., Goulding, J. & Wilson, M. L. (2014). Mapping Urban Neighbourhoods from Internet Derived Data. *Proceedings of GISRUK 2014*: 355-364.

Buliung, R.N. & Remmel, T.K. (2008). Open source, spatial analysis, and activity-travel behaviour research: capabilities of the aspace package. *Journal of Geographical Systems* 10: 191-216.

Cope, M. (2003) Coding Transcripts and Diaries. In Clifford, N., French, S. & Valentine, G. (Editors) *Key Methods in Geography*. Chapter 27: 440-452. London: Sage.

Cope, M. & Elwood, M. (2009). *Qualitative GIS. A mixed methods approach*: London: Sage.

Coulton, C.J., Jennings, M.Z. & Chan, T. (2012). How Big is My Neighbourhood? Individual and Contextual Effects on Perceptions of Neighbourhood Scale. *American Journal of Community Psychology,* 51 (1-2): 140-150.

Crooks, A., Croitoru, A., Stefanidis, A. & Radzikowski, J. (2013). #Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS,* 17 (1): 124-147.

De Longueville, B., Smith, R.S. & Luraschi, G. (2009). 'OMG, from Here, I Can See the Flames!': A Use case of Mining Location Based Social Networks to Acquire Spatiotemporal Data on Forest Fires. *Proceedings of the 2009 International Workshop on Location Based Social Networks (Association for Computing Machinery),* Nov. 3, 2009, Seattle, WA, USA: 73-80.

Doran, B. & Young, M. (2013). Assessing the impact of a remote area casino: A mixed-methods approach using cognitive mapping and GIS. *Rural Society*, 23 (1): 20-31.

Elwood, M. & Cope, M. (2009). Introduction: Qualitative GIS: Forging Mixed Methods through Representations, Analytical Innovations, and Conceptual Engagements. In Cope, M. & Elwood, M. (Editors). *Qualitative GIS. A mixed methods approach*: (Chapter 1): 1-12: London: SAGE.

Feick, R. & Robertson, C. (2015). A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems,* 53: 96-109.

Ferrari, L., Rosi, A., Mamei, M. & Zambonelli, F. (2011). Extracting urban patterns from location-based social networks. *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, 1 November 2011, Chicago, IL, USA: 9-16.

Field, K. & O'Brien. (2010). Cartoblography: Experiments in Using and Organising the Spatial Context of Micro-blogging. *Transactions in GIS* 14 (s1): 5-23.

Fletcher, J. (1844) The Metropolis; its Boundaries, Extent, and Divisions for Local Government. *Journal of the Statistical Society of London* 7 (2): 103-143.

Galster, G. (2001). On the Nature of Neighbourhood. *Urban Studies,* 38 (12): 2111-2124.

Goodchild, M.F. (2011). Formalizing place in geographic information systems. In Burton, L.M, Kemp, S.P., Leung, M., Matthews, S.A. & Takeuchi, D.T. (Editors), *Communities, Neighbourhoods and Health: Expanding the Boundaries of Place* (Chapter 2): 21-33. New York: Springer.

Hahmann, S., Purves, R.S. & Burghardt, D. (2014). Twitter location (sometimes) matters: Exploring the relationship between georeferenced tweet content and nearby feature classes. *Journal of Spatial Information Science,* 9: 1-36.

Hollenstein, L. & Purves, R.S. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, 1: 21-48.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W. & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems,* 54: 240-254.

Huck, J.J., Whyatt, J.D. & Coulton, P. (2014). Spraycan: A PPGIS for capturing imprecise notions of place. *Applied Geography*, 55: 229-237.

Jung, J.K. (2015). Code clouds: Qualitative geovisualization of geotweets. *The Canadian Geographer,* 59 (1): 52-68.

Lansley, G. & Longley, P.A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems,* 58: 85-96.

Lécué, F., Tucker, R., Bicer, V., Tommasi, P., Tallevi-Diotallevi, S. & Sbodio, M. (2014). Predicting Severity of Road Traffic Congestion Using Semantic Web Technologies. *Proceedings 11th International Conference, ESWC 2014*, Anissaras, Crete, Greece, May 25-29, 2014: 611-627

Liesch, M., Dunklee, L.M., Legg, R.J., Feig, A.D. & Krause, A.J. (2015). Use of Business-Naming Practices to Delineate Vernacular Regions: A Michigan Example, *Journal of Geography*, 114 (5): 188-196.

Longley, P.A., Adnan, M. & Lansley, G. (2015). The geotemporal demographies of Twitter usage. *Environment and Planning*, 47: 465-484.

Lovelace, R., Birkin, M., Cross, P. & Clarke, M. (2016). From Big Noise to Big Data: Towards the verification of Large Data sets for Understanding Region Retails Flows. *Geographical Analysis,* 48: 59-81.

Lovelace, R., Birkin, M. & Malleson, N. (2014). Can social media data be useful in spatial modelling? A case study of 'museum Tweets' and visitor flows. *GISRUK 2014*, 16-18 April 2014, University of Glasgow, Scotland.

Lynch, K. (1960). *The Image of the City*. Cambridge (MA): The MIT Press.

Montello, D.R., Goodchild, M.F., Gottsegen, J, & Fohl, P. (2003). Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation*, 3 (2 & 3): 185-204.

Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhang, Z., Gao, Y., Soltani, K. & Liu, Y. (2014). FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. *Concurrency and computation: Practice and Experience,* 26: 2253-2265.

Raanan, M.G. & Shoval, N. (2014). Mental maps compared to actual spatial behaviour using GPS data: A new method for investigating segregation in cities. *Cities*, 36: 28-40.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H., Milcinski, G., Niksic, M., Painho, M., Podor, A., Olteanu-Raimond, A. & Rutzinger, M. (2016) Crowdsourcing, Citizen Science or Volunteered Geographic Information?

The Current State of Crowdsourced Geographic Information. *International Journal of Geo-Information,* 5 (55): 1-23.

Shelton, T., Poorthuis, A. & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142: 198-211.

Stanton Fraser, D., Jay, T., O'Neill, E. & Penn, A. (2013). My neighbourhood; Studying perceptions of urban space and neighbourhood with moblogging. *Pervasive and Mobile Computing*, 9: 722-737.

Steiger, E., Westerholt, R., Resch. B. & Zipf, A. (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*, 54: 255-265.

Stephens, M. & Poorthuis, A. (2015). Follow thy neighbourhood: Connecting the social and the spatial networks on Twitter. *Computers, Environment and Urban Systems,* 53: 87-95.

Sui, D. & Goodchild, M. (2011). The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science,* 25 (11): 1737-1748.

Tonkiss, F. (2013) *Cities by design: the social life of urban form*. Cambridge: Polity.

Twaroch, F.A, Jones, C.B, Abdelmoty, A.I. (2008). Acquisition of a Vernacular Gazetteer from Web Sources, pp. 61-64, Susanne Boll, Erik Wilde (Editor), *First International Workshop on Location and the Web* (LocWeb 2008), Beijing, China, April 2008.

Twitter (2017). Twitter Usage/Company Facts (Last updated June 30, 2016). Available from: https://about.twitter.com/company [Accessed 12th March 2017].

Vallée, J., Le Roux, G., Chaix, B., Kestens, Y. & Chauvin, P. (2015). The 'constant size neighbourhood trap' in accessibility and health studies. *Urban Studies,* 52 (2): 338-357.

Worboys, M.F. (1994). Object-oriented approaches to geo-referenced information. *International Journal of Geographical Information Systems,* 8 (4): 385-399.

Zimmer, M & Proferes, N.J. (2014). A topology of Twitter research: disciplines, methods, and ethics. *Aslib Journal of Information Management*, 66 (3): 250-261.