RESEARCH ARTICLE

Delineate urban boundaries in Great Britain from the network of Twitter user spatial interactions

Urban regions are discrete urban areas with various socio-economic relationships and citizen commute patterns. Defining the boundaries of urban regions is important to city planning, traffic management and resource allocation. Existing urban boundaries are usually defined by government political and administrative purposes, however, it is not clear whether the boundaries truly reflect people's daily interactions with the urban space in their intra- and inter regional activities. In this study, we presented a method to redraw urban boundaries based upon human interaction with the physical space. Specifically, we delineated the urban boundaries of Great Britain using a spatial Twitter user interaction network that was inferred from over 69 million Twitter messages. Our study connects human mobility research with the delineation of non-administrative anthropographic urban boundaries based on Twitter user spatial interactions, which provides a new understanding of the interactions between urban structure and human activity. We redrew the non-administrative anthropographic boundaries in a hierarchical fashion based on different physical movement ranges of users inferred from the collective mobility patterns of Twitter users in Great Britain. The result of strongly connected urban regions in the form of communities in the network space yield geographically cohesive, non-overlapping urban areas that provides a clear delineation of the non-administrative anthropographic urban boundaries of Great Britain. The technique was applied to both national scale (Great Britain) and municipal scale (the London metropolis). While our results corresponded well to the administrative boundaries, many unexpected and interesting boundaries were identified.

Keywords: mobility pattern, urban boundary, spatial interaction, spatial network, community structure

1. Introduction

Official urban boundaries are defined by government agencies for political and administrative purposes. Urban environments are conceptualized as spaces that are recreated and formed by human activities (schliephake 2014). A fundamental question when using the administrative, "top-down", approach to defining urban boundaries is whether the outcome reflects the spatial interactions of humans. These interactions can take the

form of trade, commerce, social connections, and political activity across the borders. [7] Urban boundaries that respect the human interaction space can be very import to city [8] planning, traffic management and resource allocation (Lynch 1960, Jiang and Miao 2015, [9] Liu *et al.* 2015, Long *et al.* 2015). Many studies adopt a "bottom-up" approach to urban [10] boundary delineation, where the geographic space is partitioned into small units and each [11] unit is modeled as a node within a network structure. A suitable community detection [12] algorithm is applied to partition the network and associated geographic space based on [13] the strength of human interaction among the nodes Lancichinetti and Fortunato (2009). [14] Different social and physical human interactions were considered to establish the edges [15] of the network connecting the nodes. For example, a large set of telephone call records [16] were used to represent the network of human interaction across space to delineate urban [17] boundaries in Great Britain Ratti *et al.* (2010). Extending the previous method to differ- [18] ent countries Sobolevsky *et al.* (2013), the authors argue that this method yields cohesive [19] geographic divisions that follow the socio-economic boundaries. While other researchers [20] use social ties of Twitter users to identify cohesive regions for different countries across [21] the world Kallus *et al.* (2015), they found evidence for dividing the urban space due to [22] local conflicts and cross-country unifying trends that further support the "bottom-up" [23] approach to mapping non-administrative anthropographic boundaries. [24]

In this study, we describe a novel approach to delineating urban boundaries based on [25] the movements of Twitter users extracted from more than 69 million Twitter messages [26] from June 1st to December 31st, 2014. Geo-located Twitter data is proven to be a useful [27] source for studying human mobility patterns at large spatial scales (e.g. the country [28] level) Hawelka *et al.* (2014), Jurdak *et al.* (2015). We argue here that Twitter user [29] mobility patterns provide a different view of non-administrative units based on physical [30] commutes rather than social ties or phone call initiation. A unique advantage is that [31] non-administrative anthropographic urban boundaries can be delineated in a hierarchical [32] fashion based upon different ranges of physical movement, which are inferred from the [33] collective mobility patterns of Twitter users in Great Britain. In addition, Twitter data [34] are not as sensitive to user privacy issues and do not exhibit spatial granularity that is [35] limited to the postal code level Thiemann *et al.* (2010). [36]

We delineated the geography of urban boundaries in Great Britain by imposing a [37] virtual fishnet over the islands of Great Britain. Twitter user movements were used to [38] establish the connections between the fishnet's cells to form a connectivity network, where [39] each cell acts as a node within the network. We applied the map equation algorithm (De [40] Domenico *et al.* 2015) to partition the network and associate geographic regions. The map [41] equation algorithm was selected to avoid the inherent resolution problem Fortunato and [42] Barthlemy (2007) of the common modularity maximization method (Newman 2006). We [43] found that the collective mobility patterns of Twitter users in Great Britain are divided [44] into several distance ranges starting from short, intra- to inter-city movements with clear [45] distinction points. The identification of connected regions at each of these distance ranges [46] yielded hierarchical boundaries of the urban space in Great Britain. Our study provides a [47] first step in connecting human mobility research with the definition of non-administrative [48] anthropographic urban boundaries based on Twitter user spatial interaction. It provides [49] a new understanding of the interactions between human activity and urban structures. [50]

2.   Materials and Methods

2.1.   *Overview*

We use large-scale Twitter user spatial interactions inferred from the collective mobility patterns of Twitter users to delineate urban boundaries in Great Britain. A Twitter user's movement is defined here as the individual's geographic relocation or displacement Gonzlez *et al.* (2008). This is not equivalent to a "trip" taken by an individual, because, displacement includes situations when the time interval between two recorded locations is one month. To identify the clusters of urban regional connectedness using large-scale Twitter data, we first imposed a virtual fishnet over the islands of Great Britain. Twitter user movements were then used to establish the connectivity between fishnet cells to form a connectivity network where each cell is considered a node within the network. Finally, by detecting strongly connected clusters in the network space using community detection methods Coscia *et al.* (2011), we were able to reveal urban boundaries within the geographic space. An overview of the approach is illustrated in Fig. 1 and the details are presented in the following sections.
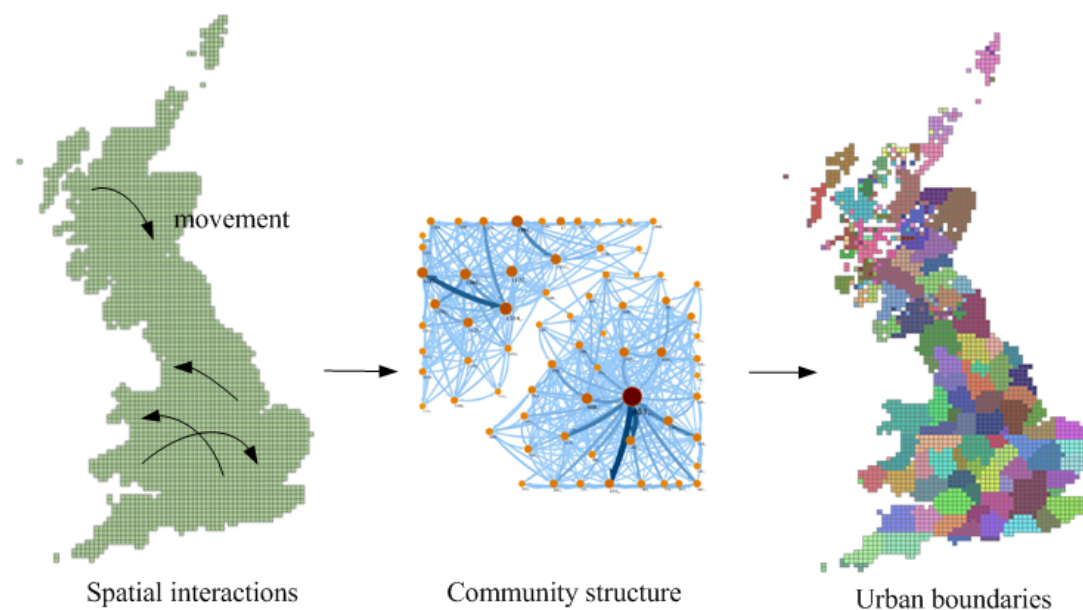

Figure 1. **An overview for delineating urban boundaries from Twitter user movements**

2.2.   *Geo-located Twitter Data*

A geo-located tweet is a Twitter message with an additional geo-tag expressed as a pair of geographical coordinates that represent the location from which the tweet was sent. For this study, geo-located tweets were collected using the Twitter Streaming API (https://dev.twitter.com/streaming/overview) by supplying a geographical bounding box to retrieve all the geo-located tweets within an area of interest. To ensure complete

coverage of Great Britain, we set the bounding box to the British Isles using the lower   72
left and upper right coordinates (49.497, -14.854), (61.186, 2.637) respectively. This does   73
include the whole of Ireland part of France. We implemented a data crawler to continu-   74
ously collect 7-months of data (June 1st – December 31st, 2014) resulting in over 101.8   75
million tweets with a total data volume of 60 GB, which we restrict to 69 million based   76
on a process outlined below. To showcase the overall spatial coverage of the collected   77
geo-located tweets, a density map of the tweet locations in British Isles for July 2014 is   78
shown in Fig. 2. The collected point visualization reveals the geography of cities. Notice   79
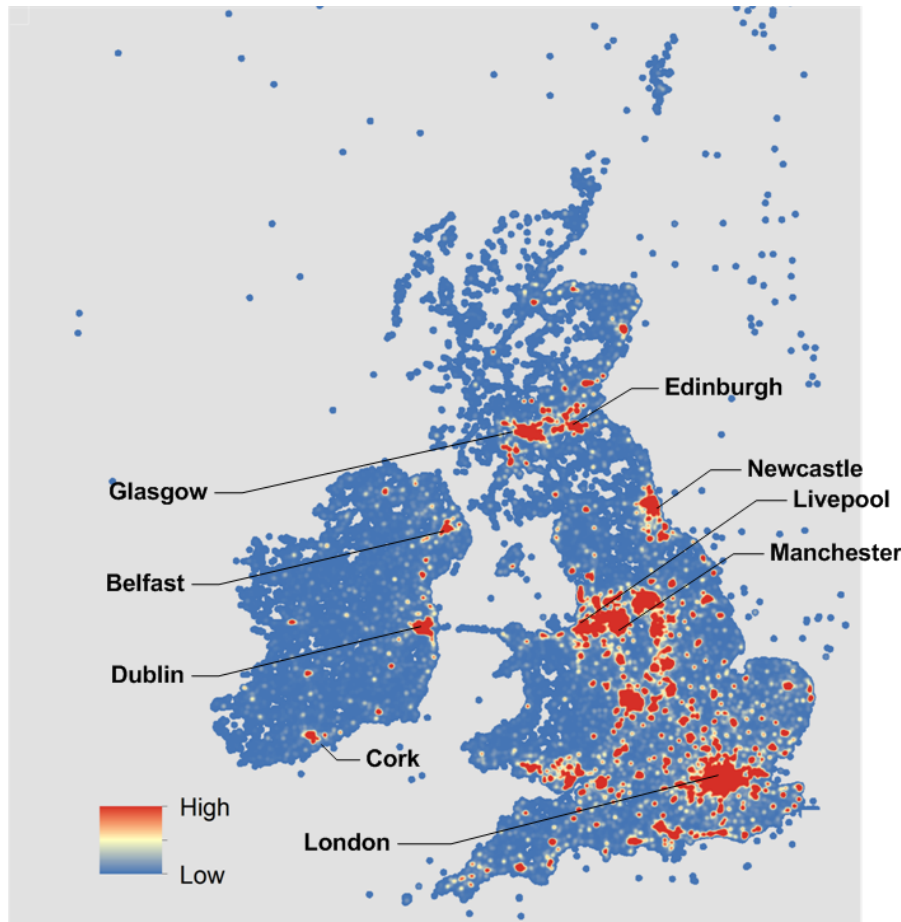the clusters with higher densities of tweets correspond to the locations of major cities.   80

Figure 2. **A density map of the Twitter user locations for July 2014**

To verify the collected tweets are from human users, the raw tweets were further fil-   81
tered using the following steps. First we removed the duplicated messages in the dataset.   82
Next, we removed non-human users based on unusual relocation speed Hawelka *et al.*   83
(2014), Jurdak *et al.* (2015). We adopted a relocation speed threshold of 240 m/s as used   84
by Jurdak *et al.* (2015). We examined all of the consecutive locations of each user and   85
excluded those with relocating speeds in excess of the threshold. Because the original   86
location information embedded in the geo-tag is given in units of latitude and longitude,   87
we projected the points into the British National Grid (EPSG: 27700) coordinate system   88
to reduce the complexity of the required distance calculations. Finally, we used the geo-   89
graphical boundary of Great Britain, which is derived from Office for National Statistics   90

(ONS) of UK (`http://www.ons.gov.uk/ons`), to further restrict the remaining tweets ₉₁
to be "domestic". Based on these restrictions, the filtered dataset for the following study ₉₂
contains 69,847,497 tweets made by 1,153,891 twitter users. ₉₃

At this stage, each geo-located tweet is represented as a tuple $\langle user\_id, loc, t, m \rangle$, where ₉₄
$user\_id$ is an anonymous Twitter user id; loc is the recorded location of the tweet as a ₉₅
coordinate pair; t is the timestamp of the tweet's post; and m is the actual content of ₉₆
the tweet. To protect Twitter user privacy, the id field was replaced with a randomly ₉₇
generated unique number. Because we are interested in collective Twitter user movement ₉₈
instead of individual user trajectories, the content of the message was removed. Our ₉₉
simplified geo-located tweet dataset can be shared with other researchers upon request. ₁₀₀

2.3.   *A network of large-scale Twitter user spatial interactions* ₁₀₁

Urban regions are discrete urban areas with various socio-economic relationships and ₁₀₂
citizen commute patterns. Some urban regions are more strongly connected to each other ₁₀₃
than other urban regions. Therefore, this study connects urban regions when a twitter ₁₀₄
user's movement begins in one region and ends in another region. These connections ₁₀₅
can be represented by an origin-destination (OD) matrix based on the collective Twitter ₁₀₆
user displacements within the dataset. This OD matrix is essentially a mathematical ₁₀₇
representation of a weighted directed graph $G \equiv \langle V, E_w \rangle$, where V is a set of spatial nodes ₁₀₈
corresponding to the underlying urban regions and E_w is a set of edges representing the ₁₀₉
connections between a pair of nodes and the corresponding weights are assigned by the ₁₁₀
accumulated volume of Twitter user movements. ₁₁₁

To build our spatial network at a national level, we had to determine the basic units ₁₁₂
to serve as spatial nodes of the connectivity network of urban regions. Previous studies ₁₁₃
have suggested equi-distant spatial tessellation to generate nodes, which uses voronoi ₁₁₄
polygons to partition the space based on the collected points Rinzivillo *et al.* (2012), ₁₁₅
Zhong *et al.* (2014). This approach demonstrates improvements for estimating the loca- ₁₁₆
tions of mobile phone records based on the cell tower triangulation Gonzlez *et al.* (2008), ₁₁₇
Qian *et al.* (2013). However, equi-distant tessellation decreases the spatial resolution of ₁₁₈
aggregated geo-located tweets, because the location information is usually derived from ₁₁₉
the embedded GPS within mobile devices and tends to provide greater accuracy Sakaki ₁₂₀
et al. (2010), Zandbergen (2009). Another popular approach is to partition the space into ₁₂₁
a grid of spatial pixels Liu *et al.* (2014), Ratti *et al.* (2010). However, the size of the cell ₁₂₂
can potentially lead to biases due to the Modifiable Area Unit Problem (MAUP) Open- ₁₂₃
shaw (1984), Wong (2009), where different choices of unit size can lead to significant ₁₂₄
variant findings. To compare our investigation with the findings of similar studies, and ₁₂₅
avoid subjectively deciding the cell size, we performed statistical analysis of Twitter user ₁₂₆
mobility patterns in Great Britain and measured the distribution of collective Twitter ₁₂₇
user displacements and the radius of gyrations of individuals Gonzlez *et al.* (2008), Jur- ₁₂₈
dak *et al.* (2015). The radius of gyration is a metric to distinguish mobility patterns of ₁₂₉
individuals Gonzlez *et al.* (2008), which is defined as Eq. (1): ₁₃₀

$$r_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - p_{centroid})^2}, \text{where } p_{centroid} = \frac{1}{n}\sum_{i=1}^{n}p_i \qquad (1)$$

It measures the accumulated distances of deviation from the center of mass of an indi- ₁₃₁

vidual user's trajectory, where p_i is one of the user's locations and $p_{centroid}$ is the center $\quad$ 132
of mass of the user's trajectory. By examining the probability distributions of the radius $\quad$ 133
of gyration, also known as the spatial dispersal kernel $P(r_g)$ Brockmann *et al.* (2006), $\quad$ 134
we chose 10 km as the cell size to quantify the spatial coverage of the majority of Twit- $\quad$ 135
ter users in Great Britain (Fig. 4 - c, with details shown in the next section). Thus, $\quad$ 136
we created a fishnet with 2784 10-km size cells. The cells of the fishnet act as proxies $\quad$ 137
to represent individuals' spatial coverage areas to focus more on the inter-connectivity $\quad$ 138
among cells and identify strongly connected cell clusters. $\quad$ 139

2.4.   *Community structure of the spatial networks* $\qquad$ 140

Based on the derived spatial network, which is a directed weighted graph, we further $\quad$ 141
determined clusters of strongly connected spatial nodes, known as communities, in the $\quad$ 142
graph space Coscia *et al.* (2011). There are a variety of community detection algorithms $\quad$ 143
that produce different results depending the definition of community within the net- $\quad$ 144
work Coscia *et al.* (2011). A common community detection method is based on modular- $\quad$ 145
ity maximization Newman (2006), seen in previous studies Hawelka *et al.* (2014), Ratti $\quad$ 146
et al. (2010), Song *et al.* (2012). However, such an approach is often problematic: it is $\quad$ 147
found to have an inherent resolution problem, where small communities are either ig- $\quad$ 148
nored Fortunato and Barthlemy (2007) or assigned with high modularity scores Guimer $\quad$ 149
et al. (2004); and it is found to produce less informative partitions in many empirical $\quad$ 150
networks Good *et al.* (2010). Since our graph is a directed weighted graph, the alternative $\quad$ 151
community detection library documented in the literature is Infomap De Domenico *et al.* $\quad$ 152
(2015), Rosvall and Bergstrom (2008), which is considered to produce better community $\quad$ 153
detection results Lancichinetti and Fortunato (2009). $\quad$ 154

$\quad$ Infomap uses the map equation Rosvall *et al.* (2010) to represent the probability of flow $\quad$ 155
of random walks within information systems Rosvall and Bergstrom (2008). It identifies $\quad$ 156
communities by minimizing the expected description length of the trajectory of a random $\quad$ 157
walker, which is shown below: $\quad$ 158

$$L(M) = qH(Q) + \sum_{i=1}^{m} p_i H(p_i) \qquad (2)$$

$\quad$ In Eq. (2) , $L(M)$ consists of two terms: qH(Q) is the entropy of the movement among $\quad$ 159
clusters and $\sum_{i=1}^{m} p_i H(p_i)$ is the entropy of movement within clusters. To be more spe- $\quad$ 160
cific, q is the probability that a random walker jumps from one cluster to another, while $\quad$ 161
p_i is the probability of the in-cluster movement of cluster i. In this regard, this algorithm $\quad$ 162
can be intuitively tailored to describe strongly connected clusters of urban regions based $\quad$ 163
on Twitter user movement. The detailed literatures and implementations of Infomap can $\quad$ 164
be found on this website (`http://mapequation.org`). Note that Infomap is capable of $\quad$ 165
performing multi-level community detection De Domenico *et al.* (2015), but we only use $\quad$ 166
the algorithm to produce our most detailed community structures to examine groups of $\quad$ 167
strongly connected urban regions. $\quad$ 168

3.   Results and Discussion

3.1.   *Collective Mobility Patterns of Twitter Users in Great Britain*

We first modeled different aspects of the collective mobility patterns of Twitter users. These patterns include: the number of visited locations per user, the collective user displacements, and the radius of gyration of individuals in order to identify natural breaks in user travel patterns. We then used these natural breaks within the mobility patterns to partition the geographic space of Great Britain into fine-grained cells and established the connectivity among these cells to redraw non-administrative anthropographic urban boundaries.

We found that the cumulative distribution function Clauset *et al.* (2009) of the number of locations visited by each Twitter user follows a two-tier power law distribution (shown in Fig. 3). The majority of the data distributed follows a truncated power-law distribution $P(X \geq x) \sim x^{-\alpha}e^{-\lambda x}$, where $\alpha = 1.24$, $\lambda = 0.00132$; and the tail part (less than 2% of the whole population) follows a power-law distribution $P(X \geq x) \sim x^{-\alpha}$ with α value is 3.2. The distribution was found to be consistent over all individual months examined (June to December, 2014), which has a slight offset in the truncated power-law distribution (the mean α value is $1.26 \pm$ with a 0.05 σ and the mean λ value as $0.00134 \pm$ with a 0.0002 σ).

The two-tier power law distribution indicates that the collective behavior of Twitter users visiting different locations can be well approximated with a (truncated) Lévy Walk (a random walk) model Rhee *et al.* (2011), Reynolds (2012), which has also been identified in many human mobility studies using different mobility data Zhao *et al.* (2015). The similarity among the distributions suggests that the mobility data collected from geo-located tweets is temporally stable, at least at the monthly interval, which indicates that our approach using Twitter user mobility to delineate urban boundaries is viable. In addition, the Lévy Walk model reveals the diversity regarding the number of visited locations per user, which indicates a level of "randomness" in Twitter user movement across space. It, in turn, justifies our choice of using the map equation community detection algorithm Rosvall and Bergstrom (2008) to identify the clusters of urban regional connectedness using large-scale Twitter user movement data.

We further studied two aspects of the Twitter user mobility patterns: the distribution of Twitter user displacement and the radius of gyration. Twitter user displacement refers to the distance between two consecutive locations in a user's trajectory using a straight-line distance metric. The radius of gyration describes the deviation of distance from the center of mass in a user's trajectory. The probability distributions of the collective user displacement $P(d)$ and radius of gyration $P(r_g)$ are presented in Fig. 4, where the fitting method for identifying different distance ranges is derived from Jurdak *et al.* (2015). The probability distribution of the collective displacements can be approximated by $P(d) \sim \lambda_1 e^{-\lambda_1(d-d_{min})}$, $d_{min} = 10m$ from [10m, 70m] (accounting for 3 % of the population), $P(d) \sim \beta\lambda_1 d^{\beta-1}e^{-\lambda^1(d^\beta-d_{min}^\beta)}$, $d_{min} = 100m$ from [70m, 70km] (93 % of the population), and $P(d) \sim d^{-\alpha}$ [> 70km] (4 % of the population). The displacement distance between 70m and 100km can be further approximated by two power law distributions with a cut-off point at 4km (55% distances are less than 4km and 40% distances between 4km and 100km), which indicates the urban movement captured by the geo-located Twitter data to reveal two different modes: inter-city and intra-city movement. In short, these fitting functions suggest the existence of multi-scale or multi-modal urban movements captured from Twitter users in Great Britain, which means the geographically cohesive, non-overlapping urban areas identified in the next section are not just a result of short
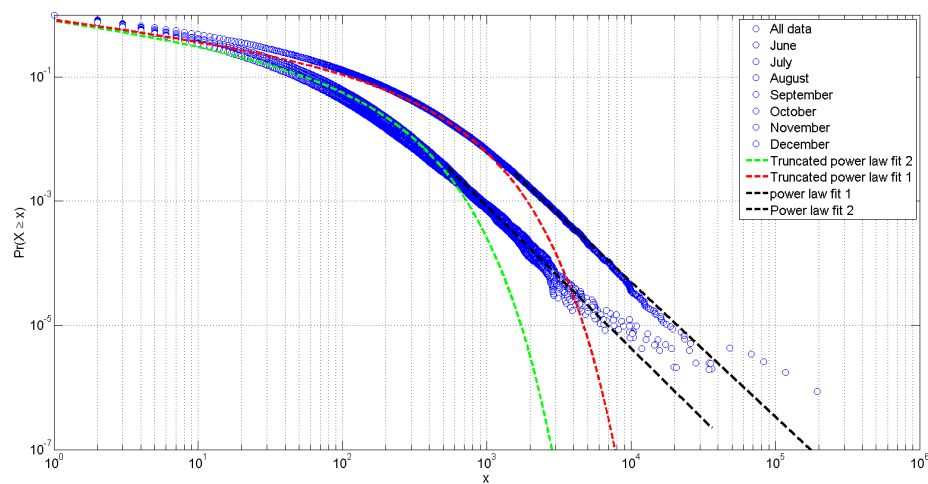
Figure 3. **Cumulative distribution of the number of locations visited by each Twitter user during different timespan**

distance movement but emerge naturally from the broader Twitter user mobility pattern. ₂₁₇
Note that a similar multiphase pattern was observed in Twitter user displacements in ₂₁₈
Australia, but with slightly different distance rangesJurdak *et al.* (2015). ₂₁₉

Further, we analyzed the distribution of radius of gyration to understand the movement ₂₂₀
from the point of view of individual Twitter users rather than separate displacements. ₂₂₁
The distribution of the radius of gyration can be approximated through a combination ₂₂₂
of three functions: $P(r_g) \sim \lambda_2 e^{-\lambda_2(r_g - r_{g_{min}})}, r_{g_{min}} = 10m$ from [10 m, 30m], $P(r_g) \sim$ ₂₂₃
$\lambda_2 e^{-\lambda_2(r_g - r_{g_{min}})}$ from [50m, 10km], and $P(r_g) \sim r_g^{-\alpha}$ [10km, 100km], where these three ₂₂₄
functions account for 92% of all the users. This suggests that there are three primary ₂₂₅
types of users that: (1) tend to stay at one location or at nearby locations when they ₂₂₆
tweet, or (2) tend to move at the intra-city scale when they tweet, or (3) tend to exhibit a ₂₂₇
large spatial coverage. (1) and (2) account for approximately 53% of all users. The radius ₂₂₈
of gyration between 30m and 100km can be further approximated by two power law ₂₂₉
distributions with a cut-off point at 2.5km (35 % users with radius of gyration less than ₂₃₀
2.5km and 57% users with radius of gyration between 3km and 100km) that suggests two ₂₃₁
main types of spatial coverage of Twitter user movement within Great Britain. Note that ₂₃₂
the accuracy of these values for defining the distance bound depends upon the accuracy ₂₃₃
of the location information of each geo-located tweet. These findings are consistent with ₂₃₄
the findings in the literature on human mobility, where the radius of gyration of human ₂₃₅
movement is bounded to different distance ranges Brockmann *et al.* (2006), Gonzlez *et al.* ₂₃₆
(2008). The distance-decay effects found in both user displacements and the radius of ₂₃₇
gyration shows evidence of spatial proximity in Twitter user movement. It explains that ₂₃₈
the communities of urban regions within the graph space are geographically close, but ₂₃₉
are able to be separated from other groups, which results in the delineation of urban ₂₄₀
boundaries based on the spatial interactions of Twitter users. ₂₄₁

3.2.   *Redrawing Great Britains Urban Boundaries*   ₂₄₂

The network of Twitter user spatial interaction was constructed by nodes representing ₂₄₃
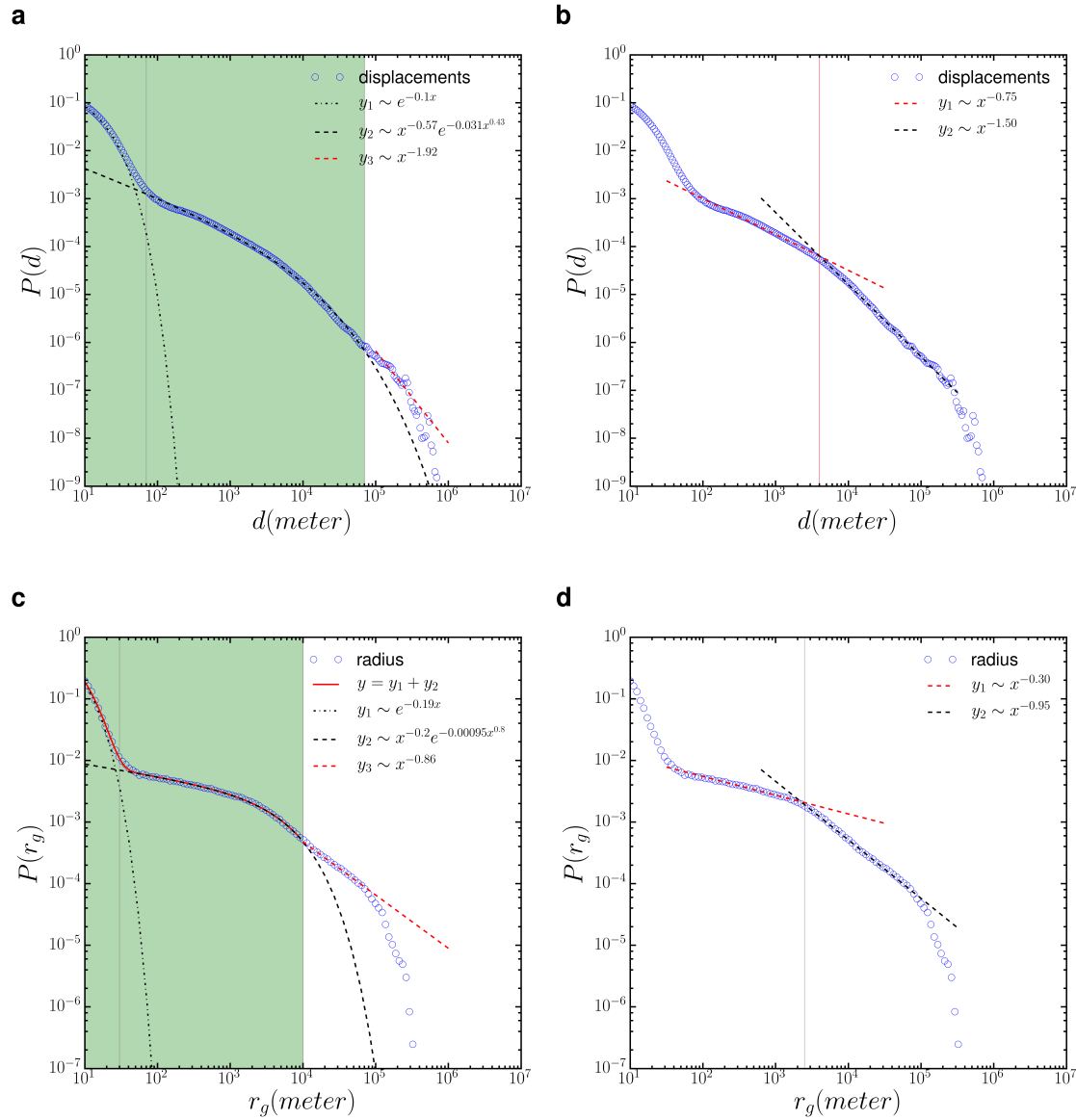10 km by 10 km fishnet cells that provide an adequate resolution for a country wide in- ₂₄₄

Figure 4. **The probability distribution of Twitter user displacements and radius of gyration:** (a) The displacement distribution $P(d)$ is approximated by an exponential, a stretched-exponential and a power-law function (b) the distance between [70m, 70km] is approximated by a double power-law functions (c) The distribution of radius of gyration $P(r_g)$ is approximated by the combination of an exponential, a stretched-exponential and a power-law function (d) the distance between [30m, 100 km] is approximated by a double power-law functions

vestigation Ratti *et al.* (2010). The edges of this network were derived from the number of directed Twitter user displacements between each pair of cells. We used this connectivity network as a proxy to partition the space associated with its nodes. Coherent geographic regions were identified as individual fishnet cells showing more internal user movement compared to user movements across the cell boundaries to neighboring cells. Fig. 5 presents the delineated urban boundaries based on Twitter user displacement distance less than 4 km, greater than 4km, greater than 10 km, and using all available displacements together compared to the administrative boundaries of Great Britain. One clear observation across the coarse and fine delineations is that most of the geographic

divisions are centered around big urban cores with relatively high populations. These ²⁵⁴
results are expected given that most of the tweets originate in urban centers. However, ²⁵⁵
what is remarkable is the performance of this approach in dividing the remaining space ²⁵⁶
between cities. We found that restricting the trip distance results in different delineations ²⁵⁷
of the catchment area around these centers. For example, one could explain these effects ²⁵⁸
as a manifestation of the underlying gravity law Simini *et al.* (2012) and the distance ²⁵⁹
decay effect on attracting movers Gonzlez *et al.* (2008). Remarkably, our approach per- ²⁶⁰
formed well in terms of dividing the entire space with minimum gaps. Empty cells were ²⁶¹
found in regions where no, or few, Twitter users had visited (e.g. forests, agriculture) ²⁶²
especially when restricting the analysis to short distance Twitter user displacements. ²⁶³

Regional boundaries inferred from short distance Twitter user displacements (less than ²⁶⁴
4 km) exhibit very small and fragmented regions, which is probably related to daily com- ²⁶⁵
muting around a user's home location. Redrawing the boundaries based on longer dis- ²⁶⁶
tance displacements produces more cohesive, large regions. For example, by partitioning ²⁶⁷
the space based on displacements greater than 10 km created regions that are comparable ²⁶⁸
to the NUTS (Nomenclature of Territorial Units for Statistics - 1) regions (Fig. 6 - a). ²⁶⁹
However, the power of this novel mapping technique is not to reproduce the partitions ²⁷⁰
already known, rather it is to point out some of the unexpected boundaries. For example ²⁷¹
the boundaries between England and Wales were found to be more diffuse compared ²⁷²
to the abrupt boundary of England and Scotland. Moreover, the city of London has a ²⁷³
wider visitor catchment area that extends beyond the authoritative boundaries of the ²⁷⁴
city. Increasing the displacement distance results in revealing the large region connected ²⁷⁵
to London (Fig. 6). ²⁷⁶

Using Twitter user mobility to delineate non-administrative anthropographic bound- ²⁷⁷
aries enables the researcher to redraw the city at different mobility ranges inferred ob- ²⁷⁸
jectively from the user's collective distribution. In addition, the distance range of the ²⁷⁹
movements is usually explained by local socio-economic factors (e.g., work commuting) ²⁸⁰
that provide for a specific interpretation of the apparent patterns. The patterns obtained ²⁸¹
from Twitter user mobility are comparable to the patterns produced by those of the net- ²⁸²
work of landline phone calls Ratti *et al.* (2010). For example, the region of Wales appears ²⁸³
to consist of three communities as found in the connectivity of both phone calls and long ²⁸⁴
distance movements. However, the regions extracted from the mobility network seems ²⁸⁵
to be more spatially consistent with minimal spatial gaps compared to the partitions ²⁸⁶
extracted from land-line call networks in Great Britain Ratti *et al.* (2010). ²⁸⁷

A more detailed study was conducted over the greater London region revealing the ²⁸⁸
intra-city spatial interaction patterns. The spatial partitions derived from a fine grid of ²⁸⁹
1km using all available Twitter user trips without any restriction on trip distances yields ²⁹⁰
geographic boundaries comparable to some of London's boroughs (Fig. 7). However, some ²⁹¹
areas are shown to be more cohesive and display greater spatial interactions across the ²⁹²
administrative boundaries, for instance, central London. Although, these results suggest ²⁹³
that travelers seem to be localized over certain areas of the city most of the time, some re- ²⁹⁴
gions do exhibit long distance interaction patterns. For example, the separate geographic ²⁹⁵
areas in the south of Hillingdon which includes Heathrow Airport exhibits more con- ²⁹⁶
nectivity to central London than its surrounding areas, which is explained by the usual ²⁹⁷
flight passenger routes. The technique also reveals some of the emerging communities ²⁹⁸
around the borders due to the spatial intermingling of both communities. For example, ²⁹⁹
East Barnet and West Enfield seem to have higher interactions than those resulted from ³⁰⁰
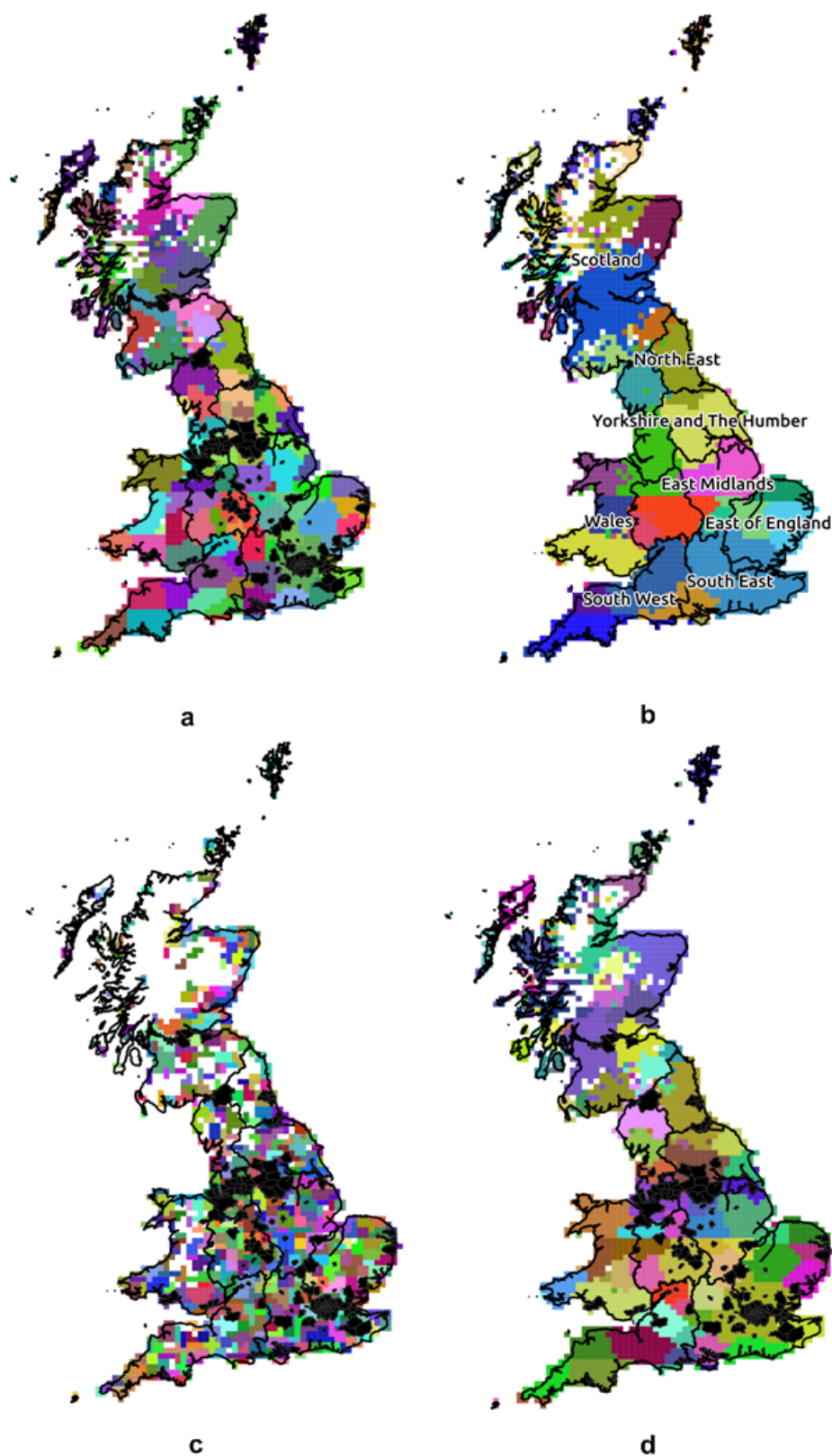in the emerging cohesive zone between the two boroughs. ³⁰¹

Figure 5. **The community structure from collective Twitter user displacements reveals non-administrative anthropographic urban boundaries** (a) displacements longer than 10 km (b) displacements shorter than 4 km (c) and displacements longer than 4 km (d) The partition of space was done using a 10 km fishnet to map the directed displacements from and to each cell. Each color represents a unique community with more Twitter users displacements among the cells compared to others. Major cities (urban audit functional areas) and NUTS are displayed in black.

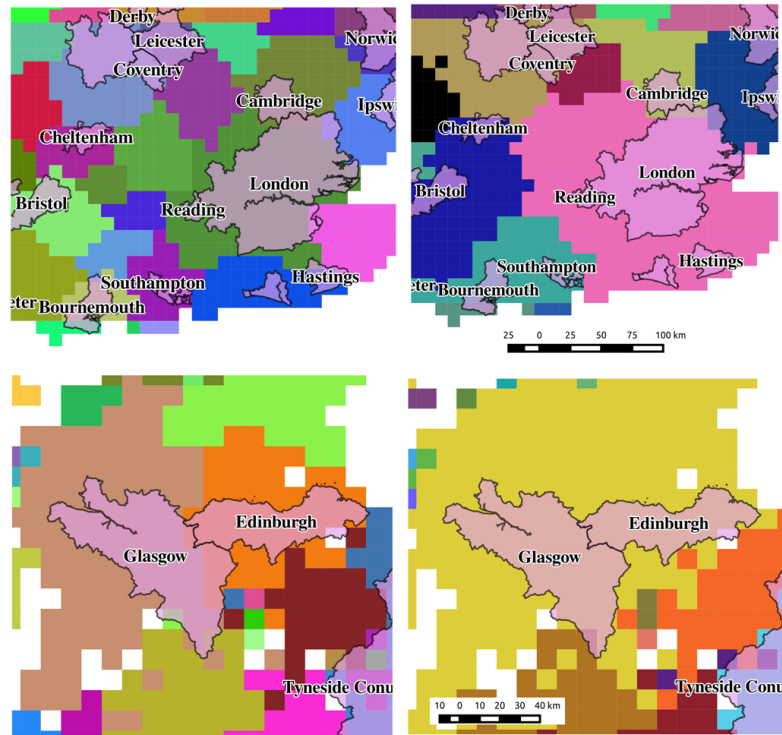12   *International Journal of Geographical Information Science*

Figure 6. **The non-administrative anthropographic regions inferred from Twitter user displacements greater than 4 km (left) and 10 km (right) in comparison with major cities in England (upper figures) and Scotland (bottom figures).** Each color represents a unique community. Including short distance movements has increased the power to differentiate the influence of nearby cities such as Glasgow and Edinburgh (lower left), while restricting the analysis to longer distance movements grouped travelers from the two previous cities into the same community (lower right).
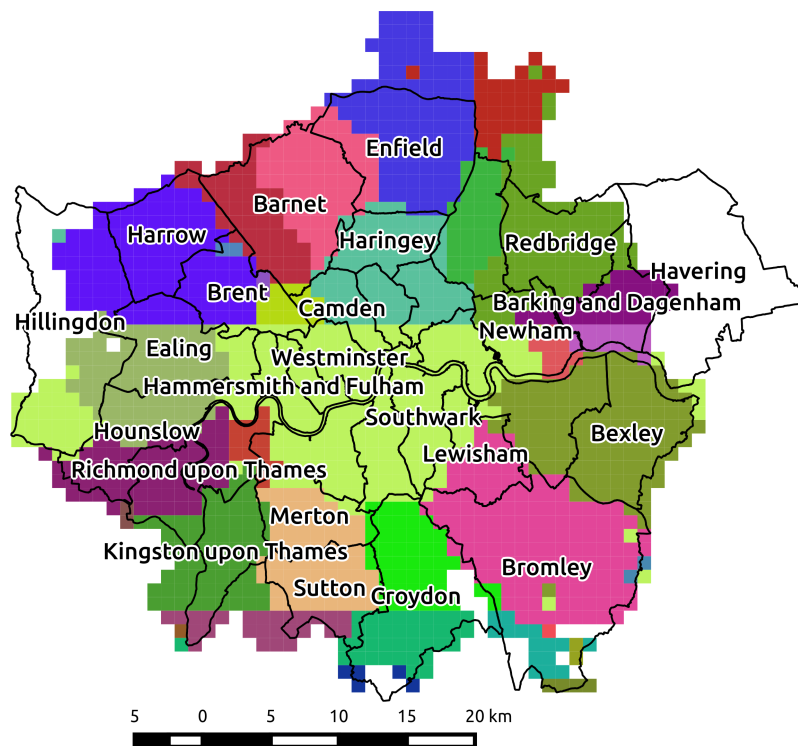
Figure 7. **Non-administrative anthropographic boundaries inferred from collective Twitter user displacements in the city of London compared to the boundaries of London boroughs** A fine fishnet of 1 km cells were used to identify the detailed connectivity patterns based on all the Twitter user displacements in the area. Each unique color represents a different non-administrative anthropographic region. Notice that some remote regions like the airport (light green region in south of Hillingdon) share the same class with downtown because it is well connected despite the geographic separation.

4.   Conclusion

In this study, we delineated the non-administrative anthropographic urban boundaries in Great Britain using a network of Twitter user spatial interactions, which were inferred from the collective Twitter user movement data extracted from more than 69 million Twitter messages. In contrast to administrative urban boundaries, our "bottom-up" approach imposes a virtual fishnet over the islands of Great Britain to partition the space. By studying the probability distributions of the radius of gyrations of individual Twitters users, we selected a cell size of 10 km to quantify the spatial coverage of the majority of Twitter users in Great Britain. Twitter user movements were used to establish a connectivity network of the fishnet cells. Then we applied the map equation algorithm to partition the network and associated geographic regions. The strongly connected communities within the network space yields geographically cohesive, non-overlapping urban areas that provides a clear delineation of the urban boundaries in Great Britain. By performing a statistical analysis of Twitter user mobility patterns in Great Britain, in particular the distribution of collective Twitter user displacements, we found multi-scale and multi-modal urban movements that were divided into several distance ranges starting from short intra-city to inter-city movements with clear destination points. Identifying the connected regions at each of these distance ranges yielded hierarchical boundaries of the urban space in Great Britain.

The power of using Twitter user mobility to delineate non-administrative anthropographic boundaries is the ability to redraw the city at different mobility ranges inferred objectively from the collective mobility patterns. Urban boundaries redrawn based on Twitter user movement represent physical commutes rather than social ties or phone call initiation to reflect the human interaction space. This study provides a first step in connecting human mobility research with defining non-administrative anthropographic boundaries, which could assist in resource allocation, political campaigns and urban planning. However, the geo-located twitter data is not able to generalize to the entire population; therefore, the urban regions that do not have any, or limited, Twitter coverage can be missed during the delineation process. Yet, our approach is still applicable when more detailed mobility data is available.

References

Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. Nature 439, 4625. doi:10.1038/nature04292

Clauset, A., Shalizi, C., Newman, M., 2009. Power-Law Distributions in Empirical Data. Society for Industrial and Applied Mathematics Review. 51, 661703. doi:10.1137/070710111

Coscia, M., Giannotti, F., Pedreschi, D., 2011. A Classification for Community Discovery Methods in Complex Networks. Statistical Analysis and Data Mining, 4, 512546. doi:10.1002/sam.10133

De Domenico, M., Lancichinetti, A., Arenas, A., Rosvall, M., 2015. Identifying Modular Flows on Multilayer Networks Reveals Highly Overlapping Organization in Interconnected Systems. Physical Review X, 5. doi:10.1103/PhysRevX.5.011027

Fortunato, S., Barthlemy, M., 2007. Resolution limit in community detection. Proceedings of the National Academy of Sciences, 104, 3641. doi:10.1073/pnas.0605965104

Gonzlez, M.C., Hidalgo, C.A., Barabsi, A.-L., 2008. Understanding individual human

mobility patterns. Nature, 453, 77982. doi:10.1038/nature06958

Good, B.H., de Montjoye, Y.-A., Clauset, A., 2010. Performance of modularity maximization in practical contexts. Physical Review E, 81, 046106. doi:10.1103/PhysRevE.81.046106

Guimer, R., Sales-Pardo, M., Amaral, L.A.N., 2004. Modularity from fluctuations in random graphs and complex networks. Physical Review E, 70, 025101. doi:10.1103/PhysRevE.70.025101

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geolocated Twitter as proxy for global mobility patterns. Cartography and Geographic Information Science, 41, 260271. doi:10.1080/15230406.2014.890072

Jiang, B., Miao, Y., 2015. The evolution of natural cities from the perspective of location-based social media. The Professional Geographer, 67(2), 295-306.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D., 2015. Understanding Human Mobility from Twitter. PLoS ONE 10, e0131469. doi:10.1371/journal.pone.0131469

Kallus Z, Barankai N, Szle J, Vattay G., 2015. Spatial Fingerprints of Community Structure in Human Interaction Network for an Extensive Set of Large-Scale Regions. PLoS ONE 10(5): e0126713. doi: 10.1371/journal.pone.0126713

Lancichinetti, A., Fortunato, S., 2009. Community detection algorithms: A comparative analysis. Physical Review E, 80, 056117. doi:10.1103/PhysRevE.80.056117

Liu, X., Gong, L., Gong, Y., Liu, Y., 2015. Revealing travel patterns and city structure with taxi trip data. J. Transportation Geography, 43, 7890. doi:10.1016/j.jtrangeo.2015.01.016

Liu, J., Zhao, K., Khan, S., Cameron, M., Jurdak, R., 2014. Multi-scale Population and Mobility Estimation with Geo-tagged Tweets. ArXiv14120327 Phys.

Long, Y., Han, H., Tu, Y., Shu, X., 2015. Evaluating the effectiveness of urban growth boundaries using human mobility and activity records. Cities, 46, 76-84.

Lynch, K., 1960. The image of the city. MIT press.

Newman, M.E.J., 2006. Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103, 857782. doi:10.1073/pnas.0601602103

Openshaw, S., 1984. The modifiable areal unit problem. Geo Abstracts University of East Anglia.

Qian, W., Stanley, K. G., Osgood, N. D., 2013. The impact of spatial resolution and representation on human mobility predictability. In Web and Wireless Geographical Information Systems, pp. 25-40, Springer Berlin Heidelberg.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., Strogatz, S.H., 2010. Redrawing the Map of Great Britain from a Network of Human Interactions. PLoS ONE 5, e14248. doi:10.1371/journal.pone.0014248

Reynolds, A., 2012. Truncated levy walks are expected beyond the scale of data collection when correlated random walks embody observed movement patterns. Journal of The Royal Society Interface, 9(68), 528-534.

Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S. J., and Chong, S., 2011. On the levy-walk nature of human mobility. IEEE/ACM transactions on networking (TON), 19(3), 630-643.

Rinzivillo, S., Mainardi, S., Pezzoni, F., Coscia, M., Pedreschi, D., Giannotti, F., 2012. Discovering the Geographical Borders of Human Mobility. KI - Knstl. Intell. 26, 253260. doi:10.1007/s13218-012-0181-8

Rosvall, M., Axelsson, D., Bergstrom, C.T., 2010. The map equation. The European Physical Journal Special Topics, 178, 1323. doi:10.1140/epjst/e2010-01179-1

Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, 105, 11181123.

Sakaki, T., Okazaki, M., Matsuo, Y., 2010. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors, in: Proceedings of the 19th International Conference on World Wide Web, WWW 10. ACM, New York, NY, USA, pp. 851860. doi:10.1145/1772690.1772777

Schliephake, C., 2014. Urban Ecologies: City Space, Material Agency, and Environmental Politics in Contemporary Culture. Lexington Books.

Simini, F., Gonzlez, M. C., Maritan, A., Barabsi, A. L., 2012. A universal model for mobility and migration patterns. Nature, 484(7392), 96-100.

Sobolevsky S, Szell M, Campari R, Couronn T, Smoreda Z, Ratti, C., 2013. Delineating Geographical Regions with Networks of Human Interactions in an Extensive Set of Countries. PLoS ONE 8(12): e81707. doi: 10.1371/journal.pone.0081707

Song, C., Wang, D., Barabsi, A.-L., 2012. Connections between human dynamics and network science. ArXiv Prepr. ArXiv12091411.

Thiemann, C., Theis, F., Grady, D., Brune, R., Brockmann, D., 2010. The Structure of Borders in a Small World. PLoS ONE 5, e15422. doi:10.1371/journal.pone.0015422

Wong, D., 2009. The modifiable areal unit problem (MAUP). SAGE Publications: London, UK.

Zandbergen, P.A., 2009. Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning. Transactions in GIS, 13, 525. doi:10.1111/j.1467-9671.2009.01152.x

Zhao, K., Musolesi, M., Hui, P., Rao, W., and Tarkoma, S., 2015. Explaining the power-law distribution of human mobility through transportation modality decomposition. Scientific reports, 5.

Zhong, C., Arisona, S.M., Huang, X., Batty, M., Schmitt, G., 2014. Detecting the dynamics of urban structure through spatial network analysis. International Journal of Geographical Information Science, 28, 21782199. doi:10.1080/13658816.2014.914521