Measuring Urban Social Diversity Using Interconnected Geo-Social Networks

Desislava Hristova
Computer Laboratory
University of Cambridge, UK
first.last@cl.cam.ac.uk

Matthew J. Williams
School of Computer Science
University of Birmingham, UK
m.j.williams@cs.bham.ac.uk

Mirco Musolesi
Dept. of Geography
University College London, UK
m.musolesi@ucl.ac.uk

Pietro Panzarasa
School of Business and Management, Queen Mary University of London, UK
p.panzarasa@qmul.ac.uk

Cecilia Mascolo
Computer Laboratory
University of Cambridge, UK
first.last@cl.cam.ac.uk

ABSTRACT

Large metropolitan cities bring together diverse individuals, creating opportunities for cultural and intellectual exchanges, which can ultimately lead to social and economic enrichment. In this work, we present a novel network perspective on the interconnected nature of people and places, allowing us to capture *the social diversity of urban locations through the social network and mobility patterns of their visitors*. We use a dataset of approximately 37K users and 42K venues in London to build a network of Foursquare places and the parallel Twitter social network of visitors through check-ins. We define four metrics of the social diversity of places which relate to their social brokerage role, their entropy, the homogeneity of their visitors and the amount of serendipitous encounters they are able to induce. This allows us to distinguish between places that bring together strangers versus those which tend to bring together friends, as well as places that attract diverse individuals as opposed to those which attract regulars. We correlate these properties with wellbeing indicators for London neighbourhoods and discover signals of gentrification in deprived areas with high entropy and brokerage, where an influx of more affluent and diverse visitors points to an overall improvement of their rank according to the UK Index of Multiple Deprivation for the area over the five-year census period. Our analysis sheds light on the relationship between the prosperity of people and places, distinguishing between different categories and urban geographies of consequence to the development of urban policy and the next generation of socially-aware location-based applications.

1. INTRODUCTION

People and places are interconnected in an organic way [3], and more intensely so in the urban context. With more than half of the world's population living in urban areas,[1] understanding how *human mobility enhances the social diversity of places* is an important question for urban planners and system designers alike. While the fundamental role of urban geography in human interactions, relationships and social capital is relatively well understood [12, 17, 32], the role of people in the success of places has been empirically understudied. Nevertheless, much of the success of cities can be attributed to their multiculturalism and the synergy of diverse attitudes within a relatively small geography that is brought on by its inhabitants and measurable through their human mobility network and social network properties [18].

Urban activist Jane Jacobs wrote "[Cities] differ from towns and suburbs in basic ways, and one of these is that cities are, by definition, full of strangers" [21]. In this work we study places that bring together strangers among other types of urban social diversity and measure the relationship with the prosperity of an area in an interconnected model of people and places. In social networks, the diversity of one's social ties is associated with the amount of social capital they have at their disposal [11, 10]. Social ties can be classified as bridging or bonding where bonding ties are those within homogeneous groups while bridging ties are those which transcend groups and are associated with diversity [34]. *However, it is not yet understood how the diversity of the social network of visitors defines the social role and affluence of a place.*

At present, government census studies are the most widely used measures of urban socioeconomic wellbeing at the neighbourhood level. However, recent works proposing the use of user-generated content and social media for measurement have emerged [26, 44, 35, 42] as part of a new science of cities, based on the availability of these new forms of data [2]. Using interim real-time measures of the urban pulse is appealing from a temporal and cost perspective but can be challenging due to the demographic biases of digital media users [16]. It is precisely these biases, however, that could provide insight into some of the most difficult to quantify and predict processes such as gentrification, which is associated with the displacement of residents of a deprived area by an influx of an (economically and digitally) affluent population. The link between urban social diversity and deprivation as measured by social media has yet not been studied.

With the goal of informing the advancement of location-based services and urban development in terms of deprivation and gentrification indices, in this work we measure the bridging and bonding potential of various types of places and their visitor diversity using an interconnected network model of people and places. Recently, multilayer networks have become increasingly popular in modelling complex interacting systems such as the power grid [8], biological networks [4] and transportation in the urban context [19, 15]. Drawing inspiration from this, we present an *interconnected*

[1] United Nations, 2014 Revision of World Urbanization Prospects. http://esa.un.org

WWW 2016, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4143-1/16/04.
http://dx.doi.org/10.1145/2872427.2883065.

geo-social network, composed of a network of places and people and the links between them. As opposed to a classic network model, this approach allows for social network projections of the spatial network and spatial network projections of the social network, enabling the measurement of the social properties of places and the geographical properties of people based on their place (spatial) network of visits. We use geo-social online data in Europe's largest metropolitan city – London – and introduce an interconnected people-place network paradigm, which aims to more realistically model urban social diversity. Using the Twitter social network of visitors to define the diversity of Foursquare venues, we are able to distinguish between categories, geographies and socioeconomic factors across London's neighbourhoods. We make the following main contributions:

- We introduce a new paradigm for the interpretation of social characteristics of places, which we call the *interconnected geo-social network model*. This model is defined by coupling the social network of individuals and the network of places, by linking people to places if they visited them. We also introduce the concepts of a social neighbourhood of a place as well as the place neighbourhood of a person.
- We define four distinct measures of urban social diversity, which employ the interconnected geo-social network to derive the *brokerage* role of a place, the potential *serendipity* of encounters, the *entropy* of visitors and their *homogeneity*. In other words, these metrics are able to summarise the intercorrelated geographic dynamics of the city in terms of socio-spatial diversity.
- Our analysis reveals the relationships between structural (i.e., based on the presence of different individuals) and characteristic (i.e., based on the actual preferences of the individuals) diversity of the social mix of visitors. The preferences of individuals are investigated by looking at the categories of venues they normally visit. In our analysis we found that *Arts* and *Nightlife* venues tend to exhibit greater social diversity than *Residences* and play different roles in terms of bridging and bonding.
- Our diversity measures are able to describe dynamics that would otherwise be difficult to capture such as the gentrification of London boroughs and their deprivation, allowing for an innovative analysis of the interactions between people and places, by linking the social dynamics of cities with their spatial network of places.

The remainder of this work details these contributions, and summarises related work, concluding with a discussion of the implications, limitations, and applications of the proposed framework.

2. RELATED WORK

Our work draws on previous literature on network diversity in social networks, urban studies on mobility, geo-social dynamics and deprivation. In this section, we will describe the state of the art in these domains.

2.1 Network Diversity

Geography plays an important role in the diversity of social networks where individuals with more geographically and structurally diverse networks are found to have higher social capital and come from more well-off areas in the UK [17]. The competitive advantage of an individual in a social network has been defined as a function of the structural holes that provide the individual with a brokerage position between otherwise disconnected others. [10]. Network brokerage has been studied in a variety of contexts from organisations [11] to online social networks [27], and most recently in geo-social networks [20]. The diversity of locations with respect to visitors has also been explored in the context of location-based social networks [14, 37]. We combine these approaches to consider the social network diversity of places and to the best of our knowledge, this is the first time that network diversity metrics have been defined for interconnected networks [8, 4].

2.2 Urban Mobility

Human mobility in urban environments has received much attention in recent years, enabled in part by the increasing availability of individual-level data via online location-sharing services and mobile phone records. Human movement in urban areas differs from other geographic regions in their strong dependence on the spatial distribution of places within the city [30]. Location-sharing services have also been used to explore urban mobility through its impact on the place network [31]. While such studies focus on place, other work has explored the interplay between space and social ties, which are known to show strong inter-dependence [41]. In the context of location-based social networks, a variety of geo-social phenomena have been analysed, including tie formation [1], co-location patterns [6, 12], homophily [45], and community structure [7]. The relationship between human mobility and the urban social diversity of places, however, is yet to be studied.

2.3 Urban Deprivation

Previous work using curated national statistical data sources has shown that the morphology of urban environments plays a key role in urban deprivation [43, 28], and that socioeconomic prosperity can be linked with other neighbourhood-level features such as human travel patterns [39], access to local facilities [28], and the prevalence of fast-food outlets [33]. Identifying and understanding associations such as these is of great interest to national policymakers, social reformers, and city planners. More recently, there has been interest in using signals from technological systems to predict urban wellbeing and deprivation. Many signals have proved useful in predicting deprivation indices, from passenger transits recorded by automated fare collection systems [38, 24] to crowdsourced data such as OpenStreetMap [44]. In the context of location-based networks, deprivation has been studied in terms of Foursquare's crowdsourced venue database [44, 36], but there has so far been little analysis from a joint geo-social perspective.

3. INTERCONNECTED GEO-SOCIAL NETWORK MODEL

Many real-world systems can be represented through a number of unique yet interconnected networks. One such system results from the interaction between geography and people. Although the properties of social and geographic networks have long been studied independently, such view does not consider the dynamics between the two. Here we present a model of interconnected geo-social networks, where projections of one carry rich information about the other.

The spatial graph $G_L = (V_L, E_L)$ has a set of nodes $l \in V_L$ which are geographical locations and can be described by a set of coordinates, and a set of edges E_L that can be described in terms of the user transitions between them with a weight equal to their number. The neighbourhood of a location l can be denoted as $N_L^h(l)$ and includes all its adjacent associated locations in terms of transitions up to h hops.

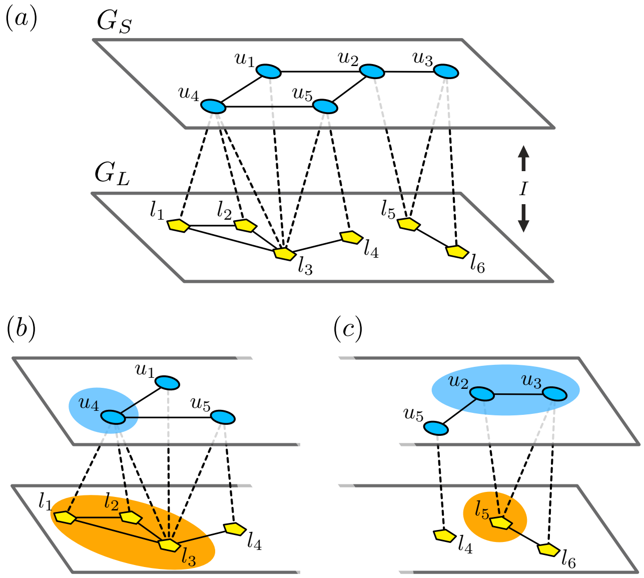


Figure 1: Panel (a): Interconnected network model, where G^L and G^S are composed by different entities coupled through interlayer edges I. Panel (b) shows the place neighbourhood of user u_4, as indicated by the shaded place nodes in the lower layer. Panel (c) illustrates the social neighbourhood of visitors to place l_5, indicated by the shaded user nodes in the upper social layer.

The social graph $G_S = (V_S, E_S)$ includes nodes which represent users (denoted as V_S) and undirected edges which are the friendship relationships between them (denoted as E_S). The neighbourhood of a user u can be denoted as $N_S^h(u)$ and includes its contacts up to h hops in the social network. The interconnected network $G_M = (G_S, G_L, I)$ contains the geographic and social graph layers along with the interlayer edge set I. We associate a weight $w_{u,l}$ to each edge which equals the number of visits (check-ins) that a user u has made to a location l.

The social network of the individuals linked to a location l at distance h can be described as its social neighbourhood, denoted by $N_S^h(l)$. For example, the 1-hop social neighborhood of location l would be composed of the individuals that visit location l, the 2-hop social neighborhood would be composed of the individuals that visit location l and their friends and so on. On the other hand, the place network of an individual u at distance h can be denoted by $N_L^h(u)$. It is a subgraph of the spatial network layer, where each place $l \in N_L^h(u)$ is at distance less or equal than h from user u. In the case of $h = 1$, it is the set of places user u has visited. For $h = 2$ it contains the places visited by u and the places connected to those visited by the user in the place network and so on.

Figure 1 illustrates the interconnected geo-social network model in Panel (a). Locations are connected based on their common visitors, and interlayer edges I represent visits made by users to locations. Panel (b) illustrates the place neighbourhood of a user node u_4 as a projection on to layer G_L, while Panel (c) shows the reverse projection of a location l_5 on to the social network G_S. Both of these projections are used to construct measures of urban social diversity in the following section.

4. SOCIAL DIVERSITY MEASURES

Here we define four measures of the social diversity associated with a place through its social network of visitors. *Brokerage* relates to the potential of a place to bring together strangers as opposed to friends, *serendipity* measures the probability that the set of visitors happened to visit that place, while *entropy* measures the diversity of visits with respect to visitors. We also measure the diversity of visitors themselves by comparing their characteristics in terms of venues.

4.1 Brokerage

The brokerage potential of a person expresses his or her ability to connect otherwise disconnected others. Extensively described by Burt, it measures the extent to which an individual's ego network is non-redudntant, which in turn reflects the individual's potential of brokering between otherwise disconnected contacts [9]. Within the context of geography, a place can possess brokerage potential with respect to the social network of its visitors, if it can bring together otherwise disconnected individuals in physical space. Based on the interpretation by Borgatti [5], the brokerage potential B of node l at distance h can be expressed as:

$$B(l) = |N_S^h(l)| - \frac{\sum_{u,v \in N_S^h(l)} e_{u,v}}{|N_S^h(l)|} \tag{1}$$

where we subtract the redundant portion of a network, equivalent to the connectedness (average number of edges $e_{u,v}$) in the social neighbourhood of l, from its size $|N_S^h(l)|$. We can then normalise this value by $|N_S^h(l)|$, resulting in the fraction of non-redundant contacts of l's social neighbourhood: $B(l)/|N_S^h(l)|$. In this work we use a hop parameter $h = 2$, which enables us to capture second-hand redundancy resulting from connections among friends of visitors. If all visitors to a place are connected, the place has no brokerage power ($B(l) = 0$), whereas if none of the visitors are connected, the place has high brokerage power which results in a brokerage value of 1.

4.2 Serendipity

The serendipity of a place is the extent to which it can induce chance encounters between its visitors. We measure this as the average probability of an edge $w_{u,l}$, given the network of places u has visited prior to venue l. This expresses the idea that all visitors to l have ended up there with a certain probability based on the network of places they have visited in the past. The lower the probability, the higher serendipity a place can provide. More formally, we can define the serendipity D of a place l as:

$$D(l) = 1 - \frac{\sum_{u \in N_S^h(l)} p_l^t(u)}{|N_S^h(l)|} \tag{2}$$

where:

$$p_l^t(u) = \frac{\sum_{v \in N_L^h(u)^{<t}} w_{v,l}}{\sum_{v \in N_L^h(l)^{<t}} w_{v,l}} \tag{3}$$

is the probability of user u checking into place l based on their place neighbourhood $N_L^h(u)^{<t}$, and t represents the first check-in to venue l made by user u. The probability is measured as the sum of weights of the number of venues v with edges to l visited by u prior to time t, over the weighted degree of l in the spatial network. The average probability of a user u visiting location l provides a

measure of what role chance plays in the composition of the social neighbourhood of l. Places with a higher serendipity value are more likely to induce chance encounters since the composition of their visitors is more unexpected.

4.3 Entropy

The entropy of a place describes the extent to which it is diverse with respect to visits. We measure its value as the Shannon entropy of a location:

$$H(l) = -\sum_{u \in N_S^h(l)} p_l(u)\, log p_l(u) \tag{4}$$

where $p_l(u)$ is the probability that a given check-in in place l is made by user u. We use this measure in a similar way to the authors in [14], where it is used to quantify the diversity of visitors to a location. We adapt this to our definition of the the social neighbourhood of a place, where places with highly entropic neighbourhoods are frequented by many diverse visitors and vice versa.

4.4 Homogeneity

Another important measure of the social diversity of a place is the extent to which its visitors are homogeneous in their characteristics. We use the mean cosine similarity between the place preferences of all pairs of visitors to a particular location to measure its overall social homogeneity as:

$$S(l) = \frac{\sum_{u,v \in N_S^h(l)} sim(u,v)}{\lvert N_S^h(l)\rvert(\lvert N_S^h(l)\rvert - 1)} \tag{5}$$

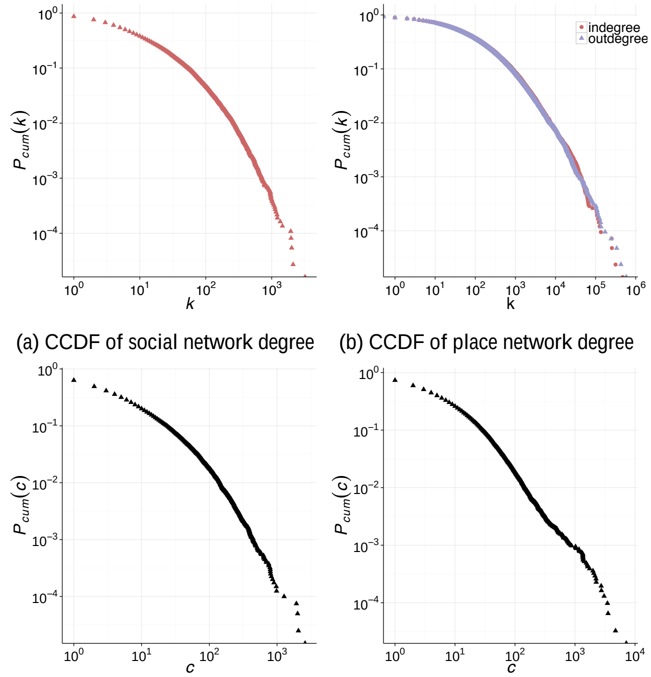
where $sim(u,v)$ is the cosine similarity of the frequency vectors of the visits to locations of a given category of user u and user v respectively. There are nine top categories in Foursquare for which we build a frequency vector for each user and then compare in a pairwise manner for all visitors of a venue. These categories are further described in the Dataset section of the paper. We derive homogeneity between users in a similar way to the authors in [13] in that we consider the cosine similarity of user activity.

This value is between 0 and 1 and indicates the extent to which the mobility patterns of a pair of users in terms of categorical venue visits are the same (1) or completely different (0). By averaging these values across the social neighbourhood of a place, we can derive an estimate of the homogeneity of its visitors in terms of venue preferences. We will next describe the data which we use to explore these measures.

5. DATASET

One of the greatest challenges of performing multilayer network analysis is the lack of suitable data, therefore most work in the field is theoretical with relatively few empirical studies [22]. Fortunately, an increasingly large amount of data is becoming available as online social networks become more integrated and many users cross-post their activity. We were able to collect a dataset of Twitter social network information and Foursquare location information through Twitter where many users link their Foursquare accounts to automatically post updates about their check-ins.

Our dataset consists of the check-ins and links connecting $37,722$ active users of the location-based social network Foursquare and micro-blogging platform Twitter in London, UK. We have downloaded $549,797$ check-ins, each representing a visit made by a user to a certain venue at a given point in time. These check-ins have been made to $42,080$ venues, and have been posted to Twitter by the users in the period between December 2010 and September



(a) CCDF of social network degree (b) CCDF of place network degree

(c) CCDF of check-ins per user (d) CCDF of check-ins per venue

Figure 2: Cumulative degree distributions of the social and place graphs as well as of interlayer edges (check-ins).

2011, with their respective social networks downloaded at the end of that period.

5.1 Online Social Network

We build an undirected *social network* from the directed Twitter network where user u follows v and v follows u back for all Foursquare users who have shared their check-ins on Twitter. This procedure is often regarded as appropriate for representing online friendship in Twitter and as consistent with other undirected online social networks [23] and we therefore refer to two users with a reciprocal edge as *friends* in this work. The undirected social network of London users consists of $432,929$ unweighted reciprocal links between $36,926$ users. In Figure 2a we can observe the cumulative degree distribution of the social network, having a long-tail with a minority of very well connected users and the majority of users with less than 100 friends.

5.2 Place Network

The spatial network of places is constructed by the transition flows of users going between locations. We refer to these spatial locations as *places* and their network as a *place network*. In terms of Foursquare terminology, these places are referred to also as *venues*, which we also use interchangeably. If a user has transitioned between two places in their history of check-ins, we draw an edge between them. The weight of the edge is proportional to the number of transitions made by all users between two places and edges are directed. In total, there are $3,151,741$ directed edges between the $42,080$ venues. The degree distribution is shown in Figure 2b.

Each venue in Foursquare is also characterised by a lower level category such as *coffee shop* and a higher level category such as *Food*. There are nine top-level categories: *Arts & Entertainment, Colleges & Universities, Food, Nightlife Spots, Outdoors & Recreation, Professional & Other Places, Residences, Shops & Services,* *Travel & Transport*, which we refer to for short in this work as *Arts, Study, Food, Nightlife, Outdoors, Professional, Residences, Shops* and *Travel*. In addition, each venue falls within a geographic administrative boundary called a borough. Each borough consists of wards, which are sectioned by population density and the natural landscape of the city. We use categories and geographical boundaries to distinguish between measurement effects in our results and government statistics of deprivation.

5.3 Geo-social Interaction

The interaction between people and places in our dataset is represented by Foursquare check-ins. There are more than half a million check-ins that we have recorded in the area of London over a little less than a year. Figures 2c and 2d plot the distribution of check-ins per user and check-ins per venue respectively. A very small fraction of users have made an exceptional number of check-ins over the time period and similarly most venues have a low number of check-ins with the exception of some highly popular venues. Heathrow Airport is the most popular venue in London with over 10K check-ins in our dataset.

5.4 Index of Multiple Deprivation

To quantify socioeconomic conditions within regions of London we use the Index of Multiple Deprivation (IMD), an official statistical exercise conducted by the UK Department of Communities and Local Government to assess the relative prosperity of neighbourhoods across England. The overall IMD for an area is a composite of seven deprivation indices. In particular, a neighbourhood is assessed according to the following domains: deprivation relating to low income (*Income*); deprivation due to lack of employment among working-age inhabitants (*Employment*); lack of education and skills among young persons and adults (*Education*); impaired quality of life due to ill health and disability (*Health*); risks of crime at a local level (*Crime*); limited provision of local services and lack of access to affordable housing (*Housing*); deprivation relating to the local environment, including quality of housing and air quality (*Living Environment*). The composite IMD and seven domain indices for each neighbourhood are publicly available[2], and provide a rich source of curated socioeconomic indicators across London. The higher the score of an index, the more deprived the neighbourhood. In this paper, we consider the indices released with the two most recent reports (2010 and 2015).

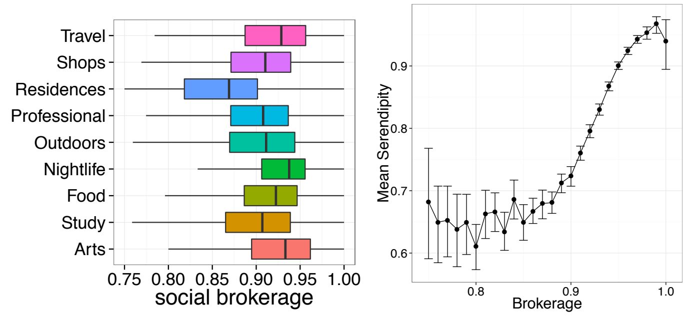
6. URBAN SOCIAL DIVERSITY

The results of our urban social diversity measurements are presented in this section. We first distinguish between the *bridging* (bringing together strangers) and *bonding* (bringing together friends) qualities on a per venue basis and distinguish between categories. We then study the *diversity of visitors* to those venues in terms of their characteristics. Ultimately, we relate these observations to *neighbourhood deprivation* and describe the differences between central and peripheral boroughs of London with regards to social diversity.

6.1 Brokerage Role of Places

One of the fundamental social roles of places is to bring people together. Just like people in social networks, some places can act as *bonding hubs*, bringing together friends to socialise and interact with each other, while others are more likely to gather strangers and therefore act as *bridging hubs*, bringing together otherwise disconnected individuals. We measure the bridging or bonding role of a Foursquare venue as its brokerage $B(l)$ using Equation 1.

[2]English Indices of Deprivation, 2015.
https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015



(a) Brokerage of urban places by category (b) Brokerage vs mean serendipity

Figure 3: Brokerage per category and brokerage vs. serendipity. Outliers below brokerage values of 0.75 (3% of values) were excluded for readability.

The role of a place to either bring together friends or strangers can be dependent on its type. Figure 3a shows the distribution of brokerage values across categories. In the box-and-whisker plot, the distribution of brokerage is split into quartiles. Each box represents the mid-quartile range with the black line in the middle being the median of the distribution. In a megacity such as London, it is unsurprising that most locations are frequented by many diverse individuals who do not know each other. There are, however, notable variations between some of the categories. *Residences* tend to be bonding hubs with 50% of values below 0.87, followed by *Study, Professional, Shops* and *Outdoors* categories where people are more likely to be with friends. Places with relatively high brokerage are in the *Arts, Nightlife* and *Travel* categories where most places in these categories have a bridging role in bringing together strangers.

While the structure of the social network of visitors to a place can determine its brokerage role, serendipity further explains how probable its composition of strangers or friends is and to what extent it can foster encounters, which may lead to new social interactions rather than pre-determined ones. It measures the average probability that a person visited the location given their prior history of locations (Equation 2). Figure 3b plots brokerage against mean serendipity. While serendipity varies for low values of brokerage, the relationship between the two is positively strong for higher values of brokerage. This suggests that bonding hubs which are more socially cohesive may have a lower ability to induce chance encounters, while high bridging places will have high serendipity.

We take a closer look at the sub-categories of places and their brokerage and serendipity roles in Table 1 where we list the top bridging (highest value) and bonding (lowest value) types of places and their serendipity values. Firstly, within the *Arts* category there is a clear distinction between the types of places that bridge which seem to be associated with public spaces, while bonding places tend to be predominantly sports/team oriented. Similarly in the *Study* category it is interesting to observe that academic buildings tend to be bridging, while specific departments and classrooms are places that friends tend to have visited together. Within the *Food* category, interestingly fast foods appear as having a greater bonding role than international cuisines such as Australian and German. In the case of *Nightlife*, however, more generic sub-categories such as *Bar* or *Pub* have greater bridging roles than more specific nightlife venues

Category	*Top Bridging*	$<B>$	$<D>$	*Top Bonding*	$<B>$	$<D>$
	Aquarium	0.98	0.98	Basketball	0.85	0.72
	Art Museum	0.95	0.84	Billiards	0.88	0.85
Arts	Opera House	0.96	0.97	Football	0.87	0.52
	Cricket	0.94	0.75	Track	0.87	0.74
	Theatre	0.94	0.87	Water Park	0.9	0.78
	Auditorium	0.92	0.9	Classroom	0.86	0.83
	University	0.91	0.82	Communications	0.89	0.93
Study	Lab	0.91	0.88	Engineering	0.85	0.56
	Rec Center	0.88	0.84	Math	0.69	0.45
	Bookstore	0.9	0.79	Medical School	0.84	0.7
	South American	0.92	0.78	Eastern European	0.88	0.75
	Scandinavian	0.94	0.83	Wings	0.8	0.59
Food	German	0.95	0.91	Indian	0.88	0.71
	Dumplings	0.93	0.88	Friend Chicken	0.87	0.62
	Australian	0.95	0.72	Felafel	0.89	0.76
	Lounge	0.93	0.81	Hookah Bar	0.88	0.83
	Gay Bar	0.92	0.86	Strip Club	0.89	0.77
Nightlife	Pub	0.93	0.85	Hotel Bar	0.89	0.7
	Cocktail	0.92	0.84	Dive Bar	0.87	0.75
	Bar	0.92	0.83	Whiskey Bar	0.88	0.83
	Bridge	0.89	0.87	Athletics & Sports	0.85	0.6
	Neighbourhood	0.9	0.83	Baseball Field	0.75	0.72
Outdoors	River	0.9	0.76	Campground	0.85	0.85
	Park	0.9	0.79	Vineyard	0.69	0.63
	Cemetery	0.9	0.8	Soccer Field	0.82	0.77
	Hospital	0.91	0.87	Emergency Room	0.81	0.82
	Landmark	0.91	0.81	Synagogue	0.83	0.79
Professional	Courthouse	0.9	0.8	Mosque	0.87	0.63
	Convention Centre	0.91	0.77	Elementary School	0.68	0.88
	Animal Shelter	0.93	0.93	Doctor's Office	0.84	0.65
	Residence	0.84	0.46	Housing Development	0.83	0.52
Residences	Apartment Building	0.86	0.75	Home	0.82	0.69
	Photography Lab	0.96	0.9	Yoga Studio	0.88	0.79
	Antiques	0.92	0.82	Laundry	0.72	0.83
Shops	Mall	0.93	0.9	Video Store	0.72	0.71
	Gift Shop	0.93	0.74	Gaming Cafe	0.86	0.51
	Travel Agency	0.95	0.3	Tanning Salon	0.84	0.6
	Motel	0.91	0.81	Resort	0.88	0.69
	Pier	0.94	0.84	B&B	0.87	0.61
Travel	Subway	0.95	0.93	Taxi	0.82	0.41
	Light Rail	0.93	0.88	Plane	0.86	0.63
	Platform	0.94	0.91	Bus	0.86	0.72

Table 1: Top bridging and bonding subcategories by category where $<B>$ is the average brokerage value for the subcategory, while $<D>$ is its serendipity value.

such as *Hookah Bar* or *Strip Club*, which tend to bring together friends.

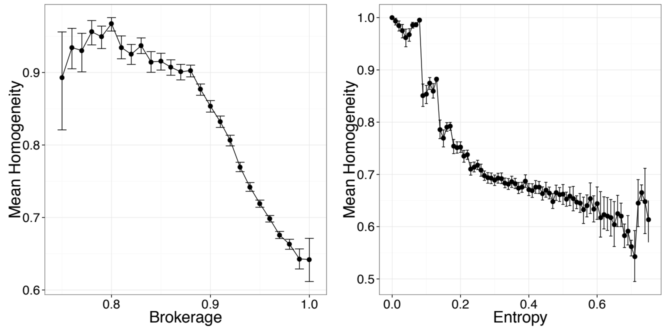
Similar to the *Arts* category, the *Outdoors* category seems to be split between sports-related activities and public spaces. In contrast to *Arts*, however, here sports are being played rather than watched and brokerage values are generally lower possibly because the team activity is more bonding than the viewing activity. In the *Professional* category we observe some public service places such as *Courthouse* and *Hospital* to be bridging hubs, where mostly strangers visit with relatively high serendipity. On the other hand, *Doctor's Office* and *Elementary Schools* are venues where more friends check-in, similar to religious venues such as *Mosques* and *Synagogues*. Because visitors to such venues are likely to be locals, it is understandable that they may know each other and be regulars at these venues.

The *Residences* category has only four sub-categories, which all have relatively low brokerage values, but those which play more of a bonding role are *Home* and *Housing Development*, while *Residence* and *Housing Development* seem to have higher bridging roles. It is interesting to also consider the serendipity values for these sub-categories where *Residence* and *Housing Development* are more predictable because visitors have arrived there with higher probability than in the other two sub-categories. *Shops* have a more unexpected mixture of top bridging places except for the *Mall* sub-category where it is likely that many strangers cross paths. Bonding shops are mainly hobby-related and further analysis of this category can reveal potentially intriguing business insight. Finally, the *Travel* category has a definitive split between transportation (bridging), where people tend to commute and travel, a bonding role where people tend to journey with friends. There is, however, the notable exception of *Motels*, which have a bridging role and *Buses*, which play a more bonding role.

The brokerage role of places and their serendipity are related but differ across categories. In this section, we have taken a close look at these differences and will compare them to the diversity of visitors and their characteristics in the following section.

6.2 Visitor Diversity

While bringing together strangers can ultimately lead to new social exchanges, these might not be truly diverse if the population of visitors is homogeneous. We measure the social composition of a place's visitors with respect to their venue preferences derived



(a) Brokerage vs mean homogeneity (b) Entropy vs mean homogeneity

Figure 4: Visitor diversity with respect to brokerage and entropy.

from mobility patterns through check-ins. For every user who has checked into a venue, we construct a vector of the frequency of their visits to the nine top-level venue categories. By averaging the cosine similarity of vectors between all pairs of visitors to a place, we obtain the *homogeneity* of a place as per Equation 5. In mobility studies entropy has been used to observe the geographical diversity of contacts that people have and the probability of co-location at diverse venues [17, 37]. However, *a location with high entropy is not necessarily one with visitors with diverse characteristics.*

In Figure 4 we compare the mean homogeneity per venue to entropy and brokerage. In both graphs, we can see that homogeneity decreases as brokerage and entropy increase. This relationship implies that the more diverse the visitors to a place are in terms of their composition and social network connectivity, the more characteristically diverse they are too as measured by homogeneity. Figure 5 further shows the distribution of homogeneity over different values of entropy across categories. The strongest relationships are present in the *Food* and *Nightlife* categories where the highest frequency of values are those with low entropy and high similarity, gradually becoming more spread out as entropy grows. Most other categories exhibit similar patterns, with the notable exception of the *Residences* category for which entropy and brokerage are in general low, yet the trend of decreasing homogeneity with entropy is still present. Overall, *venues, which exhibit high diversity in terms of entropy and brokerage also have a less homogeneous composition of visitors*.

The diversity measures that we introduce take into account different projections of subgraphs in the interconnected network such as the place network of users in the serendipity measure and the social network of a venue's visitors in the other measures. Some measures like entropy and serendipity are based on probability while brokerage is purely structural. These differences could be of interest to the designers and developers of mobile systems in improving recommendations for location-based services. In the following subsection, we focus further on the urban development applications of our research in terms of identifying deprivation and neighbourhoods undergoing gentrification.

6.3 Social Diversity & Deprivation

In social network analysis it has been suggested that individuals who act as brokers often have higher social capital at their disposal [9]. In terms of human geography, it is known that those who communicate with geographically diverse others tend to come from less deprived areas [17]. However, *it is not yet understood whether within the urban context, places which act as bridging hubs are in fact within more well-off areas.* In this section, we observe the

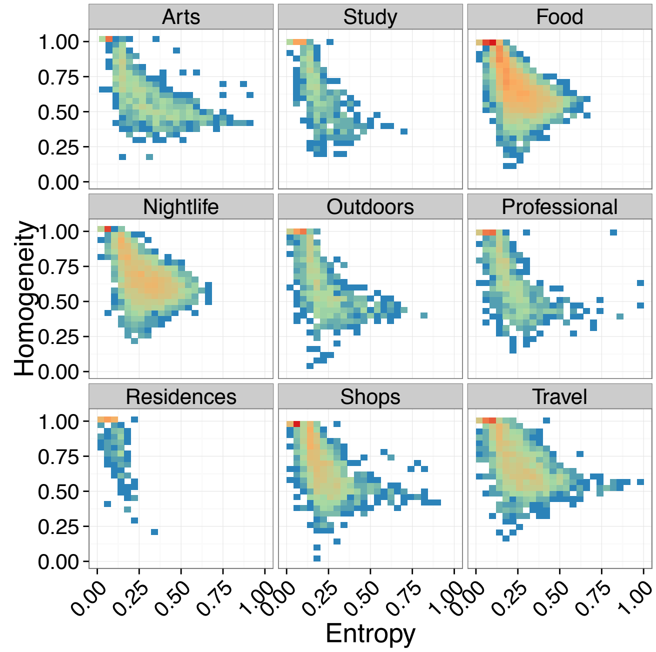
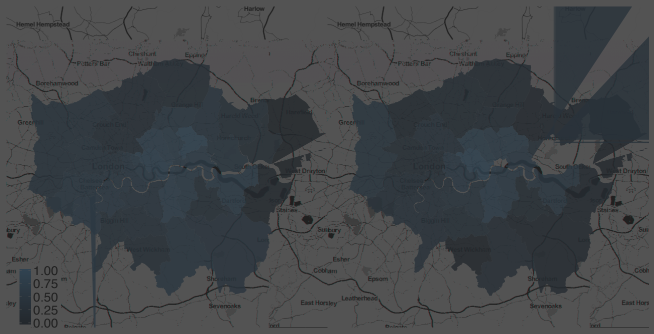


Figure 5: Entropy vs homogeneity distribution per category. Colour gradient reflects the frequency of observation with red being high and blue being low.

geographical distribution of the four diversity measures across the 32 London boroughs. To address the above question, we correlate these measures on a per-borough basis with the eight indicators of socioeconomic wellbeing included in the IMD.

Figure 6 shows the mean values of diversity measures aggregated per area. What becomes immediately apparent is that there is a clear distinction to be made between inner and outer boroughs in terms of diversity. Figures 6a to 6c show notably higher diversity within central boroughs and lower diversity in the periphery. This suggests that *the social diversity of places is highly dependent on geographic factors*. Venues within the central areas of London have higher entropy (Figure 6b) and bring together more strangers (Figure 6a) who are less homogeneous (Figure 6c). Geography, however, is not a big factor for serendipity (Figure 6d) and high serendipity can be high or low in both inner and outer boroughs. Because our measures are normalised on a per user basis this does not reflect the general popularity of an area on Foursquare.

We now study the ranked correlation between the four diversity measures and deprivation. While, with regards to social networks, studies have found a positive relationship between diversity and prosperity, in our analysis of the social diversity properties of places, however, *we find that there is a positive relationship between deprivation and diversity in London*. In Figure 7, we note that brokerage has the strongest relationship with deprivation indicators, especially the Living Environment Deprivation score ($\rho = 0.71$). This sub-domain of the IMD is made up of the following indicators: housing in poor condition, housing without central heating, number of road traffic accidents involving injuring pedestrians or cyclists and low air quality. This sub-domain is especially high for central boroughs where there is more pollution, traffic and strains on housing and health services as indicated also by the relationship with Housing and Health sub-domains ($\rho = 0.44$ and $\rho = 0.4$ respectively). As an example, Westminster, one of



(a) Brokerage

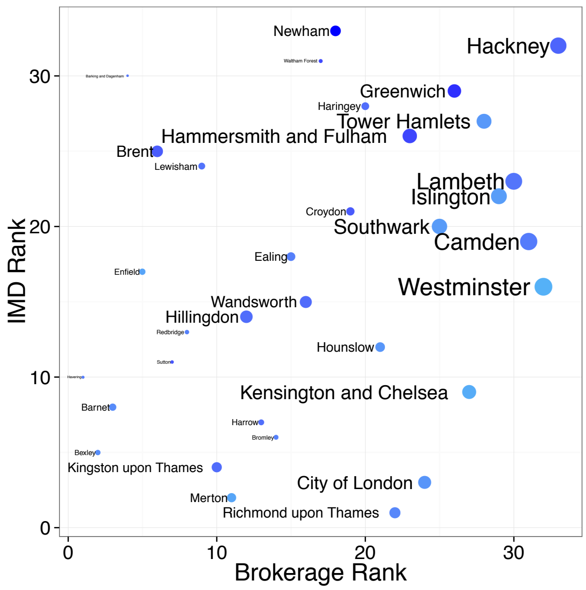


Figure 8: Scatterplot of IMD vs Brokerage rank. Size of each node is proportional to its entropy value and colour represents the change in IMD between 2010 and 2015 from low change (light blue) to high change (dark blue).

popularity or rating of a place are easily accessible through location-based services, the social role a place plays within the urban context is normally only well known by locals through experience. Whether a place is touristy or quiet, artsy or mainstream can be integrated into mobile system design for empowering newcomers or visitors to feel like locals. The role of serendipity and brokerage as described in our work can be particularly impactful when applied to location-based dating applications, where recommendations of places with new and diverse people play a fundamental role. Situation and mood dependent queries such as 'I want to have *drinks alone* at a place where it is *socially acceptable* and I can *meet new people*' will become increasingly popular given the social challenges cities present, in particular for the multitude of newcomers. Despite the challenges of integrating data from different sources and running such metrics in real-time, we believe that the conceptual framework and urban social diversity metrics presented in this work can greatly benefit local businesses as well as innovative location-based discovery applications.

Another important implication of our work is its ability to underpin novel analysis of urban dynamics. The distribution of deprivation across neighbourhoods in relation to diversity is a topic, which affects local governments and policymakers. In particular, our finding that more socially cohesive and homogeneous communities tend to be either very wealthy or very poor but neighbourhoods with both high entropy and deprivation are the ones which are currently undergoing processes of gentrification is in agreement with previous literature on homophily where tightly knit communities are more resistant to change and resources remain within the community [29]. This suggests that affluent communities remain affluent and poor communities remain poor through isolation. On the other hand, areas where there is high diversity and deprivation, communities are undergoing change and what can be described as a gentrification process. This is confirmed by the sharp improvement of their IMD scores between 2010 and 2015. Although further investigation is needed into confounding factors and generalisability to other areas of the UK and the world, the inherent biases of social media demographics [16] work to the advantage of identifying a sudden rise in the affluence of visitors to a particularly deprived neighbourhood. Diversity metrics applied to social media data as in the present analysis can act as good predictors of gentrification when measured through indices of deprivation. Furthermore, we were able to distinguish between different categories and their diversity, pointing to concrete considerations for the measurement and enhancement of social capital across London boroughs. Building applications that not only serve users and businesses but are also conscious of their impact on urban life in the longer run can become detrimental to urban development.

We believe that this paper presents a novel perspective on networks, geo-social data and the challenges of the urban metropolis. We hope that our model of interconnected geo-social networks can aid other interdisciplinary studies of the social roles of places and help the advancement of location-based applications as well as inform urban planners who are at the frontline of unprecedented urbanisation.

Acknowledgements. This work was supported by the Project LASAGNE, Contract No. 318132 (STREP), funded by the European Commission and EPSRC through Grant GALE (EP/K019392). The authors are grateful to Anastasios Noulas for his insightful comments on the manuscript.

Data Statement. Data used in this research was obtained through the openly available Twitter API: `https://dev.twitter.com/overview/api`. IMD data is available here: `https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015`. The authors may share data used in this project for academic research purposes upon request.

8. REFERENCES

[1] Lars Backstrom, Eric Sun, and Cameron Marlow. Find Me if You Can: Improving Geographical Prediction with Social and Spatial Proximity. In *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*, pages 61–70, 2010.

[2] Michael Batty. *The New Science of Cities*. MIT Press, 2013.

[3] Michael Batty and Yichun Xie. From Cells to Cities. *Environment and Planning B*, 21:s31–s31, 1994.

[4] Stefano Boccaletti, G Bianconi, R Criado, Charo I Del Genio, J Gómez-Gardeñes, M Romance, I Sendina-Nadal, Z Wang, and M Zanin. The Structure and Dynamics of Multilayer Networks. *Physics Reports*, 544(1):1–122, 2014.

[5] Stephen P. Borgatti. Structural Holes: Unpacking Burt's Redundancy Measures. *Connections*, 20:35–38, 1997.

[6] Chloë Brown, Neal Lathia, Cecilia Mascolo, Anastasios Noulas, and Vincent Blondel. Group Colocation Behavior in Technological Social Networks. *PLoS ONE*, 9(8), August 2014.

[7] Chloë Brown, Vincenzo Nicosia, Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. Social and Place-focused Communities in Location-based Online Social Networks. *The European Physical Journal B*, 86(6):1–10, June 2013.

[8] Sergey V Buldyrev, Roni Parshani, Gerald Paul, H Eugene Stanley, and Shlomo Havlin. Catastrophic Cascade of Failures in Interdependent Networks. *Nature*, 464(7291):1025–1028, 2010.

[9] Ronald S Burt. *Structural Holes*. Harvard University Press, 1995.

[10] Ronald S Burt. The Network Structure of Social Capital. *Research in Organizational Behavior*, 22:345–423, 2000.

[11] Ronald S Burt. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2):349–399, 2004.

[12] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and Mobility: User Movement in Location-based Social Networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, pages 1082–1090. ACM, 2011.

[13] David Crandall, Dan Cosley, Daniel Huttenlocher, Jon Kleinberg, and Siddharth Suri. Feedback Effects between Similarity and Social Influence in Online Communities. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, pages 160–168. ACM, 2008.

[14] Justin Cranshaw, Eran Toch, Jason Hong, Aniket Kittur, and Norman Sadeh. Bridging the Gap between Physical Location and Online Social Networks. In *Proceedings of the 12th ACM International Conference on Ubiquitous computing (UbiComp'10)*, pages 119–128. ACM, 2010.

[15] Manlio De Domenico, Albert Solé-Ribalta, Sergio Gómez, and Alex Arenas. Navigability of Interconnected Networks Under Random Failures. *Proceedings of the National Academy of Sciences*, 111(23):8351–8356, 2014.

[16] Maeve Duggan, Nicole B. Ellison, Cliff Lampe, Amanda Lenhart, and Mary Madden. Social Media Update 2014. *Pew Research Center*, 2015.

[17] Nathan Eagle, Michael Macy, and Rob Claxton. Network Diversity and Economic Development. *Science*, 328(5981):1029–1031, 2010.

[18] Richard L. Florida. *The Rise of the Creative Class: Revisited*. Basic Books, 2012.

[19] Riccardo Gallotti and Marc Barthelemy. The Multilayer Temporal Network of Public Transport in Great Britain. *Scientific data*, 2, 2015.

[20] Desislava Hristova, Pietro Panzarasa, and Cecilia Mascolo. Multilayer Brokerage in Geo-Social Networks. In *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM'15)*, 2015.

[21] Jane Jacobs. *The Death and Life of Great American Cities*. Vintage International. Random House, 1961.

[22] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. Porter. Multilayer Networks. *Journal of Complex Networks*, 2(3):203–271, 2014.

[23] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is Twitter, a Social Network or a News Media? In *Proceedings of the 19th International Conference on World Wide Web (WWW'10)*, pages 591–600, 2010.

[24] Neal Lathia, Daniele Quercia, and Jon Crowcroft. The Hidden Image of the City: Sensing Community Well-Being from Urban Mobility. In *Pervasive Computing*, number 7319 in Lecture Notes in Computer Science, pages 91–98. Springer Berlin Heidelberg, June 2012.

[25] Loretta Lees. Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance? *Urban Studies*, 45(12):2449–2470, 2008.

[26] Alejandro Llorente, Manuel Garcia-Herranz, Manuel Cebrian, and Esteban Moro. Social Media Fingerprints of Unemployment. *PLoS ONE*, 10, 05 2015.

[27] Tiancheng Lou and Jie Tang. Mining Structural Hole Spanners through Information Diffusion in Social Networks. In *Proceedings of the 22nd International Conference on World Wide Web (WWW'13)*, pages 825–836, 2013.

[28] Sally Macintyre, Laura Macdonald, and Anne Ellaway. Do Poorer People have Poorer Access to Local Resources and Facilities? The distribution of local resources by area deprivation in Glasgow, Scotland. *Social Science & Medicine*, 67(6):900–914, September 2008.

[29] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:215–444, 2001.

[30] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PLoS ONE*, 7(5):e37027, May 2012.

[31] Anastasios Noulas, Blake Shaw, Renaud Lambiotte, and Cecilia Mascolo. Topological Properties and Temporal Dynamics of Place Networks in Urban Environments. In *Proceedings of the 24th International Conference on World Wide Web (WWW'15)*, pages 431–441, 2015.

[32] Jukka-Pekka Onnela, Samuel Arbesman, Marta C González, Albert-László Barabási, and Nicholas A Christakis. Geographic Constraints on Social Network Groups. *PLoS ONE*, 6(4):e16939, 2011.

[33] Jamie Pearce, Tony Blakely, Karen Witten, and Phil Bartie. Neighborhood Deprivation and Access to Fast-Food Retailing: A National Study. *American Journal of Preventive Medicine*, 32(5):375–382, May 2007.

[34] Robert D. Putnam. *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York, 2000.

[35] Daniele Quercia, Jonathan Ellis, Licia Capra, and Jon Crowcroft. Tracking Gross Community Happiness from Tweets. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW'12)*, pages 965–968. ACM, 2012.

[36] Daniele Quercia and Diego Saez. Mining Urban Deprivation from Foursquare: Implicit Crowdsourcing of City Land Use. *IEEE Pervasive Computing*, 13(2):30–36, April 2014.

[37] Salvatore Scellato, Anastasios Noulas, and Cecilia Mascolo. Exploiting Place Features in Link Prediction on Location-based Social Networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, pages 1046–1054, 2011.

[38] Chris Smith, Daniele Quercia, and Licia Capra. Finger on the Pulse: Identifying Deprivation Using Transit Flow Analysis. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW'13)*, CSCW '13, pages 683–692, 2013.

[39] Dominic Stead. Relationships between Land Use, Socioeconomic Factors, and Travel Patterns in Britain. *Environment and Planning B: Planning and Design*, 28(4):499–528, August 2001.

[40] LB Hackney Policy Team. A Profile of Hackney, its People and Place. *Hackney Council*, 2014.

[41] Jameson L. Toole, Carlos Herrera-Yaqüe, Christian M. Schneider, and Marta C. González. Coupling Human Mobility and Social Ties. *Journal of The Royal Society Interface*, 12(105):20141128, April 2015.

[42] Carmen Vaca Ruiz, Daniele Quercia, Luca Maria Aiello, and Piero Fraternali. Taking Brazil's Pulse: Tracking Growing Urban Economies from Online Attention. In *Proceedings of the 23rd International Conference on World Wide Web (WWW'14)*, pages 451–456, 2014.

[43] Laura Vaughan, David L Chatford Clark, Ozlem Sahbaz, and Mordechai Haklay. Space and Exclusion: Does Urban Morphology Play a Part in Social Deprivation? *Area*, 37(4):402–412, 2005.

[44] Alessandro Venerandi, Giovanni Quattrone, Licia Capra, Daniele Quercia, and Diego Saez-Trumper. Measuring Urban Deprivation from User Generated Content. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW'15)*, pages 254–264, 2015.

[45] Ke Zhang and Konstantinos Pelechrinis. Understanding Spatial Homophily: The Case of Peer Influence and Social Selection. In *Proceedings of the 23rd International Conference on World Wide Web (WWW'14)*, pages 271–282, 2014.