Sensing the Urban

Using location-based social network data in urban analysis

Working Paper

Anil Bawa-Cavia (a.bawa-cavia@cs.ucl.ac.uk)

20/09/10

Abstract

Location-based services (LBS) are generating vast bodies of data relating to the whereabouts of their users. This is due to the ease with which modern mobile phones can communicate their precise location via the global positioning system (GPS). Online social networks have begun using LBS to aid social encounter and place discovery in cities. A spatial analysis of the aggregate activity generated by such networks can show us how social activity in a city is distributed, revealing fine-grained spatial patterns evident in the social life of cities. Large-scale data from one such network is analysed across three cities in order to produce an inter-urban analysis. Hubs are identified from activity distributions, and measures of polycentricity, fragmentation and centralisation are examined with respect to levels of social interaction. Spatial clustering tendencies are analysed to determine the characteristic logics of agglomeration in urban social activity. These comparative measures are used to discuss the spatial structure of the three cities in question. Finally, the impact of LBS technologies are discussed in the context of urban analysis.

Introduction

Large-scale datasets relating the activities of individuals to urban space are becoming increasingly available. These data trails reveal a multiplicity of *invisible cities* (Batty, 1990) which are beginning to emerge as objects of research. The instances of such urban analysis which exist (Pulselli et al., 2006; Ratti et al., 2007; Reades et al., 2009) have focused on the geo-location of mobile telecommunications traffic. Other attempts at "sensing human society" (Shoval, 2007) include the detection of bluetooth signals in urban space (O'Neill et al., 2006). By contrast, this paper presents an analysis of a more structured geo-social dataset from a location-based social network, which relates social network *interactions* to specific *venues* in a city.

'Networked urbanism' (Graham and Marvin, 2001a) has contained the promise that "the city itself is turning into a constellation of computers" (Batty, 1997) for over a decade now. During this transition towards ubiquitous computing, 3G mobile phones have begun interacting with the built landscape in ever more sophisticated ways. The aggregate data produced by location-based services (LBS) can be a powerful tool in analysing this relationship. This paper shows how the informational *space of flows* (Castells, 1989) engages the physical landscape of the city, through interactions which may themselves change the mobility patterns of groups in urban centers. If we are to move beyond the observation that cities are "parallel constructions linking both urban places and electronic spaces in complex ways" (Graham, 1997), this unfolding relationship between online social networks and the urban landscape needs to be examined using empirical analyses of large-scale datasets such as those presented here.

Location-Based Social Networking

Location-based social networks allow friends to share locative information via GPS-enabled mobile phones. This paper analyses data from a location-based social network called Foursquare (4sq) (Foursquare, 2010). Users interact with 4sq by *checking-in* at a *venue* using their 3G mobile phone. This communicates their whereabouts to their friend group on the network. Venues can be of any kind, usually comprising a site of social encounter either located in a building (such as a bar, restaurant or cafe) or a public space (such as a park or public square). The venue information is supplied by the 4sq community, such that the network's data also constitutes an instance of *crowd-sourced* mapping (Haklay et al., 2008).

Since 4sq is dedicated to transmitting presence information directly related to sites of encounter, this *geo-social* dataset distinguishes itself from that of generic social networks with location tracking (Twitter, 2010). Whilst activity on such networks can be considered "simply as the noise of human activity" (Batty, 2010) located in space, 4sq provides us with information about the popularity of *social venues* and thus a measure of social interaction density which is strongly coupled to the physical landscape.

Understanding some of the possible motivations for *checking-in* on the 4sq network is vital for interpreting any analysis of the aggregate data. In this we expand on (Ratti et al., 2006) in identifying potential benefits of LBS in a social networking context. Possible motivations for checking-in on 4sq include (a) Friend tracking - Tracking the location of friends in real-time can have social benefits by encouraging social encounter, (b) Urban Discovery

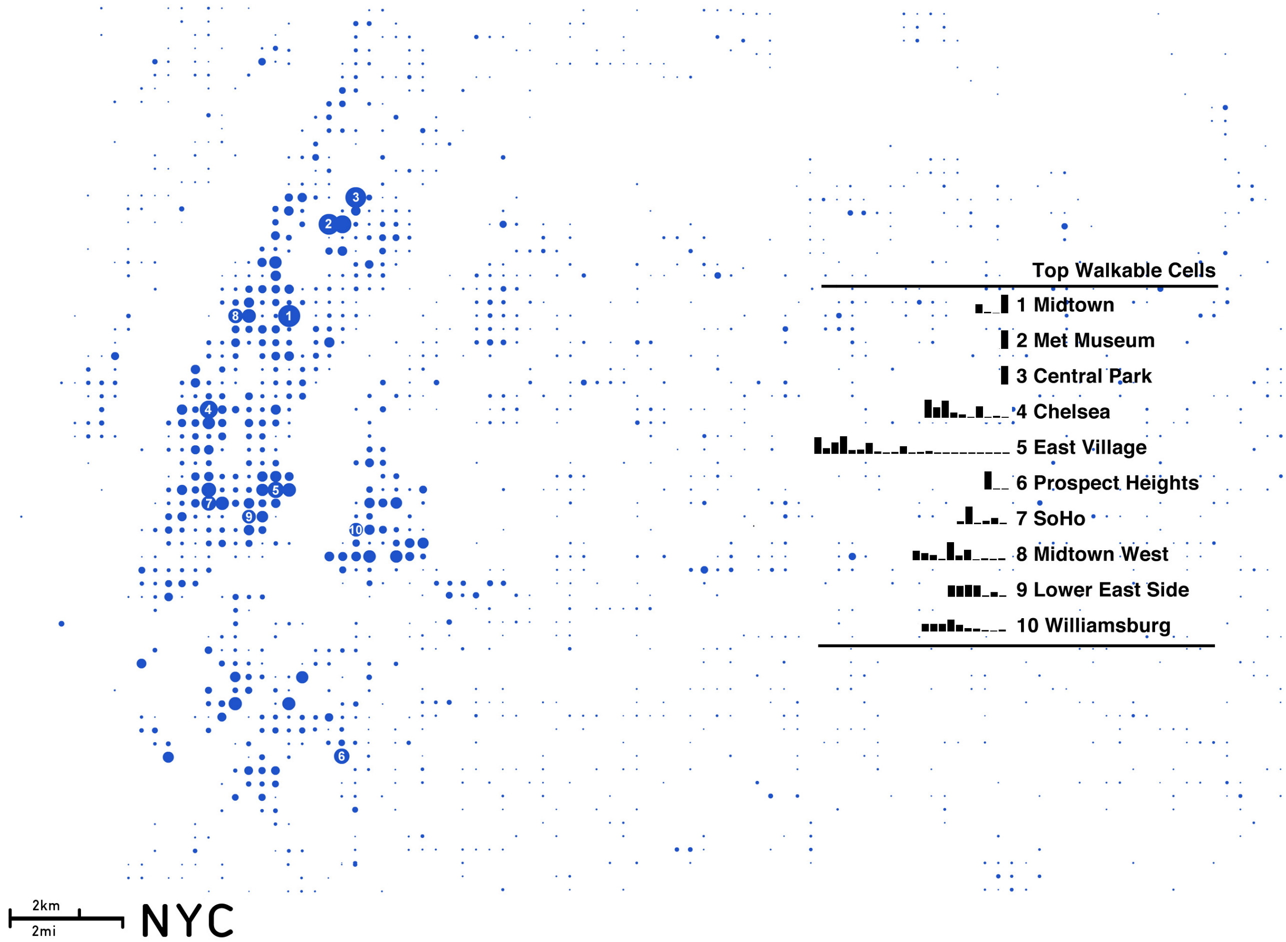

Fig. 1: Social activity fingerprint for New York City. Based on 592,062 check-ins at 7,049 venues on the Foursquare (4sq) social network. Activity is produced by GPS-enabled mobile phones. Each dot represents a walkable cell of 400x400 meters. Dot areas are proportional to cumulative 4sq activity (checkins) within the cell during the period March 2009 to July 2010. At this resolution, recognisable social centers emerge, with the top 10 walkable centers labelled according to colloquial names. The bars represent the relative contribution of different social venues to the overall activity in any given walkable cell, with the order of the bars reflecting the date at which that venue first registered activity on the 4sq network. Revealing the city in terms of social interaction density on such networks can be considered a form of 'sensing' the urban.

- As a navigation tool it can encourage exploration of urban space, (c) Place Recommendation - Users can recommend places to friends using comments that can be added as metadata to check-ins, (d) Experience Sharing - Users can share and recommend experiences to be had at certain locations by supplying 'tips' of things to do at each location, (e) Location-Based Gaming - Users can compete to become the 'mayor' of a location, (f) Listings - The network can be consulted to find out what popular venues are within walking distance of a location.

Urban Analysis

Data

This paper collects and analyses over 800,000 data records from the 4sq network. The taxonomy of the 4sq dataset consists of geo-located *venues* at which *checkins* (activity) take place. The aggregate data relates the two, by expressing the total number of checkins to date at each venue as a real number. This gives us a snapshot of the cumulative data produced by the network from March 2009 to July 2010. Using such an aggregate dataset we can see how activity on the network is distributed through a city.

The data has been collected by a systematic crawl of the 4sq public Search API, which returns upto 50 nearby venues when supplied with a geo-location. For each city, a lattice is constructed of search locations 2km apart and a search performed on each point of the grid. 2km is chosen as it consistently produces overlap in results, implying good coverage of the intervening space between search locations. This results in 200-400 searches per city, the exact number varying based on the size of the surface area covered for each city. As such, the data does not represent a comprehensive data snapshot, but sufficient venue data has been collected – in excess of 6,000 venues for each city – to assume a representative sample of 4sq data for each city.

The data has been collected for three cities: London (LDN), New York City (NYC), Paris (PAR), in order to allow for an inter-

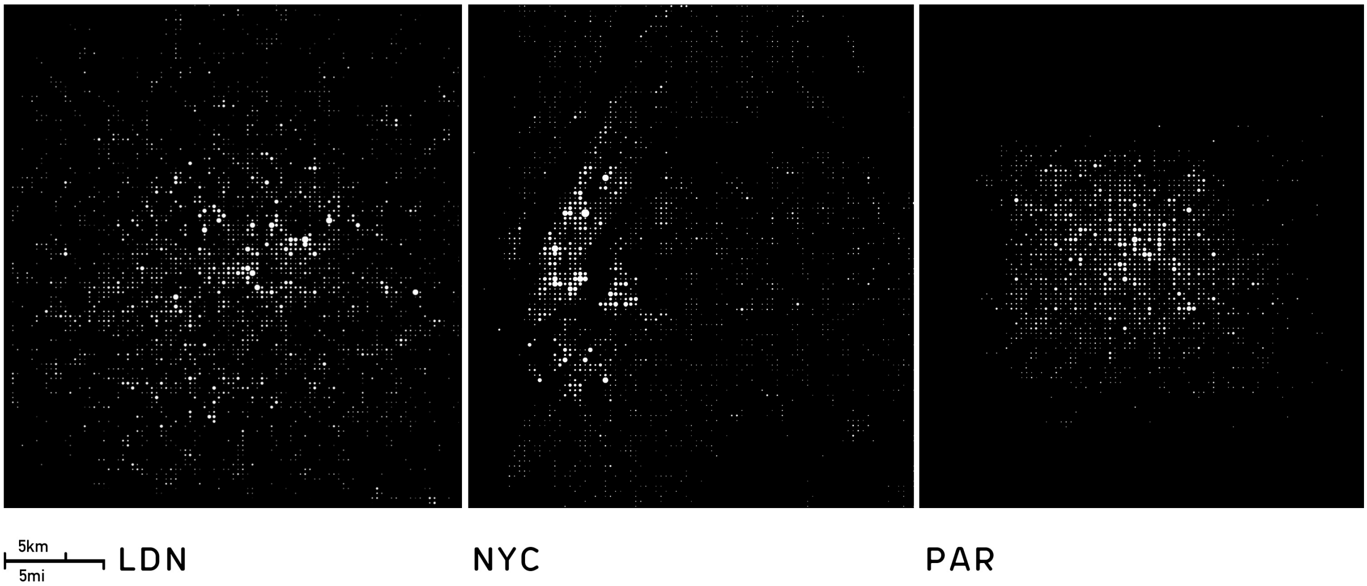

Fig. 2: Social activity fingerprints for (left to right) London, New York City and Paris. Each dot represents a walkable cell of 400x400 meters. Dot areas are proportional to cumulative 4sq activity (checkins) within the cell between March 2009 and July 2010. Activity levels are normalised across the three cities. The fingerprints thus present the cities in terms of social activity density. Parisian activity looks compact whilst London's looks dispersed and fragmented. NYC activity is highly concentrated in Manhattan, the outline of the peninsula clearly visible in the fingerprint, and is contrasted by vast areas with very little background activity. [In colour online]

urban analysis. The data, along with the parameters used for 'crawling' the three cities, is available in full online (Bawa-Cavia, 2010).

This method of data collection pertaining to social activity in space has distinct advantages over the 'gatecount' approach adopted in (O'Neill et al., 2006), which relies on electronic observation posts being set up around the urban space in question, detecting bluetooth signals emitted by the mobile phones of passers-by. Whilst that is a more direct means of 'sensing' flows of pedestrians through a point, it's a far more intensive method of data capture. The 'gatecount' method produces data relating to the proximity of individuals and therefore direct encounter, but lacks the broad scope of location-based social network data, which can be aggregated and analysed in large volumes without sacrificing a high spatial resolution, all at a low cost to the researcher.

The demographic specificity of the 4sq network should be considered when interpreting the aggregate data. In the UK, 3G mobile phone penetration is estimated at 26.5% of the population (Ofcom, 2010), whilst the global demographics of 4sq show a pronounced skew towards university educated 25-34 year olds and a small skew towards females (Alexa, 2010). Activity analysed in this paper clearly relates only to these social groups, and the exclusionary aspect of these "techno-socialities" (Sheller, 2004) should be acknowledged. This does not prevent an inter-urban analysis from proceeding. Assuming that the demographic groups are similar in each city, insights can be gained from the comparative spatial distribution of 4sq activity in each urban case.

Activity Fingerprints

The 4sq data has been used to construct social activity fingerprints for three cities (Fig. 1, Fig. 2). In these visualisations, activity on the network is aggregated onto a grid of 'walkable' cells (each one 400x400 meters in size) represented by dots. The area of each dot corresponds to the level of activity in that cell. A resolution of 400m is chosen based on research into threshold walking distances in urban areas (O'Sullivan and Morrall, 1996; Partnership and TfL, 2006) and an observation that cells of this size map closely to colloquial neighbourhood names. By this process we can see social activity centers emerge in each city. The maps yield recognisable, namable centers. The top 10 walkable centers for NYC are labelled in Fig. 1 and similar hubs emerge for London and Paris, such as Shoreditch and Le Marais. These visualisations are omitted here due to space but are available online (Bawa-Cavia, 2010).

These spatial visualisations omit the usual topographic and cartographic details present in traditional maps, abstracting the city in order to focus on its *activity fingerprint*. This presentation of the city in terms of density of social network activity can be seen as an example of 'sensing' the urban.

Social Hubs

The activity fingerprints in Figures 1 & 2 imply the presence of centers. A rank-size plot for venue popularity is produced in Fig. 3 (a), which confirms the emergence of social *hubs* in each city. Fig. 3 (a) presents a highly non-linear activity distribution, showing a handful of high popularity venues (hubs) trailed by a long tail of low activity locations. Zipf's Law is a fit for the upper portion of the distribution ($20 < n < 500$), describing a power-law decay function $C = k/r^s$ where C is checkins magnitude, k is a constant, r is the rank, and s is defined by linear regression to be ($s_{LDN}, s_{NYC}, s_{PAR} = 0.573, 0.706, 0.623$). The self-organisation of activity into such a power-law distribution is striking, and one can reasonably assume that multiplier or feedback effects are at play. Following on from this evidence of a strong concentration of activity into hubs, an examination of the spatial logics of these hubs becomes the object of this paper.

If we take *hubs* to mean high-popularity locations (venues) present in the tip of a power law distribution, we can then define *centers* as agglomerations of activity at one or more venues. Multiple definitions for these 'centers' are proposed in this paper, and we begin with a crude definition of a center as a fixed-size walkable cell (400x400m), of the type used in constructing activity fingerprints in the previous section. We look now at the urban phenomenon of *polycentricity* through an examination of the spatial dispersion of these centers.

Polycentricity

Polycentricity exists as a spatial phenomenon only with respect to *activities* organised into identifiable *centers*. Various attempts have been made to provide quantitative measures of polycentricity at the urban and regional level. Some focus on the rank-size distribution of settlement populations (Meijers, 2008a), while others have used network analysis (Green, 2007), or the distribution of flows on public transport networks (Roth et al., 2010). Here an urban-scale investigation of polycentricity is presented based on the spatial dispersion of activity centers. Unlike previous measures, this analysis expresses both *functional* and *morphological* aspects of polycentricity, as defined in (Green, 2007), since it considers both activity and its geographic distribution.

In order to highlight these two aspects of polycentricity, let us first examine the rank-size measure presented in (Meijers, 2008b), which is produced by a regression analysis of log rank-size distributions. We can frame this in terms of venues ranked by checkins, which are presented in Fig. 3(b). This clearly separates LDN and PAR, with far lower gradients and therefore greater claim to polycentricity, from NYC, which would be considered more monocentric. This is partly consistent with our visual evidence – compare Fig. 6 with Fig. 8 and NYC activity does look more monocentric. However the fingerprints in Fig. 2 imply that polycentricity varies greatly between LDN and PAR. Whilst activity in the former looks dispersed and fragmented, the latter looks more compact and contiguous. These *morphological* aspects pertaining to the spatial distribution are missing from the rank-size measure of polycentricity; it is confined to exploring solely the *functional* aspect of the phenomenon in terms of the statistical distribution of sizes.

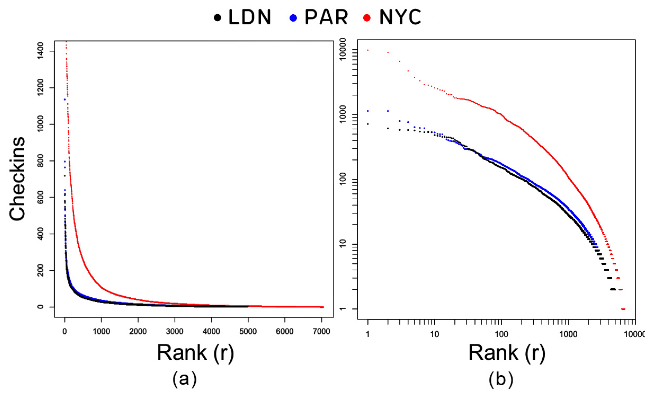

Fig. 3: Rank-size plots for 4sq venues in 3 cities, using checkins as the size parameter. (a) Normal plot, showing hubs trailed by a long tail of low activity venues, (b) Log-log plot, Zipf's law can be seen to be a fit for the upper portion of the distribution ($20 < n < 500$), describing a power-law decay function $C = k/r^s$ where C is checkins magnitude, k is a constant, r is the rank, and s is defined by linear regression to be ($s_{LDN}, s_{NYC}, s_{PAR} = 0.573, 0.706, 0.623$).

To demonstrate a fuller expression of polycentricity, we define centers morphologically as fixed-size walkable cells (400x400m). We can explore the distribution of these centers by examining the spatial dispersion of the top N walkable cells of activity, calculated according to a weighted *standard distance* measure (Bachi, 1963). Cumulative activity (check-ins) for each cell is used as a weighting mechanism and the standard distance μ is defined as,

$$\mu = \sqrt{\sum_{i=1}^{N} \frac{w_i d_i^2}{W}} \tag{1}$$

Where N is the number of top activity cells considered, w_i is the weight (checkins) for cell i, d_i is the Euclidean distance of cell i from the *mean center* and W is the sum of all weights (checkins). This gives us a dispersion statistic in distance units which can be thought of as a standard deviation in space. We produce this measure for N from 2 to 500 for each city. The activity dispersion graph is shown in Fig. 4.

The steep rise in spatial dispersion witnessed in the case of LDN as more top cells of activity are included is contrasted by linear rises for the other two cities. This steep gradient, developing into superlinear growth from $10 \leq N \leq 200$, is evidence of a greater degree of polycentricity in the social activity of LDN, as it shows how the most active cells (or 'centers') of the city are increasingly spread over a much larger area than the other two cities. The low gradient exhibited by PAR is, as I will argue later in the paper, indicative of a *decentralisation* which exists as a limiting case of polycentric space.

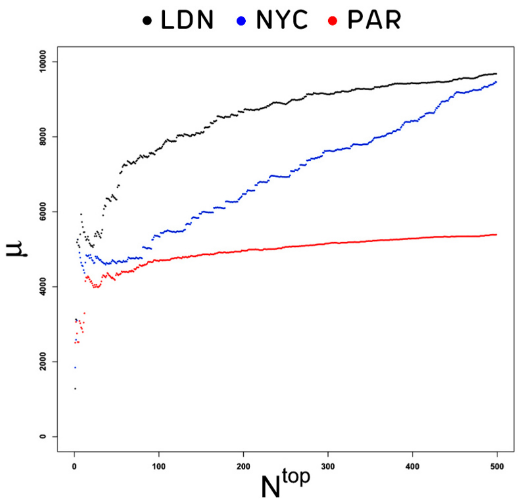

Fig. 4: (a) Spatial dispersion of activity as a weighted standard distance, μ, for N top activity cells, showing how activity in LDN tends towards a more polycentric behaviour, implied by the greater dispersion of its top centers of activity. The low slope of PAR expresses both compactness and a geographically even distribution of activity.

By the time we reach the 100 highest activity cells in each city, the standard distance for LDN (μ^{100}_{LDN}=8,657m) is almost double that of PAR (μ^{100}_{PAR}=4,965m). Taken in this form, the variance in spatial dispersion μ between cities is not a direct measure of polycentricity but rather a measure of overall compactness of the city's social functions. Indeed, if we normalise μ^{100} by the surface area considered for each city to produce $\bar{\mu}^{100}$, we find it has similar levels for all 3 cities ($\bar{\mu}^{100} \approx 19m^{-1}$). However as the magnitude of absolute μ increases, it implies an increased fragmentation of activity. This fragmentation is a necessary pre-requisite to defining a polycentric space – 'centers' themselves being defined by their distance from other sites of activity (Parr, 2010) – so we can consider absolute levels of dispersion μ as contributory to polycentricity.

By introducing a null model we can investigate the relative contribution of both aspects of polycentricity – *functional* and *morphological* – in the case of each city. Our null model preserves the geographic distribution of venues but distributes activity (checkins) evenly over all venues, eliminating variance in the *functional* dimension.

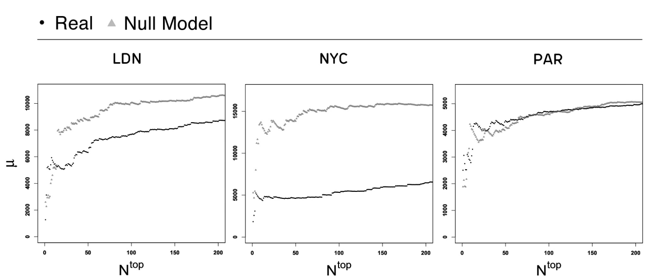

Fig. 5: Comparison of spatial dispersion μ of 4sq activity across N top activity cells in each city with a null model. The null model preserves the geographic distribution of venues but spreads activity evenly between them. NYC produces the largest difference between null model and real data, implying that a strong spatial concentration of activity is countering the geographic dispersal of venues in that city.

We run our dispersion analysis based on standard weighted distance as before, across both each city and its null model, graphing both in Fig. 5. There is a marked difference in the relation between each city and its null model – NYC's null model produces much higher levels of dispersion than the real data, more than double at μ^{200}. This implies a strong concentration of social activity in the real data which counters a high spatial dispersion of venues, the *functional* component of the activity distribution acting against *morphological* dispersal. This concentration of activity is far more evident if we map individual venues in the city, which we do in Fig. 6. PAR, by contrast, has similar levels of μ for both null model and real data, implying the real data contains an activity distribution that's evenly dispersed throughout its geographical locations. Finally LDN sits between the two, activity is more concentrated than its null model, but less so than in the case of NYC.

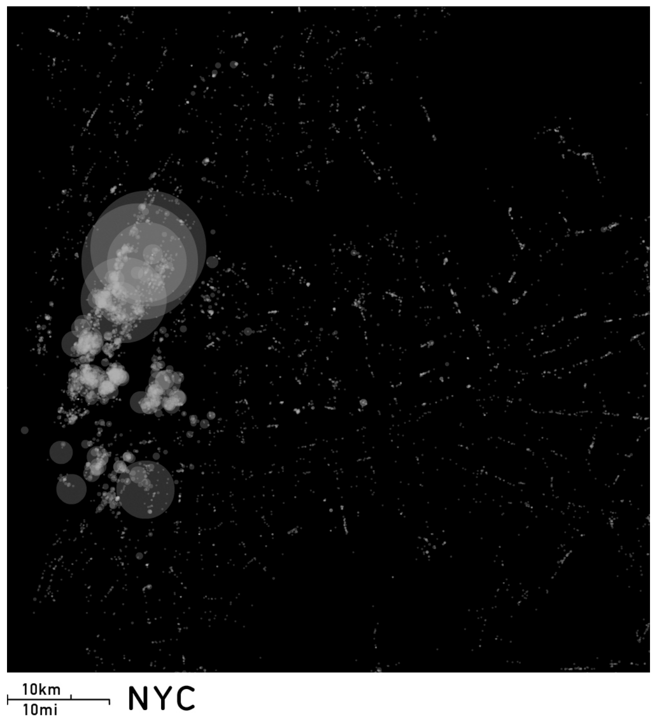

Fig. 6: Strong levels of spatial clustering in 4sq activity at venues in New York City. In this map each social venue is rendered as a translucent dot. Dot diameter is proportional to cumulative activity (checkins) at the venue. The overlapping of dots gives us a visual measure of activity intensity.

Combining observations on the statistical distribution of activity with the spatial dispersion of top activity cells provides us with a more complete understanding of functional and morphological polycentricity, clearly distinguishing tendencies between the three urban cases considered. There is clearly no one definitive measure for such a multi-scalar, multi-faceted and relational phenomena as polycentricity. We can understand it further, however, by examining its relation to a complementary spatial phenomenon: fragmentation.

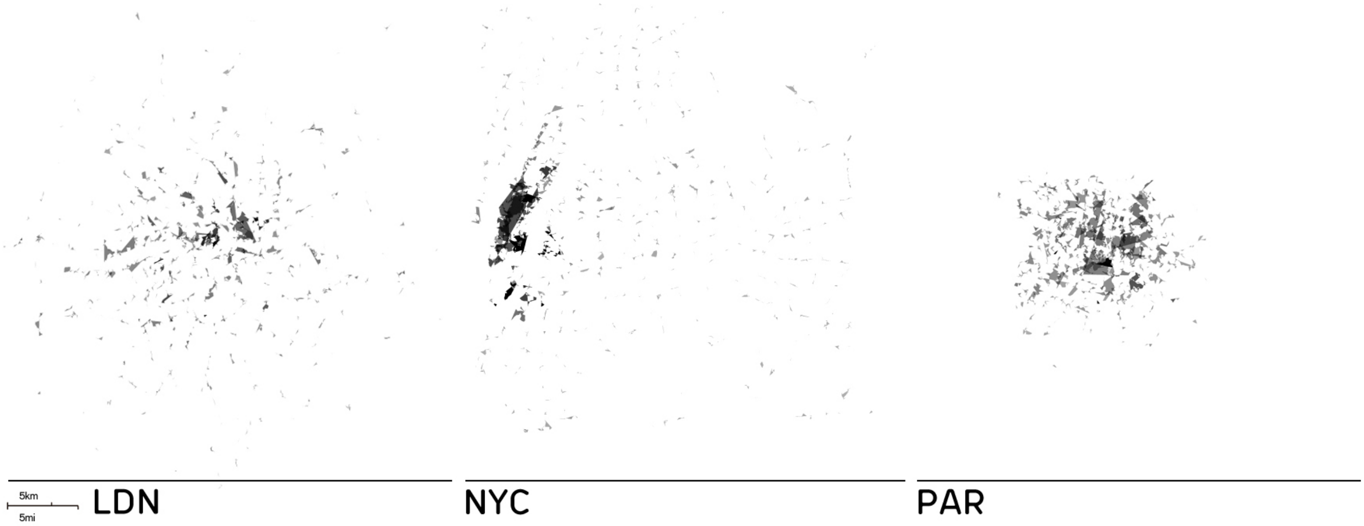

Fig. 7: Fragmentation of social activity on the Foursquare network in London, New York and Paris. The scan-based clustering technique DBScan is used to detect clusters of Foursquare venues within a walkable (400m) distance to each other. Each polygon is formed by joining venues belonging to the same cluster. The darkness of each polygon is proportional to the cluster activity density, defined as the number of Foursquare checkins divided by the number of venues in the cluster. This diagram visualises 18,805 venues arranged into 1,208 clusters. LDN and NYC show much higher levels of fragmentation than PAR, which breaks down into far fewer, larger clusters.

Fragmentation

Spatial fragmentation can only be defined with reference to a threshold distance, beyond which element A is considered distinct and separate from element B. The scan-based spatial density clustering algorithm DBScan (Ester et al., 1996) provides us with the means to examine activity with reference to such a threshold distance. This is pursued to test a hypothesis that the post-modern metropolis functions as a social 'archipelago', a fragmented set of islands characterised by high-density social activity.

DBScan detects morphologically diverse clusters of activity, providing us with a more refined definition of a 'center' than the fixed-size walkable cell. The DBScan algorithm works by iteratively aggregating geo-located points into clusters based on a threshold distance ϵ and a minimum cluster size, c_{min}. Starting with a random point, the algorithm 'scans' outwards, adding points to clusters when they are within the ϵ distance of any existing member in the cluster. Points which cannot be assigned to any cluster are marked as 'noise'. This process is iterated until every point has been examined exactly once.

We model 4sq venues as geo-located points. As DBScan is sensitive to its starting point, we run it 20 times for each city starting with a random venue each time and report here the mean results, observing a statistical variance of under 5% for all the relevant measures. We can explore clusters of venues typified by small, walkable links by defining ϵ as a Euclidean distance of 400 meters, and c_{min} as 3 venues. This threshold distance is an *embodied* measure directly related to the mobility of individuals in urban space. If we run DBScan with these parameters, Paris breaks down to far fewer, larger clusters than the other two cities (LDN,NYC,PAR = 448,496,264 clusters), generating under a quarter of the noise (LDN,NYC,PAR = 1596,1431,407). Cluster activity densities are then calculated as the aggregate 4sq activity in a cluster divided by the number of venues. The clustering results are shown in Fig. 7.

Parisian activity stands out as far less spatially fragmented than the other cities, looking less like an archipelago and more like a contiguous blanket of social activity. London has the highest noise and New York has the highest level of venue fragmentation, roughly double that of Paris. Despite the percentage variance in venues across the three cities being under 10%, the percentage variance in fragmentation is statistically large (>40%). Whilst increasing fragmentation has been thought a characteristic of the western post-metropolis (Graham and Marvin, 2001b; Soja, 2000), the analytical results show how three such urbanisations can exhibit widely differing levels of fragmentation in the context of their social activity.

Agglomeration

Figures 6 & 8 show social morphologies in terms of activity intensity at 4sq venues, each city presenting a unique agglomeration in space. Activity seems most evenly distributed in the case of Paris – often along radial and axial lines (which correspond in many cases to avenues and boulevards) – and most concentrated around large hubs in the case of New York. Within the scope of the data, Parisian activity seems the most *decentralised*. We can confirm this by examining the statistical dispersion of the activity distribution shown in Fig. 3 (a).

The Gini coefficient G_i for city i provides us with a comparative measure of statistical dispersion based on the relative mean difference of activity between any two venues in the distribution, defined as,

$$G_i = \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j| \Big/ 2\bar{\mu}_x n(n-1) \tag{2}$$

Where n is the number of venues in the distribution, $x_{i/j}$ is activity (checkins) at venue i/j as a proportion of the total activity for the city, and $\bar{\mu}_x$ is the mean city-wide fractional activity for a venue.

The Gini Coefficient is normalised by activity magnitude, allowing for a comparative analysis across urban cases, and exists in a range of $0 < G_i < 1$, where 0 indicates a completely even distribution of activity and 1 indicates a total concentration of activity at a single venue.

G_{PAR} (0.655) is significantly smaller than G_{NYC} (0.812), implying a lower statistical dispersion in its distribution. This 'flatter' activity distribution reflects less dominance by a handful of hubs; Parisian activity can thus be said to be more *decentralised*, supporting our visual evidence. We examine now whether this purely statistical analysis reflects differing patterns of spatial agglomeration.

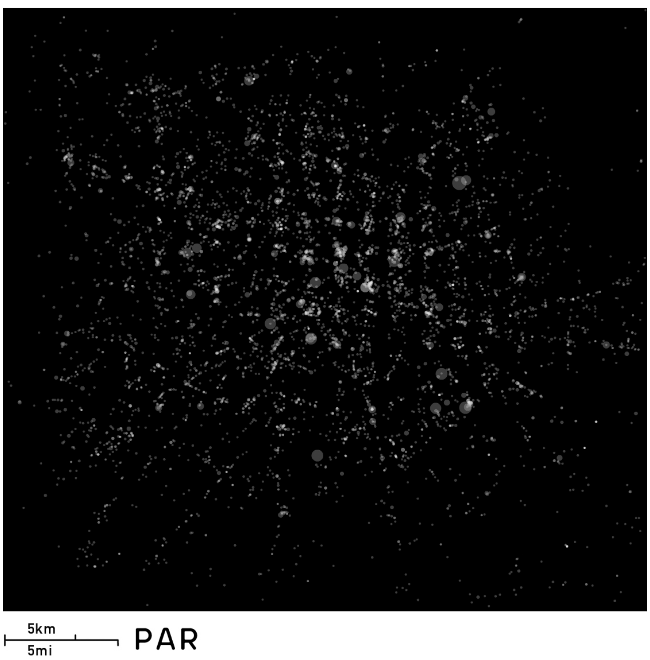

Fig. 8: Decentralised 4sq activity for venues in Paris. In this map each social venue is rendered as a translucent dot. Dot diameter is proportional to cumulative activity (checkins) at the venue. The overlapping of dots gives us a visual measure of activity intensity. Activity can be seen to settle along radial and axial lines corresponding to the larger avenues and boulevards of the city.

45% of all activity in NYC is located within a walkable 400m distance of the upper 1 percentile of activity in the city. This is a higher level of concentration than the other two cities, for which the figures are (LDN, PAR = 30%, 37%). All three figures imply a heavy agglomeration of activity with respect to the total surface area of the city. We can explore the extent of hierarchical agglomeration of activity around hubs by examining spatial auto-correlation (Cliff and Ord, 1970), defined for our data as the tendency of venues to be located near others of a similar activity level. We calculate a local version of Moran's I coefficient of spatial auto-correlation (Anselin, 1995) for each venue i, with I_i defined as,

$$I_i = (x_i - \bar{\mu}_x) \sum_{j=1}^{n} w_{ij}(x_j - \bar{\mu}_x) \Big/ m_2 \tag{3}$$

with the same notation as for G_i, where w_{ij} is the weight element of the standardised weight matrix **W**, indicating the proximity of venue i to j in terms of Euclidean distance. Elements in **W** are normalised such that they sum to 1 in each row. m_2 is the second moment of the distribution, defined as,

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{\mu}_x)^2 \tag{4}$$

As an urban-scale indicator of spatial auto-correlation, I_i expresses the locational proximity of similarly sized venues for each individual venue. The expected I for a random model, $E(I)$, is $-1/(n-1)$, and we subtract this from I_i when expressing Moran measures graphically, producing a scale in which I_i>0 indicates positive spatial auto-correlation. We plot this I_i for the upper 10 percentile of the activity distribution in each city.

The Moran scatterplot (Fig. 9) produces a positive correlation with $r = 0.569$, suggesting a process of activity clustering is in effect, with high activity locations more likely to be spaced closer together than lower activity locations. This correlation is evident in all three urban cases. This pattern contradicts that produced by canonical *central place theory* (Christaller, 1966), the spatial model which produces a hierarchy of increasingly dispersed centers, larger hubs spaced further apart than smaller ones.

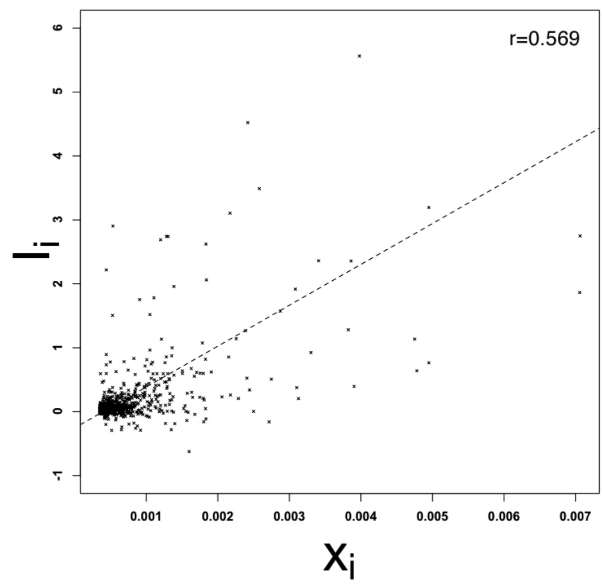

Fig. 9: Urban-scale Moran measure of spatial auto-correlation, I_i, for each venue i, plotted against the proportional activity (checkins) at each venue, x_i, for the top 10 percentile of the LDN activity distribution. Linear regression produces a positive correlation with $r = 0.569$, showing that hubs tend to be spaced closer together than lower activity venues

are from other venues with equivalent activity levels. The correlation coefficient is included on the plot.

We can now explore the hypothesis that agglomerations of venues or activity are more likely to form around high activity hubs, in other words, that activity attracts further activity by a *spatial multiplier* effect (Anselin, 2003), a hypothesis present in some spatial interaction models (Wilson, 2008). We do this by constraining our urban-scale I_i measure (Eq. 3) to an 800m radius r around the venue in question. This means only j neighbouring venues are taken into account within the radius $r = 800m$ when computing each I_i, giving us a local indicator of spatial auto-correlation (LISA).

The results are shown in Fig. 10. The weak positive linear correlation shows how at the local level, higher activity venues have a greater tendency to co-locate than low activity venues, as values of $I_i>0$ indicate a positive spatial auto-correlation. This implies the presence of a *spatial multiplier* effect either produced by, or simply reflected in, this form of social networking activity. The same trend is observed for all three cities.

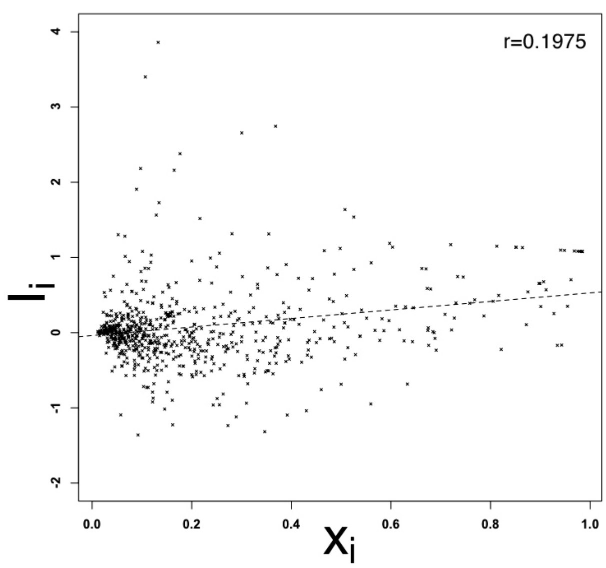

Fig. 10: Local Moran measure of spatial auto-correlation, I_i, plotted against the proportional activity (checkins) at each venue, x_i, for the top 10 percentile of the LDN activity distribution. This local measure is constrained to a radius $r = 800m$ around the venue i in question, with $I_i>0$ implying positive spatial auto-correlation. The line of best fit shows a weak positive linear correlation, with the correlation coefficient marked on the plot. This implies that at the local level, higher activity venues have a tendency to co-locate, implying a *spatial multiplier* effect in which activity attracts further activity.

Discussion

The distribution of social activity in cities has been found to be strongly non-linear in the case of the 4sq network, adopting a power-law, with relatively small quantities of hubs emerging from a long-tail of low activity locations. This self-organisation into social hubs appears to be a marker of urban activity and is true for the three cities examined here.

The 4sq data reveals spatial tendencies in social activity, allowing us to clearly distinguish each city by its characteristic spatial distribution. There is significant variance in the spatial dispersion of activity 'centers' across the three cities. We have shown how London socially functions more polycentrically than the other two cities, and that New York appears to have the most spatially concentrated activity distribution. Polycentricity is presented as a relational and multi-scalar phenomenon with functional and morphological components, which cannot be captured by any single measure.

Fragmentation is complementary to polycentricity. We have explored how the city can be thought to function as a fragmented space of social interaction, and how the cities differ greatly in this respect, the compact form of Paris countering this tendency to fragment. *Embodied measures* have allowed these types of comparative inter-urban analyses to proceed in spite of the many specificities that exist when discussing a particular metropolis.

Decentralisation, as evident in Paris, can be seen as a limiting case of polycentricity, a state beyond the polycentric in which spatial centers cannot be clearly distinguished from an activity distribution. This is characterised by a low slope coefficient in the log form of its rank-size activity function (Fig. 3(a)), coupled with a low slope coefficient in the spatial dispersion function (Fig. 4), as well as a lower urban-scale Gini coefficient for its activity distribution.

We have examined the tendency of venues to agglomerate around activity hubs. At both the urban and local scale, activity appears to attract activity, inducing co-location and implying a *spatial multiplier* effect. The power-law distribution shown in Figure 3(a), supports these notions of hierarchical agglomeration, implying the emergence of hubs via a feedback mechanism.

It is reasonable to assume that 4sq could itself exert a reinforcing influence on this multiplier effect. By making popularity data and personalised recommendations for venues easily accessible on location to users, 4sq has the capacity to produce additional feedbacks of its own that tilt the distribution further in favour of hubs. Conversely, 4sq could also promote the growth of new hubs if recommendations spread through friend networks. It seems clear that with enough adoption, location-based social networks will begin to significantly influence mobility patterns in this manner.

Online social network data is potentially capable of revealing socio-spatial phenomena, such as sprawl or segregation. Parisian data falls away sharply beyond the *péripherique* ring road, meaning activity is contained to under a quarter of the *unité urbaine*, implying a deep segregation between the 4sq network demographic and other groups in the wider city. Likewise sprawl can be considered in terms of low levels of interaction density or a scarcity of social venues. This is a potential direction for further analysis.

The data trails produced by locative networks offer us new snapshots of the city as a living system. Research can use this data to calibrate spatial interaction models (SIM) and provide empirical data for testing the validity of location choice theories. This type of data awaits correlation with a number of other spatial datasets, such as the location of public transport access nodes or land-use data.

Further research avenues are open in examining the spatio-temporal logic of social activity (rhythms, bursts, etc) using a richer social networking dataset, including individual GPS events, to provide insights into urban mobility. Furthermore, examining the spatial clustering of activity in thousands of individual social networks will allow us to characterise the spatial distributions of groups in the city, linking the structure of social networks with their spatial manifestation. The observation that "Social milieus have nonrandom spatial clustering tendencies" (Currid and Williams, 2010) can be expanded on by clarifying how different groups use urban space. These datasets would allow for an analysis not possible with the aggregate data snapshot examined here.

This form of research has the capacity to produce insights on the 'living' city, conceived as an assemblage of social interactions unfolding in real-time, leaving their cumulative history bound in data trails which engage in an increasingly complex relationship with the physical landscape.

Acknowledgements. I would like to acknowledge the contributions of my supervisors Dr. Shi Zhou and Prof. Michael Batty of UCL. Thanks also to Dr. Jon Reades of CASA, whose recommendations were influential, and Naveen Selvadurai of Foursquare for making the data public via the Foursquare API. This research is funded by a UK based EPSRC project entitled SCALE (Grant Ref: EP/G057737/1).

References

Alexa (2010). Foursquare.com audience statistics, http://www.alexa.com/siteinfo/foursquare.com.

Anselin, L. (1995). Local indicators of spatial association - lisa. *Geographical Analysis*, 27(2):94–115.

Anselin, L. (2003). Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review*, 26(2):153–166.

Bachi, R. (1963). Standard distance measures and related methods for spatial analysis. *Papers in Regional Science*, 10(1):83–132.

Batty, M. (1990). Invisible cities. *Environment and Planning B: Planning and Design*, 17(2):127–130.

Batty, M. (1997). The computable city. *International Planning Studies*, 2(2):155–173.

Batty, M. (2010). The pulse of the city. *Environment and Planning B*, 37(4):575–577.

Bawa-Cavia, A. (2010). The city as social archipelago, http://www.urbagram.net/archipelago.

Castells, M. (1989). *The informational city: information technology, economic restructuring and the urban-regional process*, chapter 3, pages 127–171. Blackwell, Oxford.

Christaller, W. (1933 (translated in 1966)). *The Central Places of Southern Germany*. Prentice-Hall, NJ.

Cliff, A. D. and Ord, K. (1970). Spatial autocorrelation: A review of existing and new measures with applications. *Economic Geography*, 46(269-292).

Currid, E. and Williams, S. (2010). The geography of buzz: art, culture and the social milieu in los angeles and new york. *Journal of Economic Geography*, 10(3):423–451.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of Second International Conference on Knowledge Discovery and Data Mining, Portland OR*, pages 226–231. Association for the Advancement of Artificial Intelligence, 445 Burgess Drive, Suite 100, Menlo Park, California 94025, USA.

Foursquare (2010). http://foursquare.com/.

Graham, S. (1997). Cities in the real-time age: the paradigm challenge of telecommunications to the conception and planning of urban space. *Environment and Planning A*, 29(1):105–127.

Graham, S. and Marvin, S. (2001a). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*, page 415. Routledge: London.

Graham, S. and Marvin, S. (2001b). *Splintering Urbanism: Networked Infrastructures, Technological Mobilities and the Urban Condition*, chapter 3 - The collapse of the Integrated Ideal, pages 90–135. Routledge: London.

Green, N. (2007). Functional polycentricity: A formal definition in terms of social network analysis. *Urban Studies*, 44(11):2077–2103.

Haklay, M., Singleton, A., and Parker, C. (2008). Web mapping 2.0: The neogeography of the geoweb. *Geography Compass*, 2(6):2011–2039.

Meijers, E. (2008a). Measuring polycentricity and its promises. *European Planning Studies*, 16(9):1313 – 1323.

Meijers, E. (2008b). Measuring polycentricity and its promises. *European Planning Studies*, 16(9):1320.

Ofcom (2010). The communications market: 2010, http://stakeholders.ofcom.org.uk/market-data-research/market-data/communications-market-reports/cmr10/telecoms-networks/.

O'Neill, E., Kostakos, V., Kindberg, T., Schiek, A., Penn, A., Fraser, D., and Jones, T. (2006). Instrumenting the city: Developing methods for observing and understanding the digital cityscape. In *UbiComp 2006, LNCS*, volume 4206, pages 315–332. Springer Berlin, Heidelberg.

O'Sullivan, S. and Morrall, J. (1996). Walking distances to and from light-rail transit stations. *Transportation Research Record*, 1538:19–26.

Parr, J. (2010). The polycentric urban region: A closer inspection. *Regional Studies*, 38(3):231–240.

Partnership, C. L. and TfL (2006). Legible london - a wayfinding study, http://www.tfl.gov.uk/microsites/legible-london/12.aspx.

Pulselli, R., Ratti, C., and Tiezzi, E. (2006). City out of chaos: Social patterns and organization in urban systems. *International Journal of Ecodynamics*, 1(2):125–134.

Ratti, C., Frenchman, D., Pulselli, R. M., and Williams, S. (2006). Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B*, 33(5):727–748.

Ratti, C., Sevtsuk, A., Huang, S., and Pailer, R. (2007). *Location based services and telecartography: Mobile Landscapes - Graz in Real Time*, chapter 31, pages 434–444. Springer: Berlin.

Reades, J., Calabrese, F., and Ratti, C. (2009). Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B*, 36(5):824–836.

Roth, C., Kang, S. M., Batty, M., and Barthelemy, M. (2010). Commuting in a polycentric city. *arXiv:1001.4915v2*.

Sheller, M. (2004). Mobile publics: beyond the network perspective. *Environment and Planning D*, 22(1):39–52.

Shoval, N. (2007). Sensing human society. *Environment and Planning B*, 34(2):191–195.

Soja, E. W. (2000). *Postmetropolis: Critical Studies of Cities and Regions*, page 231. Wiley-Blackwell.

Twitter (2010). http://www.twitter.com.

Wilson, A. (2008). Boltzmann, lotka and volterra and spatial structural evolution: an integrated methodology for some dynamical systems. *Journal of the Royal Society Interface*, 5(25):865–71.