Sensing Spatiotemporal Patterns in Urban Areas: Analytics and Visualizations using the Integrated Multimedia City Data Platform

Piyushimita (Vonu) Thakuriah, Katarzyna Sila-Nowicka,
and Jorge Gonzalez Paule

Urban Big Data Centre, University of Glasgow,
7 Lilybank Gardens, Glasgow G12 8RZ

Piyushimita.Thakuriah@glasgow.ac.uk,
Katarzyna.Sila-Nowicka@glasgow.ac.uk,
j.gonzalez-paule.1@research.gla.ac.uk

Thakuriah, P., Sila-Nowicka, K., and Gonzalez Paule, J. (2016) Sensing spatiotemporal patterns in urban areas: analytics and visualizations using the integrated multimedia city data platform. Built Environment, 42(3), pp. 415-429.

Abstract

Having the ability to detect emerging patterns in cities is crucial for efficient management of urban resources. Patterns that are useful in identifying and addressing future resource consumption needs include spatial changes in urban form and structure as well as temporal changes in human concentrations and activity patterns during the course of a day. Other patterns of interest are characteristics of local populations in dynamically changing neighborhoods and social-functional spaces. In this paper, we use the Integrated Multimedia City Data (iMCD) platform which brings together multiple strands of structured and unstructured data, to examine such trends in the Greater Glasgow region. We present an approach to, first, understand spatial and time-dependent changes that capture the flow of resources needed to meet demands of residents and businesses at different times and locations, and second, generate hypotheses regarding urban engagement, activity patterns and travel behaviour. We use social media data, GPS trajectories, and background data from the UK Population Census for this purpose. The approach identifies the "roughness" in activity patterns across the urban space that are indicative of different concentrations of social and functional activities. When the time dimension is added to the mix, we are able to uncover time-varying transitions from one type of use pattern into another in different parts of the region. Such transitions, particularly in mixed-use areas, allow early detection of points of excess urban metabolism, with implications for traffic congestion, waste production, energy and other resource consumption patterns. Finally, the ability to detect what citizens talk about socially may provide a way to understand whether or not the language patterns detected in different parts of the city reflect underlying uses and concerns. A preliminary step to evaluate this idea is explored by extracting context-awareness and semantic enrichment to socially-generated data.

1. Introduction

Early detection of emerging urban patterns is of critical importance for efficient management of urban resources. Knowing where there are emerging transport, energy, waste production, water, housing, retail and other needs are key to effective planning and decision-making. The objects of interest for urban pattern detection are numerous. For example, changes in the spatial distribution of population or activity concentrations may signal the need for infrastructure or social services, and are therefore important in monitoring and analysis. Temporal changes in human concentrations and activity patterns during the course of a day may signal time-dependent changes in the flow of resources needed to meet demands of residents and businesses in different locations within an urban area. Being able to adequately detect the characteristics of local populations in dynamically changing neighborhoods as a result of housing turnover and migration patterns leads to new ways of dealing with service and infrastructural needs and to signal new business opportunities targeted at the socio-demographics of the new residents.

Naturally-occurring sources of 'big data' generated through transactional, operational, planning and social activities have generated a great deal of interest in the urban and data science disciplines, and have significant potential for the timely detection of emerging urban patterns. The area of 'urban informatics' in particular focuses on the exploration and understanding of urban systems by leveraging such novel sources of data. The major potential of urban informatics research is in four areas: (1) improved strategies for dynamic urban resource management, (2) theoretical insights and knowledge discovery of urban patterns and processes, (3) strategies for urban engagement and civic participation, and (4) innovations in governance and urban planning, policy analysis, and design (Thakuriah, et al., forthcoming 2016). In particular, increasing access to a wide variety of observational data including infrastructure and moving object sensor data, social data from wearable sensors, social media data and other social data sources, administrative data or data from private business transaction processes, have the potential to allow an examination of a city's activity and land use patterns in ways that were previously not possible. As such data sources become more available, the approaches presented in the paper can be used for continuous monitoring of urban areas as a timely and low-cost addition to current monitoring systems.

The overall purpose of the paper is to present an approach to exploring spatiotemporal dynamics of urban areas using novel multi-modal data at scale for an urban area. Specific objectives are: first, to detect spatial patterns within the city particularly variations in activity clusters in the city, where unexpected or seasonal clusters exist; second; to understand different types of uses, e.g., social and functional use, in these activity clusters and particularly how the same areas transition from one use to another use; and third, to assess the extent to which language patterns detected in different parts of the city reflect underlying uses and concerns.

Towards the aforementioned objectives, the paper describes a data-driven approach to detect spatiotemporal structure in the city which uncovers the systemic and the granular aspects of the city, including local hotspots of activity and mixed-use areas. This approach allows for an early detection of points of excess urban metabolism, with implications for traffic congestion, waste production, energy and other resource consumption, which may call for significant public investment for more efficient and effective urban management, and which at the same time may lead to the generation of hypotheses regarding urban engagement, activity patterns and travel behaviour.

The spaces of interest are major focal points of out-of-home activity and nodes in the networks of human and commercial activity patterns as discerned from these multiple data sources. Using a subset of the multi-modal data, the research agenda consists of studying clusters of activity and identifying structural changes in activity patterns over the urban landscape. We also consider ways in which transitions occur over time from predominantly functional uses of space to social uses during the course of a day, and examine variations in these transitions in areas with different levels of functional-social mixing. Finally, we examine language in social media data to see whether there are discernible differences in the language use in different areas, and whether language patterns change as mixed-use areas transition from one type of use of space to another. This is motivated by the eventual objective to understand whether there is enough discriminative content in language to detect the underlying flow of activity and use of space.

The data source used here is the Integrated Multimedia City Data (iMCD) platform that we have constructed and brings together multiple strands of survey and sensor data on the same geographic area (Greater Glasgow consisting of the City of Glasgow local authority and six surrounding local authorities)) over the same period of time. As an integrated approach to studying cities, multi-modal data at scale integrates multiple sources of data in order to examine complex urban problems from the perspective of different measurement systems. This enables us to provide a more complete picture of urban processes, and also to have an approach to address gaps in individual data strands and a way to understand biases in measurement.

The paper is organized as follows: in Section 2, we provide background material motivating the research approach. Sections 3 and 4 describe the data and research approach respectively. The next sections 5, 6 and 7 focus on results on the three main objectives of the paper, and conclusions are drawn in Section 8.

2. Background and Motivation

Our primary motivation is to understand ways in which spatiotemporal structures in the urban areas can be detected using emerging forms of data, using a data-driven approach, for efficient management of urban resources. Urban spatial structures are becoming more polycentric in structure as new clusters and hotspots emerge within urban areas (Zhong, et al. 2014). Prior research has examined both 'polycentrism' and 'polycentric development', as well as on a morphological approaches centering on nodal features of urban structure, and on functional approaches focused on relations between centers (Burger, et al. 2012). Emerging forms of data have given a new impetus to detecting polycentricity (e.g. Zhong, et al. 2014, Bawa-Cavia, 2011) adding a data-driven dimension to assessing the costs and benefits of such development patterns. A related research strand has examined urban structure by focusing primarily on detecting Central Business Districts (CBDs) in large to smaller metropolitan areas, using traditional socio-economic or remote sensing data sources (e.g., Thurstain-Goodwin and Unwin, 2000; Wu and Murray, 2003;), as well as emerging sources of "movement" data (e.g., Roth et al., 2011; Sun, et al., 2015). Our focus here is on areas of large-scale and significant concentrations of mixed uses as in traditional CBDs as well as on the smaller-scale mixing of functional and social uses.

Researchers have examined ways in which certain sub-regions within cities or metropolitan areas are likely to exhibit particular interactions or activities to a greater degree compared to other sub-regions. For instance, various approaches have been put forward to identify social and functional spaces in cities. Social spaces have long been variously defined by many authors who have focused on different aspects of the connection between location and the social aspects of urban life. For example, some authors have laid greater emphasis on social space as spatial distributions of various social groups or areas that are socially homogeneous in some sense (Durkheim, 1899; Berry and Horton, 1971). Some, on the other hand, have connected the concept to physical areas, for example, to public spaces within cities such as plazas, streetscapes, playgrounds, and entire neighborhoods with the goal of understanding how and why people use these locations, thereby relating social spaces to public spaces (Whyte, 1980). Yet others have noted that urban social life takes place in, and is produced by multiple spatial scales simultaneously (Peterson, 2002). More recently, Woodcraft and Bacon (2013) considered what "everyday rhythms" of a place reveal about the functioning of urban communities, and the factors which support or disrupt everyday local life or the "ordinary, the small-scale and mundane aspects of urban life". These multiple considerations do not all associate social space to specific locations within a city and while there is no comprehensive framework to understanding social spaces, the implication is that networks define shared aspirations, social interactions, and public, entertainment and commercial spaces, where myriad social activities take place. In this paper, we use the term urban social space to mean physical spaces exhibiting spatial and temporal patterns reflecting out-of-home social interaction built on networks or other face-to-face interactions, and activities associated with recreation, entertainment and cultural events.

Functional spaces on the other hand are generally conceptualized in the context of economic interrelationships (Thierstein and Luthi, 2012), as being the result of aggregate economic activities in specific areas (Zhang and Sun, 2014). A related concept is that of functional regions, organised by "horizontal relations in space in the form of spatial flows or interactions of various kind" (Klapka, et al, 2013) and "by a high frequency of intra-regional economic interaction regarding trade of goods and services, commuting flows of labour, and household shopping" (Karlsson and Olsson, 2006). For example, Karlsson and Olsson (2006) identify functional regions using three measures: (1) the local labour market, (2) the region's commuting zone and (3) an accessibility measures using spatial interaction models. We use the term urban functional space to mean physical spaces with a high degree of work-related, economic and utilitarian use.

Mixed development patterns clearly lead to overlaps between functional space and social space. Except for purely residential areas, there are few areas that are purely mono-social or mono-functional. Examples are multi-use spaces with agglomerations of commercial and other utilitarian activities, or spaces where there are social and cultural opportunities which overlap with, or are in close proximity to areas with economic activities. While the city centre or Central Business District is the traditional example of concentrations of such multi-use space (Anas et al., 1998, Sassen, 2001), in reality, with increased population and employment decentralization, as is the case with polycentric cities worldwide, these are likely to be scattered throughout the city (Taubenböck, et al, 2013). In addition, mixed land-use has been a primary goal for Smart Growth, Compact City and New Urbanism planning policies, and areas which are largely residential have been noted to exhibit social uses (Koster and Rouwendal, 2010). Hence there is increased likelihood of functional and social activities occurring in residential areas.

Aside from city residents, social uses may be driven by visitors to cities, for example, the idea of a "second generation metropolis" consisting of a population composed of city users moving to a city in order to use its private and public services such as shopping, movies, museums, restaurants and other cultural and social venues, as well as activities stemming from global connectivity leading to "third generation metropolis". Those attracting specialized populations, termed "metropolitan businessmen" consists of convention goers, consultants and international managers, who travel to central cities to conduct business, establish professional contacts, and visit customers, as described by Martinotti (1994). Such differentiations potentially lead to spatial dynamics in the social space of city residents and city users.

Different areas within a city are also likely to exhibit temporal dynamics resulting from daily, weekly and seasonal activity patterns reflecting different underlying uses. For example, roads and transport systems in mono-functional spaces are likely to experience high levels of demand and congestion during peak hours reflecting commuting flows, and daytime populations and job location arrival and departure time patterns that result dynamically from specific work schedules. Dynamic patterns in such areas may be contrasted to patterns in areas with mixed functional and social spaces (e.g., where there are dining, recreational or entertainment spaces, in addition to utilitarian spaces) which are likely to exhibit commuting and work-related travel and activity patterns, but also transportation demands and activity during mid-day, evenings and weekends resulting from the social uses the same space. On the other hand, mono-social spaces may exhibit aggregate activity patterns reflecting leisure time use by city residents, use by city users and visitors, and the opening hours of recreational, entertainment and sporting venues located in these spaces, some of which might be aperiodic/episodic or seasonal in contrast to daily and recurring.

A voluminous literature has tried to capture the temporal aspects of urban dynamics resulting from time-varying road traffic sensors (Chang, et al, 1994; Sen et al, 1995) and travel and activity data (Schönfelder et al, 2010), electricity consumption (Bogomolov et al., 2016), energy and material flow data (Decker, et al., 2000) and other urban flow data that allow determination of urban consumption patterns and resultant rhythms and demands (Decuyper A. et al., 2014). Emerging sources of big data (from mobile phones, GPS trackers, smartcards, smart metering, social media, location-based social networks) that allow continuous active or passive levels of sensing, reveal investigations of these dynamics at levels not previously possible. The use of such data sources not only allows the detection of variations in use patterns among different parts of the city, they also allow insights regarding phase transitions from one mode of use to another in mixed used areas during the course of a day.

3. Data Sources

Against this backdrop, it becomes useful to examine cities from a multiplicity of data sources which enable the capture and linkage of different dynamics and patterns. Here we introduce and utilize the Integrated Multimedia City Data (iMCD) platform that covers the Greater Glasgow urban area, with a view to highlighting its potential for urban informatics research through a series of visualization and exploratory analyses consistent with the research agenda outlined in Section 1. The iMCD data is multi- modal in nature and consists of the following strands:

Strand 1: Participant survey: A primary survey of a stratified random sample of 1505 households in Glasgow and household members (2095 persons) who participated in an extensive questionnaire-based survey as well as a personal sensor experiment:

a. Questionnaire-based survey – including questions on (1) transport preferences and travel patterns through a travel diary, (2) energy use and sustainable consumption patterns, (3) ICT use patterns, (4) attitudes and personal preferences, (5) sociodemographic, economic and labour market factors, (6) education, cognitive and literacy levels as measured by specific skills-related instruments, and (7) political preferences, civic participation and citizen engagement behaviours;
b. Personal Sensor Survey and Activity Diary - Sensing surveys consisting of GPS movement data collection, lifelogging image data capture and an activity diary that is completed by participants of the sensor survey;

Strand 2: Internet Information Retrieval: information retrieval for a year of a variety of internet-based data for the period over which the social survey and the sensor survey were collected: text-based social media data – Twitter and Foursquare and multimedia data such as Flickr and images from news sources;

Strand 3: Remote Sensing Data: very high resolution satellite data and LiDAR to construct a dynamic digital terrain model for Glasgow;

Strand 4: Sensor Data: data captured from a wide range of urban sensors, e.g. transportation, crime detection systems, emissions, and weather;

Strand 5: Specialized Private Sector Datasets: e.g. cyclist GPS, cell phone, housing transactions, and other such data;

Strand 7: Background Data: consisting of a number of existing data sources that can help in benchmarking and quality assessment, including the census, other specialised government-sponsored surveys, as well as administrative data being published by the City of Glasgow through its Open Data Portal.

The basic idea of the iMCD draws from the concept of *Digital Mobility Information Infrastructure* (DMII) in urban areas proposed by Thakuriah and Geers (2013) which comprise three tiers of information sources, a primary tier consisting of a myriad of infrastructure-based and moving-object sensors as well as socially generated data, and a secondary tier with information supporting urban applications in various ways including digital map databases, Points of Interest (POI) databases, together with extant data sources pertaining to the overall state of the urban area and transportation network, including synthetic, model-derived information supplement this database while finally a tertiary tier consisting of background data from censuses, administrative data sources and other data programs for information on crime, health, safety, weather, emergencies, special events and related activities can assist in urban analytics.

The iMCD platform is motivated by multiple primary objectives, the primary ones relevant here being to study biases in emerging forms of data by benchmarking against a purposefully designed probability sample and other known administrative sources. These enable an understanding of the extent to which emerging unstructured data sources (eg, social media and user-generated multimedia data) can help augment existing census and sensor-based urban operations and management and generate data-driven hypothesis about urban behaviours.

4. Research Approach and Data Processing

We use three sources of data gathered as a part of the iMCD platform including Twitter, Foursquare and participant GPS movement data, and background data from the 2011 UK census. Regarding the first objective of the paper, to uncover the spatial structure of the region, for example, to understand where clusters of activity exist and where there are knot points and edges between different types of activities, we use Multivariate Adaptive Regression Spline and clustering approaches on geotagged tweets posted within Greater Glasgow for the study period.

Regarding the second objective, to investigate the type of use within areas of activity, i.e., social or functional or mixed uses, we allocate small areas within the region to social and functional use according to a scheme presented below, which is operationalized using data-driven rules. The approach taken here is as follows: first, we generate a lattice of 13733 (50metre X 50metre) grid squares across the Greater Glasgow area. The reason behind the creation of such a grid across Glasgow is to have a meaningfully fine spatial unit to aggregate and link multiple strands of geographically tagged data. Hence the unit of analysis in this application is at 2500 square metre resolution. We then cluster the grids as being highly social or highly functional or as a mix to different degrees of social and functional characteristics considering ways in which local areas transition from one phase to another; for example, how an area that is mono-functional during daytime becomes social during the evening, and ways in which these temporal changes vary in the different clusters.

Figure 1: The Framework – Dynamics of Mixed Use Spaces – Social Spaces and Functional Spaces to Detecting Urban Metabolism Levels

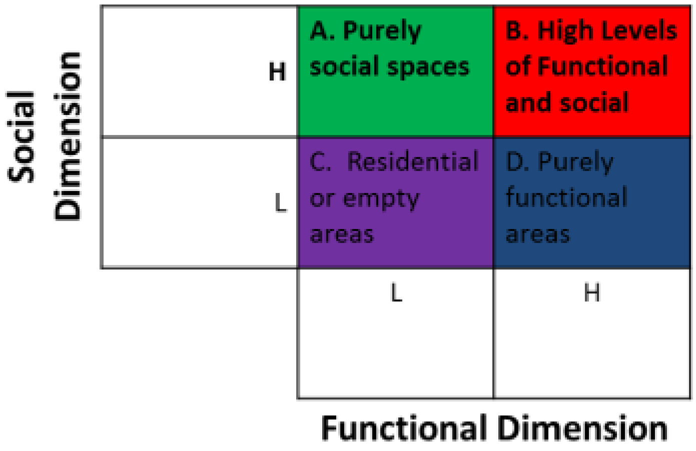


Figure 1 provides the conceptual depiction driving the research approach indicating the intensity of the social-functional mix we want to create for Glasgow using the data-driven approach. The figure has two dimensions, functional and social, giving the four groupings: Cluster A: High Social – Low Functional (H-L) or purely social places, Cluster B: High Social – High Functional (H-H) or mixed-use spaces, Cluster C: Low Social – Low Functional which are residential or empty unused areas, and Cluster D: Low Social – High Functional (S-F) which are purely functional areas. The first objective is investigate which grids belong to each cluster; this is achieved through data-driven approaches.

However, the allocation of grids to clusters is time-varying particularly in mixed-used areas, as places transition from being, say, social to functional. We determine where in the social-functional spectrum an area lies during different times of a day, and the dynamics of how mixed areas evolve over time, thereby operationalizing the dynamic aspects of the concept depicted in Figure 1. Higher levels on Twitter and Foursquare and lower levels of census commuting inflows lead to grids being tagged as purely social (Cluster A. High Social–Low Functional (H-L)) and higher levels of inflows and low levels of social activity lead to grids being tagged as purely functional (Cluster D. Low Social–High Functional (L-H). Ranking high on both dimensions, i.e., both social media as well as inflows, lead to a categorization of Cluster B (High Social–High Functional or H-H), whereas ranking low in both dimensions led to C (Low Social-Low Functional or L-L). The technical details of deriving these transition thresholds and allocation of areas to categories are not presented here but the results are clear in that they show how mixed use spaces transition from one use pattern (phase) to another.

The time at which mixed use spaces transition from one type of use (phase) to another is assessed using Foursquare data, which we examine, due to reasons of data availability, only for the centre of Glasgow and the neighboring West End area. Average stop lengths detected from GPS trajectories by time-of-day shed light on the extent and duration of movement and activity in different areas.

Regarding the third objective, to assess the extent to which language patterns detected in different parts of the city reflect underlying uses and concerns, we semantically enrich the clusters with social annotations derived from social media data which describe conversations occurring frequently in each cluster of grids. The analysis presented here is a result from three types of method: data preparation, data analytics and visualization. The data preparation steps are briefly described below whereas the analytics and visualization methods are embedded in the discussion of results.

In terms of the data processing needed for the aforementioned work, the geotagged Twitter data used consisted of 82,019 unique tweets posted by 7602 unique users. On the other hand, the original Foursquare dataset consists of 226,912 checkins generated by 2203 users in 1968 venues in Glasgow City Centre. Since we consider Foursquare as a social activities indicator, we filtered the dataset by removing all the checkins in venues not considered as social spaces, such as banks, city council offices or train stations. After the filtering process, the resulting dataset was reduced to 158,002 checkins, 1560 users and 1352 venues. The GPS dataset consists of trajectories of a subset of 121 individuals from Greater Glasgow, which processed (cleaned, segmented, classified) using a framework provided by Sila-Nowicka et al. (2015) to categorise movements of individuals into, first, a set of significant places with annotated activity (workplace, home, shopping destinations, recreation destinations, traffic stops) and, second, transportation modes for the remaining moving parts of the trajectory. The details of GPS data processing, segmenting, classifying and semantic enrichment however are beyond the scope of this paper, but readers are referred to the relevant literature (see Liao et al. 2007; Ye et al. 2009; Rodrigues et al. 2014; Umair et al. 2014; Zhou et al. 2007; Huang et al. 2013; and Sila-Nowicka et al. 2015). As we are interested in static social and functional activities for further analyses, we used only significant places of individuals moving around the area covered by the designed grid. From the initial dataset consisting of 2.5 million raw movement points, we retrieved 3801 significant places after clipping to the coverage of the study area (89000 points) and by classifying the movements.

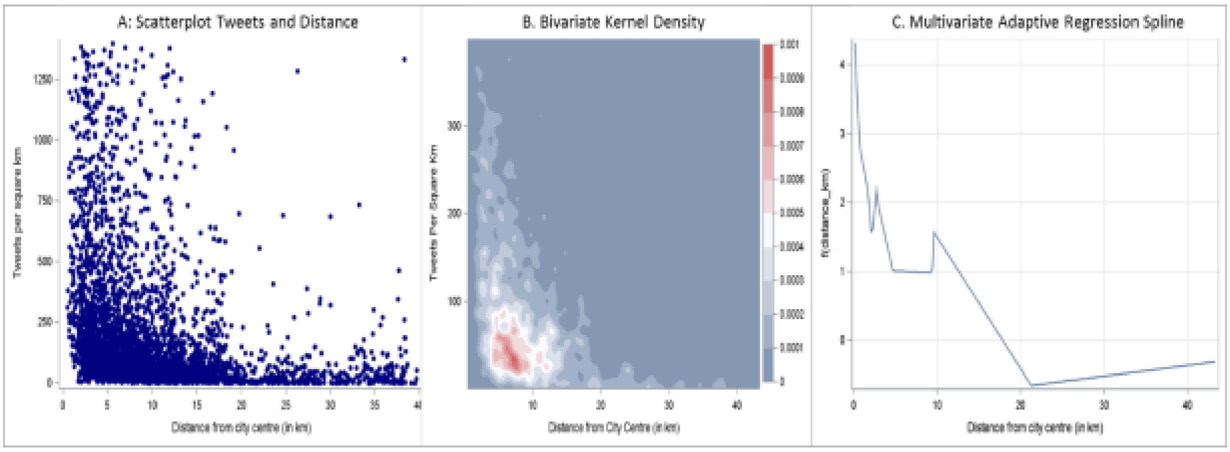
The Twitter data were used to add context-awareness and semantically enrich identified areas, for which we proceeded with extracting representative words from Tweets posted within the grid cells that represent each area (clusters). Our objective was to find terms that describe the activities taking place in those urban spaces. To accomplish the task, we computed the Inverse Document Frequency (IDF) statistic over each collection of tweets in an area (Manning et al, 2008), where the IDF is defined as the logarithmically scaled fraction of the total number of documents (tweets) divided by the number of documents containing the word. Using this measure, we can build a distribution of words over the collection ordered according to its level of importance.

To find the representative words, we created collections of tweets using the messages posted in high social areas (A), high work areas (B) and so on. To identify common words in an area, we aggregated all the tweets posted in the city which provided us a background distribution of words in Greater Glasgow. The aim was to compare the distribution of each single area against the background distribution of the whole region by computing the difference between the IDF values of a word in each distribution. The words with highest differences (peaks) are of special importance in the clusters and therefore are potentially the most representative. Since we were also interested in identifying discriminative words appearing in tweets for each of the clusters, we calculated the differences between the IDF values of a word and the background set of words, for words appearing only in particular clusters. Applying these two methods results in a rank for areas A, B, C and D, containing the most representative words or the most discriminative words.

5. Uncovering the Systemic and the Granular Structure of the City

Not surprisingly given the overall density of activity, the centre of the city has the highest concentration of tweets. These intensity of tweets decline as distance from the centre increases in a pattern that is similar to declines in the levels of concentration of population density, housing density and overall Points of Interest (POI), with distance. The scatterplot of the count of tweets per Output Area against distance from the city centre is shown in Figure 2A with a few extreme points removed. Figure 1B shows a bivariate kernel density of these aggregated tweets against distance from city centre, starting with very high levels of tweet intensity in a few locations close to the city centre, leading to a great increase in intensity of tweets around 3km; a Multivariate Adaptive Regression Spline model (MARS model) (shown in Figure 2C) detects several local knots indicating the highest levels of nonlinearity at distances less than 3 km reflecting considerable activity in the West End of the City which is a hub of mixed commercial-university activities. Moving out from there, the intensity levels continue to decline, but as seen from Figure 1C, there is a steep drop in intensity patterns around 10 km from the city centre. After the 20km mark, there is a slight rise indicating social media activity in the outlying areas.

Figure 2: Tweets and distance from the city centre from different perspectives

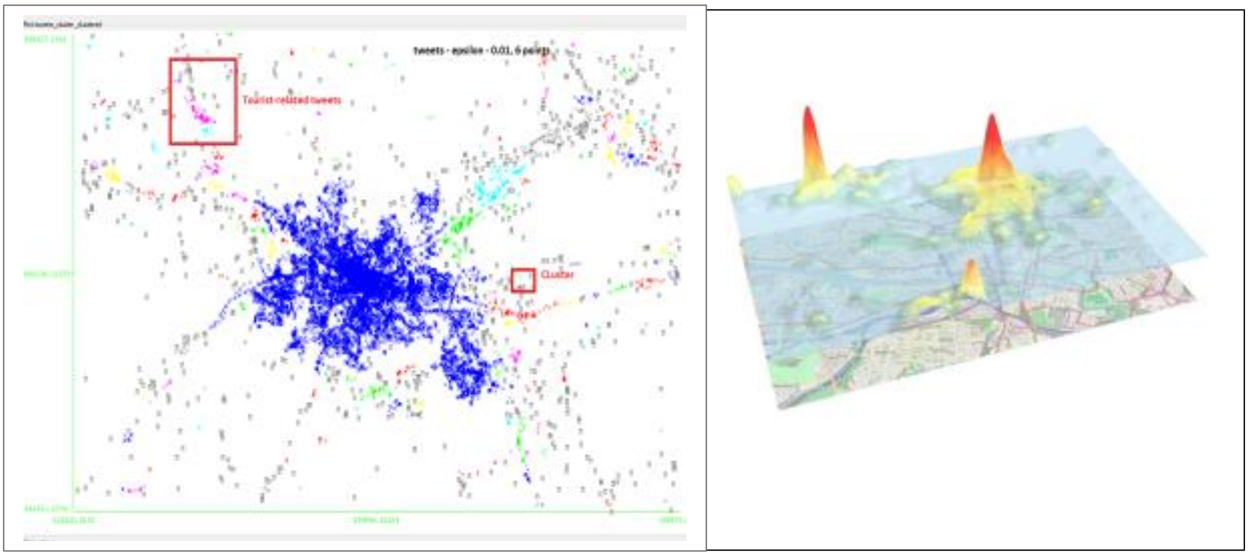


The 10km knot is an edge effect beyond which there are significant changes in the extent and direction of the relationship between the tweeting intensity and distance from the city centre. Based on regression, the magnitude of tweets increases with the density of POIs in areas within the 10 km boundary when controlling for distance from city centre, the total volume of residential and work activities, and the presence of affluent households. On the other hand, outside the 10 km region, tweets decline with an increase in the number of POIs, indicating a somewhat different relationship regarding where people tweet relative to the geography of POIs.

These results reflect larger residential and commercial land uses in the outlying areas in contrast to the centre of the city which has a greater mix of entertainment and commercial activity. The edge itself is explained to a certain degree by green space, other open or undeveloped areas, and overall far lower levels of POIs compared to areas within 10km of the city centre. A proportionately higher level of POIs may be considered to be "functional" in the outlying areas consisting of transport and public infrastructure, manufacturing and production centres, and education and health-related organisations in contrast to "social" POIs relating to sports and entertainment venues, retail and commercial locations, hotel accommodation and restaurants and other attractions. In fact, about 68% of the POIs outside the 10km distance from the city centre may be considered to be functional in contrast to some 53% within this boundary.

When studying the region as a whole, the intensity of social media activity in the central part of the city masks concentrations of regional activity patterns outside the 10km region, thereby making activity outside the centre appear as a long tail, with only a few remaining activity areas of interest. This masking effect can be seen from the MARS model in Figure 2C which shows a straight decline in the component function after the knot point at approximately 10km, with only one slight upward trend from the knot at about 22km. Figure 3A shows that there are several regional clusters of tweets in the outlying areas spatially depicting where developments past this knot are estimated. These types of activity locations can also be inferred by GPS trajectories as in Figure 3B, which shows the concentration of stops inferred for the city centre and the West End of the city, where there are considerable mix of retail, commercial, social and work-related spaces.

Figure 3A: Map of Greater Glasgow with Twitter Clusters



Overall, we find that the gradient from the centre of the city is negative but that there are a wide range of local minimum and maximum points as well as knot points that are indicative of concentrations of activity generated as a part of social and functional activities, and which change from one mode of use to another. The overall assessment is that such “roughness” across the urban space is the norm rather than any smooth spatial persistence of activity.

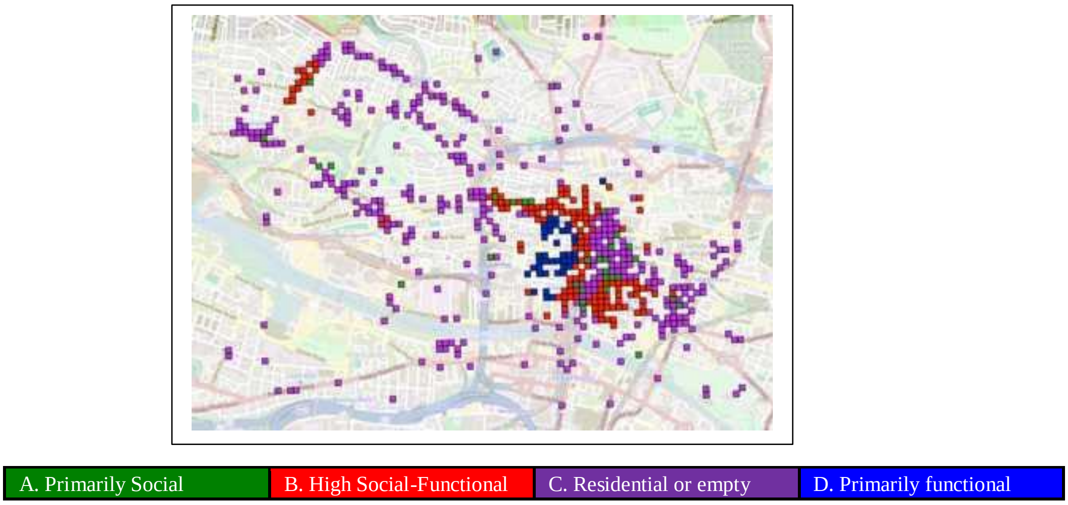
6. Adding Time to the Mix

When the time dimension is added to the mix, we are able to uncover spatial and temporal concentrations of activity which are of time-limited lifespan. For instance, we detect regional activity nodes that are not necessarily structural or persistent in nature either morphologically or functionally in the sense of Burger et al. (2012) but are rather seasonal or episodic depending on the time of the year and the offerings at different attractions. Specifically when we consider the time of tweet generation, patterns other than space and location start to appear. For example, the luster within the square in Figure 3A shows what appears to be a summertime tourist cluster which may not persist as seasons change. On the other hand, there are clusters located around the region including those outside the edge where there are short bursts of activity. Some of these “bursts” may live only for the duration of a major disruption, incident or event.

In between disruptions that last a very short period of time and more long-term patterns (such as those discussed in Section 5 or even seasonal patterns), there are diurnal patterns as places go from having everyday functional rhythms to thriving social rhythms signifying different urban metabolism and resource needs. Mixed use areas in fact are not static – they transition from being dominantly at one type of use or phase (social or functional) to being in another phase.

Figure 4A shows the spatial distribution of local areas in terms of their predominant category, i.e., the categories that the location are in for majority of the time. The map shows only those grids where there is data available for all four data strands used (GPS, Twitter, Foursquare, Census). Within the central Glasgow area using our criteria, there are comparatively fewer areas which are primarily functional (Cluster D) and primarily social places (Cluster A); most areas exhibit mixed uses based on our criteria.

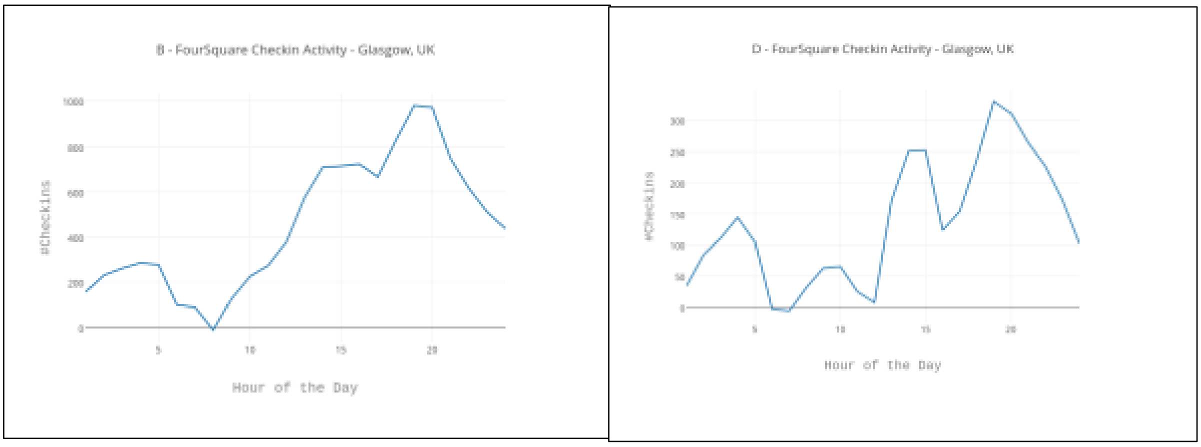
Figure 4: Placement of areas in the social-functional mix



A feature of mixed-use areas is that they transition from one category to another in time-varying ways that reflect schedules, operations and opening/closing times and overall flows of activity in that space. Night-time and evening hours are characterized by a high level of social activity. The morning hours clusters tend to capture social activities in the proximity of railway and bus stations. Lunchtime clustering shows high social and mixed activities spreading around the city centre of Glasgow, and then the phenomena fades for a few hours after lunch and increases again in the evening.

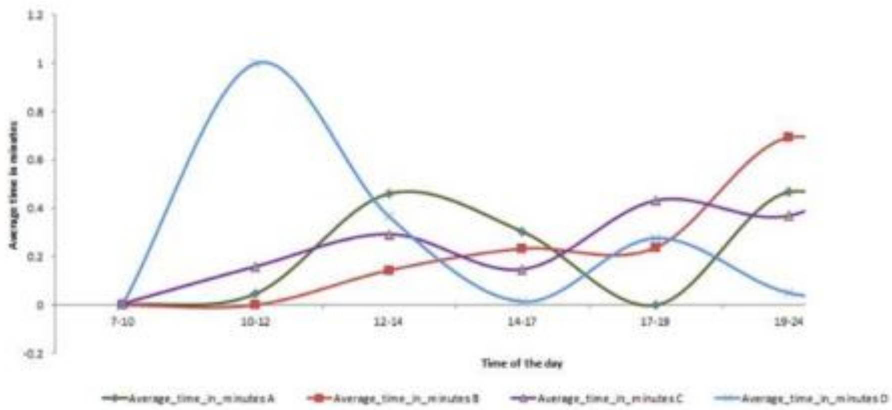
All clusters demonstrate a bimodal distribution of activity levels as discerned from Foursquare data, indicating a smaller morning peak and a larger peak later in the day. Figure 5 represents two cases where bimodal distributions of activities are seen in both [B] Highly Social-Functional Mixed Places as well as [D] more mono-functional spaces. However, the peak in [D] starts to occur later suggesting closer association with end-of-work activities whereas in B locations, the transitions from functional to social occur earlier but the peak is later.

Figure 5: Bimodal activity distributions detected from Foursquare checkins



The dynamics of urban space and time as detected from social media data can be seen in other ways. Figure 6 shows average stop lengths detected from GPS trajectories by time-of-day. The longest average stays are in D or primarily functional areas due to users who arrive early in the day for work and stay for the entire day. The longest stays in C or residential areas occur for trips arriving later in the day presumably for people staying the rest of the night (the data are censored for a 24 hour period). The shortest duration stays are early afternoon arrivals in A or in primarily social places.

Figure 6: Average GPS-inferred stop lengths in different areas by time-of-day



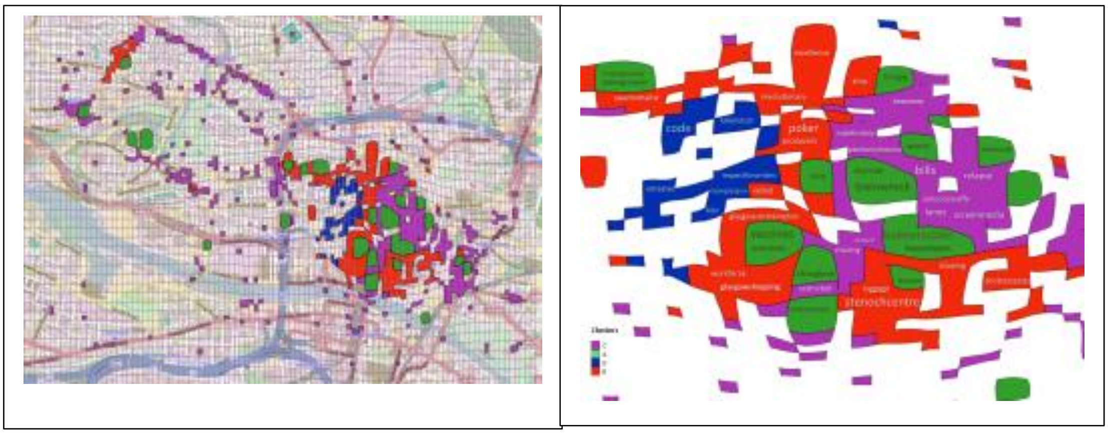
A. Primarily Social	B. High Social-Functional	C. Residential or empty	D. Primarily functional

7. Adding Semantics to Time and Space

What do people talk about in social terms in different areas of the city? Can we use language patterns detected in different parts of the city to understand underlying uses, activities, and concerns? Spatial visualizations of the semantic enrichment approach described in Section 4 allows us to see such patterns. Figure 7A shows the same map as in Figure 4 but represents a distorted cartogram of the grid cells in the city centre of Glasgow. The distortion is dependent on a number of tweets per each grid and colours represent the already mentioned clusters (A,B,C,D). Although the number of tweets have already been used to allocate grids into one of the four clusters, Figure 7 depicts rather well that the small size of primarily functional areas (D) depicted by blue represent the limited social media activity in those areas. Blue clusters represent the areas where the social activity is low and the functional activity seems to be high.

Figure 7B displays detects language patterns in each of these areas. The most common words are related to work and commuting. Grids in the green clusters or areas in A are the largest reflecting the high levels of social media activity there and are represented by words strictly connected with social activities: ‘Hotel’, ‘cup’, ‘chill’, ‘walking’, ‘champagne’ or ‘theatre’. Red clusters depict mixed-use areas and also mixed words representing social aspects of behaviour but more focused on cultural aspects of the city’s life, as well as topics connected with research, universities and traveling. Grids in Cluster C shown in purple represent areas of relatively low social and functional activity showing the sparsest group of words, whereas the main directions in conversations are hard to identify. This indicates that overall language spoken in different areas may depict underlying activities and uses in those areas, and where relevant, for the use pattern most prevalent for specific times of the day.

Figure 8: Context-awareness and semantic enrichment of identified areas



8. Conclusions

Our overall goal in this paper has been to evaluate if emerging sources of data provide a way to detect patterns in cities crucial for the efficient management of urban resources. The patterns which we focused on have implications for future resource consumption needs. These are primarily spatial patterns in urban form and structure, time-varying diurnal patterns in human concentrations and activity patterns during the course of a day, as well as language patterns detected in different parts of the city which may reflect underlying uses and concerns in different areas. We have used Twitter, Foursquare, GPS and background data from the UK Population Census for these purposes, available as part of a novel multi-modal data resource which we have built – the Integrated Multimedia City Data (iMCD) platform.

The approach identifies the "roughness" in activity patterns across the urban space that are indicative of concentrations of activity generated as a part of social and functional activities. When the time dimension is added to the mix, we are able to uncover spatial and temporal concentrations of activity which are of time-limited lifespan giving insights into the time-varying transitions in different parts of the city from one type of activity pattern to another particularly in mixed-use areas, enabling early detection of points of excess urban metabolism, with implications for traffic congestion, waste production, energy and other resource consumption. Finally, a preliminary evaluation of detecting what citizens socially talk about in different areas of the city demonstrates the potential of using language patterns to identify underlying uses and concerns.

Access to observational data of the kind used in the paper, and many other similar forms of data that are increasingly becoming available from a variety of sources have the potential to allow monitoring of a city's spatial and temporal activity and use patterns continuously. Such strategies have the potential to augment current monitoring systems at low cost, enabling targeted and timely resource management strategies to be adopted. Several major challenges that are likely to arise in varying degrees when using Big Data for urban informatics (Thakuriah, et al., forthcoming 2016); these are technological, methodological, theoretical/epistemological, and the emerging political economy of big data. These challenges have been demonstrated in various ways here, indicating that emerging forms of data can provide insights into complex challenges facing cities but that considerable effort will be needed to derive hypotheses supporting urban planning and resource management and for real-world use of novel forms of data for better urban living.

References

Anas A., Arnott, R., and Small, K. (1998) Urban Spatial Structure. Journal of Economic Literature, 36, 1426-1464.

Bawa-Cavia, A. (2011) Sensing the urban: using location-based social network data in urban analysis. In Pervasive PURBA Workshop.

Berry, B. L. and Horton, F.E. (1970) Geographic Perspectives on Urban Systems: With Integrated Readings. Prentice Hall.

Bogomolov, A., Lepri, B., Larcher, R., Antonelli, F., Pianesi, F., and Pentland, A. (2016) Energy consumption prediction using people dynamics derived from cellular network data. EPJ Data Science, 5(1), 1-15.

Burger, M. and Meijers, E. (2012). Form follows function? Linking morphological and functional polycentricity. Urban Studies, 49, 1127-1149.

Chang, G-L., and Wu, J. (1994) Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors. Transportation Research Part B: Methodological, 28(2), 141-160.

Decker, E., Elliot, S., Smith, F., Blake, D. and Rowland, F.S. (2000) Energy and material flow through the urban ecosystem. In Energy Environment, Vol. 25, pp. 685-740.

Decuyper, A., Rutherford, A., Wadhwa, A., Bauer, J. M., Krings, G., Gutierrez, T., Blondel, V.D. and Luengo-Oroz, M. A. (2014) Estimating food consumption and poverty indices with mobile phone data. Phone Data Conference: 12th International Conference on Advances in Mobile Computing and Multimedia (MoMM'14), arXiv preprint arXiv:1412.2595.

Durkheim, E. (1899) Morphologie sociale. Annee Sociologique, Vol. 2, pp. 520-521 as translated in The Rules of Sociological Method by Emile Durkheim, translated by W. D. Halls.

Huang, R., Huang, C., Shan, J., Xiong, L., and Yan, J. (2013) Evaluation of GPS trajectories on VGI and social websites. In 2013 21st International Conference on Geoinformatics, pp. 1-5, IEEE.

Karlsson, C. and Olsson, M. (2006) The identification of functional regions: theory, methods, and applications. In The Annals of Regional Science, Vol. 40: 1, pp. 1-18.

Klapka, P, M. Halas, and Tonev, P.(2013) Functional regions: concept and types. Proc. 16th International Colloquium on Regional Sciences, Brno, Masaryk University, 94–101.

Koster, H.R.A. and Rouwendal, J. (2010) The Impact of Mixed Land Use on Residential Property Values. TI 2010-105/3 Tinbergen Institute Discussion Paper.

Liao, L., Fox, D., and Kautz, H. (2007) Extracting places and activities from GPS traces using hierarchical conditional random fields. The International Journal of Robotics Research, 26(1), 119-134.

Manning, C. D., Raghavan, P., and Schütze, H. (2008) Introduction to information retrieval (Vol. 1, No. 1, p. 496). Cambridge: Cambridge university press.

Martinotti, G. (1994) The new social morphology of cities. UNESCO/MOST Wien, 10-12 February.

Peterson, M. (2002) Performing the “People's Palace” Musical Performance and the Production of Space at the Chicago Cultural Center. In Space and Culture, Vol 5(3), pp. 253-264.

Rodrigues, A., Damásio, C. and Cunha, J.E. (2014) Using GPS Logs to Identify Agronomical Activities, Connecting a Digital Europe Through Location and Place. Lecture Notes in Geoinformation and Cartography 2014: 105-121.

Roth, C., Kang, S. M., Batty, M., and Barthélemy, M. (2011) Structure of urban movements: polycentric activity and entangled hierarchical flows. PloS one, 6(1), e15923.

Sassen, S. (2001) The global city: New York, London, Tokyo (2nd ed.). Princeton: Princeton University Press.

Schönfelder, S, and Axhausen, K.W. (2010) Urban rhythms and travel behaviour: spatial and temporal phenomena of daily travel. Ashgate Publishing, Ltd.

Sen, A. and Thakuriah, P. (1995) Estimation of static travel times in a dynamic route guidance system. Mathematical and Computer Modelling, 22(4), 83-101.

Sila-Nowicka, K., Vandrol, J., Oshan, T., Long, J. A., Demšar, U., and Fotheringham, A. S. (2015) Analysis of human mobility patterns from GPS trajectories and contextual information. International Journal of Geographical Information Science, 1-26.

Sun, Y., Fan, H., Li, M., and Zipf, A. (2015) Identifying the city center using human travel flows generated from location-based social networking data. Environment and Planning B: Planning and Design, 43(3), 480-498.

Taubenböck, H., Klotz, M., Wurm, M., Schmieder, J., Wagner, B., Wooster, M. and Dech, S. (2013) Delineation of central business districts in mega city regions using remotely sensed data. Remote sensing of Environment, 136, 386-401.

Thakuriah, P., and Geers, D. G. (2013) Transportation and Information: Trends in Technology and Policy. Springer, NY.

Thakuriah, P., Tilahun, N. and Zellner, M. (forthcoming 2016) Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery. In Seeing Cities through Big Data: Research, Methods and Applications in Urban Informatics, to be published by Springer, NY.

Thierstein, A. and Luthi, S. (2012) Interlocking firm networks in the German knowledge economy: the case of the emerging mega-city region of Munich. In Marina van Geenhuizen and Peter Nijkamp (ed) Creative Knowledge Cities: Myths, Visions and Realities. New Horizons in Regional Science series, Edward Elgar Publishing.

Thurstain-Goodwin, M., and Unwin, D. (2000) Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. Transactions in GIS, 4(4), 305-317.

Umair, M., Kim, W.S., Choi, B.C. and Jung, S.Y. (2014) Discovering personal places from location traces. Proceedings of the 16th International Conference on Advanced Communication Technology (ICACT 2014), 709-713.

Whyte, W. H. (1980). The Social Life of Small Urban Spaces. Project for Public Spaces Inc.

Woodcraft, S. and Bacon, N. (2013) Introduction to The social life of cities: Stories about urban innovation. Social Life in partnership with the Urban Times. Accessed from http://www.social-life.co/blog/post/Social_life_of_cities_stories_urban_innovation/ on June 20, 2016.

Wu, C., and Murray, A. T. (2003) Estimating impervious surface distribution by spectral mixture analysis. Remote sensing of Environment, 84(4), 493-505.

Ye, Y., Zheng, Y., Chen, Y. (2009) Mining individual life pattern based on location history. In: Mobile data management: systems, services and middleware, IEEE MDM'09, Taipei, Taiwan, 18–20 May 2009, pp 1–10.

Zhang, S. and Sun, H. (2014) A Thought about Formation of Economic Function Area Based on Competition. In Proc. ICEM 2014: 16th International Conference on Economics and Management, pp. 428-436.

Zhong, C., Arisona, S.M., Huang, X., Batty, M. and Schmitt, G. (2014) Detecting the dynamics of urban structure through spatial network analysis. International Journal of Geographical Information Science, 28(11), 2178-2199.

Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., and Terveen, L. (2007) Discovering personally meaningful places: An interactive clustering approach. ACM Transactions on Information Systems (TOIS), 25(3), 12.