Beijing City Lab

Liu Y, Liu X, Gao S, Gong L, Kang C, Zhi Y, Chi G, Shi L, 2014, Social sensing: A new approach to understanding our socio-economic environments. Beijing City Lab. Working paper #47

Social sensing: A new approach to understanding our socio-economic environments

Yu Liu[1,2,3], Xi Liu[1], Song Gao[4], Li Gong[1], Chaogui Kang[1], Ye Zhi[1], Guanghua Chi[1], Li Shi[1]
[1] Institute of Remote Sensing and Geographical Information Systems, Peking University
[2] Beijing Key Lab of Spatial Information Integration and Its Applications, Peking University
[3] Shenzhen Key Laboratory of Urban Planning and Decision Making, Harbin Institute of Technology Shenzhen Graduate School
[4] Department of Geography, University of California, Santa Barbara

Abstract: The emergence of big data brings new opportunities for us to understand our socio-economic environments. We coin the term "social sensing" for such individual-level big geospatial data and the associated analysis methods. The word "sensing" suggests two natures of the data. First, they can be viewed the analogue and complement of remote sensing, since big data well capture socio-economic features for which the conventional remote sensing data do not work well. Second, in social sensing data, each individual plays the role of a sensor. This article connects social sensing with remote sensing and points out the major issues when applying social sensing data and associated analytics. We also suggest that social sensing data contain rich information about spatial interactions and place semantics, which go beyond the scope of traditional remote sensing data. In the coming big data era, GIScientists should investigate theories in using social sensing data, such as data representativeness and quality, and develop new tools to deal with social sensing data.

Keywords: social sensing, temporal activity pattern, spatial interaction, place semantics, GIScience

Introduction

The last five decades have witnessed the fast development of remote sensing techniques, of which a major objective is to reveal the physical characteristics of the Earth's surface, such as land cover features. Conventional land cover classification methods take spectral and textual properties as the major evidence (Gong and Howarth 1990). However, uncovering land uses from only remotely sensed imagery is rather difficult, since socio-economic features are not directly related to the spectral reflectance that can be detected by various sensors (Wu et al. 2009). Much literature introduces auxiliary information and domain knowledge for inferring land use (or social function) schemes (Liu, Guo, and Kelly 2008; Platt and Rapoza 2008; Wu et al. 2009; Meng et al. 2012; Hu and Wang 2013). However, these methods do not always yield ideal results. Although remote sensing data can to a certain extent capture urban/suburban landscape and infrastructure (such as buildings and street network) (Jensen and Cowen 1999), remote sensors have limited capability to extract socio-economic attributes and human dynamics such as movements and daily activities.

Recently, with the rapid development of information and communications technology (ICT), the impacts of ubiquitous big data on geography have been widely recognized (Graham and Shelton 2013), although there is not a clear and widely-accepted definition of big geospatial data (Batty 2013b). Several types of geospatial big data are available to capture the spatio-temporal patterns of human activities and thus provide an alternative approach to uncovering land uses and exploring how cities function in a fine temporal resolution. Such big data include taxi trajectories, mobile phone records, social media or social networking data[1], smart card records in public transportation systems, and so on (Lu and Liu 2012). Much research has been conducted to obtain land use characteristics using mobile phone data (Ratti et al. 2006, Toole et al. 2012, Pei et al. forthcoming), taxi data (Qi et al. 2011; Liu et al. 2012b), and smart card data (Gong et al. 2012). The primary assertion of such studies is that different land uses are associated with different temporal rhythms of activities (Sevtsuk and Ratti 2010).

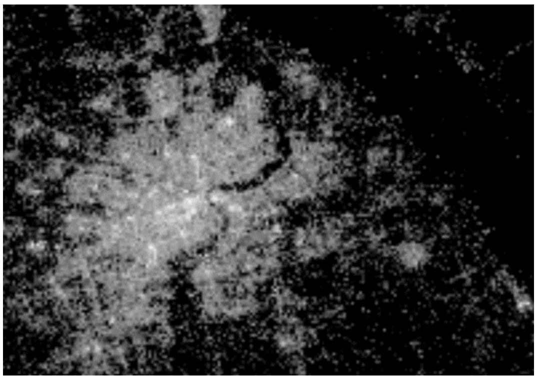
Considering that remote sensing data have been widely and successfully used to map physical features of our world, in this article, we introduce the term “social sensing” for the abovementioned geospatial big data, since such data have some features in common with the conventional remote sensing data and reveal socio-economic characteristics as a complement to remote sensing data. We use social sensing to emphasize that geospatial big data can be viewed as an analogue of remote sensing data in social science research. By adopting the methods developed in remote sensing applications, social sensing provides a promising approach to understanding our socio-economic environments, alone or integrated with remote sensing data. Additionally, social sensing data in general contain rich information, such as spatial interaction and place semantics, which go beyond the scope of traditional remote sensing data. This article summarizes the properties of social sensing data and puts forward a research agenda for applying them in geographical analyses.

Social Sensing

In this section, we introduce two data sets to demonstrate the concept of social sensing. One data set is composed of taxi trajectories and another is of social media check-in records, both of which were collected in Shanghai, China. The taxi trajectory data set covers 7 days in 2009 and includes pick-up points (PUPs) and drop-off points (DOPs) (a description of this data set can be found in Liu et al. 2012a, b). The check-in data contain about 100,000 check-in records collected over the course of one year. A check-in record is a spatio-temporally tagged text message posted by a user using a mobile device. From the data set, we can extract difference places such as workplaces, hotels, parks, and restaurants (see Wu et al. 2014 for details about this data set). The study area (Figure 1A) is rasterized into 28,000 (140 rows and 200 columns) 250*250 m^2 squares so that we can count the numbers of PUPs, POPs, and check-ins in each pixel.

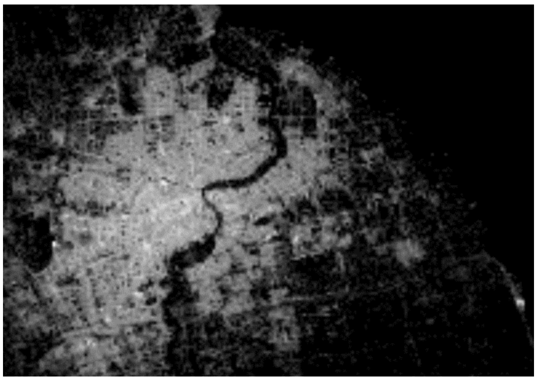
Figures 1B, 1C, and 1D depict the spatial distributions of the three activities, i.e., checking in, picking up, and dropping off. It is natural that the three distributions are positively correlated with the population distribution (Figure 1E, represented by the Landscan™ data with a spatial resolution of $1km^2$). The frequency distributions of the three activities have a heavy-tail characteristic (Figure 1F).

[1] From their original meanings, social media and social networking are distinct. In general, social media services (e.g., Twitter) focus on sharing information but social networking services (e.g., Facebook) pay much attention to connecting with others. However, both of them provide similar functions such as posting contents with locational information and maintaining relationships between users. In this article, we simply use social media for these services.



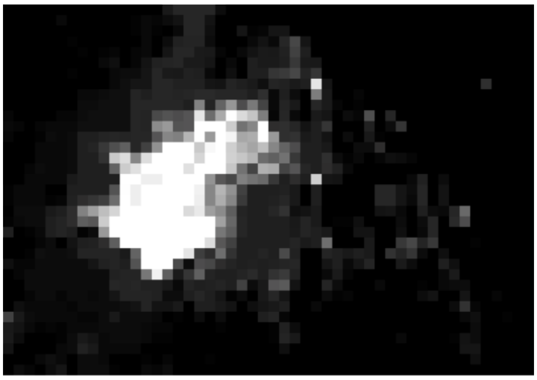
(A)

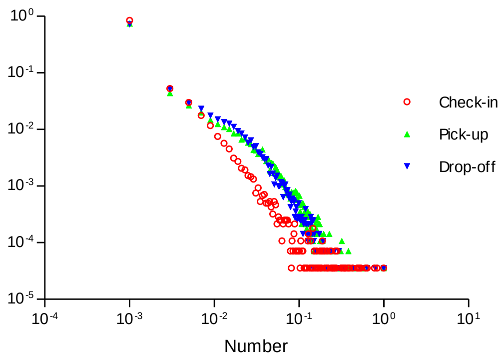
(B)



(C)

(D)





(E)

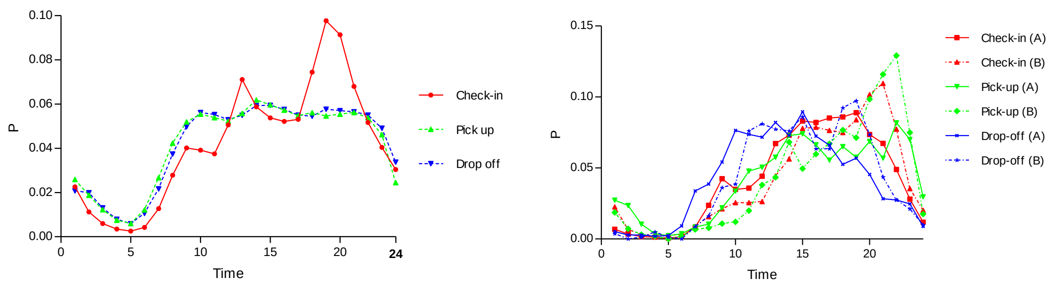
(F)

Figure 1. (A) Urban area of Shanghai; (B) Spatial distribution of check-in points in about one year; (C) Spatial distribution of pick-up points in seven days; (D) Spatial distribution of drop-off points in seven days; (E) Population density represented using Landscan data; (F) Frequency distributions of the three activities. Note that subfigures B, C, and D are obtained using logarithmic transformations.

Much literature has paid attention to the temporal activity rhythms extracted from different data sources

(Sevtsuk and Ratti 2010; Liu et al. 2012b; Kang et al. 2012b; Toole et al. 2012; Shen et al. 2013; Pei et al. forthcoming). Since human activities have a clear daily periodicity, we can aggregate the actual data set covering a long period by computing the number of activities recorded for each hour of the day. Figure 2A depicts the averaged and normalized diurnal variations of the three activities across the entire study area. The check-in curve has two clear peaks, corresponding to 1pm and 7pm. It is natural that the check-in probability is high during non-working hours, especially when people have lunch or dinner. Although the curves representing pick-up, drop-off, and check-in behavior diverge during the day, the three curves exhibit a similar trend from 10pm to 9am of the next day. Note that the temporal resolution for computing activity frequencies has been set as 1 hour. A different temporal resolution, e.g. 0.5 hours, will yield different curves but the basic trend does not change much. Hence, in the following sections, we use 1 hour as the default time interval.

Compared with the global temporal patterns, given a data source, we are more interested in local temporal patterns since different places (cf. the two triangles in Figure 1A) are associated with different temporal signatures (Figure 2B). Furthermore, these varying temporal signatures are themselves dependent on the underlying land use features. Hence, a number of studies have focused on classifying land use features from taxi data (Liu et al. 2012b) and mobile phone data (Toole et al. 2012; Pei et al. forthcoming). The basic idea of such classification research is that the temporal curves, as shown in Figure 2B, can be viewed as the signatures of various land uses. This reminds us of the foundation of photometer remote sensing image classification, which generally extracts land cover information according to the electromagnetic spectral curves of different features such as forest, water, and barren land. Conventional remotely sensed data have been successfully applied in revealing physical geographical characteristics. On the contrary, current geospatial big data well capture human activities and are thus more sensitive to our socio-economic environments. We thus coin the term "social sensing" for various spatio-temporally tagged data sources and the associated analysis methods.



(A)
(B)

Figure 2. Normalized temporal variations of different activities extracted from taxi trajectories and social media check-in records. (A) Global patterns; (B) Local patterns associated with two places.

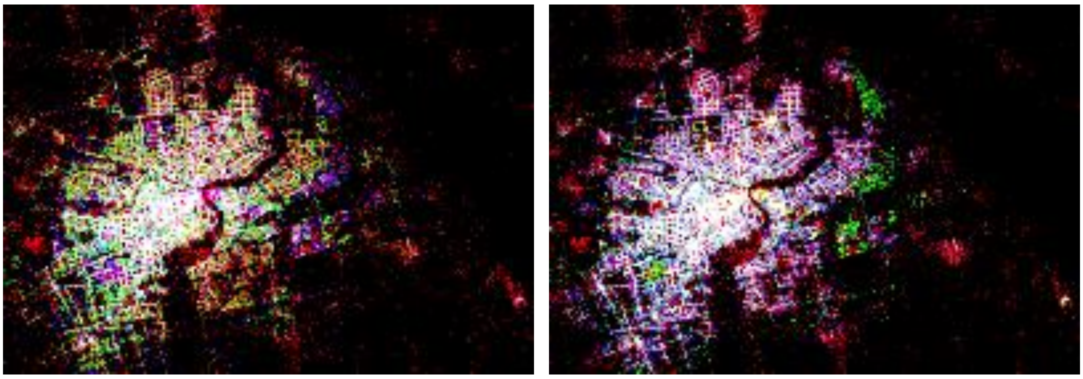
Linking social sensing with remote sensing

The temporal variations depicted in Figure 2 suggest that different places exhibit different responses to a certain activity. Meanwhile, given a place, the temporal variations of different activities captured by various social sensing data are also different. Hence, for each social sensing data, by setting the spatial and temporal resolutions, we can obtain a series of images, which are similar to different bands of remote sensing imagery. In this sense, different social sensing data can be viewed as the analogues of different remote sensing data. We compare social sensing data with remote sensing data in Table 1. It is clear that these two data sources share some common characteristics, such as containing multi-sensor, multi-resolution, multi-temporal information, but capture different aspects of a geographical environment.

Table 1. Comparisons between remote sensing and social sensing

	Remote Sensing	Social sensing
Data source	Remote sensed data collected from various sensors, such as radiometer, radar, and LiDAR	Spatio-temporally tagged data collected from different location aware devices, such as mobile phone, GPS
Major objective	Physical features of Earth surface	Socio-economic features of Earth surface
Processing method	Correction and calibration, fusion, classification, etc.	Geocoding and preprocessing, fusion, classification, etc.
Signal for Classification	Electromagnetic spectra	Temporal variation of activities

The similarities between remote sensing data and social sensing data suggest that we can introduce conventional image processing methods to analyze geospatial big data. For example, we can view the activity density maps as different bands of images and generate false color composite images. The two subfigures in Figure 3 correspond to 8:00am-9:00am and 8:00pm-9:00pm. The composition scheme is that check-ins, pick-ups, and drop-offs are represented by red, green, and blue channels respectively. From the false color images, the spatio-temporal characteristics of the three activities can be clearly identified. The region color in white is the core area of Shanghai, where the densities of the three activities are all high. For the two time intervals, the red dots in the suburban areas indicate that the activity density of check-ins is higher than the other two activities. In the first image, the green zones correspond to residential areas, where people leave home in morning. These zones are purple or blue in evening time, indicating that both check-in and drop-off activities are high. Such synthesized images tell much land use information besides the above-mentioned patterns. Hence, we can use supervised or unsupervised classifiers to extract land uses, which we suggest are more reliable than those obtained from remotely sensed data since the source data directly capture human activities. A number of studies have been conducted in this vein (Liu et al. 2012b; Toole et al. 2012, Pei et al. forthcoming), but most of these studies focus on single-source data and do not take into account the fusion of multi-source data.



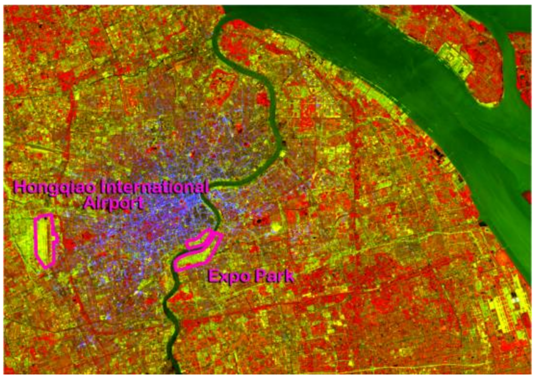
(A) (B)

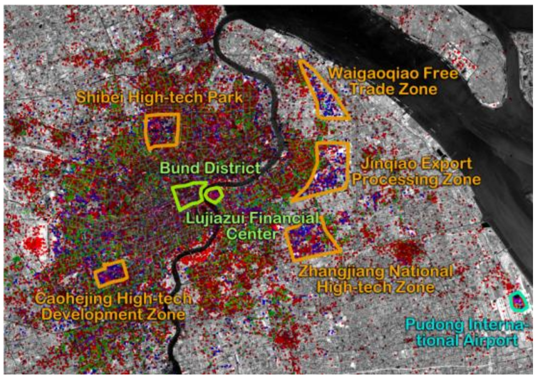
Figure 3 False color composition of three social sensing data source during different times. (A)8:00am-9:00am,

(B) 8:00pm-9:00pm.

Figure 3 demonstrates the similarity between social sensing and remote sensing. It suggests that well-developed remote sensing techniques can be applied in processing social sensing data and these two data sources can be integrated to gain a complete picture of geographical environments. For the first aspect, in addition to aforementioned classification studies, other remote sensing methods such as calibration and enhancement, feature selection, data fusion, and image segmentation have the potential to apply to social sensing data. For example, principal component analysis (or the Karhunen-Loève transform) has been used for finding the major components, which depict different land use aspects of the study area (Reades et al. 2009; Sun et al. 2011). Besides directly providing methods, remote sensing is also a source to enlighten us to conduct similar studies. For example, numerous indices such as NDVI (normalized difference vegetation index) have been proposed. Accordingly, we can estimate happiness index (Mitchell et al. 2013) and demographical properties (Li, Goodchild, and Xu 2013) from social sensing data. Here we only list a few examples and believe that more analytical methods will be developed for social sensing data.

Second, social sensing helps to solve the problem of "inferring land uses from land cover characteristics" in remote sensing applications. We can get land cover information according to the spectral characteristics from remote sensing data and human activities and movements from social sensing data. Information extracted from the two different data sources can validate each other to yield more precise results. With regard to integrating social sensing data with remote sensing data, the conventional approach for fusing remote sensing data and non-telemetric data is taking non-remote sensing data as a band of remote sensing image. Figure 4A is a false color composite image with the combination of ETM4, ETM3, and drop-offs from 8:00 to 9:00. The regions in red are covered by plants and the purple built-up areas are with high activity densities. It is interesting that the yellow zones are in general built-up areas with low activity densities denoted by drop-offs, and thus highlight certain regions with special functions, such as Hongqiao Airport and Expo Park. When we analyze social sensing data, the spatial resolutions are generally ranging from tens of meters to hundreds of meters, which are relatively lower than high-spatial resolution remote sensing data. Hence, we can merge high-resolution panchromatic imagery with social sensing data to increase their spatial resolution. An illustration is visualized in Figure 4B. We get a new RGB image based on Figure 3A with the spatial resolution of 90m, and use color transformation method to fusing the Landsat 7 panchromatic image with the RGB imagery. From the fused imagery, we can clearly find some regions with high densities of check-ins and drop-offs, whose colors are mixture of red, blue and purple. They are generally big development zones, commercial areas, and public transportation facilities.





(A)

(B)

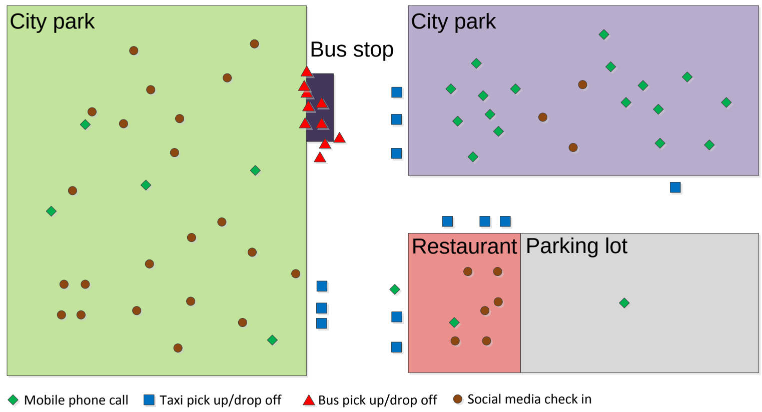
Figure 4. (A) False color composite imagery of 2010 ETM data of Landsat 7 and taxi drop-off data, where band 4, band 3, and drop-offs are represented by red, green, and blue channels, respectively. (B) Fused imagery by color transformation method: the low-resolution data is the RGB image composited of check-ins, pick-ups and drop-offs from 8:00 to 9:00 in the morning, and the high-resolution data is Landsat 7 panchromatic image.

Issues in using social sensing data

Conventional remote sensing data suffer from the limitation of capturing human factors when used in socio-economic applications. Such a limitation can be compensated for by integrating remote sensing data with social sensing data. However, a number of issues should be paid attention to when using social sensing data.

First, to compute the temporal variations of activities, the study area should be discretized into regular or irregular units. Although some research based on mobile phone records directly deals with Voronoi polygons generated from base towers, many existing studies use regular rasterization with a coarse spatial resolution, for example 0.25km^2 (Reades et al. 2009) and 1km^2 (Sun et al. 2011; Liu et al. 2012b; Toole et al. 2012), to investigate land uses. The resolutions are much lower than those of widely-used remote sensing data. Unfortunately, increasing the spatial resolution will bring difficulties in analyzing social sensing data. As an extreme case, Figure 5A depicts a high spatial resolution RS image of Pudong Airport and vicinity in Shanghai, China. A 1km^2 square is drawn using thick red lines, and the size of each small square is 250*250m^2. It is natural that human activities should be linked with buildings, the size of which is close to 1km^2. When using taxi data to measure activities in this area, however, both PUPs and DOPs concentrate in a relatively small area. If adopting a spatial resolution of 250m, a discontinuous result will be obtained. For pixels corresponding to the building, some contain high activity frequencies but others are blank. Figure 5B illustrates a fictitious urban environment, where four activities, including mobile phone calls, taxi pick-ups/drop-offs, bus pick-ups/drop-offs, and social media check-ins, exhibit greatly different distribution patterns. For example, both pick-ups and drop-offs can only occur in streets. Additionally, it is natural that bus pick-ups and drop-offs are more concentrated. On the contrary, mobile phone calls frequently occur within work places, and social media check-ins often occur at entertainment and dining establishments. In sum, the local spatial heterogeneity of different activities leads to sharp distribution gradients and makes the regularities unclear if a high spatial resolution is adopted. Hence, we should choose a coarser resolution to smoothen the activity distributions.

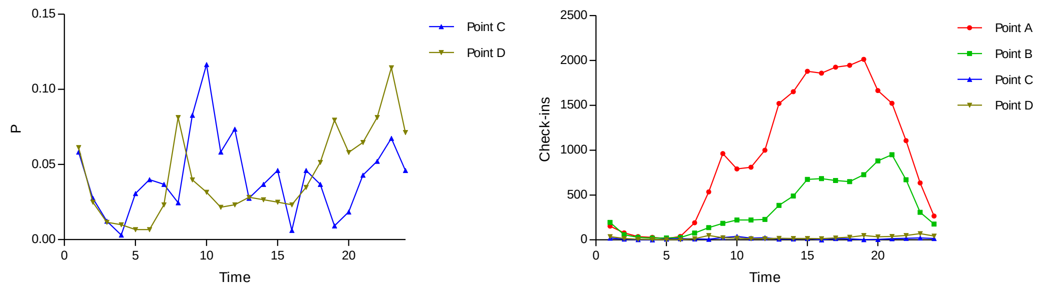
(A)



(B)

Figure 5. Spatial resolution issue of social sensing data. (A) A case of Pudong Airport, Shanghai. (B) Local spatial heterogeneity of activities captured by different social sensing data.

Second, the spatial distributions of most activities extracted from social sensing data are positively correlated with population density (Liu et al. 2012a; Kang et al. 2012b). As shown in Figure 1F, the spatial distributions of human activity frequencies suggest that most activities are concentrated in a small area, which roughly corresponds to the downtown of Shanghai. Hence, the data in Figs 1B, 1C, 1D, 3A, 3B, and 3C are all obtained using a logarithmic transform for better visualization. In urban fringe areas or rural areas, however, the activity density is relatively low and thus leads to the problem of small numbers. In such areas, the temporal variations are rather random, and we cannot find a representative pattern after normalization. Figure 6A depicts the normalized temporal variations of check-ins inside two pixels in suburban areas (cf. points C and D in Figure 1A). The two curves are quite different, so the two points are likely to be categorized into two land uses. However, when compared with pixels in urban areas, the absolute numbers of check-ins are very small, meaning that the two curves are flattened and it is difficult to find meaningful temporal patterns (Figure 6B). Figure 6 demonstrates that both the relative and absolute temporal patterns can be used to reveal the underlying land uses. Pei et al. (forthcoming) provide a good example of combining the two types of temporal variations. As a general rule, the total activity numbers reflect land use intensities and the temporal patterns depend on land use categories. Figure 6 also suggests that in rural areas, where human activity density is relatively low, social sensing data are not an effective measure of human activity. Last, considering such a spatial distribution, a varying resolution tessellation scheme may be more reasonable, although most existing studies have adopted regular rasterization. Future studies may adopt a fine resolution scheme in urban areas and a coarse one in suburban or rural areas.



(A)

(B)

Figure 6. (A) Normalized temporal variations of check-ins inside two pixels in suburban areas. (B) Absolute temporal variations check-ins inside four pixels, two in urban areas and two in suburban areas.

Third, social sensing data may also encounter temporal issues. Current land use classification research is founded on the idea that land parcels classified in the same use category exhibit similar diurnal patterns of activities. Additionally, given a place, it is assumed that the temporal curves in different days are also similar and exhibit high regularity. Hence, most studies compress the data spanning a long period into a 24-hour curve (cf. Figure 2). The assumption generally holds true. Slight differences, however, can still be found from day to day. Figure 7 plots the global temporal curves of Shanghai pick-ups and drop-offs in seven days. Some outliers exist although the periodicity is rather clear. For example, there are more taxi trips on Saturday and people begin their trips a bit late on Sunday. These two tendencies make sense given the common weekend habits of many people. On Thursday, the number of trips rapidly declines in the morning after 9:00am; while we have not found a reasonable explanation for this anomaly but suspect it may be caused by a special event. Outliers can be identified from global patterns, not to mention local patterns, which are less stable. The day-to-day changes in trip volume can be attributed to two aspects: special events and long-term dynamics. It is natural that in different seasons, the activity rhythms are different. Additionally, urban evolution, which includes sprawl and land use transition, also influences local temporal patterns. Averaging diurnal patterns can filter noise in the dataset but ultimately fails to capture these two dynamics. For many geographical applications, the latter aspect is more important: we need to decouple short-term variations and long-term variations in social sensing data. The data sets used in most current research cover a short period such as one month, and thus the problem is not serious. With the accumulation of various big data, we can reveal regional evolution in addition to land use distributions.

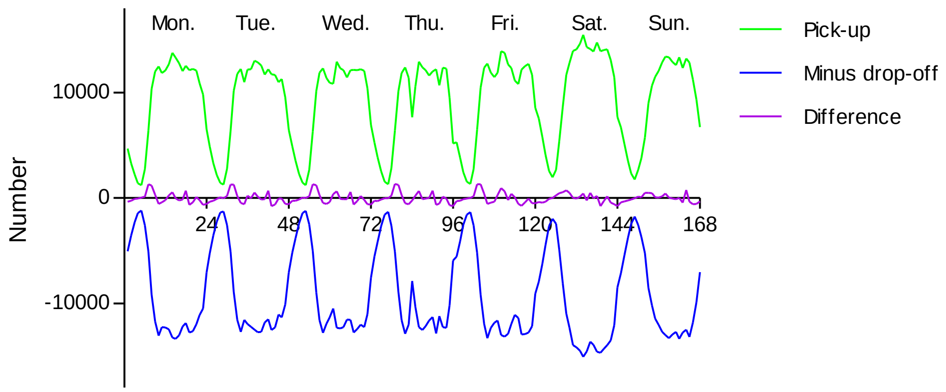


Figure 7. Global temporal variations of pick-ups and drop-offs in seven days. Since the two curves of pick-ups and drop-offs are quite similar, we use negative drop-off volumes to differentiate them.

Fourth, most existing temporal pattern studies are conducted based on the data collected in a single city such as Rome (Reades et al. 2009), Shanghai (Liu et al. 2012b), or Boston (Toole et al. 2012). Little attention has been paid to intercity comparisons. Can we set up a uniform temporal signature data base that stores the "standard" temporal curves associated with different land uses, just as we have done for remote sensing data processing? The answer is unfortunately negative. Given a city, the overall rhythm depends on its social, cultural, and economic features. The temporal signature of a particular activity is constrained by the overall rhythm. As shown in Figure 8, the diurnal activity variations of various cities are not identical, although some common patterns can be found. For example, the curves of most cities have two clear "peaks", which corresponding to 12:00pm-1:00pm and 6:00pm-7:00pm. Such a pattern has also been reported by Cheng et al. (2011). Given a land use category, the difference between two cities may be more significant than those between different land uses inside one city. This makes it difficult to extract universal classification role across cities. For this reason, unsupervised classification methods are widely preferred over supervised classifiers for social sensing data. Another difficulty for supervised classification is that delineating training areas from activity distribution maps such as Figures 1 and 2 is difficult due to the lack of "standard" temporal signatures. The differences between cities' temporal signature suggest that we should rely upon spatial distributions of activities instead of temporal variations when the study area is expanded to a region containing multiple cities. A recent good study was reported by Li, Goodchild, and Xu (2013). They introduced Twitter and Flickr data to investigate socioeconomic features in California.

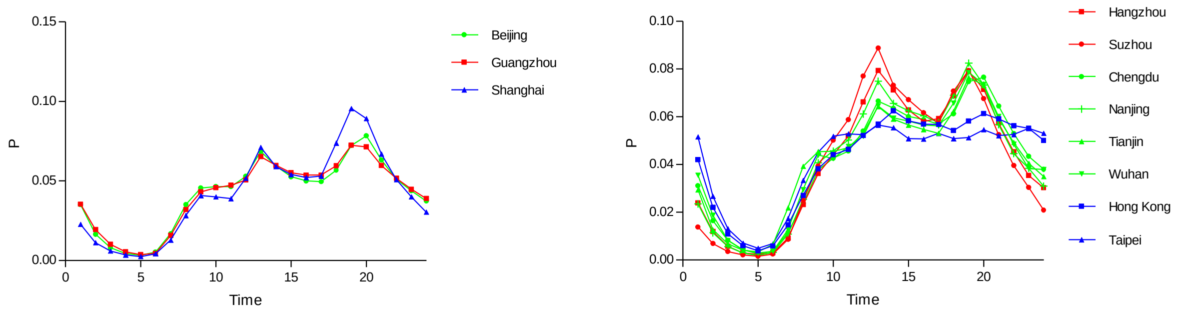


Figure 8. Diurnal variations of check-in activities in 11 top cities in China. The nine cities in mainland China exhibit similar patterns. Slight differences can still be found. For example, the evening check-in probability in Shanghai is high, indicating more night life activities. In Hangzhou and Suzhou, there are maximum check-ins activities during noon time. It is interesting and reasonable that the diurnal variations Hong Kong and Taipei are similar but different from those of mainland cities.

Fifth, people's actual activities cannot directly be acquired from most kinds of social sensing data. The need to participate in activities generates travel demands (Axhausen and Gärling 1992; Kitamura 1988), and thus detailed activity information is very important in studying human travel behaviors, traffic engineering, and urban planning. The observed activities obtained from social sensing data are "proxy activities" such as checking in on social media websites, making phone calls, or boarding a taxi. Thus, compared to conventional travel survey data, social sensing data contain much less information about people's actual activity types such as entertainment. The observed spatio-temporal patterns are composed of patterns of different actual activities, such as in-home activities, work-related activities, and entertainment. Clearly, different activities exhibit different patterns (Wu et al. 2014). Decomposing existing social sensing data into actual activities can significantly improve our understanding of human mobility and consequently the underlying socio-economic environments. Although previous studies have devised methods to infer individuals' actual activities from trajectories based on POI data (Alvares et al. 2007; Huang, Li, and Yue 2010; Phithakkitnukoon et al. 2010; Zhang, Li, and Pan 2012; Furletti et al. 2013; Schaller, Harvey, and Elsweiler 2014), uncertainty problems exist and make the related efforts challenging. On the one hand, it is difficult to ensure the destinations according to the locations where proxy activities occur, as there are generally many points of interest nearby. On the other hand, the same POI may be associated with different activities. For example, some people go to shopping malls to buy clothes and other goods, but others might want to have meals or meet friends.

Additionally, the mapping from actual activities to proxy activities is rather complex. Given a spatial unit and a time interval, suppose there are N persons with M different activities. The numbers of the M activities are denoted by A_1, A_2,…,A_m. For a proxy activity denoted by j (e.g., a mobile phone call), the occurrence number is $D_j=A_1P_{1j}+ A_2P_{2j}+…+A_mP_{mj}$, where P_{mj} denotes the probability of proxy activity j during actual activity m. P_{mj} heavily depends on the actual activity m. For example, people may make many phone calls but seldom check in during work time. Figure 9 plots the temporal curves of different activities extracted from the check-in data. People are more likely to check in during dining and entertainment, and the numbers of check-ins during other activities such as work are considerably small. Researchers have also suggested that the temporal signature characteristics of check-in data can help differentiate the POI types (Ye et al., 2011). The above equation does not take into account population heterogeneity. Even with the same actual activity, the likelihood that different individuals will complete the same proxy activity varies widely. For example, young people are more likely to check in on social media. These issues remind us to pay close attention to the representativeness of social sensing data. Different social sensing data capture different aspects of the ground truth, just as in the parable of the blind men and the elephant. We suggest that integrating multi-source data, including survey data, can lead to a better understanding of actual activity patterns.

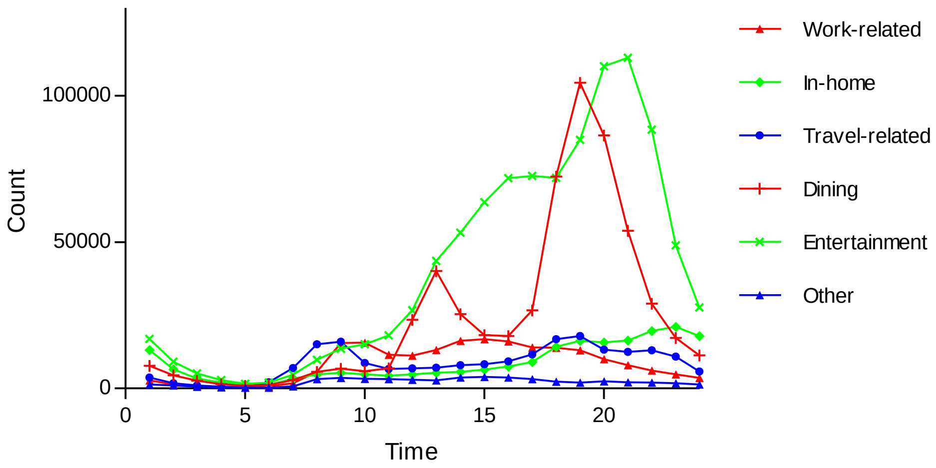


Figure 9. Temporal variations of activities extracted from check-in data in Shanghai. The check-in data explicitly record the type (e.g. restaurant, shopping mall) of place where a user checks in so that we can early infer the activity information.

Beyond capturing activities

Besides complementing remote sensing data from the temporal activity variation perspective, social sensing data can be used to extract movements, social ties, and spatial cognition (majorly from social media data) of individuals. At the collective level, social sensing data provide an approach to revealing spatial interactions and place semantics.

Sensing spatial interactions

Large volumes of spatio-temporally tagged social sensing data lead to the upsurge of human mobility research (Lu and Liu 2012; Yue et al. 2014). Different patterns have been identified from various data sources and numerous models were constructed to interpret the observed patterns (e.g., Brockmann, Hufnagel, and Geisel 2006; González, Hidalgo, and Barabási 2008; Jiang, Yin, and Zhao 2009; Liu et al. 2012a, Noulas et al. 2012). It is accepted that human mobility patterns are influenced by factors including distance decay effect, spatial heterogeneity, and population heterogeneity. At the collective level, we can aggregate individuals' or vehicles' trajectories to obtain traffic flows between places. Besides movement, when all individuals have been geo-referenced, their connections like mobile phone calls and social ties can be summed up to measure spatial interactions from a new perspective. For example, we can measure the interaction strength between two cities using the number of follower and followee pairs extracted from a social network site. Hence, ICT (information and communications technology) generated geospatial big data also have the capacity to capture spatial interactions. Such a property is different from conventional remote sensing data, in which connections between pixels are not represented. Figure 10 depicts the top 25 trip flows originating from point A and Hongqiao Airport computed using the taxi data. The flow volumes well represent the spatial interactions between places, which are 250*250 m^2 pixels.

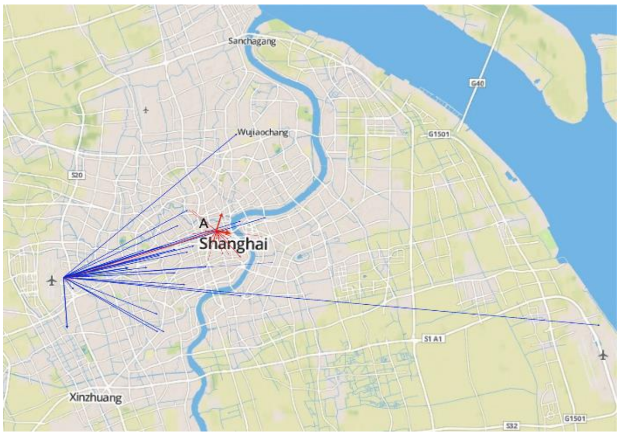
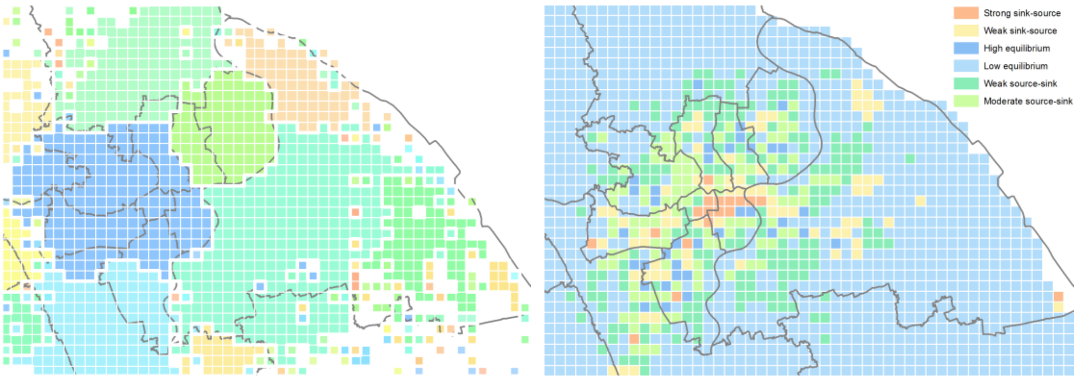


Figure 10. Spatial interactions extracted from taxi trajectories in Shanghai. The top 25 trip flows originating from two points, one is point A in Figure 1A within downtown and the other is Hongqiao Airport.

There is a long tradition of research on spatial interaction in geography. Additionally, data sources such as taxi data have been used to measure spatial interactions as early as 1970 (Goddard 1970). Large volumes of social sensing data obviously produce new opportunities for this topic. Studies aggregating individual movements to analyze regional structure from the collective level have also been boosted. Recent literature can be categorized into two spatial scales: inter-city scale (Thiemann et al. 2010; Peng et al. 2012; Kang et al. 2013; Liu et al. 2014) and intra-city scale (Roth et al. 2011; Gao et al. 2013). At the inter-city scale, the interactions between land parcels are influenced by the urban geographical environment. People travel in a city from place to place for certain objectives, suggesting that both the relationship between locations and their land uses can be revealed from spatial interactions and temporal activity variations. At the intra-city scale, interactions extracted from big data help us to uncover regional structures.

Given a data set, if we partition the study region in to areal units, a spatial interaction network can be formed (Batty 2013a). Within this network, areal units can be treated as nodes, while interactions between units are denoted by weighted edges. A number of network science methods including centrality computation and community detection have been developed and introduced to analyze spatial interaction networks. One of these methods, community detection, can identify meaningful sub-networks (sub-regions for spatial networks) with relatively dense connections. For regional or national scaled data, community detection studies have found that sub-regions that are consistent with administrative boundaries (Ratti et al. 2010; Thiemann et al. 2010; Montis, Caschili, and Chessa 2013; Liu et al. 2014). Community detection methods can also be used to detect highly interactive sub-regions of a city. Figure 11A illustrates the community detection results of the network based on discretized 1km^2 grids and the taxi trips flows between them. Most communities are spatially connected, indicating the cohesiveness of each zone with strong internal linkages. Additionally, the spatial continuity of sub-networks can be attributed to the distance decay effect (Liu et al. 2014), which exists in almost all spatial interactions.

We may compare the community detection results with the classification result (Figure 11B) using the methods mentioned in the second section. Both approaches divide the study area into sub-regions, but are conducted based on different measures: similarity and connection. The two measures capture different aspects of the relatedness between places and are thus widely used in regionalization. This case suggests that social sensing data provide more information than temporal variations. Let us revisit the discussion on land use at the beginning of this article. Spatial interactions can also help improve land use classifications. Except for temporal activity variations, land parcels of different land use types typically have different interaction patterns. Take residential land parcels, for example: they may have intense interactions with business land parcels in the morning given that people are moving to their work place. These types of spatial interactions may reduce misclassification, particularly for pixels with low activity densities or similar temporal patterns.



(A)

(B)

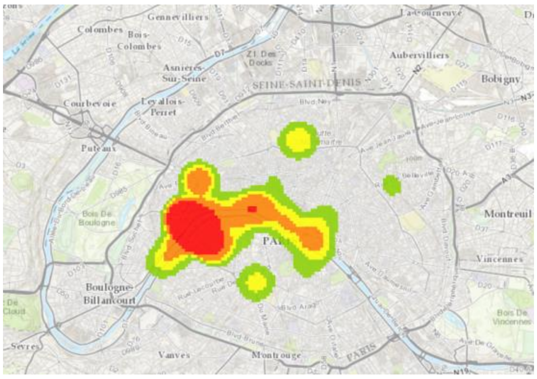
Figure 11. (A) Community detection result of network formed by all taxi flows. The sub-regions have strong internal interactions. (B) Classification result using temporally sequenced images of taxi pick-up and drop-off distributions (reproduced based on Liu et al. 2012b).

Sensing place semantics and sentiments

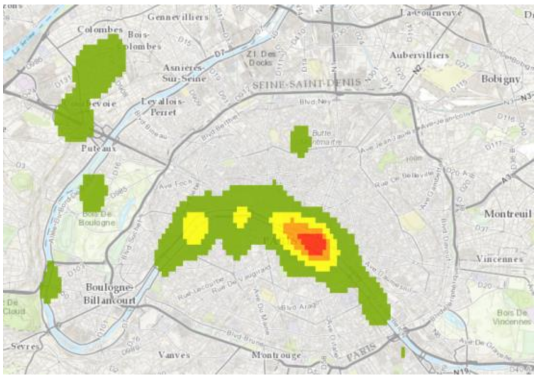
Traditionally, human geographers, anthropologists, sociologists, and urban planners have been studying on a variety of meanings of space and place to particular people (Tuan 1977, Hubbard et al. 2004). The concept “sense of place” indicates a unique identity that is deeply felt by local residents and outside tourists. With the widely use of web (documents, blogs, photos, videos), amount of user-generated social sensing data with geospatial components have been shared by millions of volunteers (Goodchild 2007). Such big data offer good opportunities for researchers to study how humans perceive, experience and describe the world and consequently to represent place semantics (Rattenbury and Naaman 2009). For instance, Crandall et al. (2009) analyzed 60 millions of geo-tagged Flickr photos to identify the top-ranked landmarks on the Earth and introduced integrated classification methods for prediction locations from visual, textual and temporal features of the photos. Adam and Mckenzie (2013) applied topic modeling on a lot of travel blogs to identify the thematic descriptions which are most associated with places around the world. The similar places can be derived based on the extracted thematic topics. A traditional challenge work of POI matching (or geospatial conflation) from different data sources can be facilitated by integrating multiple spatio-temporal-semantic attributes (McKenzie, Janowicz, and Adams 2014). Using Foursquare check-ins data, it is possible to explore the neighborhood dynamics (Cranshaw et al. 2012) and city-to-city similarity measures (Preoţiuc-Pietro, Cranshaw, and Yano 2013) analog to the use of remote sensing and landscape metrics to describe similar urban structures. Furthermore, analytics of geo-tagged tweets offer new insights on the geographical distribution of human sentimental expressions on places and the temporal changes compared with demographic and health characteristics (Dodds et al. 2011; Mitchell et al. 2013). Compared with capturing activities and spatial interactions where individuals play a passive sensor role, in sensing place semantics, each individual actively uploads and shares his (or her) environmental perceptions, which are also helpful to model both activities and social ties.

The collected geo-tagged photos provide new evidence on the social perspective that few predominant semantic tags have characterized a sense of city. Figure 12A visualizes the 200 most frequent tags extracted from about 384, 000 Flickr photos in Paris. These tags demonstrate how people perceive and annotate the characteristics of Paris, e.g., art, architecture, museum and vacation. Table 2 shows the top-20 frequent geo-tagged tags in Paris and they are grouped into geographical context, landmark names, place characteristics, urban functions and time based on their different semantic meanings. Figures 12B and 12C depict the kernel density estimation (KDE) of the geo-tagged photos associated with Eiffel Tower and Seine River. It is clear that the heat maps outline the two places' influence areas. Such an approach can help to understand the place semantics of cities, i.e., we can not only get the spatial footprints of places from geo-tagged photos, but also can derive human cognitive and hierarchal relationships between places which are captured in such social sensing data.

(A)



(B)



(C)

Figure 12. (A) A word-cloud visualization of 200 most frequent tags using Wordle tool. (B) The KDE of geo-tagged photos with Eiffel Tower. (C) The KDE of geo-tagged photos with Seine River. (The grid size is 100*100 m^2).

Table 2. The top-20 frequent tags extracted from geo-tagged Flickr photos in Paris

Groups	Tags (count)
Geographical context	France (15373), Europe (3909), Île-de-France (1862), city (1249), street (923), Disneyland (759)
Landmark names	Louvre (1174), Eiffel Tower (868), Montmartre (755), Seine (721), Notre dame (692)
Place characteristics	architecture (1486), art (1406), travel (1341), vacation (801), French (689)
Urban functions	museum (1253), concert (1097), church (795),
Time	night (864)

Goodchild (2011) discussed the idea of formalizing place in the digital world and addressed the relationship between the informal world of human discourse and the formal world of digitally represented geography. He argued that “perhaps a new field will emerge at this intersection between digital technology, social science, and digital data. If it does, the concept of place will clearly occupy a central position.” The proposed idea of social sensing might be a potential candidate.

Related concepts

At present, a number of similar but different big-data-related concepts have been proposed, such as volunteered geographical information (VGI, Goodchild 2007), crowdsourcing geographical information (Goodchild and Glennon 2010), and urban computing (Zheng et al. 2013). Aggarwal and Abdelzaher (2011, 2013) used the term social sensing for data collected from location aware devices, such as GPS (Global Positioning System)-enabled vehicles or individuals. Additionally, some scholars have coined alternative terms, such as people centric sensing (Campbell et al. 2008) and urban sensing (Lane et al. 2008), that have similar meanings. In this section, we would like to discuss these concepts to highlight the value of social sensing in the context of geographical studies.

Goodchild (2007) introduced the term VGI for geographical data “provided voluntarily by individuals” via Web 2.0 techniques. The term highlights the fact that citizens have become participants in web content contributions and play a role of sensor. In 2007, smartphones were not widely used and there was few location based apps. In the 2007 paper, the two example VGI applications are Wikimapia and OpenStreetMap, both of which are web-based. At present, seven years later, various mobile apps, such as apps for Twitter and Flickr, make individuals more convenient to contribute geographical information (Elwood, Goodchild, and Sui 2012; Li, Goodchild, and Xu 2013). When uploading VGI, an individual plays the role of an active sensor. However, for some social sensing data like mobile phone records, all individuals’ role is passive. VGI includes conventional spatial data such as street lines and points of interest (POI), which contain little human behavior information and are thus excluded from social sensing data. Compared with social sensing, VGI emphasizes more on a new data collection approach instead of mining socio-economic characteristics. With regard to crowdsourcing geographical information, we suggest that it is a subset of VGI and is always associated with particular objectives such as disaster response (Goodchild and Glennon 2010).

Urban computing is the umbrella term of Microsoft Research Asia for a series of studies that utilize data generated inside cities. According to Zheng et al. (2013), the data used in urban computing include geographical data, traffic data, mobile phone signals, commuting data, environmental monitoring data, social network data, and data about economy, energy and health care. The list obviously covers almost all possible data sets, including social sensing data, for studying urban problems. It does not focus on human behavior characteristics. As the term itself implies, urban computing pays more attention to various techniques such as data acquisition, data management, and service providing. A system implementing urban computing is actually a geographical information system for a certain city. On the contrary, social sensing roots in geography and thus supports intercity and regional studies (Ratti et al. 2010; Li, Goodchild, and Xu 2013; Liu et al. 2014), which are outside the scope of urban computing.

In the field of information technology (IT), social sensing refers to the integration of social and sensor networks (Aggarwal and Abdelzaher 2011, 2013). It pays much attention to hardware platforms for data collection techniques such as energy efficient design. From a geographical perspective, the concept social sensing extends its implications from the original IT implications to include data management, data analysts, and applications, in addition to data acquisition. The narrow sense social sensing is obviously the foundation of the broad sense social sensing due to its capacity of providing various data sources.

After comparing related concepts, we can draw some properties of social sensing. First, social sensing data compose an important sector of big data. They capture three aspects of individual level behavior characteristics: activity and movement, social tie, and emotion and perception. Hence, for a particular person, his (or her) detailed behaviors may be exposed from geospatial big data. It raises the privacy concern: “does big data mean big brother?” (Bernstein 2014). It is therefore important for researchers who use social sensing data to adopt appropriate protocols to ensure the protection of individual privacy in their studies. The three aspects affect each other (e.g. the relationship between friendship and mobility examined by Cho et al. (2011)) and are all influenced by socio-economic environments. Second, at the collective level, we can use social sensing data to uncover the geographical impacts that influence the observed patterns. Current research focuses on land uses (or social functions), spatial interactions, and place semantics. Hence, social sensing also implies a series of methods for mining different geospatial big data. In this article, we list several methods for analyzing temporal signatures, interactions, and spatial embedded networks. Third, given that social sensing data contain rich temporal information, we can monitor temporal variations from the collected data and identify particular events (Crampton et al. 2013; Tsou et al. 2013; Croitoru et al. 2013). Last, social sensing serves geographical research at different spatial scales. It indicates that some core theoretical concepts such as scale, spatial heterogeneity, and distance decay should be taken into account when dealing with social sensing data.

Figure 13 demonstrates a framework of social sensing applications. The inner ring denotes the individual level human behavior patterns, and the outer ring contains the collective level patterns. Note that the ecological fallacy should be addressed when extending collective level patterns to individual level patterns since a big data set covers large volumes of individuals (Liu et al. 2014). Collective level patterns well reflect properties of geographical environments. The linkages between the two rings suggest that social sensing may provide a new insight into human-environment interactions, the fundamental research topic of geography.

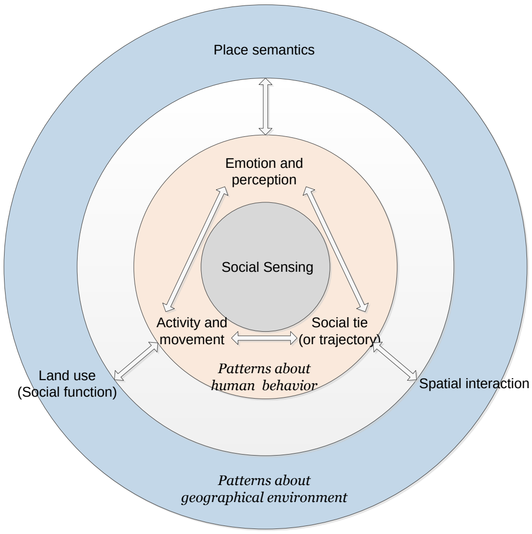


Figure 13. A framework diagram of social sensing applications

In sum, social sensing refers to a category of spatio-temporally tagged big data that provide an observatory for human behavior, as well as the methods and applications based on such big data. The major objective of social sensing is to detect socio-economic characteristics in geographical space and thus it can be viewed as a complement to remote sensing. Table 3 lists and compares four types of social sensing data that have been widely applied towards understanding behaviors. Besides the four types of data, a few studies have also introduced bank note records (Brockmann, Hufnagel, and Geisel 2006; Thiemann et al. 2010) and credit card records (Krumme et al. 2012) to extract movement information.

Table 3. Characteristics of four social sensing data sources

Data	Activity	Movement	Social tie	Emotion and perception
Mobile phone records	Mobile phone call	Long term (e.g. one month) trajectory of individuals. Stops are locations where users make calls and the sampling rate is low	Caller and callee	N/A**

	Ratti et al. 2006; Toole et al. 2012; Sun et al. 2011; Kang et al. 2012b	González, Hidalgo, and Barabási 2008; Kang et al. 2012a	Onnela et al. 2007; Kang et al. 2013; Gao et al. 2013	
Taxi trajectories	Pick up, drop off	Detailed short-term (e.g. half hour) trip trajectory of individuals	N/A*	N/A**
	Qi et al. 2011; Liu et al. 2012b	Jiang, Yin, and Zhao 2009; Liu et al. 2012a; Yue et al. 2012		
Public transportation card records	Pick up, drop off	Origin and destination of an intra-urban trip	People sharing a same bus or metro car	N/A**
	Chen, Chen, and Barry 2009; Gong et al. 2012	Roth et al. 2011; Long et al. 2012	Sun et al. 2013	
Social media check-in records	Check in	Long term trajectory of individuals. Stops are locations where a user posts geo-tagged entries and the sampling rate is very low	Friendship, follower and followee	Textual expressions that containing emotion and perception information
	Cranshaw et al. 2012; Li, Goodchild, and Xu 2013	Noulas et al. 2011; Cheng et al. 2011; Liu et al. 2014; Wu et al. 2014	Liben-Nowell et al. 2005	Rattenbury and Naaman 2009; Dodds et al. 2011; Mitchell et al. 2013; Gao et al. forthcoming

* individual-level interactions cannot be extracted from taxi data since passengers are without identifiers.

** data do not contain textual expressions

Conclusions

The emergence of big data brings new opportunities and challenges to both GIScience and geography. Kitchin (2013) categorized geospatial big data into directed, automated, and volunteered. In all types of geospatial big data, individual-level information is recorded. The term "social sensing" proposed in this article has two aspects of meaning. First, it follows the concept of volunteered geographical information, where each individual plays the role of a sensor (Goodchild 2007). Second, it can be viewed as analogue of remote sensing that excels at collectively sensing our socio-economic environments. Social sensing shares much in common with remote sensing. For a social sensing data set, after simple preprocessing, we can obtain a set of temporally sequenced images so that conventional remote sensing methods can be utilized. Additionally, since social sensing data and remote sensing data capture different aspects of geographical environments, integrating these two types of data will be an attractive research topic.

In terms of individual level behaviors, from social sensing data we can extract information about emotion and perception, and social tie in addition to activity and movement. These three aspects cover an individual's doing, feeling, and social relations. From a collective perspective, however, land uses (or social functions), semantics, and spatial interactions of places can be obtained. The three aspects interact with each other at both the individual and the collective level. Mining the underlying patterns and revealing the geographical impacts form two major directions of social sensing applications, which in consequence raise several theoretical topics. We suggest that the following aspects are of top priority: data quality and representativeness (Goodchild 2013b), location anonymization and privacy conservation, spatio-temporal scale, combining multi-source social sensing data, and linking individual versus collective level patterns.

From the perspective of geographical information systems (GIS), social sensing brings three new data types. The first is temporal images, which can be used to uncover land uses. Second, large volumes of trajectories can be extracted from social sensing data. Although this data type is not emphasized in this article, much research has been conducted for analyzing and visualizing trajectories (e.g. Kwan 2000; Lee, Han, and Whang 2007; Li and Wong 2014; Kwan, Xiao and Ding forthcoming). Last, interactions between individuals or places help us to construct spatially embedded networks so that network science methods can be borrowed to analyze them. These three data types all contain temporal semantics, and thus form a key component in the space-time integration of GIS and geography (Richardson 2013). A rich number of analytical functions as well as data management and visualization tools are in need for GIS. First, while image processing and complex network methods can be introduced to images and networks derived from social sensing data, we still lack tools supporting trajectories analyses (Goodchild 2013a). Second, even for images and networks, due to the characteristics of social sensing data, existing methods sometimes cannot be directly used. For example, current network tools seldom take into account node locations. Last, the volumes of social sensing data are large and thus raise requirements for high-performance computation (Kwan 2004). A number of emerging information technologies such as massively parallel-processing, distributed databases, and cloud based infrastructure will definitely benefit the use of social sensing data.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (Grant 41271386), Open Foundation of Shenzhen Key Laboratory of Urban Planning and Decision Making (Grant UPDMHITSZ2014A01), and State Key Laboratory of Resources and Environmental Information System. The authors would like to thank Professor M.-P. Kwan and three anonymous reviewers for their insightful comments. Professors L. Bian, G. Chi, Y. Chai, M. Goodchild, Q. Guo, B. Jiang, X. Li, L. Liu, X. Liu, Y. Lu, L. Mu, Q. Qin, S.-L. Shaw, F. Wang, and X. Yao have provided helpful input. Our thanks go to them.

References

Adams, B., and G. McKenzie. 2013. Inferring thematic places from spatially referenced natural language descriptions. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, eds. D. Sui, S. Elwood, and M. Goodchild, Springer, pp. 201–221.

Aggarwal, C. C., and T. Abdelzaher. 2011. Integrating sensors and social networks. In *Social Network Data Analytics*, ed. C. C. Aggarwal, Springer, pp. 379–412.

———. 2013. Social sensing. In *Managing and Mining Sensor Data*, ed. C. C. Aggarwal, Springer, pp. 237–297.

Alvares, L.O., V. Bogorny, B. Kuijpers, J.A.F. de Macedo, B. Moelans, and A. Vaisman. 2007. A model for enriching trajectories with semantic geographical information. In *Proceedings of the 15th annual ACM International Symposium on Advances in Geographic Information Systems, ACM*, no. 22.

Axhausen, K. W., and T. Gärling. 1992. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews* 12(4): 323–341.

Batty, M. 2013a. *The New Science of Cities*. MIT Press.

———. 2013b. Big data, smart cities and city planning. *Dialogues in Human Geography* 3(3): 274–279.

Bernstein M. F. 2014. Big data? Big brother? Big questions. *Princeton Alumni Weekly* 114(6).

Brockmann, D., L. Hufnagel, and T. Geisel. (2006) The scaling laws of human travel. *Nature* 439(7075): 462–465.

Calabrese, F., G. D. Lorenzo, L. Liu, and C. Ratti. 2011. Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Computing* 10(4): 36–44.

Campbell, A.T., S. B. Eisenman, N. D. Lane, E. Miluzzo, R.A. Peterson, H. Lu , X. Zheng, M. Musolesi, K. Fodor, and G.-S. Ahn. 2008. The rise of people centric sensing. *IEEE Internet Computing* 12(4): 12–21.

Chen, C., J. Chen, and J. Barry. 2009. Diurnal pattern of transit ridership: a case study of the New York City subway system. *Journal of Transport Geography* 17: 176–186.

Cheng, Z. Y., Caverlee, J., Lee, K., D. Z. Sui. 2011. Exploring millions of footprints in location sharing services. In *Proceeding of ICWSM 2011*, pp. 81–88.

Cho, E., S. A. Myers, and J. Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1082–1090

Crandall, D. J., L. Backstrom, D. Huttenlocher, and J. Kleinberg. 2009. Mapping the world's photos. In *Proceedings of the 18th international conference on World Wide Web*. ACM, pp. 761–770.

Cranshaw, J., R. Schwartz, J. Hong, and N. Sadeh. 2012. The Livehoods project: Utilizing social media to understand the dynamics of a city. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media* (ICWSM-12), June, Dublin, Ireland.

Crampton, J. W., M. Graham , A. Poorthuis, T. Shelton, M. Stephens, M. W. Wilson, and M. Zook. 2013. Beyond the geotag: situating ‘big data’ and leveraging the potential of the geoweb. *Cartography and*

Geographic Information Science 40(2): 130–139.

Croitoru, J., A. Crooks, J.Radzikowski, and A. Stefanidis. 2013. Geosocial gauge: a system prototype for knowledge discovery from social media. *International Journal of Geographical Information Science*, 27(12): 2483–2508.

Dodds, P. S., K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth. 2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PloS ONE* 6(12): e26752.

Eagle, N. and A. Pentland. 2006. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing* 10: 255–268.

Elwood, S., M. F. Goodchild, and D. Z. Sui. 2012. Researching volunteered geographic information: spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers* 102(3) 571–590.

Feick, R., and C. Robertson. forthingcoming. A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*. (doi: 10.1016/j.compenvurbsys.2013.11.006)

Furletti, B., P. Cintia, C. Renso, and L. Spinsanti. 2013. Inferring human activities from GPS tracks. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, ACM*, no. 5.

Gao, S., L. Li, W. Li, K. Janowicz, and Y. Zhang. forthcoming. Constructing gazetteers from volunteered Big Geo-Data based on Hadoop. *Computers, Environment and Urban Systems*. (doi: j.compenvurbsys.2014.02.004)

Gao, S., Y. Liu, Y. Wang, and X. Ma. 2013. Discovering spatial interaction communities from mobile phone data. *Transactions in GIS* 17(3): 463–481.

Goddard, J. B. 1970. Functional regions within the city centre: a study by factor analysis of taxi flows in central London. *Transactions of the Institute of British Geographers* 49: 161–182.

Gong, P., and P. Howarth. 1990. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. *Photogrammetric Engineering & Remote Sensing* 56(1): 67–73.

Gong, Y., Y. Liu, Y. Lin, J. Yang, Z. Duan, and G. Li. 2012. Exploring spatiotemporal characteristics of intra-urban trips using metro smartcard records. In *Proceedings of the 20th International Conference on Geoinformatics*, pp. 1–7.

González, M. C., C. A. Hidalgo, and A. L. Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453: 779–782.

Goodchild, M. F. 2011. Formalizing place in geographic information systems. In *Communities, Neighborhoods, and Health: Expanding the Boundaries of Place*, eds. L.M. Burton, S.P. Kemp, M.-C. Leung, S.A. Matthews, and D.T. Takeuchi, Springer, pp. 21–35.

———. 2007. Citizens as sensors: The world of volunteered geography. *GeoJournal* 69: 211–221.

———. 2013a. Prospects for a space–time GIS. *Annals of the Association of American Geographers*, 103(5): 1072–1077.

———. 2013b. The quality of big (geo)data. *Dialogues in Human Geography*, 3(3): 280–284.

Goodchild, M. F. and J. A. Glennon. 2010. Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth* 3(3): 231–241.

Graham, M., and T. Shelton. 2013. Geography and the future of big data, big data and the future of geography. *Dialogues in Human Geography* 3(3): 255–261.

Hu, S. G., and L. Wang. 2013. Automated urban land-use classification with remote sensing. *International Journal of Remote Sensing* 34(3): 790–803.

Huang, L., Q. Li, and Y. Yue. 2010. Activity identification from GPS trajectories using spatial temporal POIs' attractiveness. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, ACM, pp. 27–30.

Hubbard, P., R. Kitchin, and G. Valentine. 2004. *Key Thinkers on Space and Place*. SAGE.

Jensen, J. R., and D. C. Cowen. 1999. Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric engineering and remote sensing*, 65: 611–622.

Jiang, B., J. Yin, and S. Zhao. 2009. Characterizing human mobility patterns in a large street network. *Physical Review E* 80:021136.

Kang, C., X. Ma, D. Tong, and Y. Liu. 2012a. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications* 391(4): 1702–1717.

Kang, C., Y. Liu, X. Ma, and L. Wu. 2012b. Towards Estimating Urban Population Distributions from Mobile Call Data. *Journal of Urban Technology* 19(4): 3–21.

Kang, C., S. Sobolevsky, Y. Liu, and C. Ratti. 2013a. Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing* (UrbComp'13), *ACM*, no. 1.

Kang, C., Y. Zhang, X. Ma, and Y. Liu. 2013. Inferring properties and revealing geographical impacts of intercity mobile communication network of China using a subnet data set. *International Journal of Geographical Information Science* 27(3), 431–448.

Kitamura, R. 1988. An evaluation of activity-based travel analysis. *Transportation* 15(1): 9–34.

Kitchin, R. 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.

Krumme, C., A. Llorente, M. Cebrian, A. Pentland, and E. Moro. 2012. The predictability of consumer visitation patterns. *Scientific Reports* 3: 1645.

Kwan, M.-P. 2000. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research Part C: Emerging Technologies* 8:185–203.

———.2004. GISmethods in time-geographic research: Geocomputation and geovisualization of human activity patterns. *Geografiska Annaler B* 86(4): 267–80.

Kwan, M.-P., N. Xiao, and G. Ding. forthcoming. Assessing activity pattern similarity with multidimensional sequence alignment based on a multiobjective optimization evolutionary algorithm. *Geographical Analysis*.

Lane, N. D., S. B. Eisenman, M. Musolesi, E. Miluzzo, and A. T. Campbell. 2008. Urban sensing systems: Opportunistic or participatory? In *Proceedings of the 9th Workshop Mobile Computing Systems and Applications* (HotMobile'08), Napa, California, USA, pp. 11–16.

Lee, J.-G., J. Han, and K.-Y. Whang. 2007. Trajectory clustering: a partition-and-group framework. In *Proceedings of 2007 ACM International Conference on Management of Data (SIGMOD'07)*, ACM, pp. 593–604.

Li, L., M. F. Goodchild, and B. Xu. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science* 40(2): 61–77.

Li, J., and D. W. S. Wong. 2014. STModelViz: A 3D spatiotemporal GIS using a constraint-based approach. *Computers, Environment and Urban Systems* 45: 34–49.

Liben-Nowell, D., J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. 2005. Geographic routing in social networks. *Proceedings of the National Academy of Sciences of the USA* 102(33): 11623–11628.

Liu, Y., C. Kang, S. Gao, Y. Xiao, and Y. Tian. 2012a. Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems* 14(4): 463–483.

Liu, Y., F. Wang, Y. Xiao, and S. Gao. 2012b. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning* 106: 73–87.

Liu, Y., Q. Guo, and M. Kelly. 2008. A framework of region-based spatial relations for non-overlapping features and its application in object based image analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 63(4): 461–475.

Liu, Y., Z. Sui, C. Kang, and Y. Gao. 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE* 9(1): e86026.

Long, Y., Y. Zhang, and C. Y. Cui. 2012 . Identifying commuting pattern of Beijing using bus smart card data. *Acta Geographica Sinica*, 67(10): 1339–1352.

Lu, Y., and Y. Liu. 2012. Pervasive location acquisition technologies: Opportunities and challenges for geospatial studies. *Computers, Environment and Urban Systems* 36(2): 105–108.

McKenzie, G., K. Janowicz, and B. Adams. 2014. A weighted multi-attribute method for matching user-generated points of interest. *Cartography and Geographic Information Science* 41(2): 125–137.

Meng, X., N. Currit, L. Wang, and X. Yang. 2012. Detect residential buildings directly from Lidar and aerial photographs through object-oriented land-use classification. *Photogrammetric Engineering & Remote Sensing* 78(1): 35–44.

Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth. 2013. The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PloS ONE* 8(5): e64417.

Montis, A., S. Caschili, and A. Chessa. 2013. Commuter networks and community detection: A method for planning sub regional areas. *The European Physical Journal Special Topics* 215: 75–91.

Noulas, A., S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo. 2012. A tale of many cities: universal patterns in human urban mobility. *PloS ONE* 7(5): e37027.

Onnela, J.P., J. Saramaki, J. Hyvonen, G. Szabo, M.A. de Menezes, K. Kaski, A.-L. Barabási, and J. Kertesz. 2007. Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics* 9(6): 179.

Pei, T., S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou. forthcoming. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science* (doi:10.1080/13658816.2014.913794).

Peng, C., X. Jin, K.-C. Wong, M. Shi, P. Liò. 2012. Collective human mobility pattern from taxi trips in urban area. *PLoS ONE* 7(4): e34487.

Phithakkitnukoon, S., T. Horanont, G.D. Lorenzo, R. Shibasaki, and C. Ratti. 2010. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Human Behavior Understanding*, eds. A. A. Salah, T. Gevers, N. Sebe, and A. Vinciarelli, Springer, pp. 14–25.

Platt, R. V., and L. Rapoza L. 2008. An evaluation of an object oriented paradigm for land use/land cover classification. *The Professional Geographer* 60: 87–100.

Preoţiuc-Pietro, D., J. Cranshaw, and T. Yano. 2013. Exploring venue-based city-to-city similarity measures. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, ACM, no. 16.

Qi, G., X. Li, S. Li, G. Pan, Z. Wang, and D. Zhang. 2011. Measuring social functions of city regions from large-scale taxi behaviors. In *Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops* (PERCOM Workshops), pp. 384–388.

Rattenbury, T., and M. Naaman. 2009. Methods for Extracting Place Semantics from Flickr Tags. *ACM Transactions on the Web* 3(1): no. 1.

Ratti, C., S. Sobolevsky, F. Calabrese, C. Andris, J. Reades J, M. Martino, R. Claxton, S. H. Strogatz. 2010. Redrawing the map of Great Britain from a network of human interactions. *PLoS ONE* 5: e14248.

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

Reades, J., F. Calabrese, A. Sevtsuk, and C. Ratti. 2007. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing* 6(3): 30–38.

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

Richardson, D. B. 2013. Real-time space–time integration in Giscience and geography. *Annals of the Association of American Geographers* 103(5): 1062–1071.

Roth, C., S. M. Kang, M. Batty, and M. Barthélemy. 2011. Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLoS ONE* 6(1): e15923.

Sevtsuk, A., and C. Ratti. 2010. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology* 17: 41–60.

Schaller, R., M. Harvey, and D. Elsweiler. 2014. Detecting event visits in urban areas via smartphone GPS data. In *Advances in Information Retrieval*, eds. P. Serdyukov, et al., Springer, pp. 681–686.

Shen, Y., M.-P. Kwan, and Y. Chai. 2013. Investigating commuting flexibility with GPS data and 3D geovisualization: A case study of Beijing, China. *Journal of Transport Geography* 32(1): 1–11.

Sun, J., J. Yuan, Y. Wang, H. Si, and X. Shan. 2011. Exploring space-time structure of human mobility in urban space. *Physica A: Statistical Mechanics and its Applications* 390(5): 929–942.

Sun, L., K. W. Axhausen, D.-H. Lee, and X. Huang. 2013. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences of the USA* 110: 13774–13779.

Thiemann, C., F. Theis, D. Grady, R. Brune, and D. Brockmann. 2010. The structure of borders in a small world. *PLoS ONE* 5(11): e15422.

Toole, J., M. Ulm, D. Bauer, and M. González. 2012. Inferring land use from mobile phone activity, In *Proceedings of 1st ACM SIGKDD International Workshop on Urban Computing* (UrbComp'12), ACM, no. 1.

Tsou, M.-H., J.-A. Yang, D. Lusher, S. Han, B. Spitzberg, J. M. Gawron, D. Gupta, and L. An. 2013. Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): a case study in 2012 US Presidential Election. *Cartography and Geographic Information Science* 40(4): 337–348.

Tuan, Y. F. 1977. *Space and Place: The Perspective of Experience*. Minneapolis: University of Minnesota Press.

Wu, L., Y. Zhi, Z. Sui, and Y. Liu. 2014. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, 9(5): e97010.

Wu, S.-S., X. Qiu, E. L. Usery, and L. Wang. 2009. Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use. *Annals of the Association of American Geographers* 99(1): 76–98.

Ye, M., K. Janowicz, C. Mülligann, and W.C. Lee. 2011. What you are is when you are: the temporal dimension of feature types in location-based social networks. *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, pp. 102–111.

Yue, Y., H. Wang, B. Hu, Q. Li, Y. Li, and A.G.O. Yeh. 2012. Exploratory calibration of a spatial interaction model using taxi GPS trajectories. *Computers, Environment and Urban Systems* 36(2): 140–153.

Yue, Y., T. Lan, A.G.O. Yeh, and Q. Li. 2014. Zooming into individuals to understand the collective: A review of trajectory-based travel behaviour studies. *Travel Behaviour and Society* 1(2): 69–78.

Zhang, W., S. Li, and G. Pan. 2012. Mining the semantics of origin-destination flows using taxi traces. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing, ACM*, pp. 943–949.

Zheng, Y., L. Capra, O. Wolfson, and H. Yang. 2014. Urban computing: Concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology* 6(2): no. 9.