Applied Geography 86 (2017) 152–164

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data

Di Zhu [a], Ninghua Wang [c], Lun Wu [a], Yu Liu [a, b, *]

[a] Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, 100871, PR China
[b] Beijing Key Lab of Spatial Information Integration and Its Applications, Peking University, Beijing, 100871, PR China
[c] Vipshop US Inc, 2550 N 1st Street, San Jose, CA 95131, USA

ARTICLE INFO

Article history:
Received 1 July 2016
Received in revised form
7 June 2017
Accepted 7 July 2017
Available online 20 July 2017

Keywords:
Street
Linear unit
Spatio-temporal pattern
Big geo-data
Urban study

ABSTRACT

Quantitative research of urban geography has benefited greatly from the rapid development of big geo-data. Spatial assembly is an essential analytical step to summarize and perceive geographical environment from individual behaviours. Most research focuses on the methodology of how to utilize the big data, while the adopted spatial units for data aggregation remain areal in nature. This article conceptually proposes an idea of sensing cities from a street perspective, emphasizes the significance of street units in quantitative urban studies. Using a three-month taxi trajectory dataset and the major streets in Beijing, we explore the spatio-temporal patterns of urban mobility on streets, cluster streets into nine types based on their dynamic functions and capacities. Additionally, we discuss the differences and connections between the linear street unit and traditional areal units, investigate the possibility of uncovering urban communities using streets, and point out the complexity of streets. We conclude that street unit as a supplement to areal units, is able to effectively minify the modifiable areal unit problem (MAUP), sense urban dynamics, depict urban functions, and understand urban structures.

1. Introduction

The rise of geographic information science and big geo-data brings new opportunities for understanding urban environments (Batty, 2013a,b; Foth, Choi, & Satchell, 2011; Hao, Zhu, & Zhong, 2015). Through automated and routine movement tracking of individuals, various forms of locator devices work as sensors to collect geospatial data and characterize the activity of a city in both spatial and temporal perspectives (Batty, 2013a; Kang & Qin, 2016; Ratti, Frenchman, Pulselli, & Williams, 2006; Sun, Yuan, Wang, Si, & Shan, 2011). After a thorough review into the nature of geospatial data and its meaning to geographic studies, Liu et al. (2015) proposed a research framework entitled *social sensing* that excels at sensing the socio-economic features of urban space. From the perspective of individuals, citizens play the role of voluntary sensors and produce plenty of volunteered geographic information (Goodchild, 2010). At the collective level, the distribution of geographic phenomena such as land use (or social function) and the pattern of spatial interaction flows can be investigated after spatio-temporal aggregation of individual behaviour data. Social sensing as a combination of traditional remote sensing techniques and geospatial big data, gives us a more accurate portrait of a city's details, since the daily activities and movements of residents contained in geospatial data may indicate the socio-economic properties of urban functions from a geographic angle (Jokar Arsanjani, Helbich, Bakillah, Hagenauer, & Zipf, 2013; Gong, Lin, & Duan, 2017; Pei et al., 2014; Sevtsuk & Ratti, 2010; Toole, Ulm, González, & Bauer, 2012).

Utilizing massive amounts of geospatial data, the first-order distribution of urban attributes (e.g., economic indices, population intensity, condition of public facilities), as well as second-order interactions (e.g., human movements, flow of goods, financial flows, social ties) can be used to better understand human mobility, urban functions, and urban structures. Much literature has verified the assumption that urban spaces are associated with different temporal rhythms (or spatio-temporal patterns) of activities. Castro, Zhang, Chen, Li, and Pan (2013) and Qi et al. (2011) depicted the social dynamics and traffic dynamics based on the movements of taxi passengers. Kang, Liu, Ma, and Wu (2012) estimated

* Corresponding author. Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, 100871, PR China.
E-mail addresses: patrick.zhu@pku.edu.cn (D. Zhu), wangninghua@gmail.com (N. Wang), wulun@pku.edu.cn (L. Wu), liuyu@urban.pku.edu.cn (Y. Liu).

http://dx.doi.org/10.1016/j.apgeog.2017.07.001
0143-6228/

population distributions by comparing different mobile phone call activity patterns. Ratti et al. (2006), Toole et al. (2012), Yuan and Raubal (2012), and Pei et al. (2014) used mobile phone data to obtain human mobility patterns and detect different communities. Soto and Frías-Martínez (2011), Frías-Martínez, Soto, Hohwald, and Frías-Martínez (2012), and Liu, Wang, Xiao, and Gao (2012) adopted similar methods based on feature vectors (normalized or not) to depict temporal activity patterns, then clustering methods were used to differentiate various land types. Apart from the temporal activity pattern that can characterize urban functions, Liu, Kang, Gong, and Liu (2016) introduced taxi O-D data from a perspective of connections, focused on the spatial interaction patterns between parcels, and improved the urban land use classification. In addition, the interaction patterns in flow data are utilized in much research to study community structures through spatially embedded interaction networks (Gao, Wang, Gao, & Liu, 2013; Liu, Sui, Kang, & Gao, 2014; Ratti et al., 2010; Thiemann, Theis, Grady, Brune, & Brockmann, 2010).

Spatial assembly is an essential analytical step when perceiving our geographical environment from individual-level geospatial data. As a matter of fact, it is inevitable to confront the issue of spatial resolution (or scale) when mapping individual details onto regular or irregular units (Liu et al., 2015). Some research based on mobile phone data uses Voronoi polygons generated from base towers (Shi, Chi, Liu, & Liu, 2015), while most existing studies investigate land use based on regular grids, for example, 0.25 km^2 (Reades, Calabrese, & Ratti, 2009), 0.5 km^2 (Liu et al., 2016) and 1 km^2 (Liu et al., 2012; Sun et al., 2011; Toole et al., 2012). Zhou and Long (2016) introduces a crowdsourcing platform named *SinoGrids* that aims at the interoperation of microscale urban data by discretizing the projected extent of China into grids with multiple optional resolutions. Scientists in geo-related fields usually apply the areal units to acquire aggregated data, but the definition of discrete spatial units can be critical and arbitrary (Briant, Combes, & Lafourcade, 2010; Meentemeyer, 1989). The sensitivity of the spatial analysis results to the choice of zoning systems for which data are collected is known as *the modifiable areal unit problem* (MAUP) (Openshaw & Taylor, 1979; Openshaw, 1984). To be more precise, the word *modifiable* in MAUP contains two related but distinctive components: the scale effect and the zoning (aggregation) problem (Jelinski & Wu, 1996). The scale effect is the variations in results that may be obtained when the same areal data are combined into sets of increasingly larger areal units of analysis. The zoning problem, in contrast, is the variations in results due to alternative units of analysis where the number of units is constant (Openshaw & Taylor, 1979). In fact, the challenge of selecting proper scales or zoning systems is a long-standing issue that has been a challenge to geographers and planners. They have made great efforts in understanding and dealing with MAUP, but the puzzle of the best unit remains unsolved (Briant et al., 2010; Mu & Wang, 2008; Nellis & Briggs, 1989; Openshaw, 1977; Stehman & Wickham, 2011; Turner, Dale, & Gardner, 1989).

Although much attention has been given to spatial units when mapping socio-economic data onto geographic space, researchers are not so easy to think outside the box. The choice of areal units may be so dominant that it is taken for granted. For these reasons, we attempt to introduce a fresh new perspective in urban studies, using streets as the basic elements to characterize urban functions and understand urban structures. Currently, research on linear units is limited. Okabe, Yomono, and Kitamura (1995) conducted a series of studies on street networks, such as a market area analysis and the demand for retail stores on a street network (Okabe & Kitamura, 1996; Okabe & Okunuki, 2001), as well as some theoretical research about network-based spatial analysis approaches (Okabe, Satoh, & Sugihara, 2009; Okabe & Satoh, 2006, 2009). The emergence and development of space syntax theory is also inspiring for us to better understand the linear spatial units regarding human space organization (Hillier, 2012; Hillier, Leaman, Stansall, & Bedford, 1976; Penn, 2003; Ratti, 2004; Shen & Karimi, 2016; Turner, 2007). In fact, the street system is never an insignificant part of a city. Lynch (1960) proposed the five well-known elements of city images: paths, edges, districts, nodes, and landmarks, of which the paths are the most important and are defined as the channels (e.g., streets, walkways, railroads, transit lines) along which the observer moves. For many people, paths are the predominant elements in their image, and they observe the city while moving through them (Yin & Wang, 2016; Yin, 2017; Yin, Cheng, Wang, & Shao, 2015). It is now generally accepted that the physical movement in an urban space is usually constrained by a road network (Yu, Ai, He, & Shao, 2016) and streets interlink urban functions physically and cognitively (Shen & Karimi, 2016). Recently, Long and Liu (2017) used an online street-view service to analyse the greenery of the streets for central areas of Chinese cities. Compared with traditional areal units, the street unit is an approximate decomposition of a parcel or a block, and is capable of minifying the effect of MAUP during spatial analysis. Thus, we suggest that the street unit is a promising substitute for areal units and can help us uncover hidden knowledge concealed under areas (a discussion on this can be found in Section 4.1).

In this work, we introduce a taxi trajectory dataset in Beijing to explore the spatio-temporal mobility patterns from the street perspective and investigate the social dynamic functions on streets. We try to emphasize the significant value of the street unit in quantitative urban studies and give inspiration to other researchers and planners. The remainder of this paper is structured as follows: Section 2 provides an overview of our study area and how to prepare effective street-based trips. Section 3 introduces the social sensing methods we adopted, and shows the analytical results achieved at the street level. Section 4 discusses the comprehensiveness of the street unit compared to areal units, the potential for studying urban structures using the street unit, and the complexity of streets. Section 5 concludes and points out future directions for our research.

2. Data preparation

Beijing is the capital of China, and its spatial structure has been relatively stable. Some researchers have studied the relationship between urban forms and individuals' trip behaviours (Chai, Weng, & Shen, 2008; Liu et al., 2012, 2016), but they either used survey data with small and less representative samples or characterized the urban form with areal units such as grids and parcels. In order to report the intra-urban trips more accurately, the dataset should be representative of the general population. Even though subway transit dominates Beijing's public transportation, taxis still play an important part in ground transportation in recent years.[1] Floating taxi data is widely applied in the analysis of human mobility patterns, urban functions, and urban structures. Meanwhile, since taxi trajectories are highly consistent with roads after map-matching (Lou et al., 2009; Quddus, Ochieng, Zhao, & Noland, 2003), taxi data are of great value in reflecting traffic conditions. Thus, we consider it suitable for our work to introduce streets (or road segments) as the basic spatial unit.

We use a dataset of more than 33 000 taxis (about half the total taxis) from several anonymous taxi companies in Beijing, China, for three consecutive months (from 1 May to 31 July 2013). Each record

[1] http://www.baogaochina.com/.

contains the taxi's ID, sampling time, location, heading, velocity, and status (occupied or not) with high position accuracy and acceptable sampling intervals. In this research, we are only interested in the location, time, and status of trajectory records $R(x, y, t, status)$. After cleaning, grouping, and sorting the data, we identify the locations where pick-up and drop-off activities occurred based on the state markers of the taxis' continuous trajectories, while the trajectory details are not of interest and eliminated. Usually, the locations of these two activities are viewed as the origin and destination (OD) of a trip, except that the coordinates maybe slightly different because of the deviation of taxi service points (always along streets) and actual OD points. A record $R_i(x_i, y_i, t_i, status_i)$ is judged to be a pick-up (drop-off) event if it is occupied (vacant) and its previous record $R_{i-1}(x_{i-1}, y_{i-1}, t_{i-1}, status_{i-1})$ shows a vacant (occupied) state. Each trip is formalized to be a record pair $T = \{R_o, R_d\} = \{(x_o, y_o, t_o), (x_d, y_d, t_d)\}$ where the subscript "o" denotes a pick-up point and "d" denotes a drop-off point. The data records of a processed trip sample is shown in Table 1.

Table 1
Sample trip with pick-up and drop-off labels.

ID	time	longitude	latitude	status
33755	2013-5-1 9:11:26	116.63764	39.89415	occupied (pick-up)
33755	2013-5-1 9:48:27	116.50217	39.95583	vacant (drop-off)
33755	2013-5-1 9:53:51	116.49551	39.97067	occupied (pick-up)
33755	2013-5-1 10:3:49	116.46709	39.98861	vacant (drop-off)

Beijing has one of the largest and most complex street networks in the world, with currently six ring roads encircling the urban area,[2] several arterial highways, numerous paths, and *hutongs* (local term for alleys). For the objective of our work, we assume that a street is interior homogeneous and can be treated as an individual unit. We choose streets that are of relatively higher priority within the 5^{th} Ring Road and exclude trivial roads that are much less important according to factors such as the road's level and width. The final street collection ($\mathbb{S}$) has 1617 streets in total and is illustrated in Fig. 1.

To focus on street units and decrease the influence of GPS position errors, we match the pick-up and drop-off points onto corresponding streets using a simple map-matching method: choosing the nearest street of a GPS tick as the matched street (Quddus et al., 2003; White, Bernstein, & Kornhauser, 2000). Then, the coordinate-based trip $T = \{R_o, R_d\} = \{(x_o, y_o, t_o), (x_d, y_d, t_d)\}$ is converted into the street-based trip $T^* = \{R_o^*, R_d^*\} = \{(S_o, t_o), (S_d, t_d)\}$, where S_o (S_d) is the street where the pick-up (drop-off) activity actually took place. Given a distance searching threshold $D_{threshold}$, note that only when $\forall R \in T$, the spatial query of searching a nearest street segment of R within $D_{threshold}$ has a solution for S, can we obtain an effective street-based trip $T^* = \{(S_o, t_o), (S_d, t_d)\}$ from T for further work. In our work, by setting $D_{threshold}$ to 500 m,[3] a total of 13 076 765 effective street-based trips are extracted from the original dataset, and these trips provide us the foundation for sensing Beijing from its streets. Fig. 2 is an example of matching a coordinate-based trip onto a street-based trip.

3. Methodology and results

3.1. Spatio-temporal patterns of urban mobility in the street perspective

3.1.1. Temporal patterns of pick-ups and drop-offs

We compress the time stamps for all trips into one week (168 h) and cast the temporal pick-up and drop-off activities onto corresponding streets. The number of people getting in a taxi, as well as the number of passengers who get off at a certain time t for any street S_i, is obtained using

$$P[S_i, t] = \sum_{o} R_o^*[S_i, t] \quad (t = 0, 1, \cdots, 167) \tag{1}$$

$$D[S_i, t] = \sum_{d} R_d^*[S_i, t] \quad (t = 0, 1, \cdots, 167) \tag{2}$$

where $R^*[S_i, t]$ specifies the details of one pick-up or drop-off activity on street i at time t. For a one week duration, the dominant period length for urban mobility patterns has been verified to be 24 h by applying the discrete Fourier transform analysis on the pick-ups and drop-offs (Liu et al., 2012). Upon this premise, we calculate the hourly average activity number on each street $S_i \in \mathbb{S}$ for both weekdays and weekends, and plot the global temporal distribution of pick-ups and drop-offs in Fig. 3(a) using

$$\begin{aligned} N_\lambda^{(d)}[t] &= \sum_{S_i \in \mathbb{S}} N_\lambda^{(d)}[S_i, t] \\ &= \sum_{S_i \in \mathbb{S}} \frac{\sum_{k=0}^{4} \lambda[S_i, k \times 24 + t]}{5} \quad (t = 0, 1, \cdots, 23) \end{aligned} \tag{3}$$

$$\begin{aligned} N_\lambda^{(e)}[t] &= \sum_{S_i \in \mathbb{S}} N_\lambda^{(e)}[S_i, t] \\ &= \sum_{S_i \in \mathbb{S}} \frac{\sum_{k=5}^{6} \lambda[S_i, k \times 24 + t]}{2} \quad (t = 0, 1, \cdots, 23) \end{aligned} \tag{4}$$

where N is the count of activities, λ denotes the type of the activity (R_o for pick-ups and R_d for drop-offs), k is the ordinal number in a seven-day week, and the superscript d and e specify weekdays and weekends separately.

From Fig. 3(a), we find that Beijing residents have a tendency to travel more by taxi on weekdays than weekends, and this phenomenon is possibly caused by the traffic restriction policy established by the Beijing Municipal Government in which private cars are partly restricted on weekdays but free on weekends. In general, the trip numbers are roughly stable in the daytime with no sharp decrease, while there are three obvious traffic peaks, i.e., the morning peak (8:00–10:00), noon (12:00–14:00), and the evening peak (18:00–20:00), with a sudden rise in taxi activity. On weekends, the morning peak and noon peak are less typical, with people tending to go out later, and the evening peak is more intense because of various recreational activities. Due to our street-based trip data pre-processing, the total number of pick-ups is exactly the same as the number of drop-offs.

Additionally, Zhongguancun Street is chosen as a sample in Fig. 3(b) to show that the temporal pattern can be further investigated at the street level. Zhongguancun[4] is one of the most active commercial districts in Beijing, comprising a high-tech centre and

[2] https://en.wikipedia.org/wiki/Ring_roads_of_Beijing.

[3] Since we have eliminated some low-level streets to simplify the road network of Beijing, a relatively large $D_{threshold}$ of 500 m is selected to ensure that a taxi activity at a low-level street can be aligned onto a nearest high-level street. In fact, if we change $D_{threshold}$ from 200 m to 500 m, the extracted street-based trips remain almost unchanged.

[4] The location of Zhongguancun is marked as place B in Fig. 1.

Table 2
Detailed descriptions of the 11 marked places in Fig. 1.

Place	Name	Description
A	Peking University	A major Chinese research university located in Beijing, which is consistently ranked as the top higher learning institution in mainland China.
B	Zhongguancun	A technology hub in the northwestern part of Beijing city, and is often referred to as China's *Silicon Valley*.
C	Beijing West Railway Station	A terminal for most trains leaving the city for destinations in western and southwestern China.
D	Fengtai Science Park	A technology hub in the southwestern part of Beijing city, which is similar to Zhongguancun.
E	Xidan	A major traditional commercial area in Beijing incorporating many supermarkets and department stores.
F	Beijing South Railway Station	A large railway station that mainly serves high speed trains.
G	Beijing Railway Station	A terminal railway station located just southeast of the city centre.
H	Guomao	An area in Beijing at the centre of the Beijing central business district.
I	Santitun and the Gongti Stadium	An area containing many popular bar streets and international stores.
J	The Bell Tower and Drum Tower	Originally built for musical reasons, it was later used to announce the time and is now a tourist attraction.
K	Beijing Olympic Village	A complex of high-rise apartments, which was opened to the public in conjunction with the 2008 Summer Olympics.

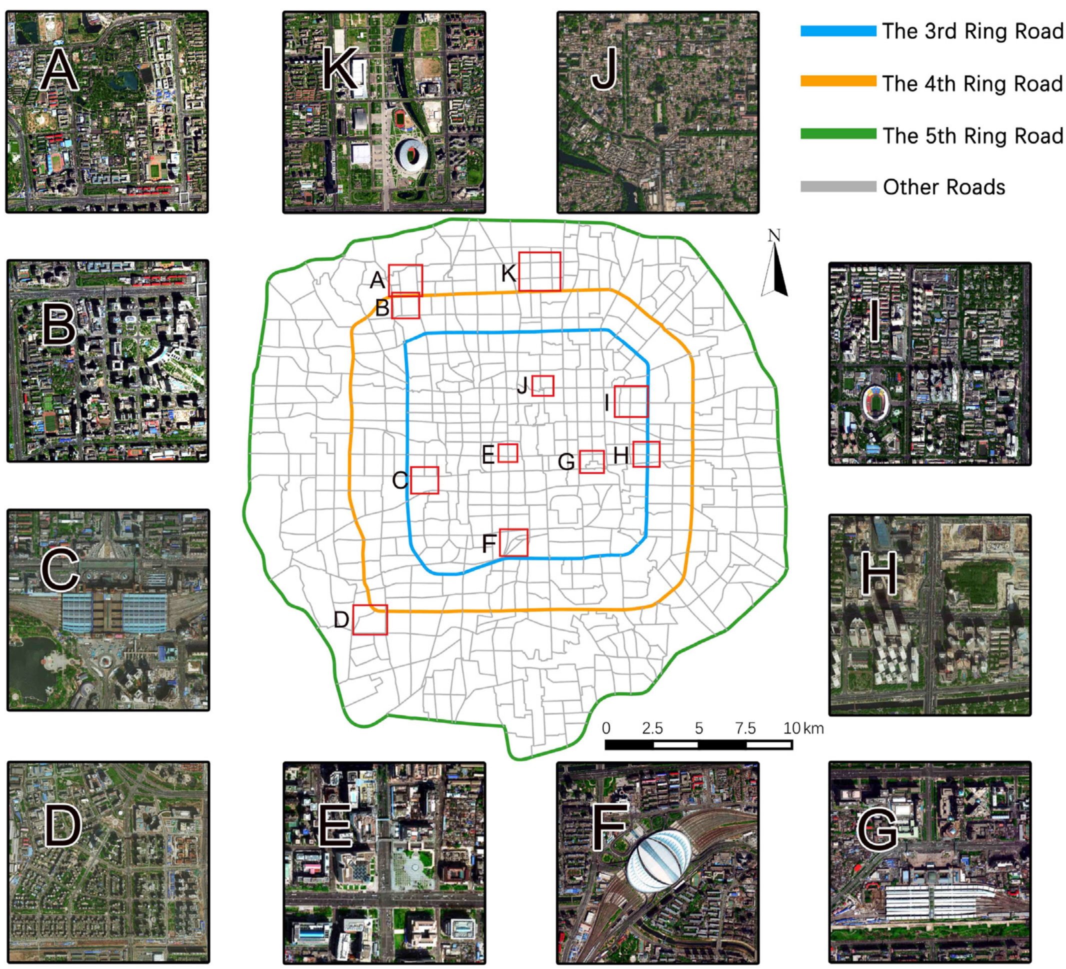


Fig. 1. Overview of the streets in our research. Blue lines indicate the 3^{rd} Ring Road, orange lines are the 4^{th} Ring Road, while the thick green lines show the 5^{th} Ring Road in Beijing. Grey lines within the 5^{th} Ring Road are the other streets of interest. The 11 highlighted subareas marked from A to J are some important places that will be further investigated or referenced in this paper. A detailed introduction of these places are shown in Table 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

entertainment industries, and offers a great number of jobs. Zhongguancun Street is the eastern road of Zhongguancun district and acts as a vital traffic line. Compared to the global pattern, the number difference between weekdays and weekends is smaller for Zhongguancun Street. On weekdays, drop-offs exceed pick-ups before noon, while on weekends the reverse trend is true. Since the temporal difference between drop-off and pick-up activities has been adopted to reveal the distinctive functions of urban areas in previous research (Liu et al., 2012), Fig. 3(b) gives us a hint that the same idea can be applied to streets.

3.1.2. *Spatio-temporal patterns of street functions*

Using Eqs. (1) and (2), the total number of pick-up or drop-off activities on each street for the entire three months are computed by

$$P[S_i] = \sum_t P[S_i, t] \quad t = 0, 1, \cdots, 167 \tag{5}$$

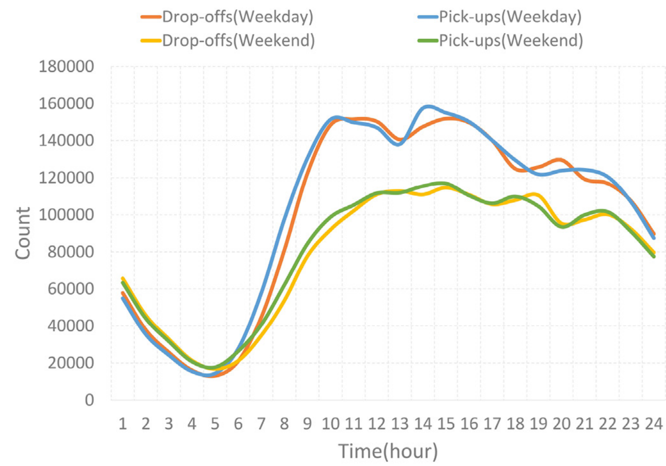
Fig. 2. Graphical representation of the map matching and extraction of effective trips. P_1 and P_2 are two pick-up points, while D_1 and D_2 are two drop-off points. There are four possible coordinate-based trips ($T = \{R_{P_1}, R_{D_1}\}$, $T = \{R_{P_2}, R_{D_1}\}$, $T = \{R_{P_1}, R_{D_2}\}$, $T = \{R_{P_2}, R_{D_2}\}$), but only $T = \{R_{P_1}, R_{D_1}\}$ can be matched as an effective street-based trip $T^* = \{R^*_{P_1}, R^*_{D_1}\}$.

$$D[S_i] = \sum_t D[S_i, t] \quad t = 0, 1, \cdots, 167 \tag{6}$$

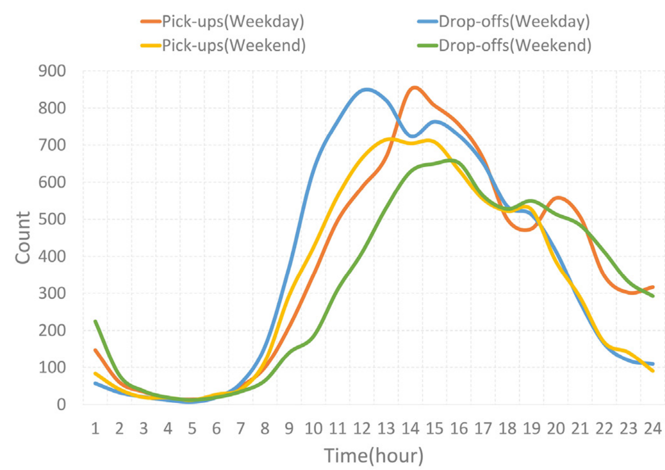
Then, we classify Beijing's major streets according to their overall activity levels into five types from inactive to active. Fig. 4 depicts the spatial distribution pattern of taxi activities in the street perspective and displays an interesting heterogeneity: streets that are spatially adjacent (especially streets that enclose the same parcel) can vary greatly, a fact which is concealed by traditional areal units.

Since experienced taxi drivers tend to pick up passengers quickly after a drop-off to maximize the profit from the next trip (Yuan, Zheng, Zhang, & Xie, 2013), a drop-off is usually followed closely by a pick-up. It is clear that the total number of pick-ups is very consistent with that of drop-offs, except for some streets at railway stations (place C,F and G in Fig. 1), where the drop-off activity level is extremely high (Fig. 4(a)), but the pick-up level is less active, or even inactive (Fig. 4(b)). This pattern may be abnormal, because of the trip purposes (Wolf, Guensler, & Bachman, 2001), as well as the GPS signal interference caused by specific architecture designs. Passengers heading for railway stations, hospitals, airports, and bar areas are more likely to choose taxis, while the preference is less significant for passengers coming out of these places. Railway stations equipped with underground taxis stands may also lead to the underestimation of trips, since the real pick-up signals are missing or treated as noise. In addition, from Fig. 4 we can see most of the streets in south part of the city are associate with very few pick-ups or drop-offs and are classified into the *inactive* street type. It maybe interesting to identify whether a street is urban or not using trajectories, but considering only the total number of activities is not enough to make a reliable judgement.

The limitation of the total pick-ups or total drop-offs in Fig. 4 lies in the insufficiency of temporal information. Some particular features of railway stations, along with Zhongguancun Street in Fig. 3(b), show that the spatio-temporal patterns of the difference between drop-offs and pick-ups reflects the imbalance of human mobility, which reveals the temporal functions of streets. By changing the time interval t in Eqs. (5) and (6), we visualize the spatio-temporal distribution of $D[S_i] - P[S_i]$ in Fig. 5. There are significantly more drop-off points than pick-ups on streets near railway stations, and the North 3rd and East 3rd Ring Roads are



(a) Total number



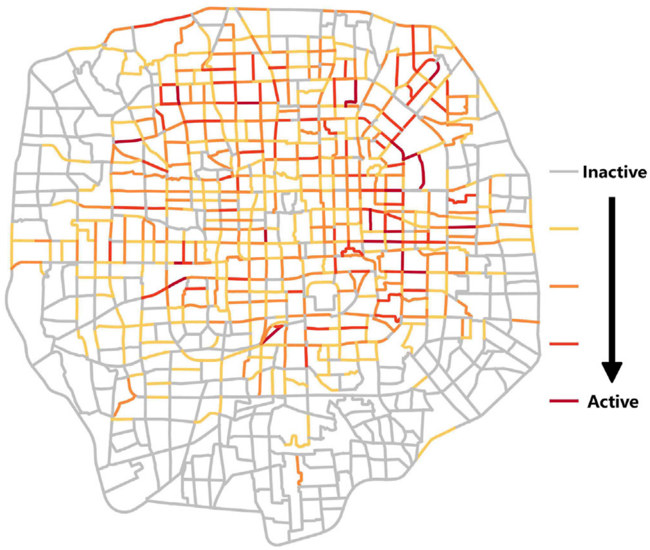
(b) Zhongguancun Street

Fig. 3. Temporal variance of the pick-up and drop-off activities. The y-axis shows the average total number of pick-up points and drop-off points, and the x-axis is defined with one-hour intervals. (a) Global temporal variance for our study area. (b) Temporal variance for Zhongguancun Street, the eastern road of place B in Fig. 1.

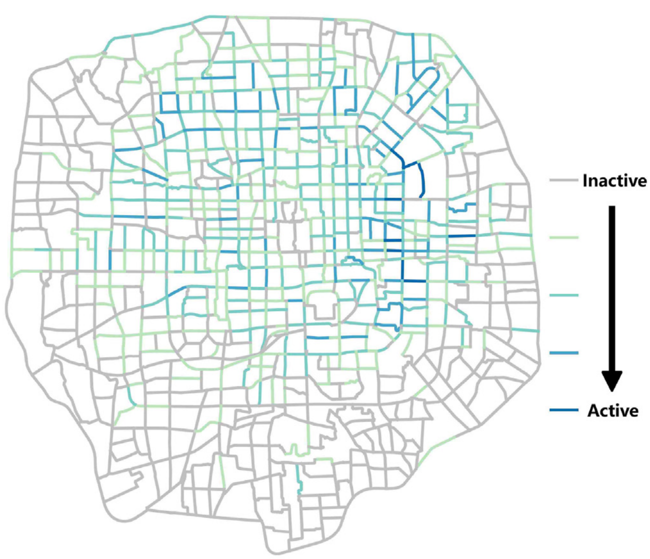
streets where passengers are picked up more and dropped off less frequently. In the morning peak (Fig. 5(b)), the heterogeneity of streets is easy to see: most streets are balanced in terms of drop-offs and pick-ups, some streets clustered at railway stations and commercial districts such as Guomao, Zhongguancun, Xidan, and Fengtai Science Park (place H, B, E, D in Fig. 1, respectively.) have more drop-offs, while streets with more pick-ups are relatively uniform spatially. This can be explained by the fact that people come out from residential districts all over the urban area in the morning and head to transportation hubs or their work places. At noon (Fig. 5(c)), the attractive and repulsive forces of streets are both weaker, and the pattern is more balanced. The evening peak shows an opposite pattern compared to the morning peak, with more people being picked up on urbanized streets and dropped off on residential streets at various places.

3.2. Association of street classifications with dynamic street patterns

Local temporal patterns are more interesting than global



(a) drop-off



(b) pick-up

Fig. 4. Spatial distribution of street activity levels measured by drop-offs and pick-ups.

temporal patterns, since different places are associated with different temporal human activity patterns (Liu et al., 2016). Street units contain more details about taxi activity than parcels, and depict more specific geographic information because the pick-ups or drop-offs often happened along street sides. Investigating the similarities and differences among streets is useful in understanding how people utilize urban areas, and vice versa, how streets are characterized according to human traffic activity. Previous research has verified that temporal signatures can be used to infer land use, and the similarity in temporal activity patterns will make similar places more likely to be classified into the same land use type. We suggest that street types can also be explored using this method. As shown in Fig. 5, the one-day dynamic $V[S_i,t] = D[S_i,t] - P[S_i,t]$ is a suitable feature of activity patterns, and the underlying usage mode reflects street types across the urban area.

3.2.1. Hierarchical clustering based on dynamic street functions and capacities

We consider both the capacity and function of a street when clustering street types based on their temporal activity patterns, which means the number of active people and the variation mode are both important when judging a street. We use a vector that consists of the absolute volumes of average drop-offs minus pick-ups in each hour $t = 0, 1, \cdots, 23$ and the total difference between drop-offs and pick-ups on street S_i to characterize the dynamic capacity of streets, which is denoted as

$$V_C = \left[A_{i,0}, A_{i,1}, \cdots, A_{i,t}, \cdots, A_{i,23}, A_{i,total}\right] \tag{7}$$

where $A_{i,t} = \sum_{k=0}^{6}(D[S_i, k \times 24 + t] - P[S_i, k \times 24 + t])/7$ and $A_{i,total} = D[S_i] - P[S_i]$. To acquire a better description of the streets' variation modes, normalization is an essential processing step (Pei et al., 2014; Toole et al., 2012) that ensure all the signature vectors to be of the same scale. We calculate the Z-score for each $A_{i,t}(t = 0, 1, \cdots, 23)$ and the normalized vector is thus constructed as

$$\begin{aligned} V_F &= \left[\frac{A_{i,0} - \mu_i}{\sigma_i}, \frac{A_{i,1} - \mu_i}{\sigma_i}, \cdots, \frac{A_{i,t} - \mu_i}{\sigma_i}, \cdots, \frac{A_{i,23} - \mu_i}{\sigma_i}\right] \\ &= \left[Z_{i,0}, Z_{i,1}, \cdots, Z_{i,t}, \cdots, Z_{i,23}\right] \end{aligned} \tag{8}$$

to represent the dynamic function of streets, where $\sigma_i = \sqrt{(\sum_{t=0}^{23}((A_{i,t} - \mu_i)^2))/23}$ and $\mu_i = \sum_{t=0}^{23} A_{i,t}/24$.

Then, an unsupervised hierarchical bisecting k-means clustering is used to classify streets into several types based on both their dynamic capacities and functions. In other words, streets with similar functions are grouped into the same basic clusters, capacities are further used to discriminate among them. A bisecting k-means algorithm is adopted to gradually split one cluster into two sub-clusters and achieve the global highest overall similarity. The advantages of bisecting k-means compared to traditional k-means can be found in Steinbach, Karypis, & Kumar (2000). Our hierarchical clustering is carried out using the following steps:

Step 1. Use the normalized dynamic street function signature V_F (see Eq. (8)) to execute the bisecting k-means algorithm, generating m 1^{st}-level street types.
Step 2. The dynamic street capacity signature V_C (see Eq. (7)) is applied to further classify the 1^{st}-level types. Run the bisecting k-means algorithm using V_C in every 1^{st}-level street type and generate n 2^{nd}-level types each.
Step 3. Synthesize 1^{st}-level and 2^{nd}-level types together and acquire the final street types (usually smaller than $m \times n$).

3.2.2. Characterizing street types by dynamic functions and capacities

In the case of classifying Beijing urban streets, the results are inspiring and interesting. We obtain four basic 1^{st}-level street types with distinctive dynamic functions after Step 1 (see Fig. 6). For each 1^{st}-level type, we identify three 2^{nd}-level types according to the street's capacity (or activity intensity) (Fig. 7). Through integration, the major streets in the Beijing urban area are finally classified into nine types with different activity patterns, as shown in Fig. 8.

The spatial distribution of the four 1^{st}-level types is visualized in Fig. 6(a), and the dynamic function signatures of the cluster centres

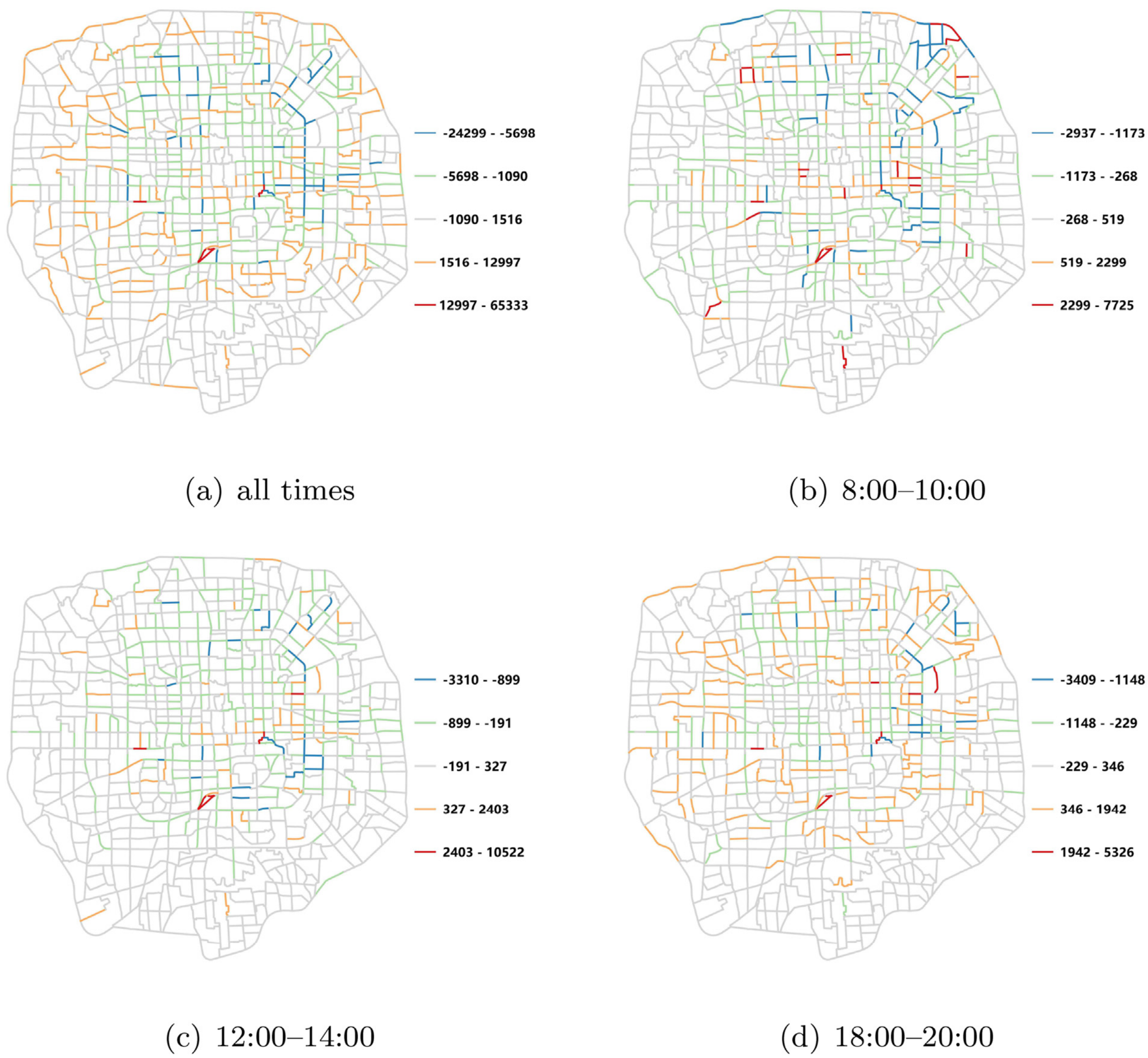
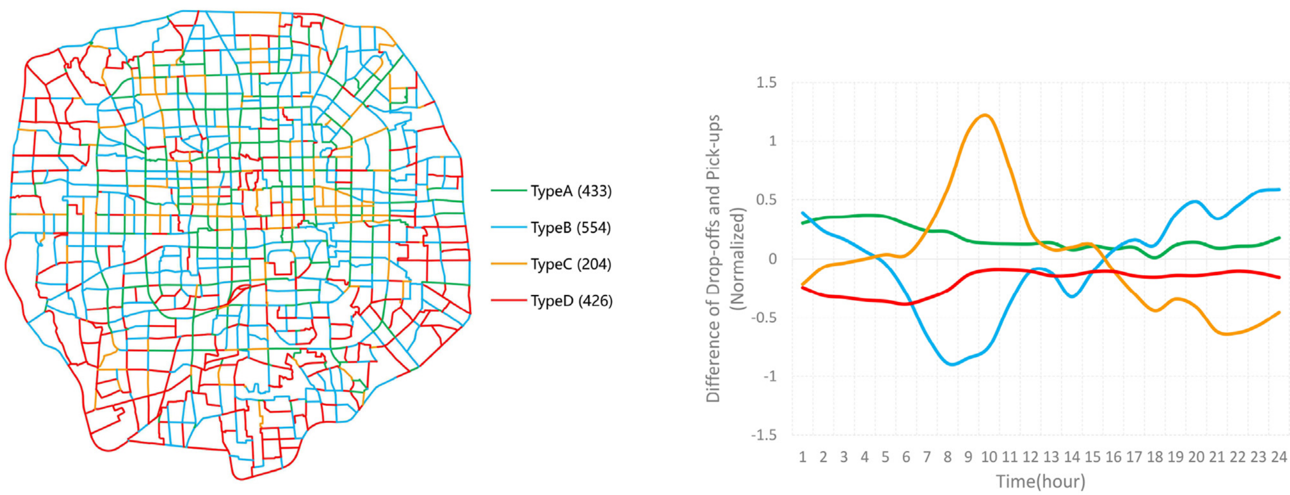


Fig. 5. Spatio-temporal patterns on streets for different time intervals. Red indicates that the drop-off number is much higher than the pick-up number on the street, while blue indicates the opposite. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are plotted in Fig. 6(b). These four street types exhibit different dynamic function patterns, and are referred to as type A (433 streets), B (554 streets), C (204 streets), and D (426 streets). Type A is plotted in green, and are mainly distributed separately within the 4^{th} Ring Road, especially along the ring roads and some traffic arteries. Type A streets are generally sources of taxi passengers where pick-up activities exceed drop-off activities in the daytime, with a weaker but similar pattern in the evening. In contrast, type D streets are shown in red, and are target streets where passengers are dropped off more than they are picked up for almost 24 h. Meanwhile, a type B street is a street where a strong pick-up peak occurs in the morning and a drop-off peak occurs in the evening; the spatial distribution pattern of type B is similar to that of type A, and is shown in blue in Fig. 6. Type C streets are typical streets where people arrive in the morning and leave at night, and are usually spatially clustered at some urban commercial and business districts (orange streets in Fig. 6(a)). What should be noted is that Step 1 clustering is based on the normalized vectors that represent the variation trend of the street's daily function, thus even if the dynamic volumes of differences between drop-offs and pick-ups are small and insignificant, streets may still be classified into a certain 1^{st}-level type, which need to be further identified in the Step 2 clustering.

Fig. 7 shows the results of the Step 2 clustering with the spatial distribution of each 2^{nd}-level type and dynamic capacity signatures V_C based on the absolute drop-off minus pick-up volumes. The brief descriptions and example streets for the final types are summarized in Table 3. Type A is decomposed into three sub-types: A1, A2, and A3, of which A1 is the principal contribution of the type A function pattern, A2's capacity is moderate, while A3 is almost insignificant (inactive). A1 streets are places where many residents choose to get into taxis, but not so many people to get off. By checking the real properties of these streets, we conclude that type A1 are mainly arterial streets in downtown areas and the ring roads of Beijing, such as the East 3^{rd} Ring Road, some segments on the North and West 3^{rd} Ring Road, and avenues such as Chengfu Road, Xuanwumen Outer Street, and Xizhimen South Street. Type A2 streets have more drop-off activities than A1 and show a stable negative small value, perhaps because of some tourist attractions nearby. Some typical A2 streets are Gulou Street (near place J) and



(a) (b)

Fig. 6. Illustration of the four basic types in the 1st-level clustering. (a) Spatial distribution of the four 1st-level street types; (b) Cluster centre of the four 1st-level types.

Yiheyuan Road (near place A). Similarly, we decompose types B, C, and D and interpret their three sub-types. Type B1 is the significant pattern of B, which has strong pick-up activity during the morning peak, and then people gradually come back in the evening. B1 streets are dispersed, and are usually residential, where people get into taxis to go to work in the morning. B2 streets are similar to B1 but the residential districts near B2 may be smaller. B3 is inactive because of the small volumes. Some special streets such as Gongti South Road and Gongti West Road (near place I) also exhibit B-type characteristics because many people are dropped off at the Gongti Stadium for sport events and concerts at night. Type C is relatively aggregated spatially compared with types A and B. C1 is the main component of C, with extremely large numbers of people in the morning, and people gradually leaving at night. C2 is weaker in capacity, but with the same function pattern, and C3 can also be treated as a noise type. C streets gather at multiple central business districts (CBD) such as Zhongguancun, Guomao, Xidan, and the National Olympic Stadium Centre (place A, H, E, K, respectively), and serve as traffic arteries for those CBDs. Type D has the most interesting pattern, with drop-offs consistently exceeding pick-ups, and drop-off peaks occurring both in the morning and evening; D1 and D2 are its significant components. Comparing Fig. 7(d) to Fig. 6(a), we can see that many red streets in Fig. 6(a) are classified into type D3, which is unimportant and can be left out. We can see that type D streets are actually small in number, and this pattern is found only for Gongti North Road (near place I) and some special streets near railway stations.

After the hierarchical clustering, the nine identified final street types in the Beijing urban area are illustrated in Fig. 8. We combine types A3, B3, C3, and D3 into type E because that their dynamic capacities and functions are weak with no obvious pattern exhibited. In fact, if we consider type E streets (amount to 901) to be not urban, it is interesting to point out that more than half of our street collection is not important in view of taxi O-D activities.

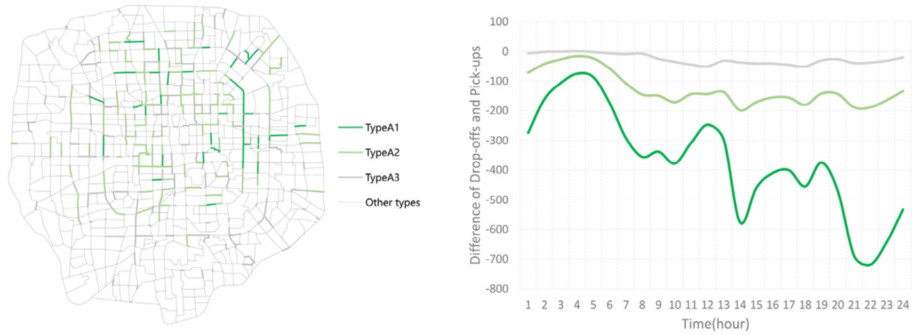
4. Discussion

4.1. Revisiting areal units in urban studies

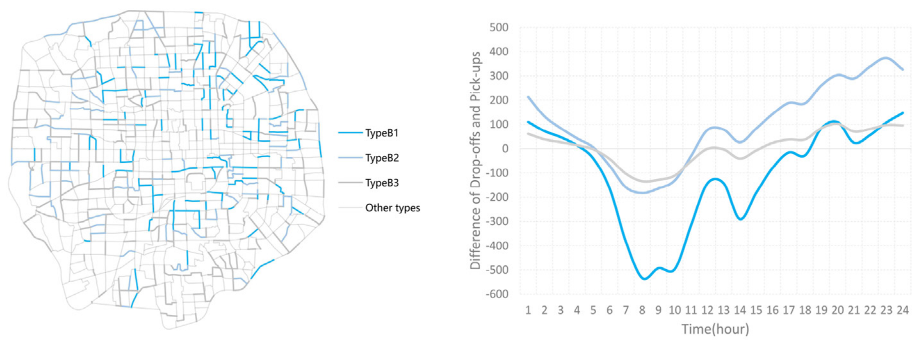
When areal units are decomposed into linear units, what happens? In Section 3.2, we analyse the spatial heterogeneity of transportation preference from the perspective of streets, and use this imbalanced feature to identify street types. For example, tourists tend to choose various transportation means such as subways, buses, and taxis to go to Peking University, but after sightseeing, a taxi might be the first choice because of fatigue. Since taxis are more accessible at Yiheyuan Road (the west side of Peking University), tourists are more likely to be picked up on this street, while other streets near Peking University do not exhibit this characteristic. In addition, people tend to take a taxi to go to Sanlitun bar street in the early evening, but leave using different means such as private cars or taxis at midnight because of specific demands. Thus, the drop-off and pick-up activity of taxis may be clustered on Gongti North Road, where the entry of the bar street lies. If we adopted traditional areal blocks as analysis units, taxi pick-ups and drop-offs at Peking University would be mapped onto one areal unit, and the variance of different block parts could be easily overlooked. Now, with the street unit, we treat a block as several streets, and each may have a specific function, thus giving us a fresh new angle to do urban research.

The example of the Sanlitun north district[5] is shown to explain how the street unit can supplement traditional areal units. Fig. 9(a) illustrates the landscape and street views for the four streets around the Sanlitun north district, and Fig. 9(b) displays the corresponding activity patterns. The four streets that encircle the district are Gongti North Road (red), Xindong Road (green), East 3rd Ring North Road (blue), and Dongzhimen Outer Street (yellow), which are the basic street units for us to understand the area. Gongti North Road is more active than the other three streets because it acts as the traffic artery for Sanlitun SOHO and Taikoo Li shopping mall (the dark grey buildings in Fig. 9(a)). In the morning, Gongti North Road is like type C1 streets in that it is a shopping/work destination. In the early evening, it exhibits a B1 pattern, in which people are dropped off, and young people shop or patronize bars until midnight or later, after which they get in a taxi or drive home (Fig. 9(b)). Different temporal activity patterns are associated

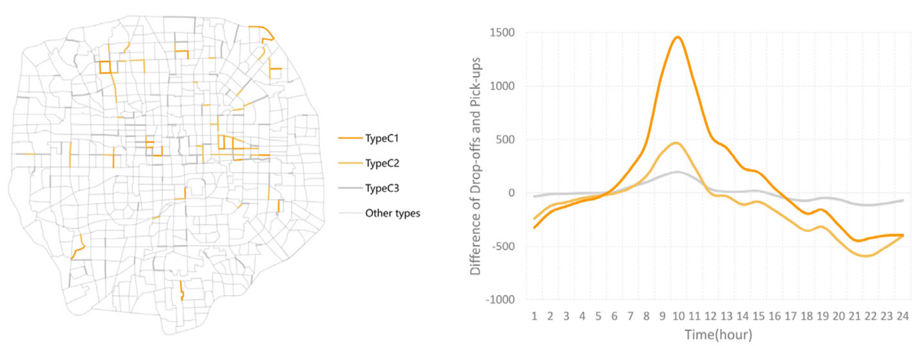
[5] The location of the Sanlitun north district is shown as the northeast block of place I in Fig. 1.



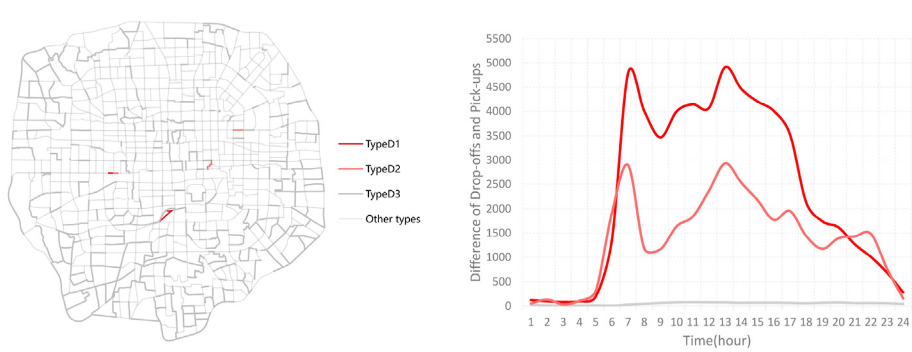
(a) Type A



(b) Type B



(c) Type C



(d) Type D

Fig. 7. Illustration of 2^{nd}-level clustering.

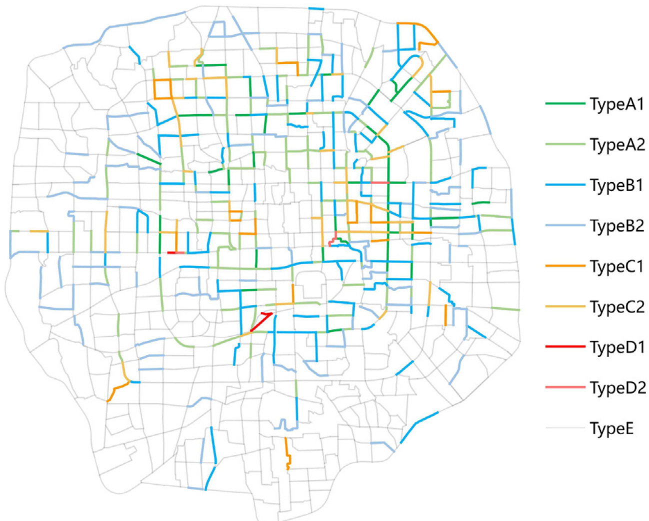
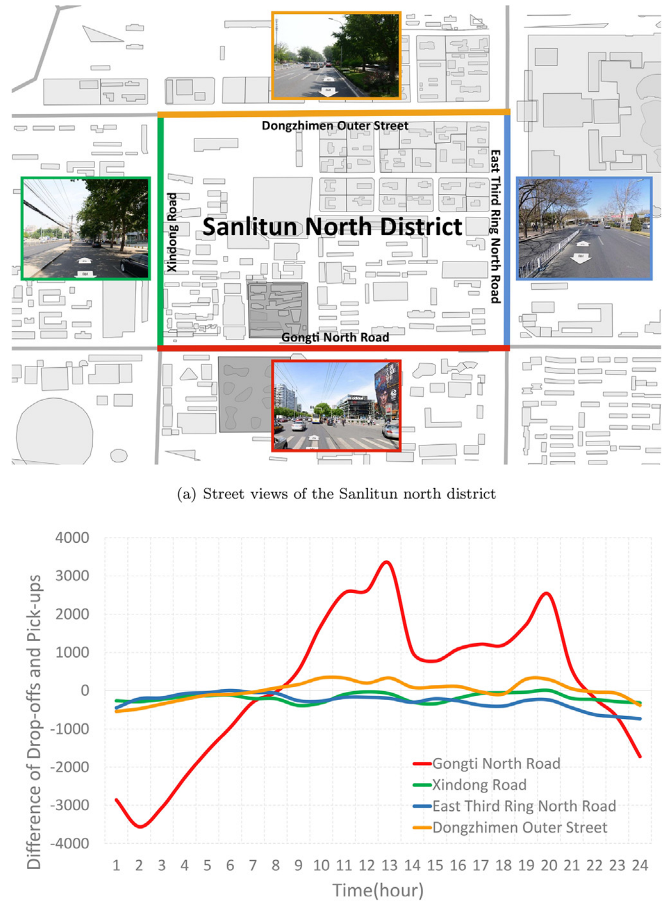


Fig. 8. Illustration of the nine final street types.



(a) Street views of the Sanlitun north district

(b) Temporal signatures of the four streets around the Sanlitun north district

Fig. 9. Decomposition of the Sanlitun north district.

with different street views, Fig. 9(a) shows similar street views for Xindong Road, East 3rd Ring North Road, and Dongzhimen Outer Street, which are wide and have green landscaping alongside them but are not so commercialized compared to Gongti North Road.

Considering that there are many if not extensive studies on the aggregation of spatial-temporal big data using areal units (Liu et al., 2012, 2015, 2016; Reades et al., 2009; Shi et al., 2015; Sun et al., 2011; Toole et al., 2012), the decomposition of the Sanlitun north district confirms our hypothesis that the street is a feasible and promising spatial unit for quantitative urban studies.

4.2. Uncovering urban structures in the street perspective

Discrete places in a city are stitched together by spatial interactions into an integrated network system (Batty, 2013b). For a spatial-embedded network, a community is a subset with relative dense connections, which means they are more spatially connected (Gao et al., 2013; Liu et al., 2014; Ratti et al., 2010; Thiemann et al., 2010). By connecting the origin and destination of a trip as an interaction flow, an interaction matrix that contains all trip flows can be built, and we can study urban structures by calculating the interaction intensity for the street-embedded interaction network.

In this street-embedded interaction network, which is different from the physical street network, each node represents a street segment, and the number of taxi trips between two streets is the weight of the edge that connects them. For a given weighted network, the modularity which is widely used for measuring how good a community division is, can be calculated as

$$Q = \frac{1}{2m}\sum_{i,j}\left(F_{ij} - \frac{k_i k_j}{2m}\delta(c_i, c_j)\right) \tag{9}$$

where m is the network's edge number, F_{ij} is the edge weight between node i and node j, and k_i and k_j are the sum weights of edges that are linked to node i and j; c_i and c_j are the community for node i and j, while $\delta(c_i, c_j)$ equals 1 when $c_i = c_j$, and 0 when $c_i \neq c_j$.

Here, we choose one of the most commonly used algorithms: the multilevel method developed by Blondel, Guillaume, Lambiotte, and Lefebvre (2008) to detect communities with the best modularity. We investigate the community structures for the Beijing urban area using interactions of taxi O-D trips (Fig. 10). The major

Table 3
Descriptions and example streets for the nine final street types.

Type	Street use description	Example streets[a]
A1	Traffic arterial streets	East 3rd Ring Road, Xuanwumen Outer Street, Xizhimen South Street
A2	Main roads near tourist attractions	Gulou Street, Yiheyuan Road
B1	Main streets for larger residential areas	Huixin West Street, Xisi North Street
B2	Main streets for smaller residential areas	Tsinghua East Road, Zhanchunyuan West Road
C1	Traffic arteries for central business districts	Zhongguancun Street, Guanghua Road, Xidan North Street
C2	Traffic arteries for secondary business districts	Zhongguancun South Street, Fuchengmen South Street
D1	Streets at railway stations	Lianhuachi East Road, Beijing South Railway Station Road
D2	Taxi drop-off streets for primary entertainment places	Gongti North Road (east part)
E	Insignificant	…

[a] Detailed information about these streets can be found on http://cn.bing.com/ditu/.

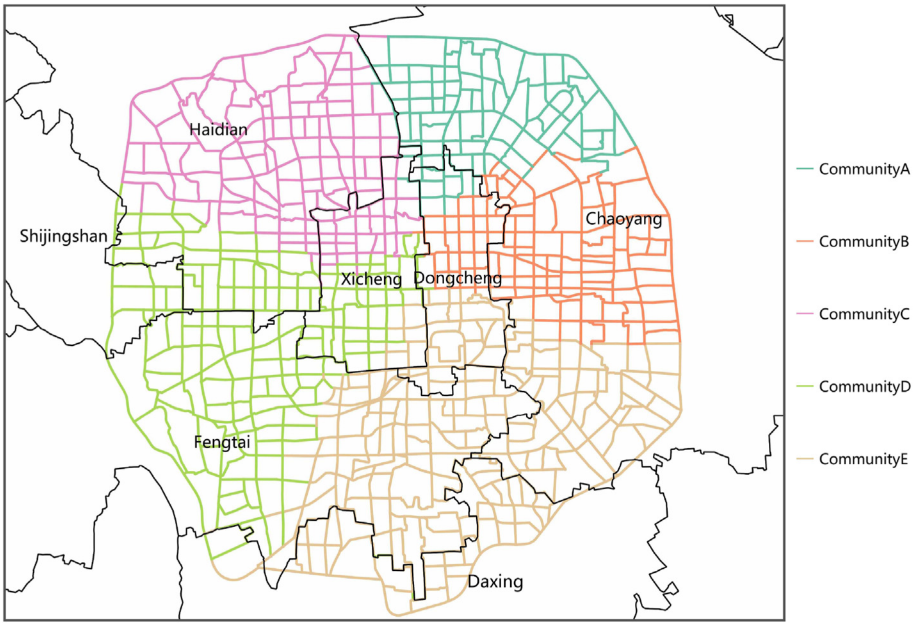


Fig. 10. Result of the community detection in Beijing urban area. We only show the detection result of taxi O-D trips for the entire 24 h (Q = 0.383) to demonstrate the feasible of uncovering urban structures with streets. In fact, we run the multilevel algorithm 30 times for each time interval (the entire 24 h, 8:00–10:00, 12:00–14:00, and 18:00–20:00) and find that the pattern of five communities is stable throughout the day, and the modularity ranges from 0.359 to 0.386.

district partitions Chaoyang, Dongcheng, Fengtai, Haidian, Shijingshan, Xicheng, and Daxing are delineated in Fig. 10 with black outline, and the communities are drawn using coloured linear unit (streets). We can see that Tiananmen Square acts as a dividing line for the western and the eastern parts of the city, and the natural separation between the north-eastern and south-western parts of the city is clear. The detected community partitions are similar to the administrative district partitions but not exactly the same, indicating the existence of inter-district connections. In the northern part of the city, Haidian has a strong connection with the northern part of Xicheng, and Chaoyang is connected with Dongcheng in the north-east. In the southern part of the city, inter-district interaction seems stronger, since Fengtai interacts with Xicheng and Haidian, and the southern part of Chaoyang is bonded

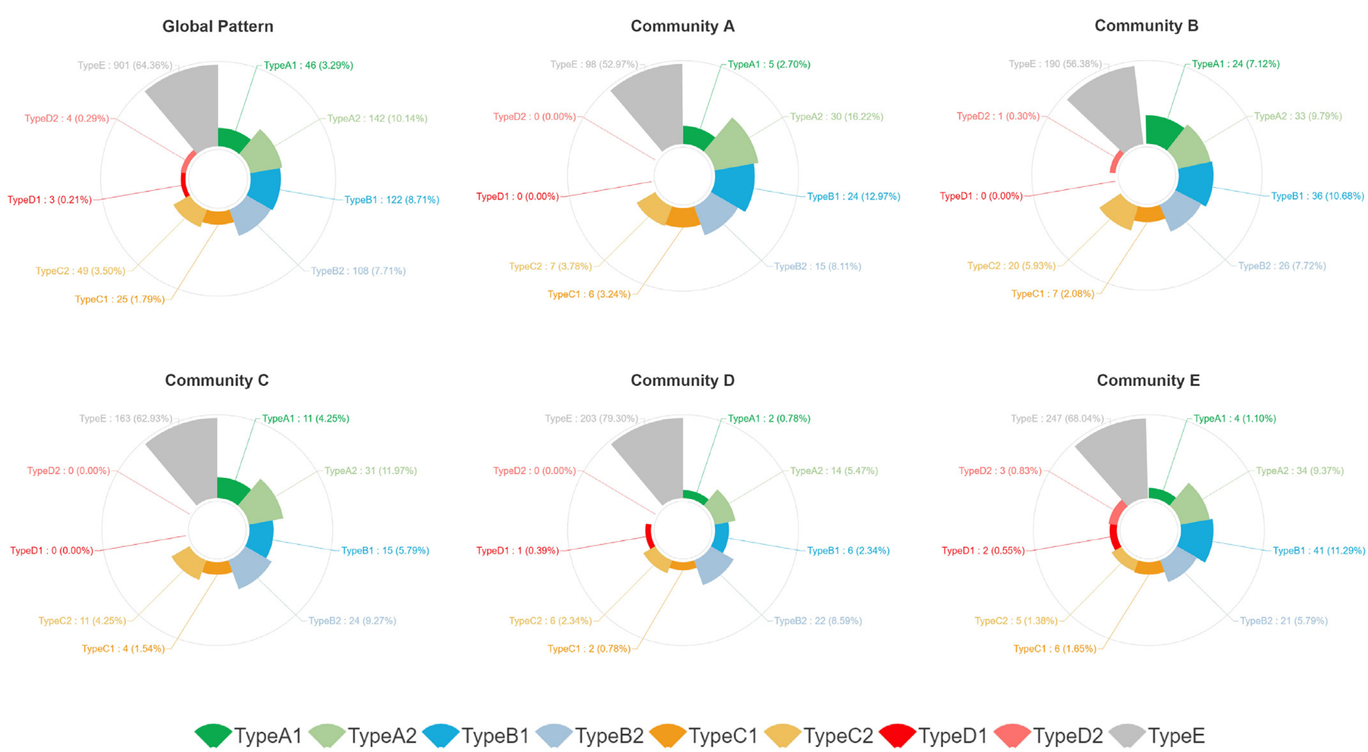


Fig. 11. The composition of street types.

with Dongcheng, Fengtai, and Daxing.

In Section 3.2.2, we associate urban streets types with different dynamic mobility, which indicate the individuals' activity. While the stable community patterns shown in Fig. 10 reflect the spatial interactions among places, imply the connections among individuals (Liu et al., 2015). To further find out the relationship between the street types and communities, we analyse the composition of street types in each detected community and the Nightinggale's Rose Diagram (Brasseur, 2005) is plotted for both global pattern and each community. In Fig. 11, the area of each sector represents the number of streets with that type. In general, the street composition patterns of the five communities are all similar to the global composition pattern, making each community an independent functional district. More than half the streets are type E, type A and type B are roughly the same proportion with 10% to 20%, type C1 and C2 add up to no more than 10% and only a few (less than 2% or even 0%) streets are type D. Using taxi O-D data, it seems that the Beijing urban area is naturally divided into five stable social communities with similar street type compositions. Inside one community, strong connections caused by the variance of street functions sustain the community's operation. While among communities, the homogeneous social functions segregate each other and lead to weak connections.

4.3. The complexity of streets

Although we assumed that streets are homogeneous in terms of their dynamic socio-economic function reflected by taxi O-D data, we have to admit the complex nature of streets. First, urban function locations, which act as the origins and destinations of human activities, comprise the complex land uses along streets (Shen & Karimi, 2016). Land use through streets is an important dimension to scrutinise in order to understand the details of street functions that are often overlooked in traditional studies on urban land use and accessibility (Geurs, Krizek, & Reggiani, 2012). For taxi data, the drop-offs and pick-ups on a given street may simply focus on some hot spots, while most parts of the street are not available for passengers to get on and off. The objective of our work requires us to regard each street as a homogeneous entity and neglect the fact that a street is composed of segments and several function locations, which can be MAUP on streets. Second, the bidirectional nature of streets is not considered in our work for simplicity. As a matter of fact, the left lane of a street segment can sometimes significantly differ from the right lane because of different planning purposes and development levels, and thus the adjacent areas may exhibit a rapid spatial change in geographic phenomena such as their socio-economic conditions (Liu, 2016).

5. Conclusions

Much existing literature about massive geo-spatial data and urban geography uses areal units for analyses, while neglecting the importance of streets. Incorporating taxi trajectory data and major streets in the Beijing urban area, we propose a street-based approach, introducing the street unit as a promising data assembly and analysis unit for quantitative urban studies. This method, on the one hand, is good at more accurately depicting human behaviours, and on the other hand, helps in solving MAUP to some extent.

We convert a coordinate-based taxi trip into a street-based trip by a simplified map-matching process so that the spatio-temporal patterns of street functions can be investigated using the pick-ups and drop-offs of taxi trips. The streets show an interesting distribution pattern in that the adjacent streets (even streets that enclose the same block) can vary greatly in socio-economic dynamic functions. Examining the similarities and differences among streets is useful for understanding how people utilize urban areas, and how streets are characterized according to human activities. We run a hierarchical bisecting k-means clustering and identify nine street types within Beijing's 5^{th} Ring Road, based on both the dynamic functions and capacities of the streets. Then we revisit traditional areal units by decomposing a typical block into four streets that encircle it, and the result shows that the street unit is more powerful in identifying human activity, since it is more compatible with public facilities and human mobility. Phenomena concealed by areal units maybe revealed if we analyse them in the street perspective. Meanwhile, as a spatial unit, streets can also be adopted to uncover urban structures. We run a multilevel community detection algorithm within the Beijing urban area and analyse the composition of street types in each community, the result is satisfactory and inspiring.

Streets are basic elements for people to recognize and utilize cities. Many important buildings are distributed along streets, and most human activities take place on streets. With the contribution of geo-spatial big data, not only taxi trajectories, other dataset such as social media check-ins and smart card data are also suitable for analysing how individuals commute and move along streets, thus enrich the properties of streets from multiple data perspectives. We suggest that street, as a novel spatial unit, is able to effectively minify MAUP, provide elaborate understanding of urban dynamic functions as well as urban structures, and is a supplement for existing research on urban big data. Sensing cities from streets is of significant value in quantitative urban studies and the street unit will also be useful for urban planners and government policy makers in optimizing public resources, promoting urban construction, relieving crowded areas, and many other socio-economic applications.

Acknowledgements

The authors would like to thank the anonymous reviewers for their constructive comments, which surely strengthened this article. This research was supported by the National Natural Science Foundation of China (Grant no. 41625003).

References

Batty, M. (2013a). Big data, smart cities and city planning. *Dialogues in Human Geography, 3*, 274–279.
Batty, M. (2013b). *The new science of cities*. MIT Press.
Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics Theory and Experiment, 2008*, 155–168.
Brasseur, L. (2005). Florence nightingale's visual rhetoric in the rose diagrams. *Technical Communication Quarterly, 14*, 161–182.
Briant, A., Combes, P.-P., & Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics, 67*, 287–302.
Castro, P. S., Zhang, D., Chen, C., Li, S., & Pan, G. (2013). From taxi GPS traces to social and community dynamics: A survey. *ACM Computing Surveys, 46*, 17.
Chai, Y. W., Weng, G. L., & Shen, J. (2008). A study on commercial structure of shanghai based on residents' shopping behavior. *Geographical Research, 27*, 897–906.
Foth, M., Choi, H. J., & Satchell, C. (2011). Urban informatics. In *Proceedings of the ACM 2011 conference on computer supported cooperative work* (pp. 1–8). ACM.
Frías-Martínez, V., Soto, V., Hohwald, H., & Frías-Martínez, E. (2012). Characterizing urban landscapes using geolocated tweets. In *Privacy, security, risk and trust (PASSAT)* (pp. 239–248). IEEE.
Gao, S., Wang, Y., Gao, Y., & Liu, Y. (2013). Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality. *Environment and Planning B: Planning and Design, 40*, 135–153.
Geurs, K. T., Krizek, K. J., & Reggiani, A. (2012). *Accessibility analysis and transport planning: Challenges for Europe and north America*. Edward Elgar Publishing.
Gong, Y., Lin, Y., & Duan, Z. (2017). Exploring the spatiotemporal structure of dynamic urban space using metro smart card records. *Computers, Environment and Urban Systems, 64*, 169–183.
Goodchild, M. F. (2010). Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0. *International Journal of Spatial Data Infrastructures*

Research, 2, 24–32.

Hao, J., Zhu, J., & Zhong, R. (2015). The rise of big data on urban studies and planning practices in China: Review and open research issues. *Journal of Urban Management, 4*, 92–124.

Hillier, B. (2012). Studying cities to learn about minds: Some possible implications of space syntax for spatial cognition. *Environment and Planning B: Planning and Design, 39*, 12–32.

Hillier, B., Leaman, A., Stansall, P., & Bedford, M. (1976). Space syntax. *Environment and Planning B: Planning and Design, 3*, 147–185.

Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology, 11*, 129–140.

Jokar Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. (2013). Toward mapping land-use patterns from volunteered geographic information. *International Journal of Geographical Information Science, 27*, 2264–2278.

Kang, C., Liu, Y., Ma, X., & Wu, L. (2012). Towards estimating urban population distributions from mobile call data. *Journal of Urban Technology, 19*, 3–21.

Kang, C., & Qin, K. (2016). Understanding operation behaviors of taxicabs in cities by matrix factorization. *Computers, Environment and Urban Systems, 60*, 79–88.

Liu, Y. (2016). Revisiting several basic geographical concepts: A social sensing perspective. *Acta Geographica Sinica, 71*, 564–575.

Liu, X., Kang, C., Gong, L., & Liu, Y. (2016). Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science, 30*, 334–350.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., et al. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers, 105*, 512–530.

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

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

Long, Y., & Liu, L. (2017). How green are the streets? An analysis for central areas of Chinese cities using tencent street view. *PLoS One, 12*, e0171110.

Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., & Huang, Y. (2009). Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 352–361). ACM.

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

Meentemeyer, V. (1989). Geographical perspectives of space, time, and scale. *Landscape Ecology, 3*, 163–173.

Mu, L., & Wang, F. (2008). A scale-space clustering method: Mitigating the effect of scale in the analysis of zone-based data. *Annals of the Association of American Geographers, 98*, 85–101.

Nellis, M. D., & Briggs, J. M. (1989). The effect of spatial scale on konza landscape classification using textural analysis. *Landscape Ecology, 2*, 93–100.

Okabe, A., & Kitamura, M. (1996). A computational method for market area analysis on a network. *Geographical Analysis, 28*, 330–349.

Okabe, A., & Okunuki, K.-i. (2001). A computational method for estimating the demand of retail stores on a street network and its implementation in GIS. *Transactions in GIS, 5*, 209–220.

Okabe, A., & Satoh, T. (2006). Uniform network transformation for points pattern analysis on a non-uniform network. *Journal of Geographical Systems, 8*, 25–37.

Okabe, A., & Satoh, T. (2009). Spatial analysis on a network. In *The SAGE handbook on spatial analysis* (pp. 443–464).

Okabe, A., Satoh, T., & Sugihara, K. (2009). A kernel density estimation method for networks, its computational method and a GIS-based tool. *International Journal of Geographical Information Science, 23*, 7–32.

Okabe, A., Yomono, H., & Kitamura, M. (1995). Statistical analysis of the distribution of points on a network. *Geographical Analysis, 27*, 152–175.

Openshaw, S. (1977). Optimal zoning systems for spatial interaction models. *Environment and Planning A, 9*, 169–184.

Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography, 38*.

Openshaw, S., & Taylor, P. J. (1979). A million or so correlation coefficients: Three experiments on the modifiable areal unit problem. *Statistical Applications in the Spatial Sciences, 21*, 127–144.

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S.-L., Li, T., & Zhou, C. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science, 28*, 1988–2007.

Penn, A. (2003). Space syntax and spatial cognition or why the axial line? *Environment and Behavior, 35*, 30–65.

Qi, G., Li, X., Li, S., Pan, G., Wang, Z., & Zhang, D. (2011). Measuring social functions of city regions from large-scale taxi behaviors. In *2011 IEEE international conference on pervasive computing and communications workshops* (pp. 384–388). IEEE.

Quddus, M. A., Ochieng, W. Y., Zhao, L., & Noland, R. B. (2003). A general map matching algorithm for transport telematics applications. *GPS Solutions, 7*, 157–167.

Ratti, C. (2004). Space syntax: Some inconsistencies. *Environment and Planning B: Planning and Design, 31*, 487–499.

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

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., et al. (2010). Redrawing the map of great britain from a network of human interactions. *PLoS One, 5*, e14248.

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

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

Shen, Y., & Karimi, K. (2016). Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities, 55*, 9–21.

Shi, L., Chi, G., Liu, X., & Liu, Y. (2015). Human mobility patterns in different communities: A mobile phone data-based social network approach. *Annals of GIS, 21*, 15–26.

Soto, V., & Frías-Martínez, E. (2011). Automated land use identification using cell-phone records. In *Proceedings of the 3rd ACM international workshop on MobiArch* (pp. 17–22). ACM.

Stehman, S. V., & Wickham, J. D. (2011). Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment. *Remote Sensing of Environment, 115*, 3044–3055.

Steinbach, M., Karypis, G., & Kumar, V. (2000). A comparison of document clustering techniques. In *Proceedings of the ACM SIGKDD international workshop on text mining* (pp. 525–526). ACM.

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

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

Toole, J. L., Ulm, M., González, M. C., & Bauer, D. (2012). Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing* (pp. 1–8). ACM.

Turner, A. (2007). From axial to road-centre lines: A new representation for space syntax and a new model of route choice for transport network analysis. *Environment and Planning B: Planning and Design, 34*, 539–555.

Turner, M. G., Dale, V. H., & Gardner, R. H. (1989). Predicting across scales: Theory development and testing. *Landscape Ecology, 3*, 245–252.

White, C. E., Bernstein, D., & Kornhauser, A. L. (2000). Some map matching algorithms for personal navigation assistants. *Transportation Research Part C Emerging Technologies, 8*, 91–108.

Wolf, J., Guensler, R., & Bachman, W. (2001). Elimination of the travel diary: Experiment to derive trip purpose from global positioning system travel data. *Transportation Research Record Journal of the Transportation Research Board, 1768*, 125–134.

Yin, L. (2017). Street level urban design qualities for walkability: Combining 2d and 3d gis measures. *Computers, Environment and Urban Systems, 64*, 288–296.

Yin, L., Cheng, Q., Wang, Z., & Shao, Z. (2015). 'Big data' for pedestrian volume: Exploring the use of google street view images for pedestrian counts. *Applied Geography, 63*, 337–345.

Yin, L., & Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and google street view imagery. *Applied Geography, 76*, 147–153.

Yu, W., Ai, T., He, Y., & Shao, S. (2016). Spatial co-location pattern mining of facility points-of-interest improved by network neighborhood and distance decay effects. *International Journal of Geographical Information Science, 31*, 280–296.

Yuan, Y., & Raubal, M. (2012). Extracting dynamic urban mobility patterns from mobile phone data. *Lecture Notes in Computer Science, 7478*, 354–367.

Yuan, N. J., Zheng, Y., Zhang, L., & Xie, X. (2013). T-finder: A recommender system for finding passengers and vacant taxis. *IEEE Transactions on Knowledge and Data Engineering, 25*, 2390–2403.

Zhou, Y., & Long, Y. (2016). Sinogrids: A practice for open urban data in China. *Cartography and Geographic Information Science, 43*, 379–392.