Environment and Planning B: Planning and Design 2013, volume 40, pages 135 – 153

doi:10.1068/b38141

Understanding urban traffic-flow characteristics: a rethinking of betweenness centrality

Song Gao, Yaoli Wang, Yong Gao, Yu Liu¶
Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China; e-mail: songgaogeo@gmail.com, wangyaoli0810@yahoo.com.cn, gaoyong@pku.edu.cn, liuyu@urban.pku.edu.cn
Received 14 August 2011; in revised form 13 February 2012

Abstract: In this study we estimate urban traffic flow using GPS-enabled taxi trajectory data in Qingdao, China, and examine the capability of the betweenness centrality of the street network to predict traffic flow. The results show that betweenness centrality is not a good predictor variable for urban traffic flow, which has, theoretically, been pointed out in existing literature. With a critique of the betweenness centrality as a predictor, we further analyze the characteristics of betweenness centrality and point out the 'gap' between this centrality measure and actual flow. Rather than considering only the topological properties of a street network, we take into account two aspects, the spatial heterogeneity of human activities and the distance-decay law, to explain the observed traffic-flow distribution. The spatial distribution of human activities is estimated using mobile phone Erlang values, and the power law distance decay is adopted. We run Monte Carlo simulations to generate trips and predict traffic-flow distributions, and use a weighted correlation coefficient to measure the goodness of fit between the observed and the simulated data. The correlation coefficient achieves the maximum (0.623) when the exponent equals 2.0, indicating that the proposed model, which incorporates geographical constraints and human mobility patterns, can interpret urban traffic flow well.

Keywords: taxi trajectory, traffic flow, betweenness centrality, mobile phone data, spatial heterogeneity, distance decay

1 Introduction

With the rapid development of information and communication technology (ICT), the availability of large amounts of GPS (Global Positioning System) data and mobile phone data over time and space has increased the capability for monitoring, visualizing, analyzing, and modeling urban dynamics (Ahas et al, 2010; Kang et al, 2012; Li et al, 2011). The GPS-enabled taxi trajectory data is one type of floating car data (FCD), which make it possible to obtain real-time traffic statuses; therefore, quantitative analysis and modeling of traffic flow has become a hot topic in both transportation research and geographical information science (Bar-Gera, 2007; de Fabritiis et al, 2008; Herrera et al, 2010; Jiang, 2009; Kerner et al, 2005; Kobayashi et al, 1999; Lorkowski et al, 2005; Rose, 2006). At the same time, since the movements of vehicles are under the restraint of the street network, much research has made great efforts in investigating the characteristics of street networks in order to understand traffic flow patterns. Traditionally, centrality measures have been instrumental in describing essential properties of networks, where the centrality determines the relative importance of a vertex within the graph (Bonacich, 1987; Freeman, 1977; Freeman et al, 1991; Newman, 2005). These measures originate from structural sociology and have also been applied widely in studying complex networks, urban structures, and transportation (Borgatti, 2005; Crucitti et al, 2006a; 2006b; Holme, 2003; Porta et al, 2006a; 2006b). From the view of the

¶ Corresponding author.

'space syntax' community, the configuration of a city's street network plays an important role in movement and vehicular flow (Hillier et al, 1993; Penn et al, 1998). The structural and morphological properties of a street network, represented in topological or geographic metric measurements, are considered to be the key factors that shape dynamic urban traffic flow (Hillier et al, 1993; Jiang and Claramunt, 2004; Jiang and Liu, 2009; Jiang et al, 2008; Kazerani and Winter, 2009a; 2009b; Turner, 2007). However, there is a wide spectrum of opinion on the use of street centrality measures to predict urban traffic flow. For instance, Hillier et al (1993) argued that integration (a topological step-distance measure similar to normalized closeness centrality) for the axial representation of a street network exhibits a higher potential for predicting human movement than the connectivity measure. Turner (2007), nevertheless, challenged this point by demonstrating that betweenness centrality, computing what percentage of shortest paths of all vertex pairs in a network pass through a particular node, in a new angular segment analysis based on street center line is a better index. Jiang and Liu (2009) further found that the street-based topological representation can predict vehicle flow better than conventional axial lines. They obtained high predictability for streets taken from small sample areas of the entire map. However, the global measures are still not ideal. In most existing research, street segments are merged according to names or continuity (Jiang and Liu, 2009). This representation may neglect the spatially inconsistent context of different street segments. For instance, a very long street with many intersections would take various traffic-flow volumes in different parts, even though they share the same name or have good continuity. Instead, we suggest that the urban street network as a geographic network should be constructed naturally from intersections, which are more easily perceived and correlate well with geographical factors such as human activities and land use intensity (Crucitti et al, 2006b; Porta et al, 2009; Wang et al, 2010). Kazerani and Winter (2009a) later argued that betweenness centrality is not a suitable measure for predicting traffic flow in static urban street networks because it neglects the dynamics of travel behavior, and thus they proposed a modified measure but unfortunately there is a lack of real dataset support (Kazerani and Winter, 2009b).

Urban traffic flow can be seen as individual trips aggregately distributed in street networks. Each trip is generated from an origin and destination (OD) pair or multiple destinations (not available in the behavior of taking a taxi) with a network path connecting them. Generally, urban planners and transportation engineers rely on household questionnaires or transportation surveys on job–housing places for traffic-demand forecasting (McFadden, 1974). Recently, sensor-based OD estimation methods have been well developed (Calabrese et al, 2011). There is much literature about the importance of urban structures and human behavior in shaping urban transportation demand (Black, 2003; Chen et al, 2009; Couclelis, 1986; Goodchild and Janelle, 1984; Goodchild et al, 1993; Ratti et al, 2006; Sevtsuk and Ratti, 2010; Wang, 2001). According to some studies, geographical constraints in urban space really matter in predicting that individuals' motion and the trip lengths follow a power law, or a truncated power law, which indicates the distance decay (Brockmann et al, 2006; González et al, 2008; Song et al, 2010). Therefore, the framework to explain the observed traffic flow should bridge the gap between static street-network centrality and dynamic traffic demands derived from human activities in real urban space.

In this research we investigate the spatial distribution of urban traffic flow based on taxi trajectories in Qingdao, China and compute the correlation between urban traffic flow and street betweenness centrality. The results confirm that purely betweenness centrality is not an ideal measure for predicting urban traffic flow. We analyze the characteristics of betweenness and point out the 'gap' between betweenness centrality and actual flow. The gap is filled in steps, by a framework taking into account the spatial heterogeneity of human activities, which is measured by the Erlang values of mobile phone calls, and the distance-decay law.

The correlation between simulated and real traffic flow shows that both aspects play important roles in urban traffic flow, thus validating the proposed method.

This paper is organized as follows. The second section introduces the study area and the datasets used in this research, and we investigate the spatial and statistical characteristics of traffic flow in this urban area. In the third section, the correlation between traffic patterns and betweenness centrality of the street networks is computed, and the gap behind the low correlation coefficient is thoroughly analyzed. In the fourth section we introduce the spatial heterogeneity of human activities and the distance-decay law to model actual traffic flow generation using Monte Carlo simulations. The proposed method can be viewed as an extension of existing betweenness measures but with a higher capability to predict urban traffic flow. In the fifth section, we discuss further the weakness of betweenness and identify several features for improving urban traffic-flow prediction. The final section contains our conclusion.

2 Characteristics of Qingdao traffic flow

2.1 Study area and data description

Qingdao, a picturesque coastal city enclosing Jiaozhou Bay on the southern tip of Shandong Province in China, is selected as the study area. We focus on its major urban area on the eastern coast of Jiaozhou Bay (figure 1). The core urban area with the densest commercial and human activity is located in the southern part of this study area.

Qingdao is a big city without a subway system. Thus, buses and taxis play an important role in its urban transportation system. Many taxi companies have equipped their taxis with GPS receivers so that they can monitor the operation of each taxi. GPS data records include device identification (ID), date, time, longitude, latitude, velocity, and driving direction (table 1), and such records are collected automatically in approximately 10 s during the day and 300 s at night. The large volumes of taxi trajectory data are valuable for departments of urban management in obtaining real-time traffic statuses and in improving transportation planning and management. In this research we employ a one-week-long dataset of 149 taxis, containing 1261 475 taxi records, for calculating the real traffic volumes in Qingdao.

The street network of Qingdao is also used for map matching of the trajectory data. For a street network, we can adopt primal or dual representations (Jiang and Claramunt, 2002; 2004; Porta et al, 2006a; 2006b). In the primal representation (table 2), as a geographic metric each street intersection is transformed into a node, while a street segment is represented by an edge, where the length of the segment can be measured as the edge weight. In the dual representation, each street segment is transformed into a node, while a street intersection, which may connect two or more street segments, establishes links between nodes with topological path steps as distance. After editing, the resulting street network contains 3179 nodes and 5095 edges for the primal approach, and 5095 nodes and 12 293 edges for the dual approach. Note that we keep the natural spatial structures without merging street segments into named streets (Jiang and Liu, 2009).

2.2 Spatial and statistical distributions of Qingdao traffic flow

In this study we adopt the line-density method (Jiang, 2009; Scheepens and Willems, 2011) to compute real traffic-flow volumes according to the GPS-enabled taxi trajectory data over the seven days. The line-density method calculates the density of taxi trajectory lines by connecting temporally neighboring GPS points into a continuous line. We use $\boldsymbol{F}_C$ for the estimated traffic flow and the spatial distribution of $\boldsymbol{F}_C$ is depicted in figure 2. It can be recognized that the southern part of Qingdao plays a more significant role with higher traffic volumes, especially on several east–west streets (in red) on the map, while a few north–south streets (in yellow) are essential for connecting the urban center to the airport in the northern part of the city (see figure 1). Additionally, the main streets (eg, expressways and avenues)

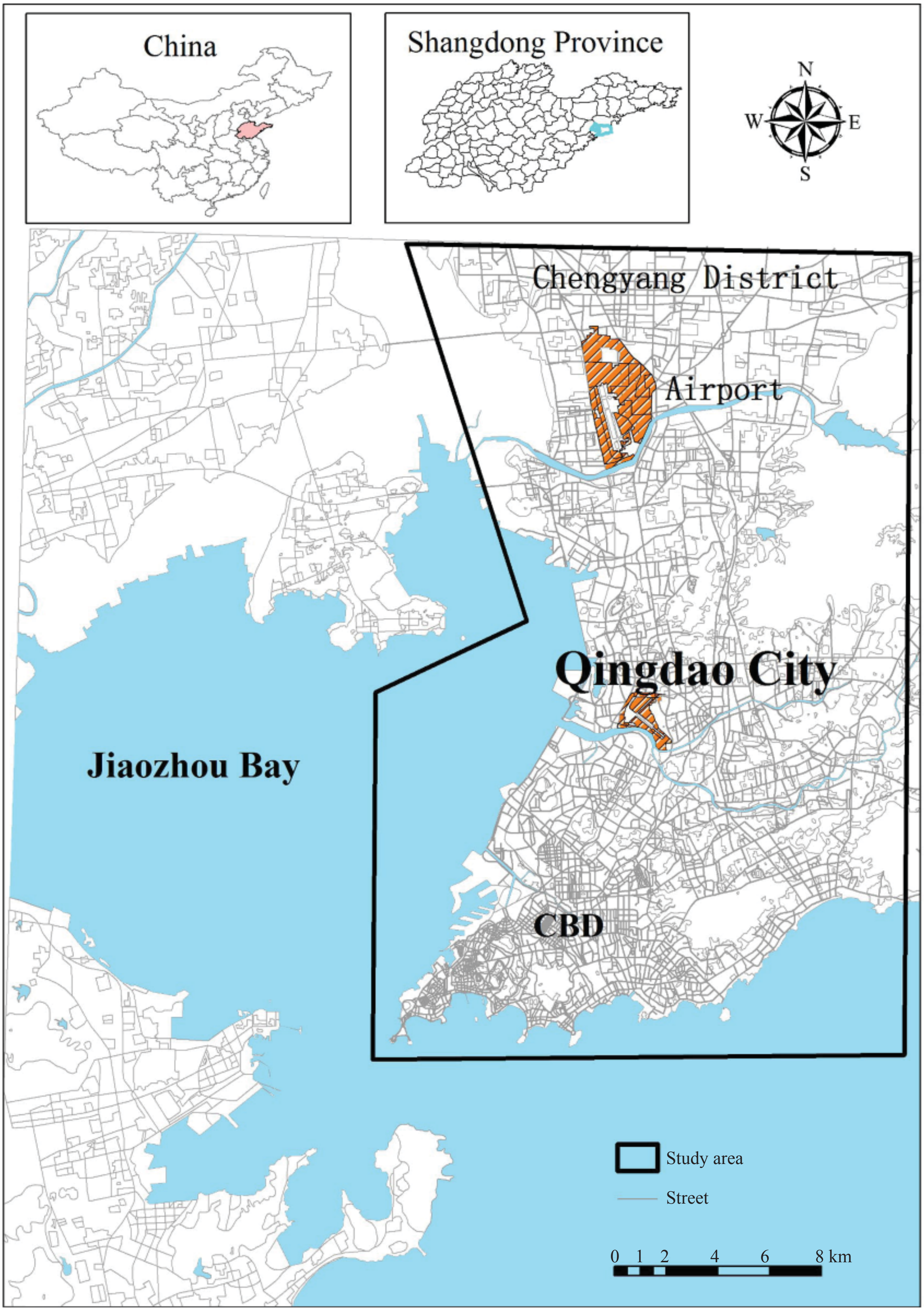


Figure 1. [In color online.] Urban area of Qingdao and the street network.

Table 1. Data schema of taxi GPS trajectories collected in Qingdao

Device identification	Date	Time	Longitude	Latitude	Velocity (km/h)	Orientation (°)
30**** 1954	2009/3/13	09:25:45	120.384	36.096	29	88
30**** 1954	2009/3/13	09:25:53	120.385	36.096	34	87

Table 2. Data schema of the primal representation of the street network.

Street	FNODE	TNODE	StartX	StartY	EndX	EndY	Length (m)
S0	N1	N2	533282.9	3994672.9	533026.3	3994532.5	292.3
S1	N3	N4	525997.2	3991700.8	526166.1	3992212.9	866.6

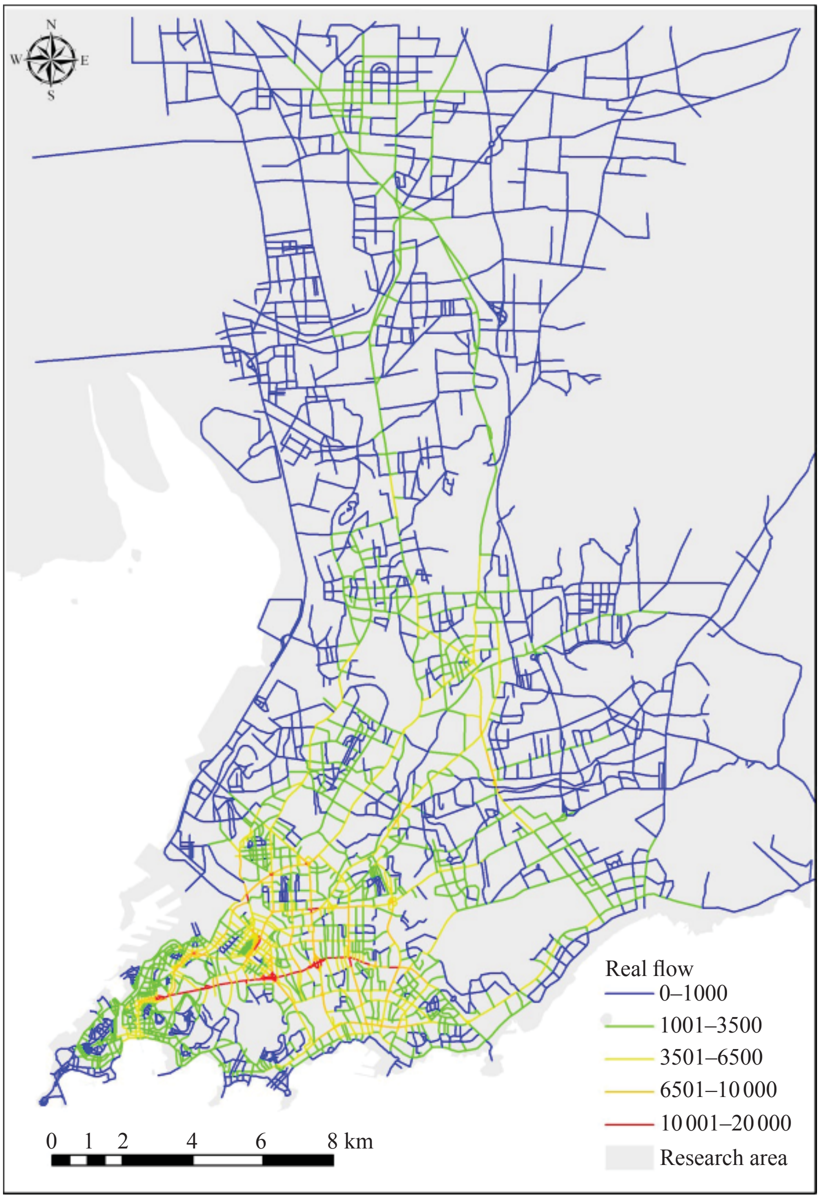


Figure 2. [In color online.] Qingdao real traffic flow over one week.

account for more traffic flow than the secondary streets (eg, alleys), indicating that both urban space and street network have a hierarchical structure, which would impact on residents' movements and drivers' pathfinding behaviors.

In previous studies, urban street networks, represented as named streets and deflection-angle threshold-based natural streets, are considered to be hierarchically selforganized (Crucitti et al, 2006a; Jiang, 2009). We also examine quantitatively the statistical properties of the observed traffic flow based on the intersection–segment representation of the street network and confirm that most streets in Qingdao have a low volume of traffic while streets with a high traffic volume make up only a small proportion of all streets. More specifically, we find that the top 20% of streets (see figure 3) account for approximately 55% of the traffic volume (and the top 40% of streets account for 80% of the traffic volume) in the cumulative distribution of observed taxi traffic flow.

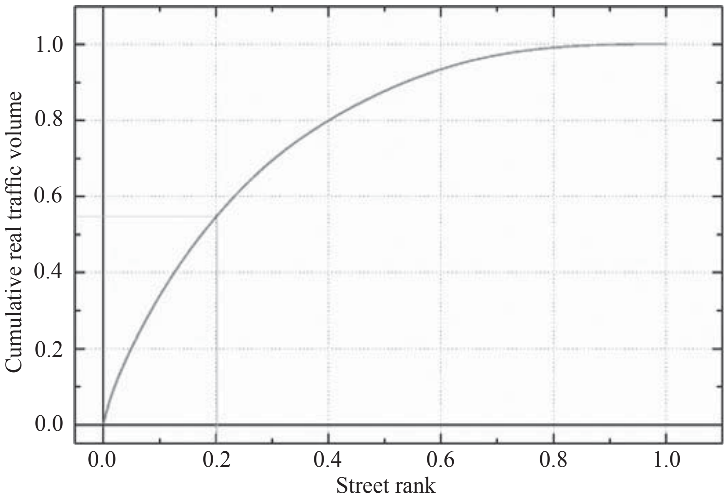


Figure 3. Cumulative distribution of the observed taxi traffic flow with ranked streets.

3 Correlation and gap between traffic flow and street betweenness centrality

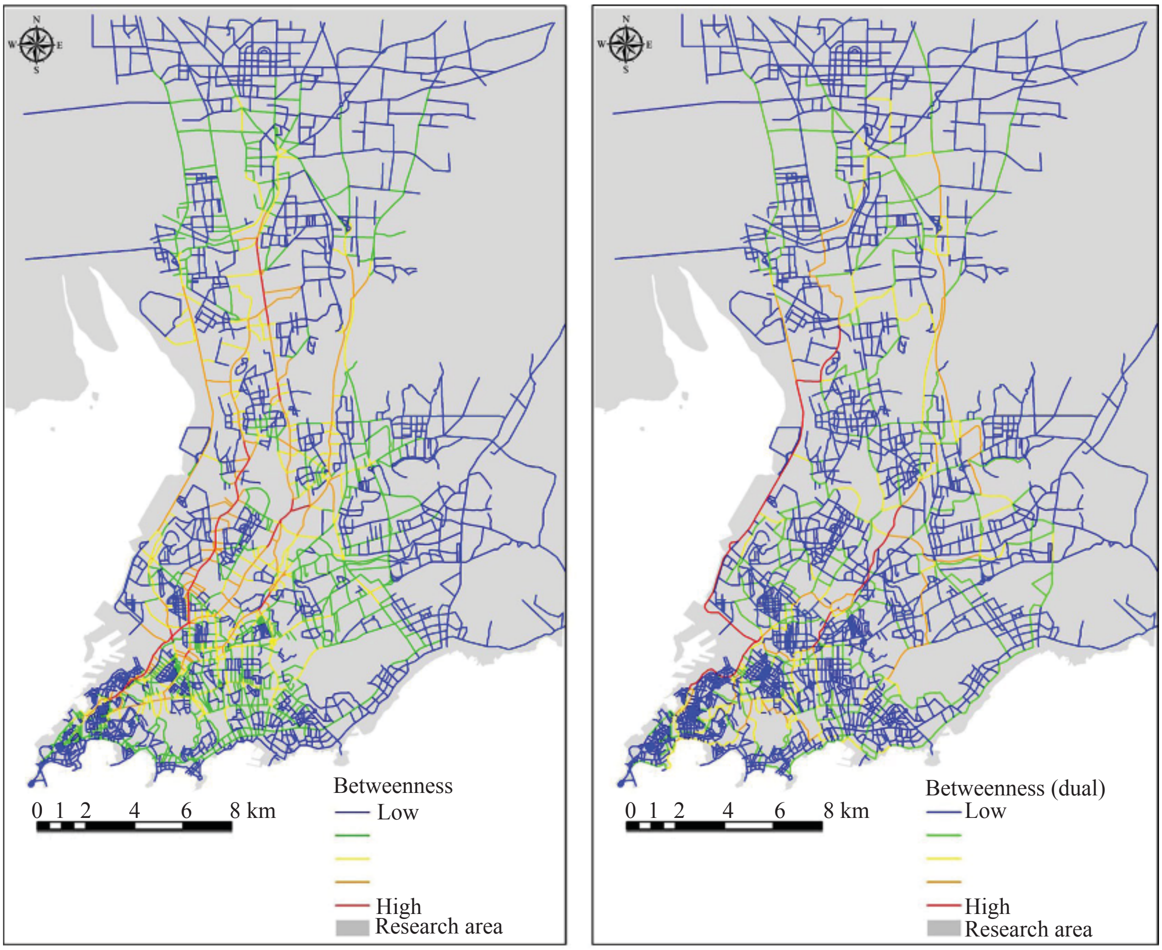
3.1 Correlation between traffic flow and betweenness

Among the collection of centrality measures, betweenness centrality represents the global importance of a node in connecting others together by through-movement, also called global choice. For a given node i in a network containing N vertices, the betweenness centrality is defined as:

$$C_i^B = \sum_{j=1;\, k=1;\, i \neq j \neq k}^{N} \frac{n_{jk}(i)}{n_{jk}} , \qquad (1)$$

where n_{jk} denotes the total number of shortest paths between nodes j and k, and $n_{jk}(i)$ is the number of shortest paths between j and k passing node i. Generally, C_i^{B} represents the load on a given node i in the network. There is much literature on the relationship between betweenness centrality and observed traffic flow (Jiang and Liu, 2009; Kazerani and Winter, 2009a; 2009b; Turner, 2007).

In this study, we try to correlate the observed flow with both the primal representation (Porta et al, 2006a) and the dual representation of street networks (Jiang and Claramunt, 2004; Porta et al, 2006b). For the primal representation, the betweenness measures ($C_{\mathrm{P}}^{\mathrm{B}}$) are calculated using geometric distances for the shortest path, and the betweenness of a street segment is the mean of the values for the two end nodes [figure 4(a)]. Meanwhile, we use the topological distances to directly compute the betweenness for each street in dual representation, denoted by $C_{\mathrm{D}}^{\mathrm{B}}$ [Figure 4(b)].



(a) (b)

Figure 4. [In color online.] Betweenness centrality of Qingdao street network. (a) Primal representation ($\boldsymbol{C}_{\mathrm{P}}^{\mathrm{B}}$); (b) dual representation ($\boldsymbol{C}_{\mathrm{D}}^{\mathrm{B}}$).

The street betweenness centrality of Qingdao is shown in Figure 4, which highlights the north–south directional streets that connect the city center in the south with a major satellite town, Chengyang, in the north (cf figure 1), where the airport is located. The through movement is well represented by betweenness centrality. Unfortunately, when computing the correlation coefficients with the observed traffic flow, betweenness centrality is not satisfactory.

Considering that the street lengths vary in urban space, we adopt a weighted correlation coefficient (WCC) to measure the goodness of fit between the observed traffic and betweenness. The WCC measure is defined as:

$$r = \frac{\mathrm{cov}_w(\boldsymbol{C}^{\mathrm{B}}, \boldsymbol{F})}{[\mathrm{cov}_w(\boldsymbol{C}^{\mathrm{B}}, \boldsymbol{C}^{\mathrm{B}})\,\mathrm{cov}_w(\boldsymbol{F}, \boldsymbol{F})]^{1/2}}\,, \tag{2}$$

where $\boldsymbol{F}$ is the observed traffic flow, $\boldsymbol{C}^{\mathrm{B}}$ is betweenness and cov_w is the weighted covariance defined as:

$$\mathrm{cov}_w(x, y) = \frac{\sum_i w_i [x_i - m_w(x)][x_i - m_w(x)]}{\sum_i w_i}\,, \tag{3}$$

and

$$m_w(x) = \frac{\sum_i w_i x_i}{\sum_i w_i} . \qquad (4)$$

In this research we represent the length of the ith street segment as w_i, because a segment with a longer length should contribute more to the total correlation than a shorter one does, despite the fact that they may have the same traffic volume. The correlation coefficients between $\boldsymbol{F}_C$ and $\boldsymbol{C}_{\mathrm{P}}^{\mathrm{B}}$ and between $\boldsymbol{F}_C$ and $\boldsymbol{C}_{\mathrm{D}}^{\mathrm{B}}$ are 0.416 and 0.186, respectively. The correlations are both significant at the 0.01 level (two-tailed). This is consistent with the finding by Jiang and Liu (2009) that the global measure of dual representation is poor. However, they did not figure out why the weakness exists; instead their further work was restricted to some small sample areas and obtained a higher correlation value. The question of how to better explain the traffic-flow distribution of the entire urban space is still not explicitly stated. When comparing the spatial distributions of real traffic flow with betweenness (figures 2 and 4), we find that it was not suitable to adopt purely betweenness as a measure to predict traffic flow. The spatial patterns are quite distinct. The real flow stresses the southern part of the city, the core urban area of Qingdao, while betweenness highlights the north–south streets linking the two major parts of Qingdao but fails to identify the core urban area with high traffic volumes. Moreover, betweenness centrality is not a stable indicator since it varies hugely with the change in shape and range of the study area (Kazerani and Winter, 2009a). For these reasons we introduce Monte Carlo simulations to find the gap between traffic flow and betweenness centrality and to understand traffic flow based on a framework with geographical background constraints.

3.2 The gap between traffic flow and betweenness centrality

The unsatisfactory correlation analysis of betweenness as a predictor for traffic flow is not surprising. It gives us cause to reflect on the gap between traffic flow and betweenness and rethink about this measure in depth. As shown in figure 5(a), the street network includes six nodes ($n_1, \ldots, n_6$) and seven segments ($e_1, \ldots, e_7$). Table 3 lists different betweenness measures for the synthetic network. We can imagine a number of trips with particular OD points distributed in the network. According to equation (1), two conditions should be met if, in reality, the observed flow is to be positively correlated with betweenness: (1) the distribution of OD points is identical to that of network nodes; and (2) the access probabilities from one node to all other nodes are equal, such that all shortest paths are traversed the same number of times. As such, two reasons may explain the gap between observed traffic flow and betweenness centrality.

The first consideration is the spatial distribution of the OD points. When calculating betweenness, start and end points are network nodes in primal representation. For a street network, OD points are associated with streets instead of nodes. Hence, much research focuses on the dual representation to interpret the observed traffic flow [figures 5(b) and 5(c)] (Jiang and Liu, 2009; Kazerani and Winter, 2009a). According to dual representation, all street segments (or natural streets) are viewed as identical units. However, this is not the case in trip generation. As shown in figure 5(d), the shortest paths from different points on the same street to the same destination are different (see S_1 to n_1, and S_2 to n_2), indicating that a street segment is not an appropriate analysis unit for predicting traffic flow. Additionally, it is natural that streets with different lengths have various trip demands and longer streets are generally associated with more trip demands. We may further extend the consideration on trip-demand distribution to construct a model more akin to real trip generation. Considering that

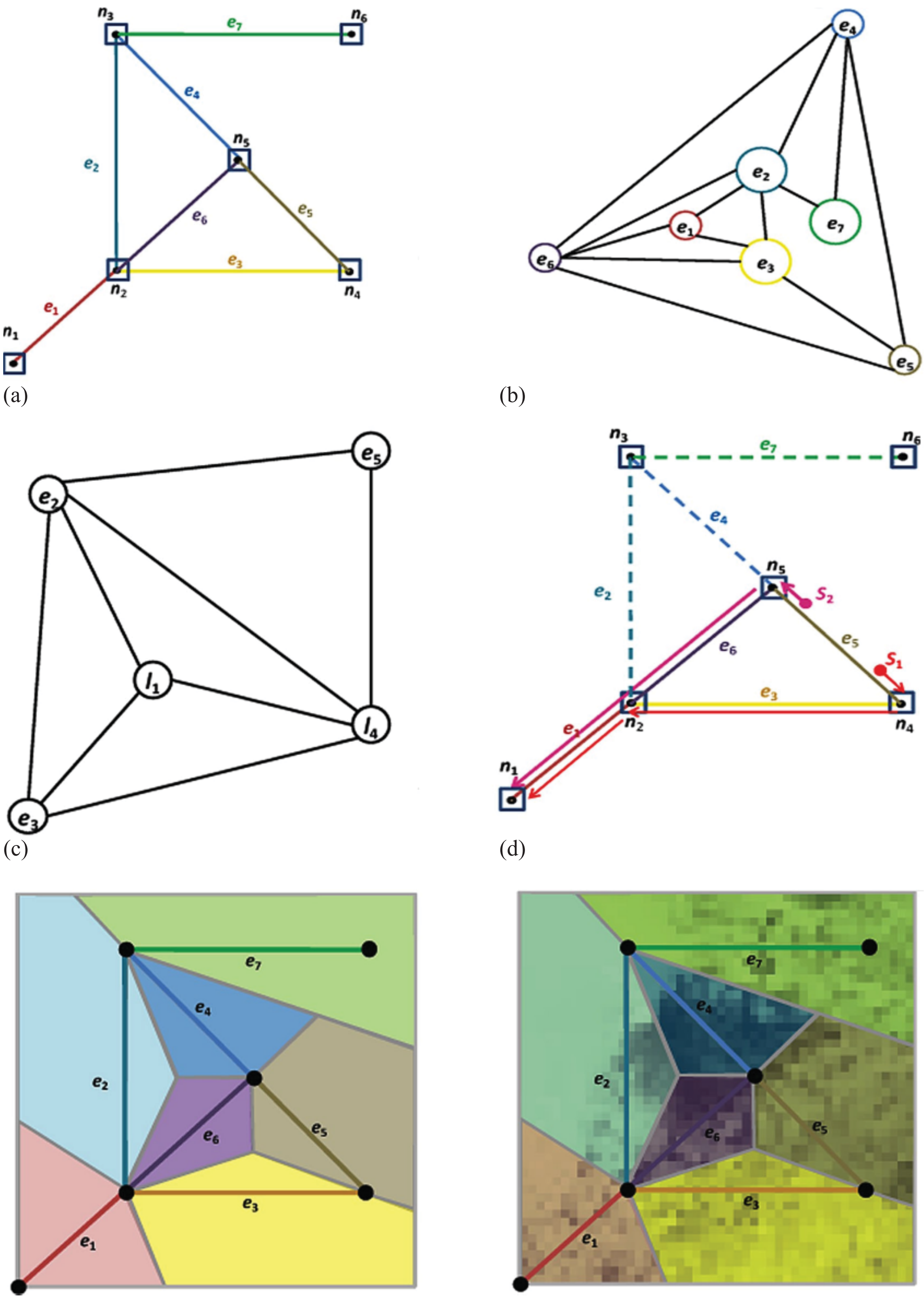


Figure 5. [In color online.] (a) Primal representation of the street network; (b) dual representation based on street segments; (c) dual representation based on natural streets, where l_1 consists of e_1 and e_6, and l_4 consists of e_4 and e_5 [where e_1, e_6, e_4, and e_5 are shown in figure 5(a)]; (d) different parts of one street play different roles in trip generation; (e) Voronoi polygons of street segments; (f) Voronoi polygons with geographical heterogeneity.

Table 3 Betweenness centrality measures for different representations of a synthetic street network.

Intersection-based street	Primal betweenness (geographic distance)	Dual betweenness (street segment)	Dual betweenness (natural street)
e_1 (n_1–n_2)	5.00	0.00	0.00[a]
e_2 (n_2–n_3)	4.00	3.33	1.00
e_3 (n_2–n_4)	2.00	0.83	0.00
e_4 (n_3–n_5)	4.00	1.83	1.00[a]
e_5 (n_4–n_5)	3.00	0.33	1.00[a]
e_6 (n_2–n_5)	2.00	1.67	0.00[a]
e_7 (n_3–n_6)	5.00	0.00	0.00

[a] Calculated using l_1 and l_4 in figure 5(c).

OD points are distributed in areas instead of along lines, a Voronoi polygon for each street segment can be obtained so that an origin (or a destination) inside a polygon should be captured by its nearest street segment [figure 5(e)]. If the OD points are uniformly distributed in the study area, the travel demand of a segment will be positively proportional to the corresponding polygon area. Additionally, the potential OD distribution is not uniform and is positively proportional to the distribution of activity density, which is influenced by many factors such as land uses (McFadden, 1974; Wang et al, 2010). Thus, we can introduce a spatial variable for the geographic heterogeneity of potential OD points, and the weighted areas can be used to simulate travel demands [figure 5(f)]. Hence, figure 5(f) depicts a model akin to real trip generation.

Following figure 5(f), we can randomly generate a number of OD pairs to simulate traffic flow. The betweenness centrality implies that the shortest paths of all OD pairs are traversed equally. Due to the distance-decay effect, however, the number of long-distance trips should be less than that of short-distance trips. Hence, the second aspect that we incorporated into trip simulations is the distance-decay law (see section 4.3 for details).

4 Monte Carlo simulations of traffic flow

4.1 Applying uniform OD distribution

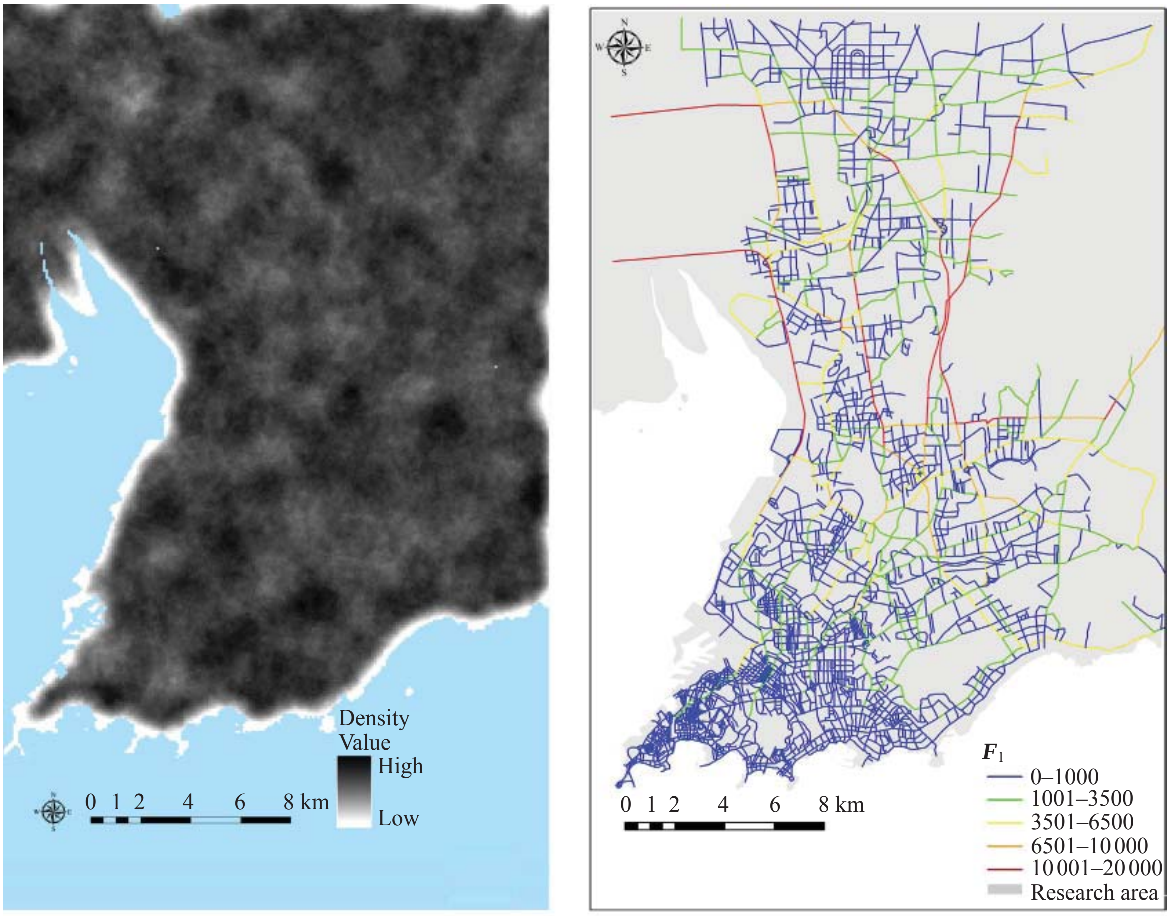
To understand the observed distribution of traffic flow, we use Monte Carlo simulations to reproduce the traffic flows of all streets. In each simulation we generate a large number (136 778 in this study) of OD pairs following a particular distribution. For each OD pair, the two points are associated with the nearest street segments such that the shortest path can be found. Finally, we add together the total number of shortest paths on all segments and acquire the simulated traffic flows.

First, we assume that the OD points are uniformly distributed in the study area [figure 6(a)]. The simulated traffic flow $\boldsymbol{F}_1$, shown in figure 6(b), cannot represent totally the observed traffic flow, with a negative WCC of −0.031 (table 4)

Table 4. Weighted correlation analysis between estimated flow and real flow.

		$\boldsymbol{F}_1$	$\boldsymbol{F}_2$	$\boldsymbol{F}_3{}^*$
Real flow	correlation coefficient	−0.031**	0.343**	0.623**
	significance (2-tailed)	0.000	0.000	0.000
	N	5095	5095	5095

** Correlation is significant at the 0.01 level.
Note. [a] $\boldsymbol{F}_1$, $\boldsymbol{F}_2$, $\boldsymbol{F}_3$ stand for the simulated traffic flow obtained in the three scenarios.



(a) (b)

Figure 6. [In color online.] (a) Uniformly distributed origin–destination points (visualized using kernal density estimation); (b) the corresponding simulated traffic flow $\boldsymbol{F}_1$.

4.2 Using mobile phone Erlang values to simulate OD distribution

It is natural that the uniform OD distribution cannot reproduce the actual traffic flow, because human activities in the study area are heterogeneous. Hence, we should find an appropriate measure for representing the spatial distribution of human activities. With the development of ICT, mobile phone data(1) have been used widely to understand human mobility and urban dynamics in various cities (Ahas et al, 2010; González et al, 2008; Ratti et al, 2006; Rose, 2006; Sevtsuk and Ratti, 2010; Song et al, 2010). Following the method proposed by Calabrese et al (2011), the mobile phone Erlang value, a measure of mobile phone call traffic, is selected as the indicator for human activity distribution to generate OD points in this research.

When a mobile phone call is made, the mobile phone is linked to the nearest base station, which has specific geographic coordinates. The base station also records the duration of each call that it routes, so that the call traffic within a period (eg, one week) can be computed. In the telecommunication industry, one Erlang is the total call-traffic volume of a base station for one hour.

We introduce the Erlang value records for all stations in Qingdao to simulate the human-activity distribution. The advantage of Erlang data is the varied geographic scale that can depict the relationship between human activities and land uses. The dataset was collected

(1) Generally, there are two types of mobile phone data available from the telecommunication operators. One is phone call record details and the other is mobile phone traffic at each base station, which is used in this research.

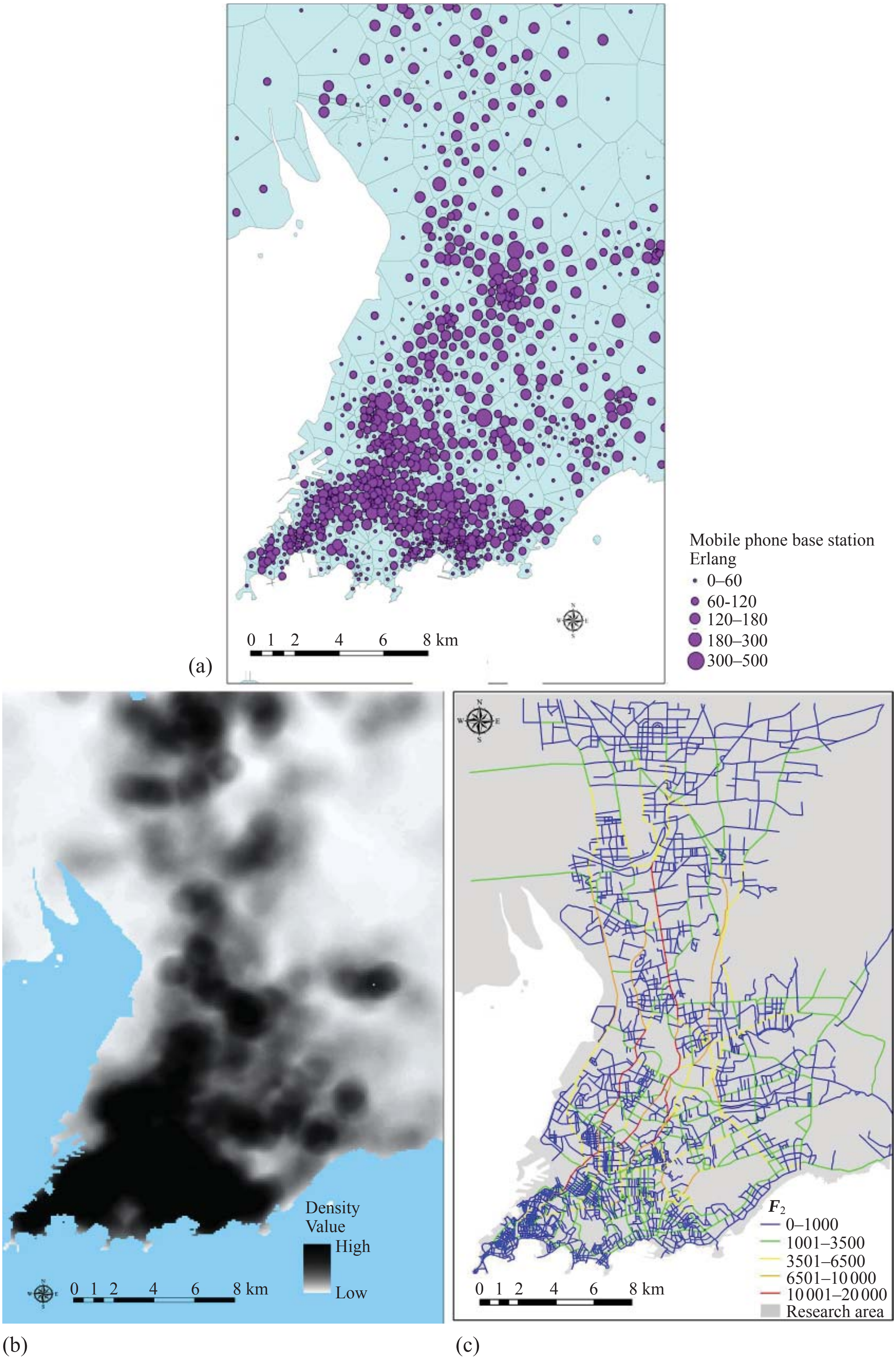


Figure 7. [In color online.] (a) Distribution of base stations and their Erlang values; (b) distribution of potential origin–destination points constrained by Erlang values; (c) simulated traffic flow $\boldsymbol{F}_2$.

Table 5. Data structure of mobile phone data recorded by two base towers.

Tower ID	Longitude[a]	Latitude[a]	Elevation[a] (m)	Erlang	Angle[b]
1	120.353	36.065	0	100	36.01
2	120.402	36.159	10	45	36.11

[a] Longitude, latitude, and elevation represent the location of a base tower.
[b]Angle means the average angle for its sector subdivisions with different orientations.

over the same period as the taxi dataset. Table 5 lists two sample records including Erlang values and locations, and figure 7(a) depicts the distribution of all base stations, with symbols representing the associated Erlang values. In figure 7(a), we use a Voronoi polygon (cell) for the region where mobile calls are routed by a base tower. With the assumption that the statistics of individuals' phone call activities in different cells are generally similar, the Erlang values of aggregated call volumes provide a relatively good approximation of the distribution of human-activity intensity, with which the travel demands are correlated positively. We follow the data processing method of Sevtsuk and Ratti (2010) to perform pretreatment. It is rational to assume that the origins and destinations have similar spatial distributions in a relatively long time interval (for example, one week, as in our study) and a particular area, according to the anchor point theory (Huang et al, 2010). Hence, we adopt the Erlang values of all towers in the study area to control the generation of both origin and destination points. The Erlang-constrained OD distribution is shown in figure 7(b). Figure 7(c) shows the simulated traffic flow $\boldsymbol{F}_2$ with a WCC of 0.343 (table 4) with respect to the observed traffic volumes.

4.3 Considering distance decay

The simulated traffic flow that considers the distribution of human activity is better than that which uses uniform OD distribution. However, the spatial patterns of simulated and real traffic volumes are different. Figure 7(c) overestimates the traffic flows of several long north–south streets but underestimates the flows in the core urban area. We use figure 8 to clarify the differences. The study area contains two major urban areas, A and B, with high population and street densities. Hence, the travel demands in both areas are high. According to the above model, from a given origin O, the probabilities that D_1 and D_2 serve as destinations are close. However, the traffic volume between D_1 and O should be greater than that between D_2 and O due to the distance-decay effect. This leads to an overestimation for the street L linking areas A and B if we do not take into account the distance-decay effect.

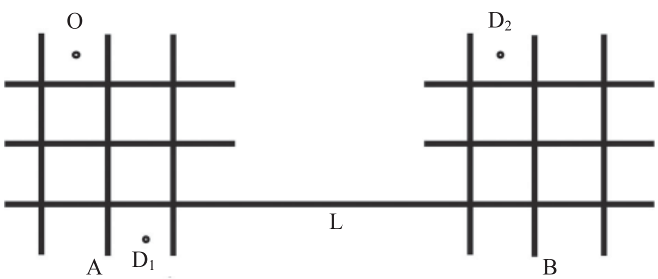


Figure 8. Trip generation in a synthetic dumbbell-shaped city, which is similar to the study area.

The distance-decay effect has been widely identified in many geographic phenomena. Tobler's first law of geography makes an informal expression of this effect (Tobler, 1970). In many applications, distance decay is defined mathematically as:

$$I \propto \frac{1}{d^{\beta}} \,, \tag{5}$$

where I denotes the spatial interaction (the flow or the intensity), d is the distance between two localities, and β is a decay coefficient (Taylor, 1971). In the field of transportation, distance decay contributes to the decision to migrate, causing more short distance travels than long distance ones. Note that a number of functions, including power functions and exponential functions, are available for representing the distance-decay effect (Wang, 2012). Among them, the power-decay function is widely applied since it is independent of the scale of d. Hence, the estimation of β is crucial and can be used to compare different spatial behaviors and spatial structures (Fotheringham, 1981; Skov-Petersen, 2001). For example, Hansen (1959) listed a number of decay parameters, varying from 0.9 to 2.0, for different types of intraurban interaction. Recently, the widespread application of location-aware devices such as the mobile phone enables us to investigate human mobility based on large volumes of individuals' trajectories. For example, the exponent is about 1.75 according to González et al (2008), and Kang et al (2012) found that the exponents of eight cities in China vary between 1.3 and 1.8.

In this research, we use equation (6) to incorporate the power-law distance decay. That is, the probability that a trip between two known points O and D follows a power-law distribution. One assumption for the simulation is that the length of each trip is longer than 1 km since very short trips usually are not made by taxi but by walking.

$$\Pr(T \mid \mathrm{O}, \mathrm{D}) \propto \frac{1}{\mathrm{dist}(\mathrm{O}, \mathrm{D})^{\beta}} \,. \tag{6}$$

In the Monte Carlo simulation, we first create two candidate points for a trip following the constraint of Erlang values and then use the acceptance–rejection method to determine whether to accept the trip according to the distance threshold (1 km) and equation (6), where dist(O, D) is the Euclidean distance between these two points. Obviously, the number of long-distance travels will be decreased when the distance-decay effect is considered.

We try different β values between 1.0 and 3.0 to search for a best simulation $\boldsymbol{F}_3$ and find that the WCC achieves a maximum of 0.623 (table 4) when $\beta = 2.0$. The corresponding traffic-flow distribution is depicted in figure 9. In addition to the high WCC value, the spatial patterns of observed and simulated traffic volumes are rather similar. For example, the simulated result highlights the importance of streets in the southern part of the study area. This indicates that the proposed method reproduces the observed traffic flow well and the power-law distance decay with an exponent of 2.0 can be used to simulate the urban-scale transportation of Qingdao. Although the exponent is estimated using the data collected in Qingdao, it is close to the values reported in other research, such as that by González et al (2008) and Kang et al (2012), suggesting that we can use such an exponent value to model intraurban motion.

5 Discussion

5.1 Correlation between simulated flow and street betweenness centrality

We compute the correlation coefficients between the betweenness centrality ($\boldsymbol{C}_{\mathrm{P}}^{\mathrm{B}}$, $\boldsymbol{C}_{\mathrm{D}}^{\mathrm{B}}$) and the simulated traffic flows in the three models (table 6). In the primal approach, betweenness centrality was found to correlate highly with the simulated traffic flow in the second scenario ($\boldsymbol{F}_2$) that considers human-activity distribution (WCC = 0.769) but less so with the simulation

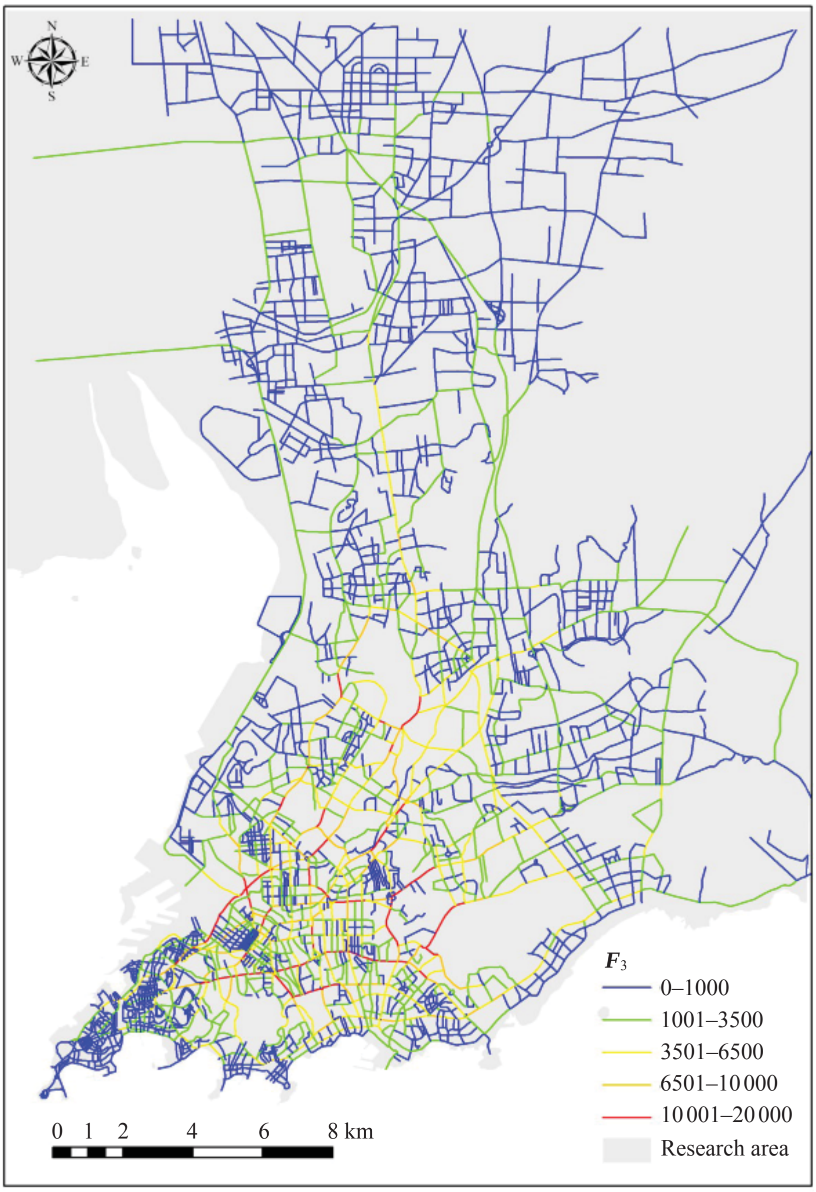


Figure 9. [In color online.] Simulated traffic flow $\boldsymbol{F}_3$, considering the distance-decay effect.

Table 6. Weighted correlation analysis between simulated traffic flow in the three models and betweenness centrality.

		$\boldsymbol{F}_1$	$\boldsymbol{F}_2$	$\boldsymbol{F}_3$
$\boldsymbol{C}_{\mathrm{P}}^{\mathrm{B}}$	correlation coefficient	0.341**	0.769**	0.595**
$\boldsymbol{C}_{\mathrm{D}}^{\mathrm{B}}$		0.043**	0.478**	0.371**
	significance	0.000	0.000	0.000
	N	5095	5095	5095

** Correlation is significant at the 0.01 level (2-tailed).

that incorporated distance decay ($\boldsymbol{F}_3$, WCC = 0.468). In the dual approach, the circumstances are similar. However, $\boldsymbol{F}_3$ estimates the actual traffic flow better than $\boldsymbol{F}_2$. If betweenness is a good indicator for the observed traffic flow, then $\boldsymbol{C}_{\mathrm{P}}^{\mathrm{B}}$ or $\boldsymbol{C}_{\mathrm{D}}^{\mathrm{B}}$ would be expected to have a higher correlation with $\boldsymbol{F}_3$. Table 6 confirms the gap between the actual traffic flow and betweenness. It is interesting that $\boldsymbol{F}_2$ has a relatively high WCC with $\boldsymbol{C}_{\mathrm{P}}^{\mathrm{B}}$. When comparing figures 1 and figure 7(b), we can see that the distribution of street intersections and mobile phone Erlang values are rather similar, as areas with dense human activity generally have higher street density. The second simulation, which corresponds to using street intersections to generate OD points, leads to a high WCC with $\boldsymbol{C}_{\mathrm{P}}^{\mathrm{B}}$.

Following the above comparison, the conclusion that betweenness centrality is not a good predictor variable for urban traffic flow can be verified by the decreased WCC from $\boldsymbol{F}_2$ to $\boldsymbol{F}_3$, since $\boldsymbol{F}_3$ estimates the actual traffic flow better than $\boldsymbol{F}_2$. The contradiction arises from the fact that betweenness does not take into account the distance-decay effect, given that the urban streets and human activities have similar spatial distributions. Due to distance decay, the probability of long-distance travels is decreased, and this is another force that causes trips and human-activity distribution to be concentrated in the core urban areas. The conventional betweenness measure, as well as the second simulation, does not consider distance decay and thus underestimates the traffic flow in the south part of Qingdao [see figures 2, 4, and 7(c)].

5.2 Further improvement of the current method

The simulation results demonstrate the impact of the heterogeneity of population activities and distance decay on urban traffic generation. However, the resulting correlation coefficient is still not very satisfactory. If we inspect some local areas in detail, the simulation results are not ideal. This deficiency may be attributed to the following three issues.

First, this research assumes that people have the same probability of choosing a taxi as their mode of travel in different places and over different time intervals, resulting in traffic flow estimated using taxi trajectories that comprise the same proportion of the whole traffic volumes for all streets. This assumption is widely adopted in practice when monitoring traffic statuses using FCD. However, biases obviously exist because the proportion of taxi trips varies for different streets.

Second, in terms of the generation of OD points, the application of mobile phone Erlang values is reasonable, that is, the travel demands and Erlang values in all mobile phone cells are positively correlated. However, biases do exist and these introduce errors to the simulation result.

Finally, the assumption that individuals rationally choose the geometrically shortest path is not necessarily the actual case (Gigerenzer, 2008; Kazerani and Winter, 2009a). The reasons for this are twofold. First, street attributes (such as hierarchy and one-way direction constraints) have not yet been considered in the simulation because of the incompleteness of the dataset; thus, we treat all streets equally. However, this is not always the case, as the expressways in the city are, in general, the preferred routes. Second, even if we had enough street information it could not ensure a better simulation result, because drivers' wayfinding behaviors are not always rational or because of the uncertainty of localities which are described by fuzzy named places and spatial relationships (Liu et al, 2009). This work is actually based on the four-step model (McNally, 2000) and much research from another perspective has adopted the agent-based model, which introduces agents' cognition and behaviors recorded by location-aware devices, cyberspace, and traffic diaries into the traffic simulation and can be applied to address this issue (Adler et al, 2005; Couclelis, 1986; Dia, 2002).

6 Conclusions

To explore the distribution of urban traffic flow from the perspective of trip generation, in this study we applied the method of Monte Carlo simulations. It is noticeable that the weighted correlation coefficient between the simulated flow and real flow rises from –0.031 to 0.343 after the human-activity distribution was incorporated; furthermore, it rises to 0.623 when the distance-decay effect is considered. These results demonstrate that the last simulated result captures well those main streets with dense traffic flow in the core urban area of Qingdao. It indicates that the Monte Carlo simulation effectively identifies the two factors that contribute to urban traffic flow: geographic heterogeneity of human-activity distribution and the law of distance decay. The former suggests the spatially varying travel demands, while the latter implies the common regularity of human travel behaviors. Meanwhile, we also point out some deficiencies in the proposed simulation, and future work may focus on solving these.

This research also confirms that using only betweenness centrality of the street network is unsuitable for predicting traffic flow. We clarify that it underestimates the traffic in the core urban area, as it does not take into account the distance-decay effect. In other words, the high traffic concentration in the core urban area can be attributed to two aspects: the dense human-activity distribution and the distance-decay effect. The betweenness measure can only represent the former aspect. Given that betweenness is an important measure in social-network studies, more attention should be paid to geographical characteristics in transportation research.

Acknowledgements. This research was supported by the National High Technology Development 863 Program of China (Grant 2011AA120303), the National Natural Science Foundation of China (Grants 41171296, 41271385, and 41271386), and the National Key Technology Research and Development Program of China (Grant 2012BAJ05B04). The authors would like to thank Dr David O'Sullivan and reviewers for their comments that helped to improve this manuscript.

References

Adler J L, Satapathy G, Manikonda V, Bowles B, Blue V J, 2005, "A multi-agent approach to cooperative traffic management and route guidance" *Transportation Research Part B* **39** 297–318

Ahas R, Aasa A, Silm S, Tiru M, 2010, "Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: case study with mobile positioning data" *Transportation Research Part C* **18** 45–54

Bar-Gera H, 2007, "Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel" *Transportation Research Part C* **15** 380–391

Black W R, 2003 *Transportation: A Geographical Analysis* (Guilford Press, New York)

Bonacich P, 1987, "Power and centrality: a family of measures" *American Journal of Sociology* **92** 1170–1183

Borgatti S P, 2005, "Centrality and network flow" *Social Networks* **27** 55–71

Brockmann D, Hufnagel L, Geisel T, 2006, "The scaling laws of human travel" *Nature* **439** 463–465

Calabrese F, Lorenzo G D, Liu L, Ratti C, 2011, "Estimating origin–destination flows using mobile phone location data" *IEEE Pervasive Computing* **10** 36–44

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

Couclelis H, 1986, "A theoretical framework for alternative models of spatial decision and behavior" *Annals of the Association of American Geographers* **76** 95–113

Crucitti P, Latora V, Porta S, 2006a, "Centrality in networks of urban streets" *Chaos* **16** 015113

Crucitti P, Latora V, Porta S, 2006b, "Centrality measures in spatial networks of urban streets" *Physical Review E* **73** 036125

de Fabritiis C, Ragona R, Valenti G, 2008, "Traffic estimation and prediction based on real time floating car data", in *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, Beijing* 12–15 October, ISBN 978-1-4244-2111-4, pp 197–203

Dia H, 2002, "An agent-based approach to modelling driver route choice behaviour under the influence of real-time information" *Transportation Research Part C* **10** 331–349

Fotheringham A S, 1981, "Spatial structure and distance-decay parameters" *Annals of the Association of American Geographers* **71** 425–436

Freeman L C, 1977, "A set of measures of centrality based on betweenness" *Sociometry* **40** 35–41

Freeman L C, Borgatti S P, White D R, 1991, "Centrality in valued graphs: a measure of betweenness based on network flow" *Social Networks* **13** 141–154

Gigerenzer G, 2008 *Rationality for Mortals: How People Cope with Uncertainty* (Oxford University Press, New York)

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

Goodchild M F, Janelle D G, 1984, "The city around the clock: space–time patterns of urban ecological structure" *Environment and Planning A* **16** 807–820

Goodchild M F, Klinkenberg B, Janelle D G, 1993, "A factorial model of aggregate spatio-temporal behavior: application to the diurnal cycle" *Geographical Analysis* **5** 277–294

Hansen W G, 1959, "How accessibility shapes land use" *Journal of the American Institute of Planners* **25** 73–76

Herrera J C, Work D, Ban J, Herring R, Jacobson Q, Ban J, Bayen A, 2010, "Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment" *Transportation Research Part C* **18** 568–583

Hillier B, Penn A, Hanson J, Grajewski T, Xu J, 1993, "Natural movement: or, configuration and attraction in urban pedestrian movement" *Environment and Planning B: Planning and Design* **20** 29–66

Holme P, 2003, "Congestion and centrality in traffic flow on complex networks" *Advances in Complex Systems* **6** 163–176

Huang W, Dong Z, Tian H, Song G, Chen G, Jiang Y, Xie K, 2010, "Anchor points seeking of large urban crowd based on the mobile billing data", *Advanced Data Mining and Applications: Lecture Notes in Computer Science* **6440** 346–357

Jiang B, 2009, "Street hierarchies: a minority of streets account for a majority of traffic flow" *International Journal of Geographical Information Science* **23** 1033–1048

Jiang B, Claramunt C, 2002, "Integration of space syntax into GIS: new perspectives for urban morphology" *Transactions in GIS* **6** 151–162

Jiang B, Claramunt C, 2004, "Topological analysis of urban street networks" *Environment and Planning B: Planning and Design* **31** 151–162

Jiang B, Liu C, 2009, "Street-based topological representations and analyses for predicting traffic flow in GIS" *International Journal of Geographical Information Science* **23** 1119–1137

Jiang B, Zhao S, Yin J, 2008, "Self-organized natural roads for predicting traffic flow: a sensitivity study" *Journal of Statistical Mechanics: Theory and Experiment* P07008

Kang C, Ma X, Tong D, Liu Y, 2012, "Intra-urban human mobility patterns: an urban morphology perspective" *Physica A* **391** 1702–1717

Kazerani A, Winter S, 2009a, "Can betweenness centrality explain traffic flow?", in *Proceedings of 12th AGILE International Conference on Geographic Information Science, Hanover, Germany, June 2-5* http://plone.itc.nl/agile_old/Conference/2009-hannover/pdfs/111.pdf

Kazerani A, Winter S, 2009b, "Modified betweenness centrality for predicting traffic flow", in *Proceedings of the 10th International Conference on GeoComputation, Sydney, Australia, November 30–December 2* http://www.science.mcmaster.ca/~igu-cmgs/publications/geocomputation/Kazerani_and_Winter.pdf

Kerner B S, Demir C, Herrtwich R G, Klenov S L, Rehborn H, Aleksiü M, Haug A, 2005, "Traffic state detection with floating car data in road networks", in *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems Vienna, Austria, September 13–16* ISBN 0-7803-9215-9, pp 44–49

Kobayashi T, Shinagawa N, Watanabe T, 1999, "Vehicle mobility characterization based on measurements and its application to cellular communication systems" *IEICE Transactions on Communications* **12** 2055–2060

Li Q, Zhang T, Wang H, Zeng Z, 2011, "Dynamic accessibility mapping using floating car data: a network-constrained density estimation approach" *Journal of Transport Geography* **19** 379–393

Liu Y, Guo Q, Wieczorek J, Goodchild M F, 2009, "Positioning localities based on spatial assertions" *International Journal of Geographical Information Science* **23** 1471–1501

Lorkowski, S, Mieth P, Schäfer R P, 2005, "New ITS applications for metropolitan areas based on floating car data", ECTRI Young Researcher Seminar, Lyon, 6 December, http://elib.dlr.de/21000/1/YRS_2005.pdf

McFadden D, 1974, "The measurement of urban travel demand" *Journal of Public Economics* **3** 303–328

McNally M, 2000, "The four step model", in *Handbook of Transport Modeling* Eds D A Hensher, K J Button (Elsevier, Amsterdam), pp 35–52
Newman M E J, 2005, "A measure of betweenness centrality based on random walks" *Social Networks* **27** 39–54
Penn A, Hillier B, Banister D, Xu J, 1998, "Configurational modeling of urban movement networks" *Environment and Planning B: Planning and Design* **25** 59-84
Porta S, Crucitti P, Latora V, 2006a, "The network analysis of urban streets: a primal approach" *Environment and Planning B: Planning and Design* **33** 705–725
Porta S, Crucitti P, Latora V, 2006b, "The network analysis of urban streets: a dual approach" *Physica A* **369** 853–866
Porta S, Latora V, Wang F, Strano E, Cardillo A, Scellato S, 2009, "Street centrality and densities of retail and services in Bologna, Italy" *Environment and Planning B: Planning and Design* **36** 450–465
Ratti C, Pulselli R M, Williams S, Frenchman D, 2006, "Mobile landscapes: using location data from cell phones for urban analysis" *Environment and Planning B: Planning and Design* **33** 727–748
Rose G, 2006, "Mobile phones as traffic probes: practices, prospects and issues" *Transport Reviews* **26** 275–291
Scheepens R, Willems N, 2011, "Interactive visualization of multivariate trajectory data with density maps", in *Proceedings of IEEE Pacific Visualization Symposium, Hong Kong, China, March 1–4* http://www.win.tue.nl/~cwillems/public/pacificvis11.pdf
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
Skov-Petersen H, 2001, "Estimation of distance-decay parameters—GIS-based indicators of recreational accessibility", in *Proceedings of ScanGIS 2001, Aas, Norway, June 24–27* http://www.scangis.org/scangis2001/papers/22.pdf
Song C, Qu Z, Blumm N, Barabási A-L, 2010, "Limits of predictability in human mobility" *Science* **327** 1018
Taylor P J, 1971, "Distance transformation and distance decay functions" *Geographical Analysis* **3** 221–238
Tobler W, 1970, "A computer movie simulating urban growth in the Detroit region" *Economic Geography* **46** 234–240
Turner A, 2007, "From axial to road-center 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
Wang F, 2001, "Explaining intraurban variations of commuting by job accessibility and workers' characteristics" *Environment and Planning B: Planning and Design* **28** 169–182
Wang F, 2012, "Measurement, optimization and impact of healthcare accessibility: a methodological review" *Annals of the Association of American Geographers* **102** 1104–1112
Wang F, Antipova A, Porta S, 2010, "Street centrality and land use intensity in Baton Rouge, Louisiana" *Journal of Transport Geography* **19** 285–293