

Received August 5, 2020, accepted August 7, 2020, date of publication August 11, 2020, date of current version August 21, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015752

Revealing Urban Traffic Demand by Constructing Dynamic Networks With Taxi Trajectory Data

HUI ZHANG¹, LELE ZHANG², FA CHE³, JIANMIN JIA¹, AND BAIYING SHI¹

¹School of Transportation Engineering, Shandong Jianzhu University, Jinan 250101, China

²Yantai Yishang Electronic Technology Company, Ltd., Yantai 264003, China

³Zibo Transportation Service Center, Zibo 255000, China

Corresponding author: Fa Che (zbsgljljyanghuke@zb.shandong.cn)

This work was supported in part by the Doctoral Funds of Shandong Jianzhu University under Grant XNBS1614, and in part by the National Natural Science Foundation of China under Grant 41901396.

ABSTRACT As a crucial travel mode, taxi plays a significant role in residents' daily travel. Uncovering taxi traffic demand has become a hotspot in transport studies. Previous researchers pay more attention to the statistical characteristics of taxi trips, while few studies focus on the dynamic features in different periods of a day. In this article, we study the taxi travel demand by constructing dynamic networks based on taxi trajectory data. In addition, relationship between travel intensity and point of interest (POI) in Xiamen, China is discussed. Firstly, the study area is divided by 1km × 1km uniform cells. The pick-up and drop-off activities of passengers are recorded for each cell. Secondly, the networks are constructed by regarding each cell as a node and regarding taxi trips from a cell to another cell as an edge. On this basis, we divide a day into 12 periods by two hours and construct the networks for different periods. Finally, correlation between travel intensity and POI intensity is detected with regression analysis. Results show that the taxi trip networks have large clustering coefficient and small shortest path length, which indicates they are 'small world' networks. Moreover, the taxi trip networks are disassortative networks that hotspot areas tend to connect with the common areas. Furthermore, the taxi trip length in a day follows a lognormal distribution and the peak hour of taxi trip appears around midnight. Finally, a cubic polynomial curve could fit the relationship between travel intensity and POI intensity. Our findings provide a new insight for understanding the traffic demand of taxi.

INDEX TERMS Urban traffic demand, dynamic spatial-interaction network, spatiotemporal characteristics of taxi trips, taxi GPS trajectory data, point of interest.

I. INTRODUCTION

Traffic demand plays an important role in urban traffic planning, traffic management and city planning. However, it is hard to grasp urban traffic demand because the structure of traffic flow is complex and changeful. Traditional traffic survey cannot provide enough information to reflect actual traffic demand for the big cities because the survey is time-consuming and expensive. Big geospatial data (e.g. mobile phone data, bike-sharing order data and taxi trajectory data) provide a large number of human positions to uncover the traffic demand. The data is more accurate, continuous, cost-effective compared with traffic survey data. In recent years, researchers have investigated traffic problems with geospatial data in several aspects such as traveler's behavior, travel

pattern and spatial-temporal characteristics [1]–[3]. To understand the traffic demand, a common way is dividing studied area into small parts [4]–[7].

Taxi is a significant component of urban transit systems, which could provide 24-hour service. It is an important supplement for other transit options, especially at night when the most transit options stop operating. Taxi trajectory data that record a series of GPS points and other information such as instantaneous velocity, status and time, which has been widely used in analyzing characteristics of taxicab movement [8]–[10], travelers' mobility [11]–[13], travel pattern [14], [15], traffic congestion [16], [17], route choice [18]–[20], taxi trip characteristics [21], [22] etc. Taxi operates without fixed line, schedule and stations, which can achieve the spatiotemporal characteristic of travelers. There have been many studies focusing on spatiotemporal characteristics of urban travel demand by using taxi trajectory

The associate editor coordinating the review of this manuscript and approving it for publication was Yilun Shang¹.

data. Cui *et al.* extracted travel demand from taxi trajectory data and identified the mismatch between travel demand and transport network services [23]. Sun *et al.* proposed a framework to derive the dynamics of the pick-up and drop-off locations of taxi [24]. To understand the travel demand, the study areas have always been divided into small grids, cells or zones. Tang *et al.* proposed an Entropy-Maximizing method to extract OD information by dividing study area into 40×60 grids [22]. Zhang *et al.* studied the taxi OD trips with different cell sizes and found trip length distribution is negative exponential across scales [25]. Ke *et al.* divided Hangzhou, China into 7×7 grid, and each grid is $4.77 \text{ km} \times 4.81 \text{ km}$. They studied the short-term forecasting of passenger demand using DiDi data [26].

Recently, complex network theory has been introduced to analyze the taxi trip structure, where small grids or cells are taken as nodes and trips from one node to another node are taken as directed edges. The network can be called spatial-interaction network, which provides a new insight to understand the taxi trip structure and cities' functions. Liu *et al.* proposed a complex network method to study the community structure of taxi trips [27]. He *et al.* partitioned studied area into grids with $100 \text{ m} \times 100 \text{ m}$ and constructed spatiotemporal traffic diagrams to extract traffic dynamics [28]. Yang *et al.* divided study area into zones by transportation planning method and found that the complex taxi network exhibited "small world" characteristic [29]. Yang *et al.* proposed a weighted k-core decomposition method incorporating complex network theory and weighted distance, and three layers have been found in spatial-interaction network: a core layer, a bridge layer and a periphery layer [30]. He studied taxi networks with different cell scales, and observed that the degree distributions follow the power law and the power law exponents decreases with the growth of square cells [31].

Urban traffic flows contain many types, such as commuter flow, leisure flow and transfer flow. Some types of urban traffic flow highly depend on the urban form such as commuter flow that moves between residence and workplace. The built environment has been widely discussed for the impact on residents' travel in three factors: density, diversity and design [32]–[35]. The aforementioned study indicates that there is a close relationship between taxi driver activities and built environment characteristics [36]. Li *et al.* used points of interest (POIs) to apply a gravity model to evaluate the interzonal commuting patterns [37]. Zhang *et al.* studied the relationship between travel intensity and POIs based on car-hailing data, and the results show that some types of POIs such as traffic facilities have great impacts on pick-up and drop-off [38]. Liu *et al.* argued that higher proportion of commercial area and public service area produced greater taxi demand, while the proportion of residential area and land use mix have negative impacts on taxi demand [39]. Based on the strong relationship between taxi traffic demand and POI, taxi trajectory data could be used to identify the urban functional regions [40].

There were many researches focused on the spatial-interaction network by dividing study area into small grids. However, there was little work to study the dynamic change of taxi networks due to random trip demands in different periods. The traffic demand fluctuates in different time period, which is very important for urban traffic management. For example, the traffic demand is high in peak time, traffic manager should provide more service to satisfy the requirement. At present, the fluctuation of traffic demand from network perspective during a day is not clear, which hinders the improvement of taxi service. To fill this gap, this article proposes a spatiotemporal analysis method by constructing taxi trip network to detect the spatiotemporal characteristic of taxi trips. Moreover, we intend to build a relationship between taxi trip demand in cells and the number of POI. This article intends to provide a new insight to understand the traffic demand by constructing dynamic taxi trip networks. The rest of paper is organized as follows. Section 2 gives the study area and data description. Section 3 introduces the method. Spatiotemporal characteristics of taxi trips are given in section 4. Section 5 describes network-based characteristics. Section 6 introduces relationship between taxi trips and POI. Conclusions are given in section 7.

II. STUDY AREA AND DATA DESCRIPTION

A. XIAMEN CITY

The case study area is Xiamen, China, which locates in the southeast of China (see Fig.1). The main study area locates in the place where longitude ranges from 117.89E to 118.35E , latitude ranges from 24.42N to 24.74N . At present, The population of Xiamen is approximate 4.29 million. The length from east to west is about 34 km and north-south length is about 44 km. The entire region is divided into 1496 uniform grids, and each grid is $1 \text{ km} \times 1 \text{ km}$.

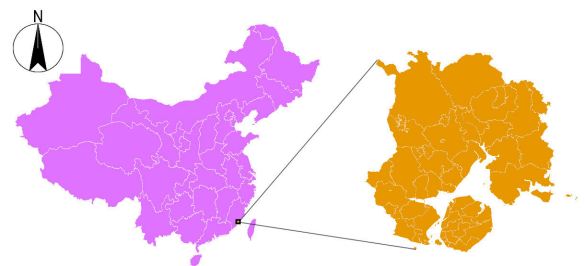


FIGURE 1. The studied city of Xiamen.

B. TAXI GPS TRAJECTORY DATA

We collected a week taxi GPS trajectory data from July 2, 2018 to July 8, 2018. Moreover, we select a weekday July 3 to study the dynamic the characteristics of taxi trips within a day. There are totally 4973 vehicles and approximate 7 million GPS trajectory records on July 3. The taxi data content contains taxi id, timestamp, longitude, latitude, instantaneous speed and the status of taxi (vacant or occupied). In the

TABLE 1. Description of data fields of GPS records.

Data field	Sample	Description
TaxiID	3400633150	The ID of taxis
Longitude	118.113291	The longitude of GPS points
Latitude	24.544415	The latitude of GPS points
Status	1	Status of taxi: '0' represents vacant; '1' represents occupied
Date	2018-7-3	The date of GPS points
Time	07:15:43	The Time of GPS points
Speed	53	The instantaneous speed of taxis



FIGURE 2. Trajectory data of a taxi during a day.

dataset, '0' represent the vacant status, while '1' represent occupied. Table 1 shows the sample of a GPS record data.

The GPS trajectory data record taxi vehicles' information, including the status of travelers. Fig. 2 illustrates the GPS trajectory data of a taxi in a day. The red points stand for occupied vehicles, while green points are with vacant status. Before the data is used, a data cleaning process is implemented. The data will be removed if the GPS point is not within the scope of studied area or the length of trips is smaller than 200 m. From the trajectory data, we can extract each trip information such as start point, end point, trip distance, trip start time and trip cost time.

C. POI DATA

In this article, we use Python to extract POI data from Baidu map (<https://map.baidu.com/>). The POI data contains 13 types such as dinning facility, shopping facility, service facility, traffic facility, etc. There are totally 83816 POIs in the study area. Each POI consists of four fields: type, name, longitude and latitude. We describe the example of POI data in Table 2. Since travelers' behaviors have high correlation with the type and scale of POI, we intend to identify the relationship between trips and the number of POIs in the grids in this article. Fig. 3 shows the distribution of POIs in Xiamen. As can be seen, most POIs concentrate in Xiamen island including Huli and Siming districts.

III. METHODS

A. NETWORK CONSTRUCTION

In urban area, travelers' movement trips connect different areas with their needs. However, it is hard to know the exact places where travelers attempt to get to and the places where

TABLE 2. Categories and example of POI data.

Type	Name	Longitude	Latitude
Dinning facility	AnAn Xiaowu	118.121653	24.429279
Shopping facility	Jiahui Market	118.138465	24.436814
Service facility	Yuanshan Gas station	118.131112	24.512183
Traffic facility	Jinwei Qipei	118.030691	24.454289
Financial facility	Bank of China	118.167487	24.474138
Science & Education	Xiamen University	118.089617	24.440991
Scenic spot	Gulang Dongtian	118.063193	24.444828
Vehicle service	Baili Xiche	118.112524	24.434086
Parking service	Haigangcheng	118.120222	24.426857
Medical facility	Guoda Yaofang	118.093986	24.473237
Government agency	Xiamen Kejiju	118.087920	24.456299
Accommodation service	Paopao Kezhan	118.121022	24.427908
Company	Yuanchang Jituan	118.095956	24.474787

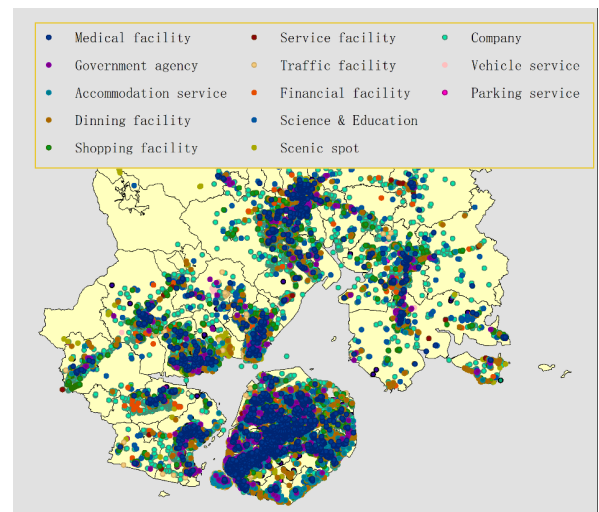


FIGURE 3. POIs of Xiamen city.

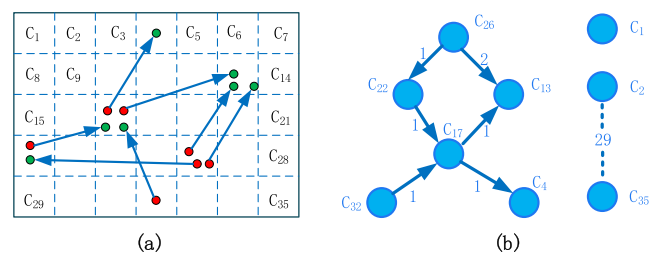


FIGURE 4. Building a network from the taxi trips: (a) dividing study area into square cells, red nodes are the starting points of taxi trips and green nodes are end points of taxi trips, (b) each cell represents a node, a directed edge connecting two nodes if there exist trips, the weight of an edge is the number of trips during the studied period.

they come from. Therefore, we adopt to divide study area into small square cells. The cells can be regarded as nodes, and the trips between nodes can be regarded as directed links or edges. Then we can construct spatiotemporal networks to detect the taxi trip patterns by taxi trajectory data.

Fig.4 shows the process of constructing a network. To build a network by taxi trips of a period T , the spatial area could

be divided into same large square cell with a side length L . An area of $ML \times NL$ can be divided into $M \times N$ cells, each cell C_i stands for a node v_i . To conveniently delineate nodes, nodes are given a consecutive series number $[1, M \times N]$. We use $G = (V, E, W)$ to represent a directed network, where V is the set of nodes, E is the set of edges, W is the set of edge weight. $e_{ij} = 1$, if there is an edge from node v_i to v_j ; otherwise, $e_{ij} = 0$. w_{ij} is the weight of the edge e_{ij} , which stands for the number of taxi trips between node i and node j .

B. MEASURES

1) DEGREE

Node degree k_i is defined as the number of nodes connect node i , a node with large degree means it is important in the network. For a directed network, node degree contains in-degree and out-degree. In-degree means the number of nodes pointed to the node i , which is calculated by

$$k_i^{in} = \sum_j e_{ji} \tag{1}$$

while out-degree is the number of nodes pointing from the node i , which is defined as

$$k_i^{out} = \sum_j e_{ij} \tag{2}$$

So, the degree of a node is sum of in-degree and out-degree:

$$k_i = k_i^{in} + k_i^{out} \tag{3}$$

2) WEIGHTED DEGREE

In this article, we defined a weighted degree of nodes to measure the number of trips. Let w_{ij} stand for the number of trips between node i and node j , similarly, the weighted in-degree \tilde{k}_i^{in} , weighted out-degree \tilde{k}_i^{out} and weighted degree \tilde{k}_i are defined as follows:

$$\tilde{k}_i^{in} = \sum_j w_{ji} \tag{4}$$

$$\tilde{k}_i^{out} = \sum_j w_{ij} \tag{5}$$

$$\tilde{k}_i = \tilde{k}_i^{in} + \tilde{k}_i^{out} \tag{6}$$

3) AVERAGE SHORTEST PATH LENGTH (ASPL)

The average shortest path length is a key role to detect the small world property of networks, which is defined as the number of steps along the shortest paths of all possible pair of network nodes. It is calculated as

$$ASPL = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \tag{7}$$

where N is the total number of nodes in the network, d_{ij} is the shortest path length between node i and node j .

4) CLUSTERING COEFFICIENT

Clustering coefficient describes the local cohesiveness of current node or the extent to which the node in the network are cluster together. For a node i with k_i connected neighbor

nodes, there are at most $k_i(k_i-1)/2$ edges among the k_i nodes. So, the clustering coefficient is defined as

$$C_i = \frac{2E_i}{k_i(k_i-1)} \tag{8}$$

where E_i is the actual number of edges among the k_i nodes. Obviously, $0 \leq C_i \leq 1$. The clustering coefficient of the network is the average value of all nodes.

5) MODULARITY

In complex networks, community structure is a common phenomenon. There are close ties in the community structure, while sparse connections between communities. Newman and Girvan [41] proposed an index Q to measure the community structures. For a division with g communities, then define a $g \times g$ matrix e whose component e_{ij} is the fraction of edges in the original network that connects nodes in community i to those in community j . The modularity is defined as

$$Q = \sum_i e_{ii} - \sum_{ijk} e_{ij}e_{ki} = Tre - \|e^2\| \tag{9}$$

where $\|x\|$ means the sum of all elements of x . It can be achieved $0 \leq Q \leq 1$. $Q = 0$ means the community structure is not stronger than random choice. The larger value of modularity indicates stronger community structure.

6) ASSORTATIVE COEFFICIENT

Assortative coefficient, which describes the tendency of connections between nodes according to node degree. It is a crucial indicator to measure network structure, which is defined as follows [42].

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \tag{10}$$

where M represents the total number of edges, j_i and k_i are the degree of nodes connected by edge i respectively, with $i = 1, \dots, M$.

The assortative coefficient is in the range $[-1, 1]$. The value is closer to 1, the more assortative the network shows, which means nodes with high degree tend to connect nodes with high degree; otherwise, the value is closer to -1 , the more disassortative the network exhibits, which means nodes with high degree tend to connect nodes with low degree.

IV. SPATIOTEMPORAL CHARACTERISTICS OF TAXI TRIPS

A. POINTS OF PICK-UP AND DROP-OFF

The density of pick-up and drop-off points of taxi reflect the traffic demand in the area. Generally, the well-developed areas will have more traffic demand. Identifying the hotspots is crucial for many fields such as city management and traffic control. To illustrate the spatiotemporal characteristic of taxi trip, we have mapped the taxi trips on a weekday (Fig. 5). Fig. 5(a) shows the distribution of pick-up points and Fig. 5(b) shows the distribution of drop-off points. It is noteworthy that

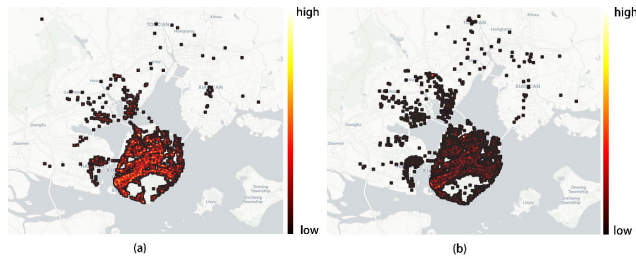


FIGURE 5. Spatial distribution of (a) pick-up points and (b) drop-off points on a weekday (July 3).

the pick-up points are the start points of trips and the dropped off points are the end of trips. We can see that the pick-up points are more concentrated than drop-off points. That is because travelers tend to find conspicuous places to call taxis for making taxi drivers easily find them. From the figure, we can also see that the taxi data are mainly distributed in Xiamen island.

B. CHARACTERISTICS OF TAXI TRIPS

In general, traffic demand fluctuates during different time period in a day. Fig. 6 shows the number of taxi trips per hour from July 2, 2018 to July 8, 2018. Note that, the smallest number of trips appear between 5:00 and 6:00. Another low point appears between 17:00-18:00. The reason is that it is the time to change shifts when some taxi drivers stop service. It is noticed that the peak values emerge around midnight. On weekdays, the peak values appear a little before midnight. While they appear a little after midnight on weekend. The results demonstrate that the public transport system stop operating at midnight and many travelers have to use taxi to travel.

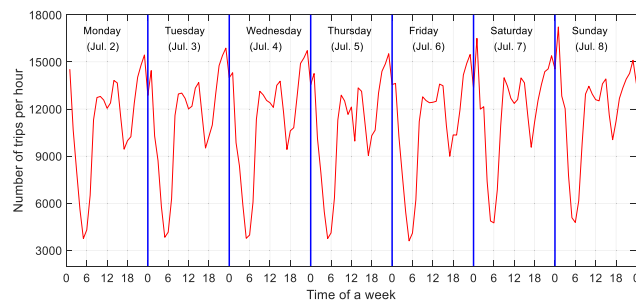


FIGURE 6. Number of taxi trips per hour from July 2, 2018 to July 8, 2018.

Fig. 7 shows the distributions of trip distance and trip time of taxi trips of Xiamen on July 3, 2018. The distribution of trip distance follows a lognormal distribution, the red line in Fig. 7(a) is the fitted curve. Fig. 7(b) shows the trip time distribution, which can also be fitted by a lognormal distribution. We can see that there is a small proportion of trips that have a long distance and spend a long time. Most trips are within 10 km and 40 minutes.

To specifically explore the occurrence and attraction of each cell, we divide a day into 12 periods: 0:00-2:00, 2:00-4:00, 4:00-6:00, 6:00-8:00, 8:00-10:00, 10:00-12:00,

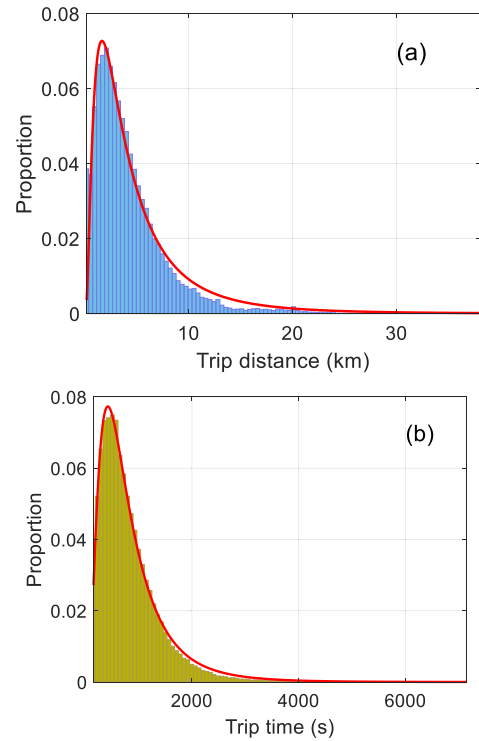


FIGURE 7. Distributions of taxi trips of Xiamen (July 3, 2018): (a) trip distance (b) trip time.

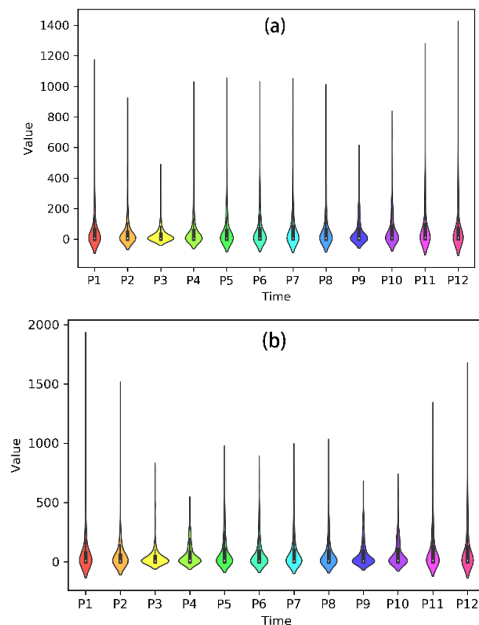


FIGURE 8. Characteristics for 12 periods on July 3, 2018 of cells: (a) number of pick-up, (b) number of drop-off.

12:00-14:00, 14:00-16:00, 16:00-18:00, 18:00-20:00, 20:00-22:00 and 22:00-24:00. In this article, we use P_1, \dots, P_{12} to represent the 12 periods. For a cell in one period, the number of pick-ups and drop-offs are different. The number of pick-up and drop-off could reflect the occurrence and attraction of cells, which have been counted respectively. Fig. 8 illustrates

the distributions of occurrence and attraction of cells. We remove the values equals 0 in the cells which means there is no trips. The figures show there is a small number of cells that have very large number of pick-up and drop-off points. These cells concentrate in the downtown areas. Take the number of pick-up points for example, most cells' value is smaller than 200 in each period of a day. There exist some hotspots in the figure. We can see that the hotspots happen around midnight, which indicates many travelers use taxi when the public transport stop operating. The hotspots in night are located in the central of city and the transportation hubs.

We construct the taxi trip network by dividing the study area into small grids. Trips that start in the same cell are regarded as starting from the central point of the cell, and trips that end in the same cell are regarded as ending from the central point of the cell. Fig. 9 shows the distributions of taxi trips in different periods. Each period is two hours long. We intend to exhibit the trips changes during different time. The trip density is the smallest during 4:00-6:00, which testifies that the traffic demand is smallest in the period. Moreover, the trips within Xiamen island are with high density.

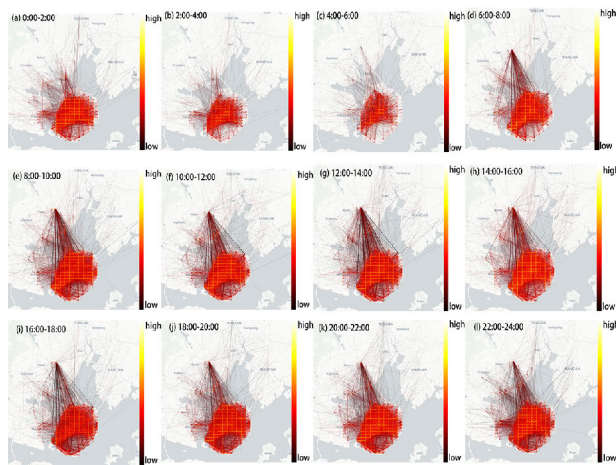


FIGURE 9. Density distributions of taxi trips for 12 periods of July 3, 2018.

V. NETWORK-BASED CHARACTERISTICS

In this part, we discuss the taxi trip network structure using common complex network indicators. In the networks, there exists some nodes that do not connected with other nodes, which are isolated nodes. We remove the isolated nodes when analyzing the network. Fig. 10 shows the taxi trip network on July 3, 2018, which contains 630 nodes and 21349 edges. The colors mean the community structures.

A. CHARACTERISTICS OF NETWORKS IN DIFFERENT PERIODS

To explore the characteristics of networks in different period, we calculate the basic measures of networks in Table 3. The average degree of taxi trip network in a day reaches 33.88, which means the network connection is tight.

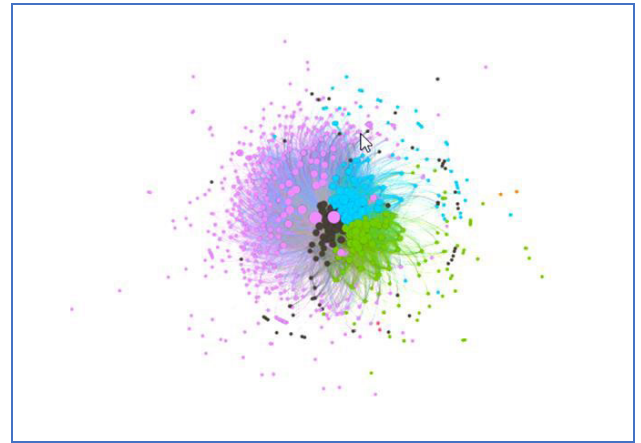


FIGURE 10. Network of taxi trips on July 3, 2018 (Drawn by Gephi).

TABLE 3. Basic network measures of networks in different periods: average degree (AK), average clustering coefficient (ACC), average shortest path length (ASPL), modularity (M), assortativity coefficient (R).

Time	Node	Edge	AK	ACC	ASPL	M	R
P1	361	6312	17.485	0.339	2.575	0.198	-0.0456
P2	318	4495	14.135	0.331	2.488	0.181	-0.1081
P3	277	2995	10.812	0.281	2.744	0.187	-0.1001
P4	329	5651	17.176	0.318	2.519	0.203	-0.0694
P5	382	7484	19.592	0.337	2.524	0.249	-0.0479
P6	383	7237	18.896	0.354	2.562	0.239	-0.0591
P7	371	7074	19.067	0.366	2.487	0.241	-0.1178
P8	381	7427	19.493	0.361	2.561	0.232	-0.0631
P9	366	6358	17.372	0.306	2.565	0.240	-0.0671
P10	356	7255	20.379	0.359	2.458	0.260	-0.0442
P11	345	7649	22.171	0.368	2.455	0.227	-0.0184
P12	368	8172	22.207	0.381	2.484	0.202	-0.0189
Whole day	630	21349	33.88	0.539	2.389	0.222	-0.2026

Furthermore, we note that all the networks in the table have large clustering coefficient and small average shortest path length, which demonstrates that the taxi trip networks are 'small world' networks. The modularity values of all networks are very small. That indicates that the community structures of taxi trip networks are not obvious. The reason is that taxi has no fixed routes. Regions that have long distances can be connected with taxi trips. Moreover, the assortative coefficients of all networks are negative, which means the taxi trip networks are disassortative networks. In reality, travelers prefer to use taxi from suburb to core urban area when the public transport cannot satisfy their travel requirement. The average degree is smallest with value 10.812 in P3, which demonstrates the traffic demand is smallest. The transportation efficiency is worst in P3 that ASPL reaches to 2.744, while it is best in P11 with value 2.455. The network of whole day outperforms the networks in different time periods.

B. DISTRIBUTIONS OF DEGREE DISEQUILIBRIUM FACTOR

Generally, the inbound and outbound traffic flows for nodes are of disequilibrium due to many reasons such as the competition from other traffic modes. The traffic flow disequilibrium could result in heavy traffic congestion and the

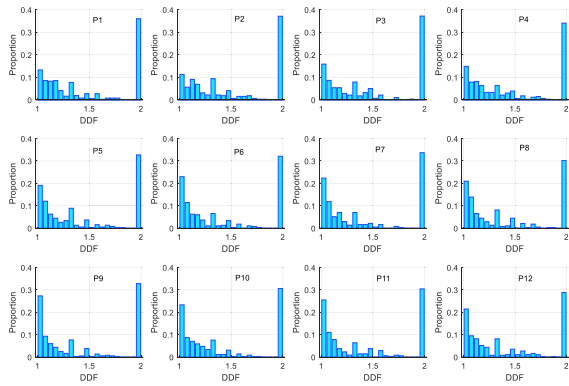


FIGURE 11. Distributions of degree disequilibrium factor in different period.

mismatch between traffic demand and supply. There have been many works to measure the disequilibrium by Gini coefficient, Theil index and complex network indicators [43]–[46]. In this article, we use degree disequilibrium factor (*DDF*) to represent the disequilibrium of taxi traffic demand for nodes. It is defined as follows.

$$DDF = \frac{2 * \max(k_{in}, k_{out})}{k_{in} + k_{out}} \quad (11)$$

Obviously, *DDF* ranges from 1 to 2. The larger value means the more unbalanced of the node. Fig. 11 shows the distributions of *DDF* in 12 periods. As the figure shows, there is a large proportion of nodes' *DDF* values under 1.5, which indicates the disequilibrium is not serious. The proportion tends to decrease between $DDF = 1$ and $DDF = 2$. However, it increases to a large proportion when *DDF* equals 2. $DDF = 2$ means the one of in-degree and out-degree of the node equals 0. We analyze the nodes with value $DDF = 2$ and find that these nodes' degrees are almost smaller than 4, which means these nodes have less connections with other nodes. Furthermore, the connections concentrated in urban core areas. Additionally, these nodes are located in suburb areas with less transit connections with core areas. The proportions of these nodes are very large (approximate 30%). The results imply that taxi plays an important role in connecting peripheral areas and urban core areas.

VI. RELATIONSHIP BETWEEN TAXI TRIPS AND POI NUMBER

Generally, the number of POI in one area reflects the prosperity of the area. In other words, a region with a large number of POIs would have more pick-ups and drop-offs, which means the development of the area is better. In this part, we attempt to study the relationship between the number of POI and the number of taxi trips. Fig. 12 illustrates the number of POI in cells from large values to small values. The cells that have very large number of POI account for small proportions. These cells mostly concentrate in the downtown area. An exponential curve could fit the value distribution.

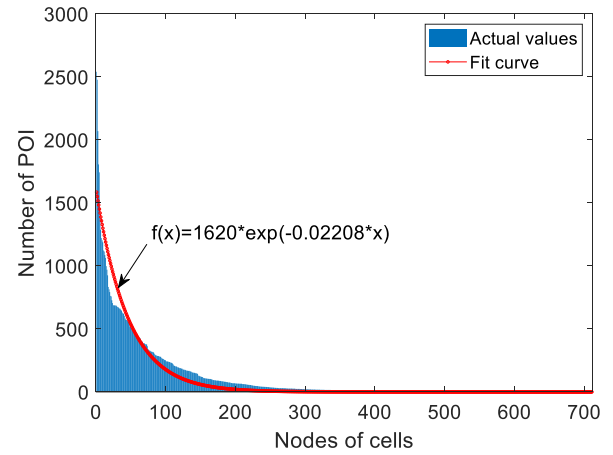


FIGURE 12. Distribution of number of POI for cells.

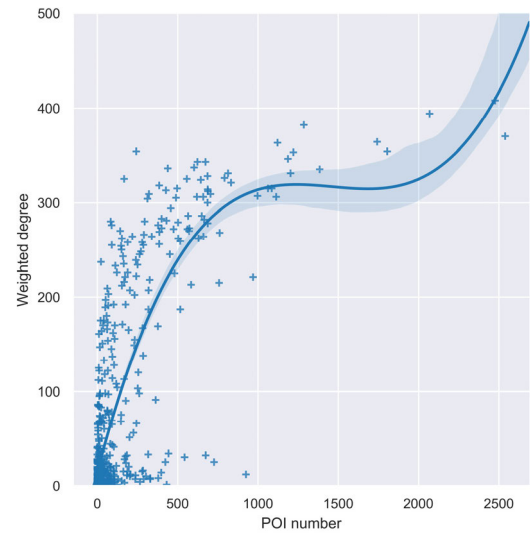


FIGURE 13. Relationship between node degree and POI number in cells.

In this part, we use weighted degree to represent the travel activity because it represents the sum of outbound and inbound flow. We plot the relationship between weighted degree and the number of POIs in cells in Fig. 13. From the figure, we can see that about 70% of cells that have weighted degree smaller than 50. There is a small number of cells that have very large weighted degree. As can be seen, the value of weighted degree increases as the POI number increases. A cubic polynomial curve could fit the relationship. The results show that there is a close relationship between POI number and traffic demand.

VII. CONCLUSION

This article proposes a framework to analyze taxi traffic demand by constructing complex networks based on taxi trajectory data. The study area is divided into small cells with size 1km × 1km. Each cell is regarded as a node and taxi trips that start from a node and end in other node are regarded

as edges. In order to study the fluctuation of traffic demand, we divide a day into 12 periods by two hours and construct networks in each period. Moreover, we count the number of POI for each cell, and build the relationship between travel intensity and the POI intensity.

The results show that the number of taxi trips fluctuate during a day and it reaches the peak around midnight, because most public transport stop operating at that time. Statistical results show that the taxi trip length in a weekday follows a lognormal distribution. The pick-up points are more concentrated than drop-off points. For the taxi trip networks, they exhibit a ‘small-world’ characteristic that having a large clustering coefficient and small shortest path length. The community structures of taxi trip networks in all periods are not obvious, while they show a disassortative characteristic. The results show that residents prefer to use taxi between hotspots and the common region, which indicates that taxi is a significant supplement for the public transportation. The network performance is worst during 4:00-6:00 according to degree and ASPL. The traffic demand in that period is smallest. The network performance becomes better as time goes on, and it reaches the peak at night. Furthermore, we count the number of POI in each cell and achieve that the cells that have large POI intensity account for a small proportion. A cubic polynomial curve could fit the relationship between travel intensity and POI intensity.

This study could provide a new sight to understand the travel demand of taxi. Future research should focus on the following aspects. First, the partition method and scale of the studied area should be discussed specifically. Second, different POI type should be given different weight. Finally, more kinds of data will be introduced to analyze the traffic demand such as bus, metro and bike-sharing to gain accurate activities of residents.

REFERENCES

- [1] G. Zhong, T. Yin, J. Zhang, S. He, and B. Ran, “Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data,” *Transportation*, vol. 46, no. 5, pp. 1713–1736, Apr. 2018.
- [2] X.-F. Xie and Z. J. Wang, “Examining travel patterns and characteristics in a bikesharing network and implications for data-driven decision supports: Case study in the Washington DC area,” *J. Transp. Geography*, vol. 71, pp. 84–102, Jul. 2018.
- [3] S. Zhang, J. Tang, H. Wang, Y. Wang, and S. An, “Revealing intra-urban travel patterns and service ranges from taxi trajectories,” *J. Transp. Geography*, vol. 61, pp. 72–86, May 2017.
- [4] C. Xu, J. Ji, and P. Liu, “The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets,” *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 47–60, Oct. 2018.
- [5] X. Yang, Z. Fang, Y. Xu, L. Yin, J. Li, and S. Lu, “Spatial heterogeneity in spatial interaction of human movements—Insights from large-scale mobile positioning data,” *J. Transp. Geography*, vol. 78, pp. 29–40, Jun. 2019.
- [6] W. Mungthanya, S. Phithakkittukoon, M. G. Demissie, L. Kattan, M. Veloso, C. Bento, and C. Ratti, “Constructing time-dependent origin-destination matrices with adaptive zoning scheme and measuring their similarities with taxi trajectory data,” *IEEE Access*, vol. 7, pp. 77723–77737, 2019.
- [7] P. Lin, J. Weng, S. Hu, D. Alivanistos, X. Li, and B. Yin, “Revealing spatio-temporal patterns and influencing factors of dockless bike sharing demand,” *IEEE Access*, vol. 8, pp. 66139–66149, 2020.
- [8] W. Zhang, S. V. Ukkusuri, and J. J. Lu, “Impacts of urban built environment on empty taxi trips using limited geolocation data,” *Transportation*, vol. 44, no. 6, pp. 1445–1473, May 2016.
- [9] T. Babicheva, W. Burghout, I. Andreasson, and N. Faul, “Empty vehicle redistribution and fleet size in autonomous taxi systems,” *IET Intell. Transp. Syst.*, vol. 13, no. 4, pp. 677–682, Apr. 2019.
- [10] F. Zong, T. Wu, and H. Jia, “Taxi drivers’ cruising patterns—Insights from taxi GPS traces,” *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 571–582, Feb. 2019.
- [11] X. Liang, X. Zheng, W. Lv, T. Zhu, and K. Xu, “The scaling of human mobility by taxis is exponential,” *Phys. A, Stat. Mech. Appl.*, vol. 391, no. 5, pp. 2135–2144, Mar. 2012.
- [12] C.-Z. Yao and J.-N. Lin, “A study of human mobility behavior dynamics: A perspective of a single vehicle with taxi,” *Transp. Res. A, Policy Pract.*, vol. 87, pp. 51–58, May 2016.
- [13] J. Yang, Y. Sun, B. Shang, L. Wang, and J. Zhu, “Understanding collective human mobility spatiotemporal patterns on weekdays from taxi origin-destination point data,” *Sensors*, vol. 19, no. 12, p. 2812, Jun. 2019.
- [14] H. Zhang, B. Shi, C. Zhuge, and W. Wang, “Detecting taxi travel patterns using GPS trajectory data: A case study of Beijing,” *KSCSE J. Civil Eng.*, vol. 23, no. 4, pp. 1797–1805, Jan. 2019.
- [15] P. Zhao, X. Liu, M.-P. Kwan, and W. Shi, “Unveiling cabdrivers’ dining behavior patterns for site selection of ‘taxi canteen’ using taxi trajectory data,” *Transportmetrica A, Transp. Sci.*, vol. 16, no. 1, pp. 137–160, Aug. 2018.
- [16] S. An, H. Yang, and J. Wang, “Revealing recurrent urban congestion evolution patterns with taxi trajectories,” *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 4, p. 128, Mar. 2018.
- [17] Z. Kan, L. Tang, M.-P. Kwan, C. Ren, D. Liu, and Q. Li, “Traffic congestion analysis at the turn level using taxis’ GPS trajectory data,” *Comput. Environ. Urban Syst.*, vol. 74, pp. 229–243, Mar. 2019.
- [18] J. Tang, J. Liang, S. Zhang, H. Huang, and F. Liu, “Inferring driving trajectories based on probabilistic model from large scale taxi GPS data,” *Phys. A, Stat. Mech. Appl.*, vol. 506, pp. 566–577, Sep. 2018.
- [19] J. Tang, Y. Wang, W. Hao, F. Liu, H. Huang, and Y. Wang, “A mixed path size logit-based taxi customer-search model considering spatio-temporal factors in route choice,” *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1347–1358, Apr. 2020.
- [20] F. Zong, Y. Tian, Y. He, J. Tang, and J. Lv, “Trip destination prediction based on multi-day GPS data,” *Phys. A, Stat. Mech. Appl.*, vol. 515, pp. 258–269, Feb. 2019.
- [21] R. Bian, C. G. Wilmot, and L. Wang, “Estimating spatio-temporal variations of taxi ridership caused by hurricanes Irene and sandy: A case study of New York city,” *Transp. Res. D, Transp. Environ.*, vol. 77, pp. 627–638, Dec. 2019.
- [22] J. Tang, S. Zhang, X. Chen, F. Liu, and Y. Zou, “Taxi trips distribution modeling based on entropy-maximizing theory: A case study in Harbin city—China,” *Phys. A, Stat. Mech. Appl.*, vol. 493, pp. 430–443, Mar. 2018.
- [23] J. Cui, F. Liu, J. Hu, D. Janssens, G. Wets, and M. Cools, “Identifying mismatch between urban travel demand and transport network services using GPS data: A case study in the fast growing Chinese city of Harbin,” *Neurocomputing*, vol. 181, pp. 4–18, Mar. 2016.
- [24] R. Su, Z. Fang, N. Luo, and J. Zhu, “Understanding the dynamics of the pick-up and drop-off locations of taxicabs in the context of a subsidy war among E-hailing apps,” *Sustainability*, vol. 10, no. 4, p. 1256, Apr. 2018.
- [25] S. Zhang, D. Zhu, X. Yao, X. Cheng, H. He, and Y. Liu, “The scale effect on spatial interaction patterns: An empirical study using taxi O-D data of Beijing and Shanghai,” *IEEE Access*, vol. 6, pp. 51994–52003, 2018.
- [26] J. Ke, H. Zheng, H. Yang, and X. Chen, “Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach,” *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.
- [27] X. Liu, L. Gong, Y. Gong, and Y. Liu, “Revealing travel patterns and city structure with taxi trip data,” *J. Transp. Geography*, vol. 43, pp. 78–90, Feb. 2015.
- [28] Z. He, L. Zheng, P. Chen, and W. Guan, “Mapping to cells: A simple method to extract traffic dynamics from probe vehicle data,” *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 3, pp. 252–267, Jan. 2017.
- [29] Y. Yang, Z. He, Z. Song, X. Fu, and J. Wang, “Investigation on structural and spatial characteristics of taxi trip trajectory network in Xi’an, China,” *Phys. A, Stat. Mech. Appl.*, vol. 506, pp. 755–766, Sep. 2018.
- [30] J. Yang, D. Yi, B. Qiao, and J. Zhang, “Spatio-temporal change characteristics of spatial-interaction networks: Case study within the sixth ring road of Beijing, China,” *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 6, p. 273, Jun. 2019.
- [31] Z. He, “Spatial-temporal fractal of urban agglomeration travel demand,” *Phys. A, Stat. Mech. Appl.*, vol. 549, Jul. 2020, Art. no. 124503.

[32] R. Cervero, "Built environments and mode choice: Toward a normative framework," *Transp. Res. D, Transp. Environ.*, vol. 7, no. 4, pp. 265–284, Jun. 2002.

[33] C. Ding, Y. Lin, and C. Liu, "Exploring the influence of built environment on tour-based commuter mode choice: A cross-classified multilevel modeling approach," *Transp. Res. D, Transp. Environ.*, vol. 32, pp. 230–238, Oct. 2014.

[34] K. Sila-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. Demšar, and A. S. Fotheringham, "Analysis of human mobility patterns from GPS trajectories and contextual information," *Int. J. Geographical Inf. Sci.*, vol. 30, no. 5, pp. 881–906, Oct. 2015.

[35] Y. Liu, Y. Zhang, S. T. Jin, and Y. Liu, "Spatial pattern of leisure activities among residents in Beijing, China: Exploring the impacts of urban environment," *Sustain. Cities Soc.*, vol. 52, Jan. 2020, Art. no. 101806.

[36] P. Zhao, Y. Xu, X. Liu, and M.-P. Kwan, "Space-time dynamics of cab drivers' stay behaviors and their relationships with built environment characteristics," *Cities*, vol. 101, Jun. 2020, Art. no. 102689.

[37] M. Li, M.-P. Kwan, F. Wang, and J. Wang, "Using points-of-interest data to estimate commuting patterns in central Shanghai, China," *J. Transp. Geography*, vol. 72, pp. 201–210, Oct. 2018.

[38] B. Zhang, S. Chen, Y. Ma, T. Li, and K. Tang, "Analysis on spatiotemporal urban mobility based on online car-hailing data," *J. Transp. Geography*, vol. 82, Jan. 2020, Art. no. 102568.

[39] X. M. Liu, L. Sun, Q. X. Sun, and G. Gao, "Spatial variation of taxi demand using GPS trajectories and POI data," *J. Adv. Transp.*, vol. 2020, Jan. 2020, Art. no. 7621576.

[40] X. Liu, Y. Tian, X. Zhang, and Z. Wan, "Identification of urban functional regions in Chengdu based on taxi trajectory time series data," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 3, p. 158, Mar. 2020.

[41] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 2, Feb. 2004, Art. no. 026113.

[42] M. E. J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, no. 20, Oct. 2002, Art. no. 208701.

[43] S. Goswami, C. A. Murthy, and A. K. Das, "Sparsity measure of a network graph: Gini index," *Inf. Sci.*, vol. 462, pp. 16–39, Sep. 2018.

[44] F. Bickenbach, A. Mbelu, and P. Nunnenkamp, "Is foreign aid concentrated increasingly on needy and deserving recipient countries? An analysis of Theil indices, 1995–2015," *World Develop.*, vol. 115, pp. 1–16, Mar. 2019.

[45] M. Bellingeri, D. Bevacqua, F. Scotognella, and D. Cassi, "The heterogeneity in link weights may decrease the robustness of real-world complex weighted networks," *Sci. Rep.*, vol. 9, no. 1, p. 10692, Jul. 2019.

[46] F. Su, Y. Qin, X. Zou, and H. Su, "Disequilibrium measurement and analysis of a road traffic network based on section influence degree and flow betweenness," *IEEE Access*, vol. 8, pp. 24428–24442, 2020.



HUI ZHANG received the B.S. degree in applied mathematics from Qingdao Agricultural University, in 2009, and the M.S. degree in system science and the Ph.D. degree in transportation planning and management from Beijing Jiaotong University, in 2012 and 2016, respectively. From 2016 to 2019, he was a Lecturer with the School of Transportation Engineering, Shandong Jianzhu University, where he has been an Associate Professor, since 2019. His research interests include transportation engineering, big data analysis, complex networks, and engineering management.



LELE ZHANG received the B.S. degree in electronic and information engineering from Naval Aeronautical and Astronautical University, in 2008. His research interests include data mining, system optimization, communication and information systems, and intelligent transportation.



FA CHE received the Ph.D. degree in highway and railway engineering from Chang'an University, in 2011. His research interests include traffic safety, highway operation management, and traffic engineering.



JIANMIN JIA was born in Heze, Shandong, China, in 1988. He received the B.S. degree in information and computing science and the M.S. degree in transportation engineering from Shandong University, in 2010 and 2013, respectively, and the Ph.D. degree in transportation engineering from Florida International University, Miami, FL, USA, in 2017. From 2013 to 2017, he was a Research Assistant with the Lehman Center for Transportation Research (LCTR), Florida International University. Since 2018, he has been a Lecturer with the Shandong Jianzhu University, China. He is the author of more than ten articles. His research interests include traffic simulation and modeling, big data analysis, and travel demand forecasting.



BAIYING SHI received the B.S. degree in traffic and civil engineering from Shandong Jianzhu University, in 1999, the M.S. degree in traffic information and control from Lanzhou Jiaotong University, in 2004, and the Ph.D. degree in traffic information engineering and control from Tongji University. He is currently an Associate Professor with the School of Transportation Engineering, Shandong Jianzhu University. His research interests include big data in transportation, public transportation, and traffic control.

...