

Research Article

Analysis of Spatial-Temporal Characteristics of Operations in Public Transport Networks Based on Multisource Data

Hui Zhang,¹ Yanjun Liu,¹ Baiying Shi ,¹ Jianmin Jia,¹ Wei Wang ,² and Xiang Zhao³

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

²School of Economics, Ocean University of China, Qingdao 266100, China

³Institute of Computing Technology, China Academy of Railway Sciences, Beijing 100044, China

Correspondence should be addressed to Baiying Shi; shibaiying@sdjzu.edu.cn

Received 26 June 2021; Accepted 30 August 2021; Published 9 September 2021

Academic Editor: Yong Wang

Copyright © 2021 Hui Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Operational efficiency and stability are two critical aspects to measure bus systems. Influenced by many stochastic factors, buses always suffer from delay and bunching. Traditional studies focus on a single route and lack research on the systematic evaluation of bus network. In this paper, we propose a data-driven framework to analyze the efficiency and stability based on small granularity GPS trajectory data from the perspective of entire bus network. The IC card data and route data are used to extract the boarding passenger number and topological structure, respectively. The results show that the average headway of stations follows a lognormal distribution. Moreover, the distribution of arrival efficiency of stations is inhomogeneous and a small number of stations have large values. In addition, the relationships among average headway of stations, boarding passenger number, bus number, and complex network indicators are revealed. It is found that the average headway of station is negatively correlated with other indicators, which implies that complex network connections and more passenger flows could weaken the efficiency of bus operations. This paper provides a way to evaluate the operational performance of bus networks and could give help for monitoring and optimizing the daily operation of bus systems.

1. Introduction

Nowadays, public transport plays a growing role in alleviating traffic congestion and reducing greenhouse gas emission. Building effective, convenient, and stable public transport has become a crucial step to solve urban traffic problems in many countries. The efficiency and stability of public transport are two core problems that are concerned by both travelers and operators. Affected by many stochastic factors such as weather, congestion, passenger flow, and drivers' behavior, bus delay and bunching occur in many routes during the operation time [1–4]. Bus bunching is that the adjacent buses belonging to the same route are too close to each other. Bus bunching could impact the uniformity of buses and lead to big interval from the other buses, which enhance passengers' waiting time and result in low efficiency. In some bus systems with dense stations and large passenger flow, a small disturbance of buses in a route may

spread all over the network. So, it is imperative to construct a robust public transport system to provide better service.

The development of public transport information technology provides a powerful tool to monitor and manage the transportation system. The Automatic Vehicle Location (AVL) system and Automatic Passenger Count (APC) system could record the travel trajectory and passenger information. In recent years, many studies focused on travel time, delay, and reliability of buses based on AVL data [5–7]. To address these problems, numerous control strategies have been proposed, like speed control [8], holding strategies [9, 10], and skip-stop strategies [11]. The prediction methods give great help to improve the control effects, which is better than traditional models without predictions [12]. High-resolution bus GPS data could also be used to identify congestion hotspots in the urban street [13]. The APC data are commonly used to extract OD information, estimate waiting time, and find missing transfers [14–16]. Recently,

some researchers use APC data to identify public transit corridors and transit network flow characteristics [17, 18]. Typically, the bus operational status can be achieved by AVL data incorporating APC data [19].

Essentially, the delay and bunching of bus are caused by small disturbances such as more boarding passengers and bad traffic condition of road. These small disturbances may spread, superpose, and amplify, which results in heavy disorder of the entire bus system. The public transit system is a complex spatiotemporal network embedded in complicated urban surroundings [5]. Understanding the mechanism of bus operation is conducive to enhance the efficiency of transit service. Most researches have focused on a single route [8, 20], and few studies pay attention to the whole transit network's operational stability. The structures and dynamics of bus networks are so complicated. In the past years, researchers studied the bus networks from complex network perspective using line information and IC card data [21–23]. They pay main attention to the structure characteristics of bus such as community structure and “small world.”

Recently, there are many researches focusing on extract bus network characteristics by merging AVL data and APC data. Chen et al. used IC card data and GPS data to estimate passenger boarding and alighting station of the entire bus and found that passenger flow is mainly distributed in an east-west belt-shaped downtown area [19]. Sui et al. constructed a layered network model to depict public transport network, OD flows, and transfer flows [24]. The data-driven methods appear to analyze the whole bus network performance in recent years. Zhang et al. studied the average headway and headway deviation of the entire bus network in Jinan, China, and found that the two indicators follow lognormal distributions [25]. Iliopoulou et al. used AVL data to identify spatiotemporal patterns of bus bunching by clustering method. However, they did not study the passenger flow factor due to the lack of APC data [26]. Nowadays, network-based studies using big data have been successfully applied in many transportation systems [27–31].

Most aforementioned studies focus on the operational performance of a single route using AVL data. However, there is limited research for the whole network. Actually, the operational performance of bus network is more significant in bus network design and optimization of operation. For example, planners consider passenger flow, network structure, accessibility, and transfer when designing or adding new routes. They rarely consider the operation performance systematically. Therefore, there exist many stations having very bad operation performance. Passengers should wait a long time or suffer from bus bunching at these stations. This is a big problem for bus systems. At present, the network-based studies of bus network concentrated on the topological structure rather than operational performance. How to evaluate the bus operation performance and find the hub stations is the key to solve the problem. To fill this gap, this paper designs proper indicators based on GPS trajectory data to evaluate the operational performance of the entire bus network.

In this study, we conduct a data-driven framework to measure the operational status of bus network-based incorporating GPS trajectory data, IC card data, and route data. First, we propose average headway of station and average arrival rate of station to assess the stability and efficiency of bus network. Second, the relationships among the two proposed indicators, boarding passengers, bus supply, and topological structure are studied. This paper aims to construct a framework based on multisource data to evaluate the operation of bus systems. The rest paper is organized as follows. Section 2 gives the literature review. Section 3 proposes evaluation indicators of bus network performance. Section 4 introduces the data used in this paper. Section 5 shows the results. Section 6 concludes the paper.

2. Literature Review

2.1. Application of Locator Data in Transportation Area. The GPS trajectory data are space-time continuum, which is the foundation to study bus operation status and residents' activity. In transportation area, it has been widely applied in travelers' behavior, estimation of traffic demand, traffic status, and traffic model optimization. The GPS trajectory data are playing a more and more important role in traffic planning and management. In some areas, it has become an alternative to traditional traffic survey.

At present, the studies of travelers' behavior and estimation of traffic demand concentrated on taxi trajectory data and mobile phone data. Tang et al. divided the studied area into small cells and estimated the distribution of OD [32]. Zhang et al. utilized complex network theory to reveal the urban traffic demand based on taxi trajectory data [33]. Mobile data are a good source to study travelers' behavior, which is a research hotspot [34, 35]. Dockless bike sharing is an emerging traffic mode, which plays a significant role in connecting with other traffic modes. The trip data can be used to analyze travelers' behavior, traffic demand estimation, and bicycle rebalance [36–38]. In the transit area, there are many researches that focus on travel time prediction and delay [5, 39]. The GPS trajectory data could reflect the operational status of bus, which is crucial for operation management. Besides that, the trajectory data can be used to detect the travelers' behavior incorporating IC card. Tu et al. studied the dynamic characteristics of multimode travel by trajectory data, IC card data, and taxi trajectory data [40]. Tang et al. used GPS trajectory data and smart card data to optimize the timetable of the bus line [41]. The GPS trajectory data could also be used to identify transportation mode by GIS information and machine learning methods [42, 43].

2.2. Operational Stability of Public Transport. The operational stability of public transport has drawn much attention since the AVL and APC devices were applied. Chepuri et al. pointed out that the travel times during peak hours followed normal distributions [44]. Fan and Machemehl verified the relationship between the waiting

time of passengers and headway deviation and they found a positive relationship between them [45]. The travel time of bus contains running time and stopping time. Studies show that the running time is affected by traffic conditions of roads and traffic signals [46, 47]. Kieu et al. analyzed the distribution of transit travel time used transit signal priority data and found that they followed lognormal distributions [48]. The stopping time comprises door opening, door closing, and the time of passengers boarding and alighting. Research shows that the time of stopping could account for 26% of total travel time [49]. Generally, the main factor that impacts the stopping time is the number of boarding and alighting passengers [50]. Ji et al. claimed that enlarging the platform areas and installed guide guardrails could reduce the variation of dwell time, but not the time [51]. Passengers' arrival times are related to the headway. When the headway is smaller than 12 minutes, passengers arrive randomly [45]. Chepuri et al. proposed new reliability indicators to measure the bus route by trajectory data from both route level and segment level [52]. Paudel revealed that high volumes of bus ridership could cause a significant increase in the variance of bus service reliability [53].

2.3. Evaluation of Transit System. The service quality of the transit system is influenced by many components such as transit network structure and management level. The transit systems are fragile when they meet periodic passenger flow fluctuations and other stochastic factors. It is necessary to grasp the operational status of the transit system. Traditional studies focused on a single route rather than the entire network. From the evaluation of transit network perspective, there have been various researches in the past years. Zhang et al. constructed the evaluation framework which includes convenience, comfort, security, reliability, and facility level according to survey data [54]. Lots of studies analyzed transit network topology structure using complex network theory [55–57]. Sun et al. estimated the transit ridership by points of interest [58]. Chen et al. used GPS trajectory data and IC card data to estimate the number of boarding and alighting passengers and the distributions of passenger flow [19]. Illiopoulou et al. used the clustering method to distinguish different types of bus bunching [26]. Zuo et al. proposed a holistic accessibility measurement considering land use, timetable, and individual factors to evaluate the bus system [59]. Zhang et al. identified the bus station patterns based on network structure, operational status, and POI data [60]. Wei et al. highlighted the bus lines value and proposed a “line-line” network to examine the spatial characteristics of cross-administration bus lines [61]. Bree et al. studied the relationship between transit ridership and local accessibility and found that it was more closely to predict public transit ridership when including gravity-based accessibility in the model [62]. All in all, there were many studies on the evaluation of transit system, but there is little study focus on the operational status of the entire transit network.

3. Evaluation Indicators of Bus Network Performance

The performance of a bus network is significant in daily operations. Reliability and efficiency are two key indicators to measure the performance of bus systems. Headway is an important indicator to study the reliability. The headway between adjacent buses varies, affected by many stochastic factors such as traffic congestion and bad weather. The GPS trajectory data are a feasible source to detect the operational status of buses. Figure 1 shows the GPS trajectory data of bus routes. Traditional studies mainly focus on the operational status of a single route. In this paper, we use the average headway of stations and the average arrival rate of stations to evaluate the operational performance of bus networks. We intend to extract the macroscopic operation status of bus networks by mining the GPS trajectory data.

3.1. Average Headway of Stations. The bus run is considered as a series of events that comprises section running, arrivals, and departures. Headway is the time difference between two successive vehicles belonging to the same routes. It is defined as [25]

$$\Delta H_k^{i,j} = \begin{cases} 0, & j = 1, \\ |h_k^{i,j+1} - h_k^{i,j}|, & \text{otherwise,} \end{cases} \quad (1)$$

where $\Delta H_k^{i,j}$ is the headway between vehicle j and the former vehicle of line i at station k ; set the headway for the first vehicle as 0; $h_k^{i,j}$ is the arrival time of vehicle j of line i at station k .

Generally, the headways of buses are even according to their schedule timetable when they depart at original stations. The headways fluctuate due to the complicated external environment and drivers' factors. Sometimes, the headway between two consecutive buses is smaller than the plan headway. When the headway is too small that two consecutive buses arrive at the same time, the bus bunching appears. On the contrary, there is larger headway. Figure 2 shows the four kinds of bus operation status. Bus bunching and large headway disturb the balance of operation, which could result in long waiting time for passengers.

In this paper, the average headway of stations during time period t is defined as

$$\overline{\Delta H}_k^t = \sum_{i=1}^{m_t} \sum_{j=1}^{n_t^i} \frac{\Delta H_k^{i,j}}{s_k^t}, \quad (2)$$

where m_t is the number of routes that stop at station k , n_t^i is the number of buses of route i during time period t , and s_k^t is the total number of buses that stop at station k during time period t .

3.2. Average Arrival Rate of Stations. Bus operational efficiency is crucial for passengers, which could reflect the quality of service. In this paper, we introduce the average arrival rate of stations to measure the operational efficiency of the entire bus network. It is defined as the mean value of

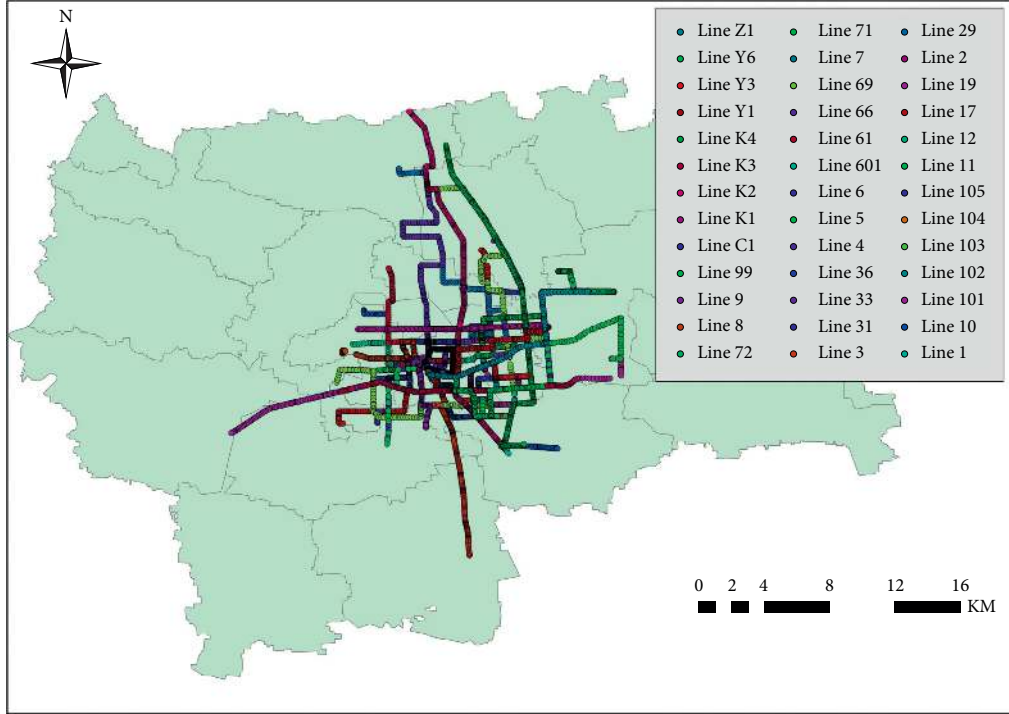


FIGURE 1: GPS trajectory points of bus lines.

time of the nearest bus that runs to a station at a certain time. This index is correlated with the position of the buses. When a bus delays, the operational efficiency decreases. It is known that the time of sending GPS message is out of synchronization. To solve the problem, we introduce a small time period and choose the last GPS message in the time period.

The average arrival rate of stations is defined as

$$\lambda_k^t = \begin{cases} \frac{1}{(1/s_{\Delta t}^k) \sum_{i=1}^L \sum_{j=1}^{O_i} (t_{ij}^k - t_{ij}^c + \Delta t) \gamma_{ijk}^{\Delta t}}, & t_{ij}^k - t_{ij}^c \geq 0, \\ 0, & t_{ij}^k - t_{ij}^c < 0, \end{cases} \quad (3)$$

where L is the number of bus routes, O_i is the number of buses of i -th route that running to k station, Δt is the given time period, $\Delta t \geq \max(t_{\text{send}})$, t_{send} is the time interval of sending message, t_{ij}^k is the arrival time of i -th route of j -th vehicle at station k , t_{ij}^c is the last message time of i -th route of j -th vehicle in the time period $t + \Delta t$, and $s_{\Delta t}^k$ is the number of buses that runs to station k . $\gamma_{ijk}^{\Delta t} = 1$ if i -th route of j -th vehicle is running to station k ; otherwise, $\gamma_{ijk}^{\Delta t} = 0$. Figure 3 shows the sketch map of arrival efficiency of bus station.

3.3. Structural Measures. In many complex networks, the structure of network has an important impact on system functions. Some nodes with a large degree or betweenness play a critical role in system dynamics. During the past two decades, there are many studies focusing on the topological structure of public transport networks such as bus and metro. In this part, we introduce four main network-based

indicators to study the relationship between bus network structure and operational performance.

3.3.1. Node Degree. Node degree is defined as the number of nodes that are connected with the node. Typically, nodes with a very large degree account for a very small proportion. It is defined as

$$k_i = \sum_{j \neq i}^N e_{ij}, \quad (4)$$

where k_i denotes the degree of node i , e_{ij} is the connection status between node i and node j , and $e_{ij} = 1$ means that there exist connections between node i and node j ; otherwise $e_{ij} = 0$.

3.3.2. Betweenness. In transportation networks, transport efficiency is very crucial. Betweenness is another index to evaluate the importance of nodes in propagation. It is calculated as follows:

$$C_B(v) = \sum_{s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \quad (5)$$

where $C_B(v)$ is the betweenness value of node v , σ_{st} is the number of shortest paths going from s to t , and $\sigma_{st}(v)$ is the number of the shortest paths going from s to t through the node v .

3.3.3. Clustering Coefficient. Clustering coefficient reflects the connections among the neighbors of a node. It is defined as

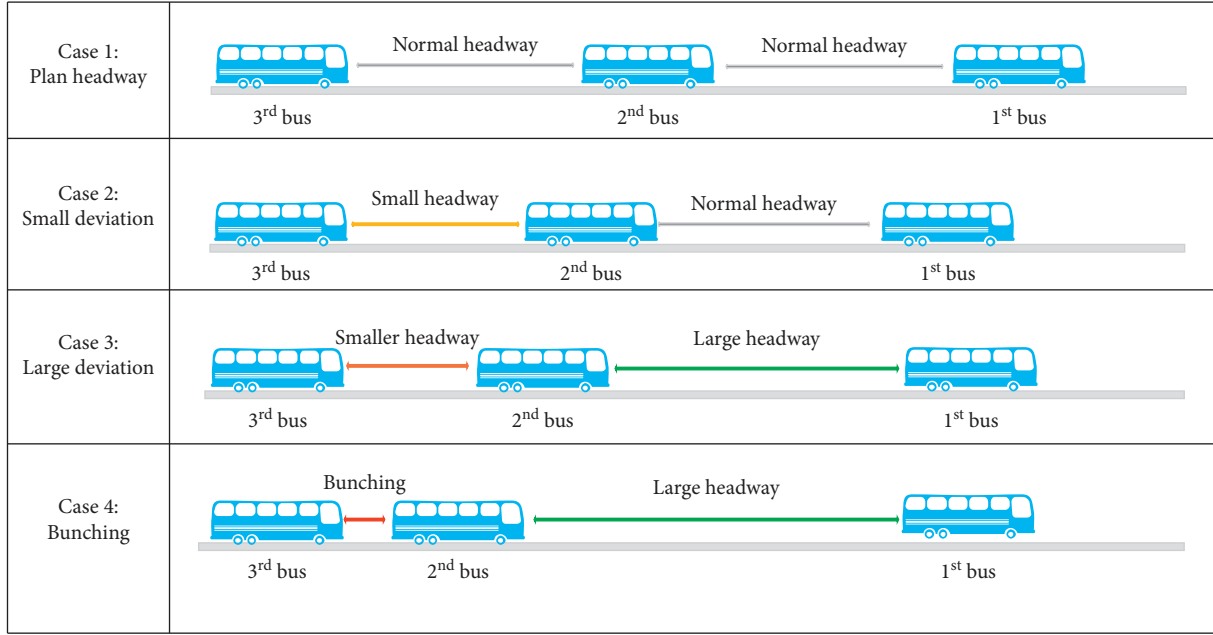


FIGURE 2: Four kinds of bus operation status.

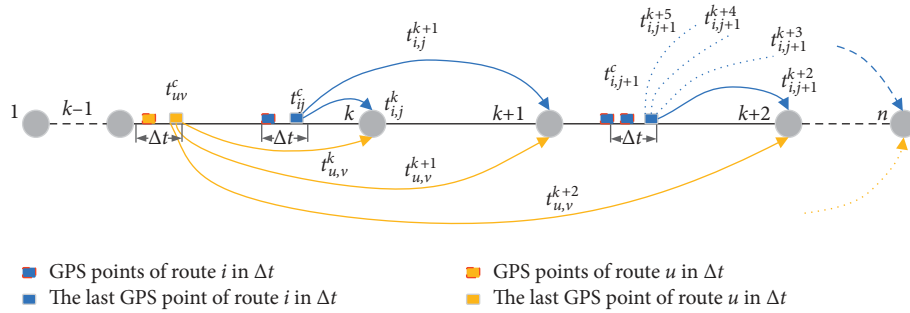


FIGURE 3: Sketch map of arrival efficiency of bus station.

$$C(v_i) = \frac{2e_i}{m_i(m_i - 1)}, \quad (6)$$

where $C(v_i)$ is the clustering coefficient of node v_i , e_i is the number of edges among local neighbors of node v_i , and m_i is the connection degree of neighbors of node v_i .

3.3.4. *PageRank*. PageRank is widely used to measure node importance in many networks. PageRank could grasp the global topological information. Thus, we introduce the PageRank algorithm to achieve the importance of nodes, which is defined as

$$p_i = \frac{1}{N} (1 - p) + p * \sum_{j=1}^N \frac{a_{ji}}{k_j^{\text{out}}} * p_j, \quad (7)$$

where p_i denotes the influence score of i th node, p is the damping coefficient, k_j^{out} means the out-degree of j th node, and a_{ji} is the adjacency matrix.

4. Data Description

The data were collected on October 25, 2018, in Xuchang, China. Xuchang is a famous historical city that is located in the central Henan Province. The studied bus network of Xuchang comprises 39 routes and 629 stations. Figure 4 exhibits the bus network of Xuchang, China. There are three kinds of data that contain GPS trajectory data, IC card data, and bus routes data.

Table 1 shows the basic information of bus route data, which includes line number, direction, station index, station name, longitude, and latitude. Direction “0” represents the upstream route and “1” represents the downstream route. The bus network can be constructed by the bus route information. It is noted that the station index in the table is given for a route. In this paper, each station will be given a unique index when generating the bus network. Table 2 gives the main information of GPS trajectory data, which contains line number, bus number, date, time, longitude, latitude, station index, and direction. The GPS devices of bus send a message every 10–15 seconds. The station index “0” means

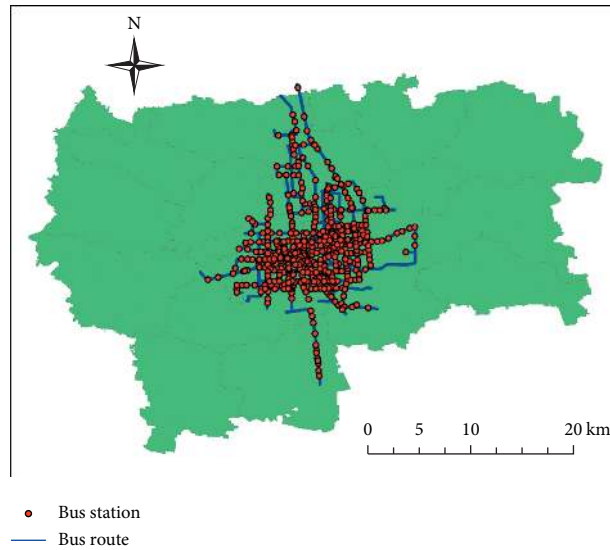


FIGURE 4: Bus network of Xuchang, China.

TABLE 1: Illustration of bus routes data of Xuchang.

Line no.	Direction	Station index	Station name	Longitude	Latitude
1	0	1	Huilong Guoji	113.7723	34.0141
1	0	2	Shiba Zhongxue	113.7761	34.0150
.....
1	0	36	Shangmaocheng	113.8607	33.9769
2	1	1	Gongjiao Gongsi	113.8808	34.0675
2	1	2	Dazhang Shequ	113.8777	34.0648
.....

Direction: “0” represents upstream route and “1” represents downstream route.

TABLE 2: Main information of GPS trajectory data.

Line no.	Bus no.	Date	Time	Longitude	Latitude	Station index	Direction
1	1511	2018/10/25	6:38:17	113.7723	34.0141	0	0
1	1511	2018/10/25	6:38:31	113.7725	34.0142	0	0
.....
1	1511	2018/10/25	6:40:27	113.7761	34.0150	2	0
.....
1	113.7812	34.0161	0	0
.....

Station index: “0” represents nonstation.

that the buses are running in the sections between stations. The running status of each bus can be achieved by the GPS trajectory data. In addition, we can get the departure and arrival information of stations for each bus. Table 3 introduces the main information of IC card, which comprises card number, line number, bus number, date, time, and direction. We can obtain the boarding number of passengers by joining with GPS trajectory data.

The data were preprocessed to eliminate the outliers. The duplicated and missing values are removed. Moreover, the longitudes and latitudes out of the studied area are removed. The bus network will be constructed by the bus routes data.

TABLE 3: Information of IC card data.

Card no.	Line no.	Bus no.	Date	Time	Direction
EB3F10	1	1511	2018/10/25	7:27:16	0
000F2A	1	1511	2018/10/25	7:36:28	0
EB278D	1	1511	2018/10/25	8:28:06	0
.....

The performance of the bus network could be calculated by the proposed indicators with given data. Then the statistical analysis and visualization will be done. In this paper, we use ArcGIS 10.3 to visualize the data.

5. Results

5.1. Traffic Demand and Bus Vehicle. Passenger flow and bus supply of bus play critical roles in daily management. On one hand, bus will delay at stations if there are a large number of boarding and alighting passengers. On the other hand, bus operational efficiency will enhance if the operational company provides more buses. Figure 5 shows the number of boarding passengers and bus vehicles during different hours in a day. We can see that the distributions of boarding passengers exhibit two obvious peaks in morning and afternoon. During the morning peak (7:00–9:00) and evening peak (16:00–18:00), there are many commuters. The peaks are not sharp; that is because some elder people prefer to travel to avoid peak hours [63]. The number of bus distributions is smooth and the large value appears in the morning peak hours.

Figure 6 shows the distributions of bus number and boarding passenger number of stations in a day. As we can see that the bus number of stations follows a lognormal distribution and the boarding passenger number of stations follows an exponential distribution. The results indicate that the stations that have a large value of bus number and boarding passenger number account for small proportions. In bus networks, the lognormal and exponential distributions are ubiquitous. It is reported that the average dwell time and headway deviation of stations follow lognormal distributions [25]. Bus network structure could have some impacts on the results. A small proportion of stations in the bus network have more connections and most stations have a few connections. The degree of stations follows power-law distributions or exponential distributions [55].

Figure 7 shows the spatial distributions of bus number and boarding passenger numbers. We can see that the stations with large values are located in the central city. The stations with a large number of boarding passengers are more concentrated. Operators should pay more attention to the balance of operational status of the bus network.

5.2. Average Headway of Stations. Bus headway is an important index to measure the operational efficiency and reliability. The headway of a route is changing according to many stochastic factors such as the number of boarding and alighting passengers, traffic signal, and weather conditions. For passengers, a stable and small headway is expected. Unstable headway will disturb passengers' travel plan and large headway will enhance the waiting time. For operators, stable headway could enhance the efficiency of bus routes. There are a large number of studies focusing on strategies to keep the headway stable by holding bus, skipping stations, and controlling signals [64–66]. Bus bunching is a serious problem about headway stability in daily operation, which means two or more buses of one route arrive at the same station simultaneously. Figure 8 exhibits the running maps of route 9 and route 16. As can be seen, there are many buses running with small headways in route 16 during the morning and evening peak hours.

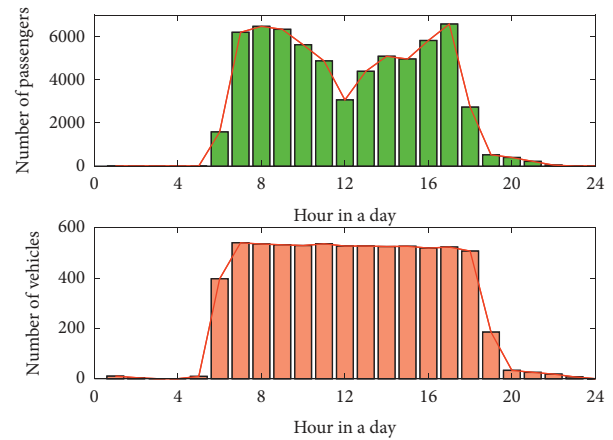


FIGURE 5: Number of passengers and vehicles of bus system per hour in a day.

Traditional headway studies mostly focus on single routes. The situation of the entire network headway is not very clear. Understanding the headway condition of bus networks could provide a macroscopic view for planner and manager when making schedules and strategies. In this paper, we use the proposed average headway of stations to evaluate the whole network headway conditions. Figure 9 shows the distributions of average headway of stations in different time periods: (a) 8:00–9:00, (b) 10:00–11:00, (c) 12:00–13:00, (d) 14:00–15:00, (e) 16:00–17:00, and (f) 18:00–19:00. The red curve is the fitted curves of lognormal. We can see that the average station headways follow lognormal distributions with obvious right tails. Most values are concentrated between 100 seconds and 1000 seconds. Stations with large values more than 1000 seconds account for a small proportion.

Figure 10 shows the average headway of station values of all stations in a day. We can see that there are some stations that have very large values. The top five stations are Ruixianglu-Gongnonglu, Xuchang Shiyuan School, Jianglijijie, Yangguangdao Dongkou, and Nongji Wuliuyuan. In addition, there is a large difference in the deviation of the values in different hours in a day among the stations. The standard deviations range from 6.32 to 2780.

5.3. Average Arrival Rate of Station. The average arrival rate of station is an index to measure the instantaneous operational efficiency of stations, which can be achieved by the high dense GPS trajectory data. In this part, we mainly consider the buses that run to the nearest station. Figure 11 shows the average arrival rate of station that there exit buses running to them at 8:00, 10:00, 12:00, 14:00, 16:00, and 18:00. As we can see, the distributions of the values are inhomogeneous. A small number of stations have large values, while most stations have small values. In the morning and evening peak hours, there are many stations that have a large value. The reason is that there are more buses running on the road in those time periods. For one station, the average arrival rate is changing as time goes. The indicator could measure the station service quality instantaneously.

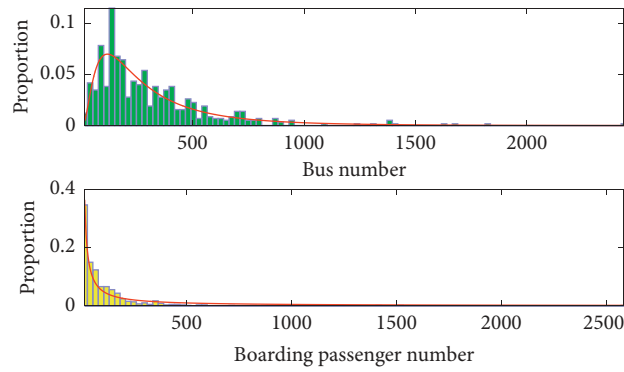


FIGURE 6: Distributions of bus number and boarding passenger number.

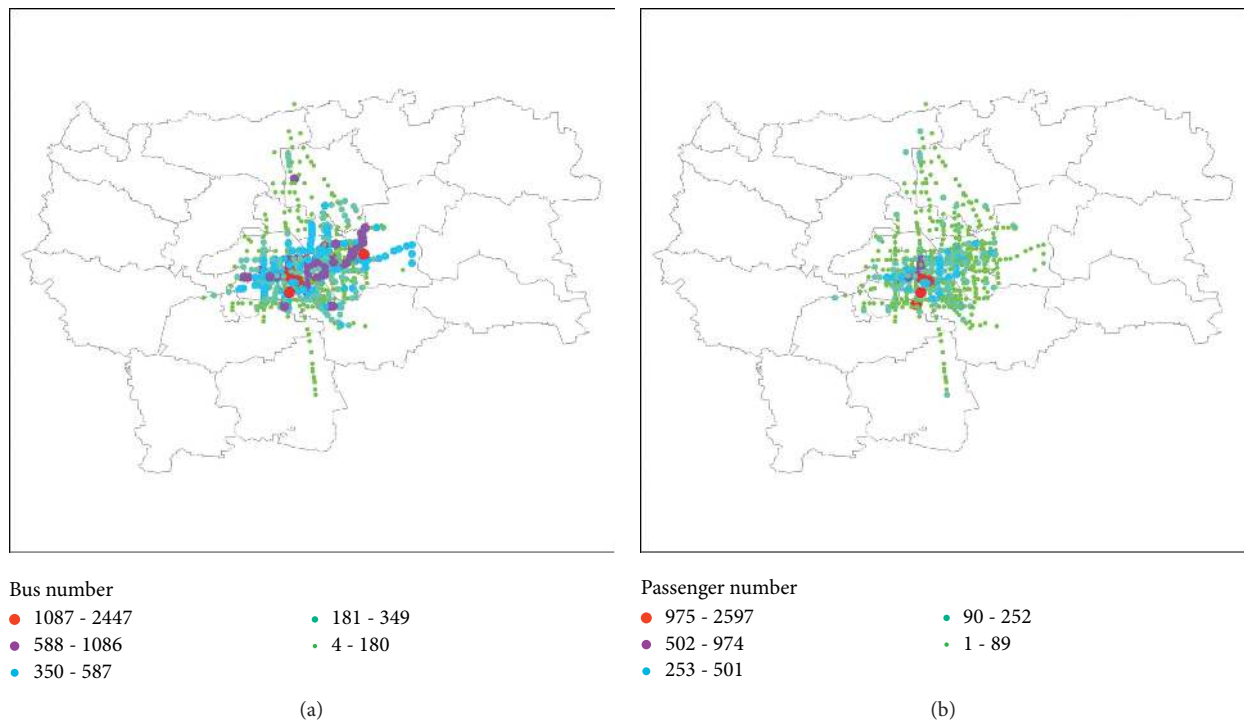


FIGURE 7: Maps of spatial distributions of (a) bus number and (b) passenger number.

Operators could make the targeted policy to enhance the arrival efficiency of stations by providing more buses.

Figure 12 shows the mean value of the average arrival rate of station at different hours of a day. We can see that the values are large in the peak hours and small in the off-peak hours. The value is very small after 20:00, because most routes stop running on the road. It is noted that there are some time periods that have larger values. Due to some stochastic factors, the headways are affected and lose homogeneity. Operators should provide a more robust timetable to keep the values stable. As we know, the stability of headway plays an important role in residents' travel mode choice. High operation efficiency and stable headway could attract more

people to use the public transport. The headway control could not pay attention to a single route but the entire bus network.

5.4. Topological Structural Characteristics. In many complex networks, the topology structure is of significance in system function. We introduce node degree, betweenness, clustering coefficient, and PageRank index to evaluate the structural characteristics of the studied bus network. Figure 13 shows the spatial distributions of the four indicators. There is a small proportion of stations that have very large values of the four indicators, which are concentrated on the center of the city. Take node degree, for example; there is

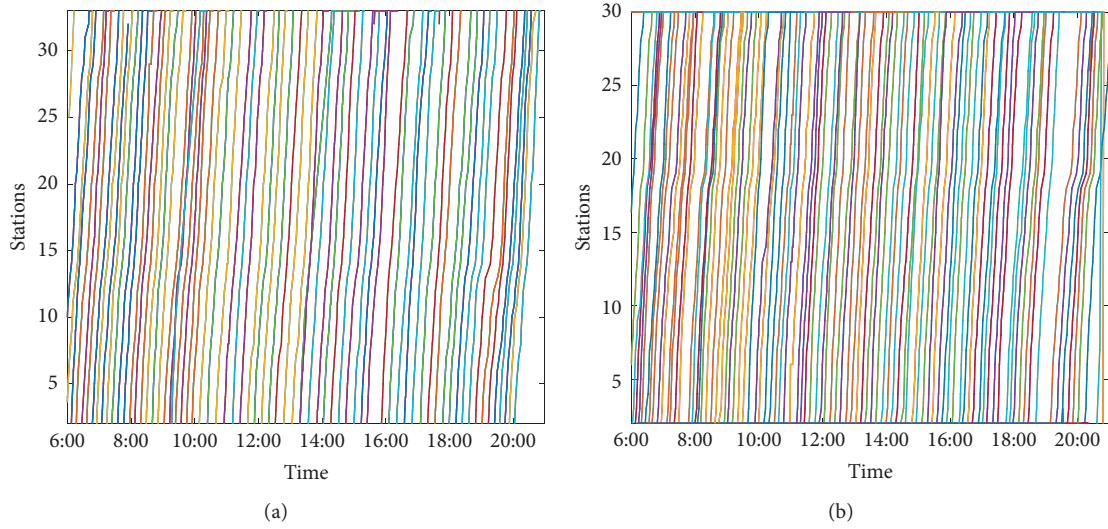


FIGURE 8: Running maps of (a) route 9 and (b) route 16.

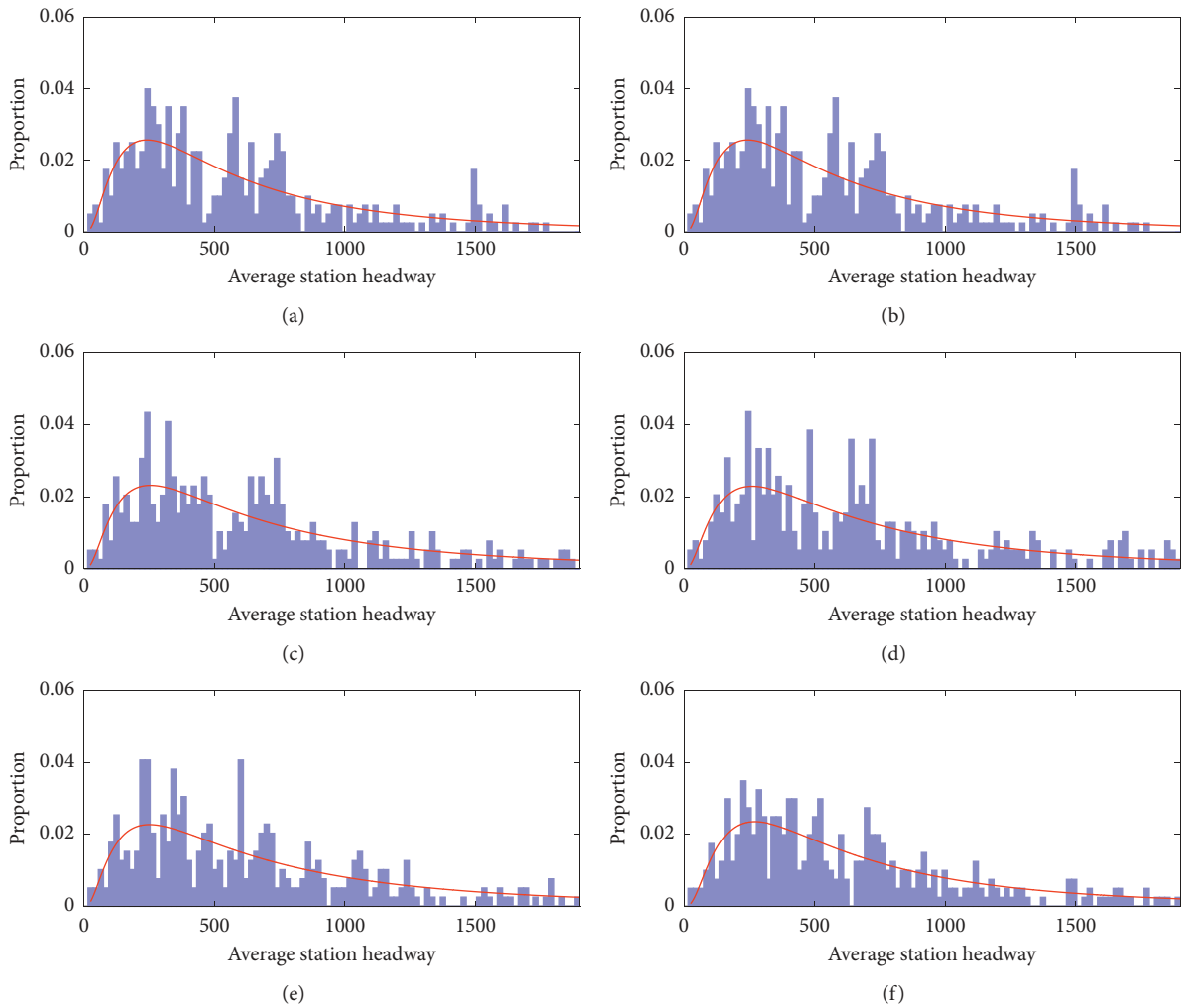


FIGURE 9: Average station headway distributions in different time periods: (a) 8:00–9:00, (b) 10:00–11:00, (c) 12:00–13:00, (d) 14:00–15:00, (e) 16:00–17:00, and (f) 18:00–19:00.

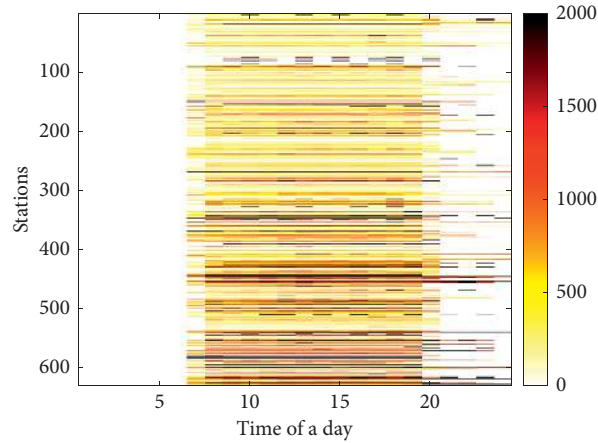


FIGURE 10: Heatmap of average headway of stations in a day.

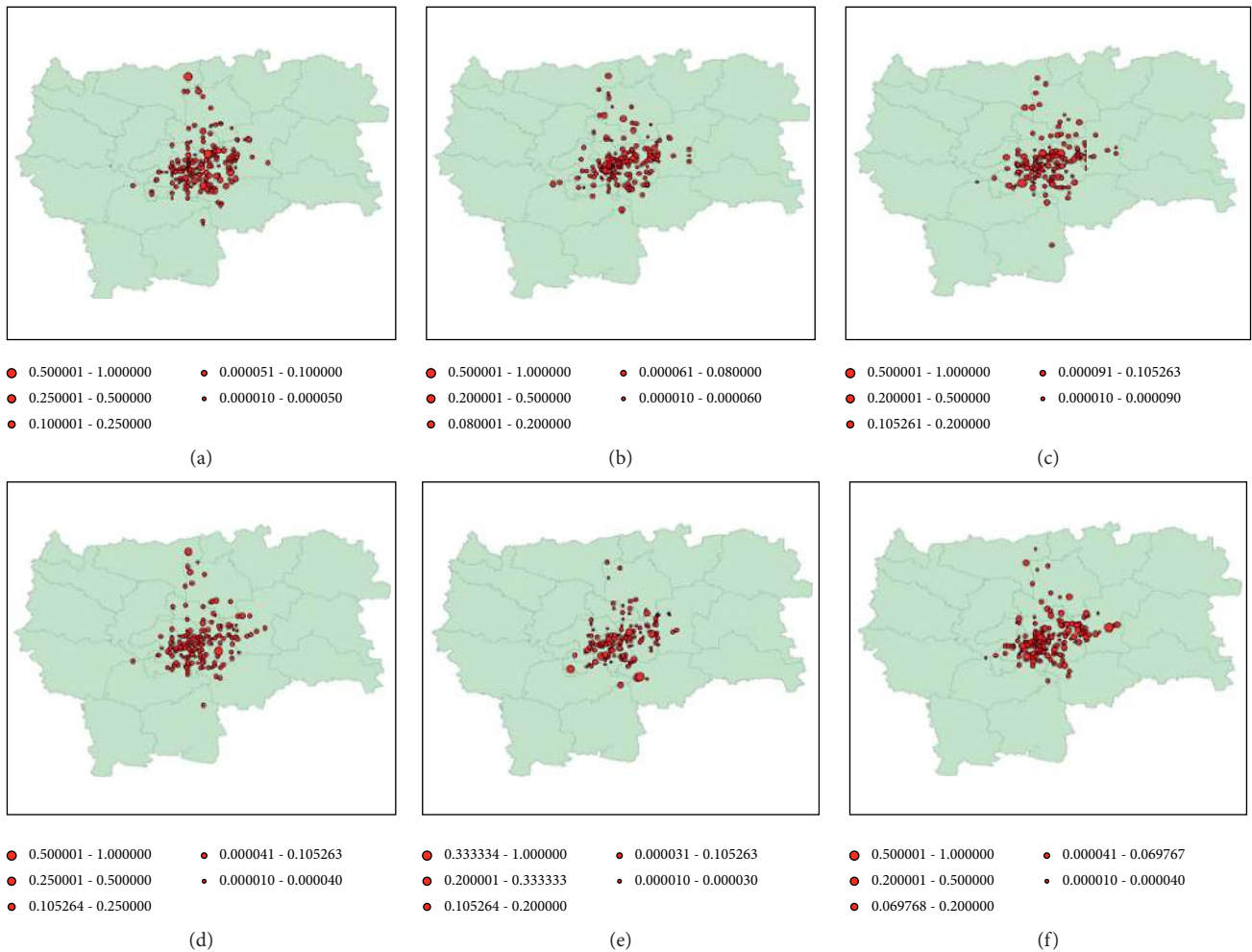


FIGURE 11: Average arrival rate of stations: (a) 8:00, (b) 10:00, (c) 12:00, (d) 14:00, (e) 16:00, and (f) 18:00.

only 5.2% of stations that the degree is larger than 8. The top five stations are Gaotiedong Zhan, Keyunbei Zhan, Beihai Park, Railway station, and Jianan Hospital. Those stations are transport hub which attract numerous passengers.

5.5. Relationship between Indicators. To explore the relationship among passengers, operational performance, and the connections of stations, Figure 14 shows the heatmap of the Pearson correlation coefficient among these indicators.

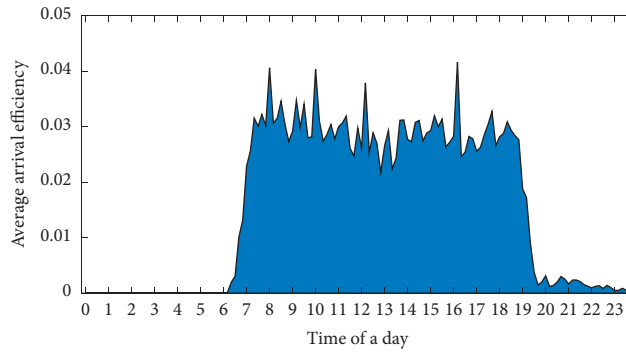


FIGURE 12: Mean value of average arrival rate of stations in a day.

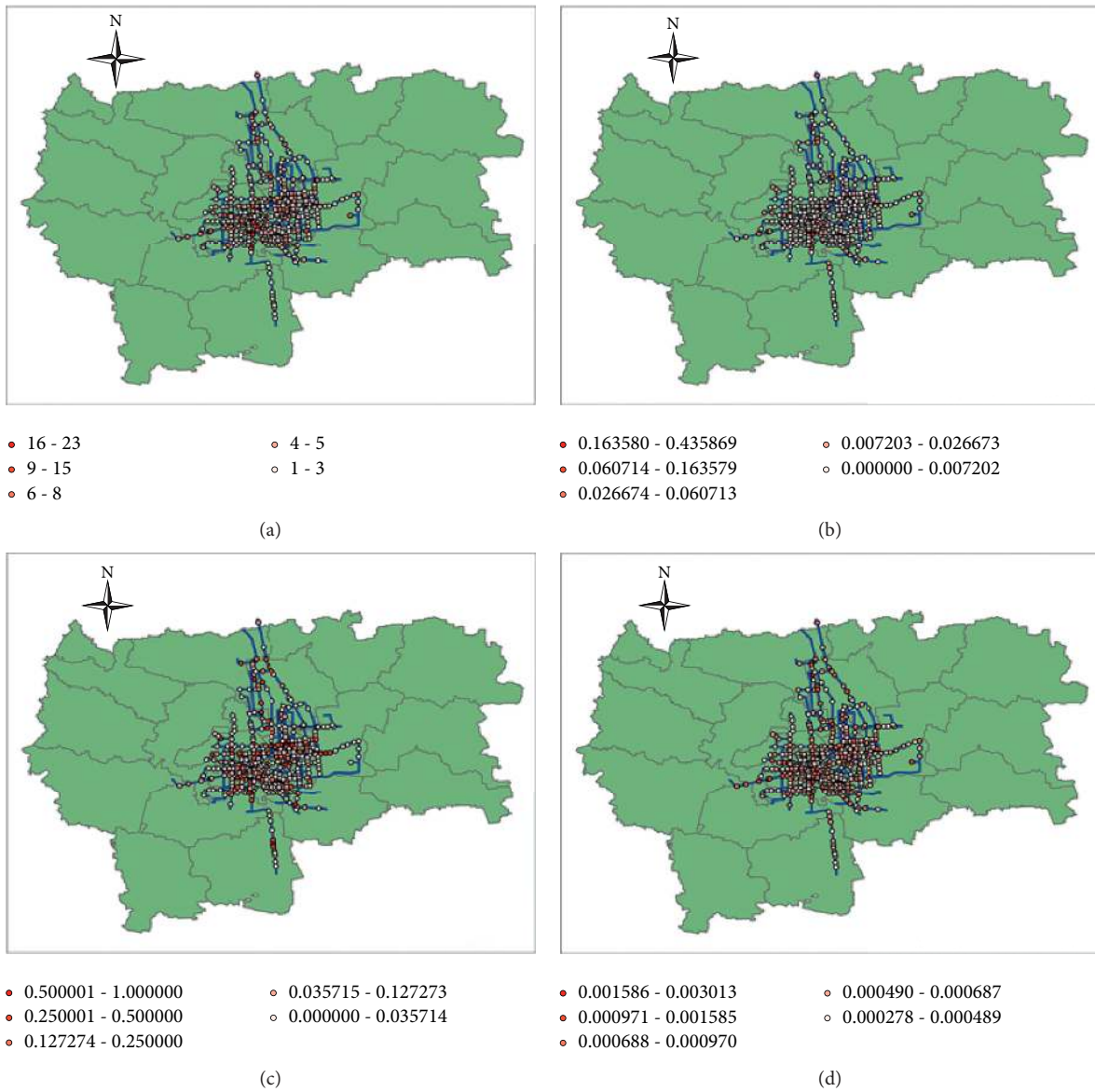


FIGURE 13: Distributions of bus topological structure properties: (a) degree, (b) betweenness, (c) clustering coefficient, and (d) PageRank.

As can be seen, the boarding number is strongly correlated with the bus number. Moreover, the boarding number is positively correlated with topology structure indicators,

which means that a better network connection could attract more passenger flows. It is remarkable that the average headway of station is negative with boarding passenger

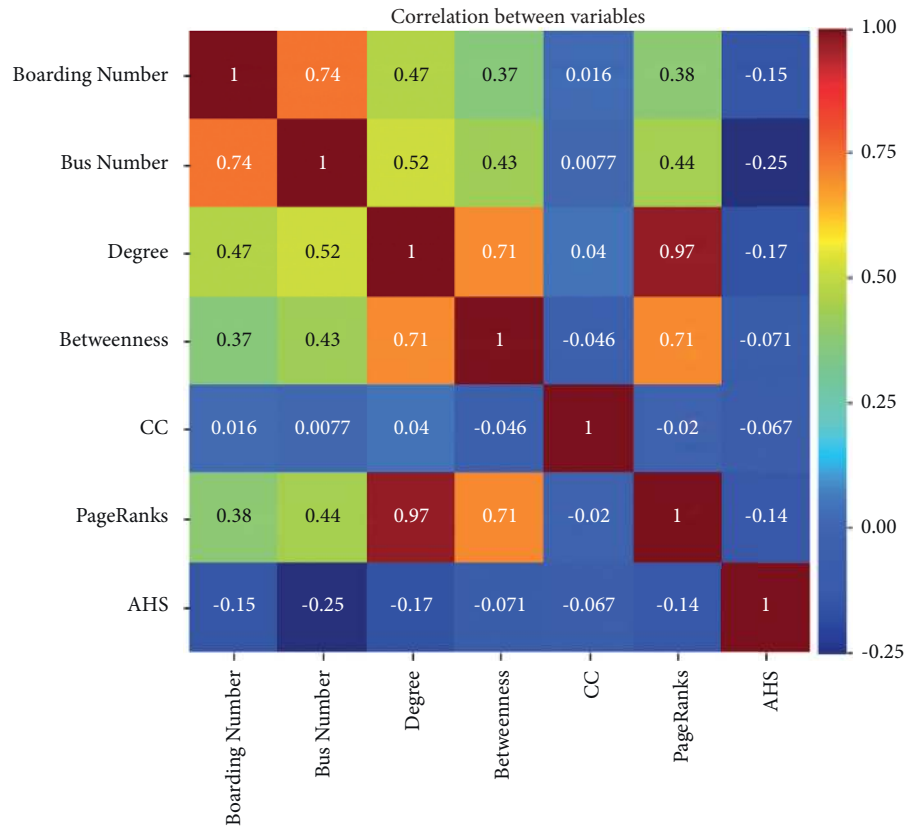


FIGURE 14: Correlations of main indicators.

number, which demonstrates that the passenger number is a critical factor to disturb the stability of operation. For the network structure, the PageRank index has a strong positive correlation with node degree. Clustering coefficient could reflect the local connections. The figure shows clustering coefficient nearly has no correlations with other indicators, which indicates that local connections hardly impact the operation. Operators should pay more attention to the entire network. We can see that the average headway of station is negatively correlated with other indicators, which implies that complex network connections and more passenger flows could weaken the efficiency of bus operation.

6. Conclusions

Bus system is a complicated spatiotemporal network, which plays a key role in alleviating traffic congestions. Understanding the operational status of the whole network is of significance for improving the service quality. The bus system often suffers from delay and bunching due to many factors such as traffic condition of road, weather, traffic light, and number of boarding and alighting passengers. Any disturbance of bus operation could cause the cascade reaction from stops to a route, even to the entire network. The efficiency and stability are two key indicators to measure the bus operation.

To evaluate the operational performance of the bus network, this paper proposes a spatiotemporal analysis of the bus network based on GPS trajectory data, IC card data, and

route data. We build average headway of station and arrival efficiency of station to evaluate the bus network operation. The results show that the bus number and boarding passenger number of bus network follow lognormal and exponential distributions. Moreover, the average headway of stations follows a lognormal distribution. There exist some stations, where the average headway of the station is very large during the operation time. Managers should arrange more buses for the routes that serve these stations. The distributions of the average arrival rate of stations are inhomogeneous. A small number of stations have large values, while most stations have small values. We test the relationships among passengers, operational performance, and the connections of stations. It is found that the average headway of station is negatively correlated with other indicators, which implies that complex network connections and large passenger flows could weaken the efficiency of bus operation. In addition, the boarding number is strongly correlated with bus number. The boarding number is positively correlated with topology structure indicators, which means that a better network connection could attract more passengers to use public transport.

This paper will promote the study of stability and efficiency of bus system from a single route to the entire network, which has important theoretical and practical meaning for bus systematic management and control. The limitations of this study are as follows. This paper did not involve the number of alighting passengers due to data limitation. The external factors, such as traffic condition and

weather, have not been considered in the paper. In the future study, more factors will be considered and prediction model will be studied.

Data Availability

The data are available by contacting the corresponding author.

Conflicts of Interest

The authors declare there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (42001396 and 41901396) and the Beijing Jingwei Information Technology Co., Ltd. Scientific Research Project (DZYF20-02).

References

- [1] C. R. Sekhar, T. Iryo, and Y. Asakura, "Analysis of travel-time variation over multiple sections of Hanshin expressway in Japan," *Current Science*, vol. 102, no. 11, pp. 1527–1533, 2012.
- [2] N. Julio, R. Giesen, and P. Lizana, "Real-time prediction of bus travel speeds using traffic shockwaves and machine learning algorithms," *Research in Transportation Economics*, vol. 59, pp. 250–257, 2016.
- [3] A. H. F. Chow, S. Li, and R. Zhong, "Multi-objective optimal control formulations for bus service reliability with traffic signals," *Transportation Research Part B: Methodological*, vol. 103, pp. 248–268, 2017.
- [4] S. M. H. Moosavi, A. Ismail, and A. Balali, "Evaluating bus dwell time at key stops using automatic data collection systems," *ITE Journal-Institute of Transportation Engineers*, vol. 87, no. 11, pp. 45–49, 2017.
- [5] Y. Park, J. Mount, L. Liu, N. Xiao, and H. J. Miller, "Assessing public transit performance using real-time data: spatiotemporal patterns of bus operation delays in Columbus, Ohio, USA," *International Journal of Geographical Information Science*, vol. 34, no. 2, pp. 367–392, 2020.
- [6] Y. J. Deng, X. Luo, X. B. Hu, Y. F. Ma, and R. Ma, "Modeling and prediction of bus operation states for bunching analysis," *Journal of Transportation Engineering Part A*, vol. 146, no. 9, Article ID 04020106, 2020.
- [7] A. Khadhir, B. A. Kumar, and L. D. Vanajakshi, "Analysis of global positioning system based bus travel time data and its use for advanced public transportation system applications," *Journal of Intelligent Transportation Systems*, vol. 25, no. 1, pp. 58–76, 2021.
- [8] Y. J. Deng, X. H. Liu, X. B. Hu, and M. Zhang, "Reduce bus bunching with a real-time speed control algorithm considering heterogeneous roadway conditions and intersection delays," *Journal of Transportation Engineering Part A*, vol. 146, no. 7, Article ID 04020048, 2020.
- [9] J. Argote-Cabanero, C. F. Daganzo, and J. W. Lynn, "Dynamic control of complex transit systems," *Transportation Research Part B: Methodological*, vol. 81, pp. 146–160, 2015.
- [10] J.-D. Schmöcker, W. Sun, A. Fonzone, and R. Liu, "Bus bunching along a corridor served by two lines," *Transportation Research Part B: Methodological*, vol. 93, pp. 300–317, 2016.
- [11] Q. Huang, B. Jia, R. Jiang, and S. Qiang, "Simulation-based optimization in a bidirectional \$A/B\$ skip-stop bus service," *IEEE Access*, vol. 5, pp. 15478–15489, 2017.
- [12] S. J. Berrebi, E. Hans, N. Chiabaut, J. A. Laval, L. Leclercq, and K. E. Watkins, "Comparing bus holding methods with and without real-time predictions," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 197–211, 2018.
- [13] N. B. Stoll, T. Glick, and M. A. Figliozzi, "Using high-resolution bus GPS data to visualize and identify congestion hot spots in urban arterials," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2539, no. 1, pp. 20–29, 2016.
- [14] J. Seo, S. H. Cho, D. K. Kim, and P. Y. J. Park, "Analysis of overlapping origin-destination pairs between bus stations to enhance the efficiency of bus operations," *IET Intelligent Transport Systems*, vol. 14, no. 6, pp. 545–553, 2020.
- [15] A. Webb, P. Kumar, and A. Khani, "Estimation of passenger waiting time using automatically collected transit data," *Public Transport*, vol. 12, no. 2, pp. 299–311, 2020.
- [16] L. Liu and H. J. Miller, "Measuring risk of missing transfers in public transit systems using high-resolution schedule and real-time bus location data," *Urban Studies*, 2020.
- [17] T. Zhang, Y. Li, H. Yang, C. Cui, J. Li, and Q. Qiao, "Identifying primary public transit corridors using multi-source big transit data," *International Journal of Geographical Information Science*, vol. 34, no. 6, pp. 1137–1161, 2020.
- [18] J. Li, P. Zheng, and W. Zhang, "Identifying the spatial distribution of public transportation trips by node and community characteristics," *Transportation Planning and Technology*, vol. 43, no. 3, pp. 325–340, 2020.
- [19] F. Chen, J. Zhang, Z. Wang, S. Shi, and H. Liu, "Passenger travel characteristics and bus operational states: a study based on IC card and GPS data in Yinchuan, China," *Transportation Planning and Technology*, vol. 42, no. 8, pp. 825–847, 2019.
- [20] H. Zhang, B. Y. Shi, S. G. Song, Q. M. Zhao, X. M. Yao, and W. Wang, "Statistical analysis of the stability of bus vehicles based on GPS trajectory data," *Modern Physics Letters B*, vol. 33, no. 3, Article ID 1950015, 2019.
- [21] H. Zhang, C. Zhuge, and X. Yu, "Identifying hub stations and important lines of bus networks: a case study in Xiamen, China," *Physica A: Statistical Mechanics and Its Applications*, vol. 502, pp. 394–402, 2018.
- [22] Z. Neal, "Is the urban world small? The evidence for small world structure in urban networks," *Networks and Spatial Economics*, vol. 18, no. 3, pp. 615–631, 2018.
- [23] T. N. Maeda, J. Mori, I. Hayashi, T. Sakimoto, and I. Sakata, "Comparative examination of network clustering methods for extracting community structures of a city from public transportation Smart Card Data," *IEEE Access*, vol. 7, pp. 53377–53391, 2019.
- [24] Y. Sui, F. Shao, X. Yu, R. Sun, and S. Li, "Public transport network model based on layer operations," *Physica A: Statistical Mechanics and Its Applications*, vol. 523, pp. 984–995, 2019.
- [25] H. Zhang, H. Cui, and B. Shi, "A data-driven analysis for operational vehicle performance of public transport network," *IEEE Access*, vol. 7, pp. 96404–96413, 2019.
- [26] C. A. Iliopoulou, C. P. Milioti, E. I. Vlahogianni, and K. L. Kepaptsoglou, "Identifying spatio-temporal patterns of bus bunching in urban networks," *Journal of Intelligent Transportation Systems*, vol. 24, no. 4, pp. 365–382, 2020.
- [27] J. J. Tang, Z. T. Li, F. Gao, and F. Zong, "Identifying critical metro stations in multiplex network based on D-S evidence theory," *Physica A*, vol. 574, Article ID 126018, 2021.

- [28] A. A. De Bona, M. D. Rosa, K. V. O. Fonseca, and R. Luders, "A reduced model for complex network analysis of public transportation systems," *Physica A*, vol. 567, Article ID 125715, 2021.
- [29] Y. Wang, Y. Y. Yuan, X. Y. Guan, M. Z. Xu, L. Wang, and H. Z. Wang, "Collaborative two-echelon multicenter vehicle routing optimization based on state-space-time network representation," *Journal of Cleaner Production*, vol. 258, Article ID 120590, 2020.
- [30] Y. Wang, S. G. Peng, X. S. Zhou, M. Mahmoudi, and L. Zhen, "Green logistics location-routing problem with eco-packages," *Transportation Research Part E*, vol. 143, Article ID 102118, 2020.
- [31] T. Li and L. Rong, "Impacts of service feature on vulnerability analysis of high-speed rail network," *Transport Policy*, vol. 110, pp. 238–253, 2021.
- [32] 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," *Physica A: Statistical Mechanics and Its Applications*, vol. 493, pp. 430–443, 2018.
- [33] H. Zhang, L. Zhang, F. Che, J. Jia, and B. Shi, "Revealing urban traffic demand by constructing dynamic networks with taxi trajectory data," *IEEE Access*, vol. 8, pp. 147673–147681, 2020.
- [34] L. Ni, X. Wang, and X. Chen, "A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 510–526, 2018.
- [35] M. G. Demissie, S. Phithakitnukoon, and L. Kattan, "Trip distribution modeling using mobile phone data: emphasis on intra-zonal trips," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 7, pp. 2605–2617, 2019.
- [36] H. Zhang, C. X. Zhuge, J. M. Jia, B. Y. Shi, and W. Wang, "Green travel mobility of dockless bike-sharing based on trip data in big cities: a spatial network analysis," *Journal of Cleaner Production*, vol. 313, Article ID 127930, 2021.
- [37] Z. Tian, J. Zhou, and M. Wang, "Dynamic evolution of demand fluctuation in bike-sharing systems for green travel," *Journal of Cleaner Production*, vol. 231, pp. 1364–1374, 2019.
- [38] Y. L. Lu, U. Benlic, and Q. H. Wu, "An effective memetic algorithm for the generalized bike-sharing rebalancing problem," *Engineering Applications of Artificial Intelligence*, vol. 95, Article ID 103890, 2020.
- [39] Z. Gurmu and W. Fan, "Artificial neural network travel time prediction model for buses using only GPS data," *Journal of Public Transportation*, vol. 17, no. 2, pp. 45–65, 2014.
- [40] W. Tu, R. Cao, Y. Yue, B. Zhou, Q. Li, and Q. Li, "Spatial variations in urban public ridership derived from GPS trajectories and smart card data," *Journal of Transport Geography*, vol. 69, pp. 45–57, 2018.
- [41] J. Tang, Y. Yang, W. Hao, F. Liu, and Y. Wang, "A data-driven timetable optimization of urban bus line based on multi-objective genetic algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 4, pp. 2417–2429, 2021.
- [42] R. Zhang, P. Xie, C. Wang, G. Y. Liu, and S. H. Wang, "Classifying transportation mode and speed from trajectory data via deep multi-scale learning," *Computer Networks*, vol. 162, Article ID 106861, 2019.
- [43] J. Li, X. Pei, X. Wang, D. Yao, Y. Zhang, and Y. Yue, "Transportation mode identification with GPS trajectory data and GIS information," *Tsinghua Science and Technology*, vol. 26, no. 4, pp. 403–416, 2021.
- [44] A. Chepuri, J. Ramakrishnan, S. Arkatkar, G. Joshi, and S. S. Pulugurtha, "Examining travel time reliability-based performance indicators for bus routes using GPS-based bus trajectory data in India," *Journal of Transportation Engineering Part A*, vol. 144, no. 5, Article ID 04018012, 2018.
- [45] W. Fan and R. B. Machemehl, "Do transit users just wait for buses or wait with strategies?" *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2111, no. 1, pp. 169–176, 2009.
- [46] S. S. Moghaddam, R. Noroozi, J. M. Casello, and B. Hellenga, "Predicting the mean and variance of transit segment and route travel times," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2217, no. 1, pp. 30–37, 2011.
- [47] A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, "Traveling time prediction in scheduled transportation with journey segments," *Information Systems*, vol. 64, pp. 266–280, 2017.
- [48] L. M. Kieu, A. Bhaskar, and E. Chung, "Public transport travel-time variability definitions and monitoring," *Journal of Transportation Engineering*, vol. 141, no. 1, Article ID 04014068, 2015.
- [49] R. Rajbhandari, S. I. Chien, and J. R. Daniel, "Estimation of bus dwell times with automatic passenger counter information," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1841, no. 1, pp. 120–127, 2003.
- [50] A. Tirachini, "Bus dwell time: the effect of different fare collection systems, bus floor level and age of passengers," *Transportmetrica: Transportation Science*, vol. 9, no. 1, pp. 28–49, 2013.
- [51] Y. Ji, L. Gao, D. Chen, X. Ma, and R. Zhang, "How does a static measure influence passengers' boarding behaviors and bus dwell time? Simulated evidence from Nanjing bus stations," *Transportation Research Part A: Policy and Practice*, vol. 110, pp. 13–25, 2018.
- [52] A. Chepuri, S. Joshi, S. Arkatkar, G. Joshi, and A. Bhaskar, "Development of new reliability measure for bus routes using trajectory data," *Transportation Letters*, vol. 12, no. 6, pp. 363–374, 2020.
- [53] J. Paudel, "Bus ridership and service reliability: the case of public transportation in Western Massachusetts," *Transport Policy*, vol. 100, pp. 98–107, 2021.
- [54] C. Zhang, Z. Juan, Q. Luo, and G. Xiao, "Performance evaluation of public transit systems using a combined evaluation method," *Transport Policy*, vol. 45, pp. 156–167, 2016.
- [55] X. P. Yang, S. W. Lu, W. F. Zhao, and Z. Y. Zhao, "Exploring the characteristics of an intra-urban bus service network: a case study of Shenzhen, China," *ISPRS International Journal of Geo-Information*, vol. 8, no. 11, Article ID 486, 2019.
- [56] G. L. Jia, R. G. Ma, and Z. H. Hu, "Urban transit network properties evaluation and optimization based on complex network theory," *Sustainability*, vol. 11, p. 2007, 2019.
- [57] Y. J. Wang, Y. Deng, F. Ren et al., "Analysis the spatial configuration of urban bus networks based on the geospatial network analysis method," *Cities*, vol. 96, Article ID 102406, 2020.
- [58] L.-S. Sun, S.-W. Wang, L.-Y. Yao, J. Rong, and J.-M. Ma, "Estimate of transit ridership based on spatial analysis and precise land use data," *Transportation Letters*, vol. 8, no. 3, pp. 140–147, 2016.
- [59] Y. Zuo, Z. Liu, and X. Fu, "Measuring accessibility of bus system based on multi-source traffic data," *Geo-Spatial Information Science*, vol. 23, no. 3, pp. 248–257, 2020.
- [60] H. Zhang, X. Li, L. Zhang, W. Wang, J. Jia, and B. Shi, "Discovering station patterns of urban transit network with multisource data: empirical evidence in Jinan, China," *KSCSE Journal of Civil Engineering*, vol. 25, no. 2, pp. 680–691, 2021.

- [61] S. Wei, W. Zheng, and L. Wang, "Understanding the configuration of bus networks in urban China from the perspective of network types and administrative division effect," *Transport Policy*, vol. 104, pp. 1–17, 2021.
- [62] S. Bree, D. Fuller, and E. Diab, "Access to transit? Validating local transit accessibility measures using transit ridership," *Transportation Research Part A: Policy and Practice*, vol. 141, pp. 430–442, 2020.
- [63] W. Y. Szeto, L. Yang, R. C. P. Wong, Y. C. Li, and S. C. Wong, "Spatio-temporal travel characteristics of the elderly in an ageing society," *Travel Behaviour and Society*, vol. 9, pp. 10–20, 2017.
- [64] K. Gkiotsalitis and E. C. Van Berkum, "An analytic solution for real-time bus holding subject to vehicle capacity limits," *Transportation Research Part C*, vol. 121, Article ID 102815, 2020.
- [65] X. M. Chen, X. M. Han, L. Yu, and C. H. Wei, "Does operation scheduling make a difference: tapping the potential of optimized design for skipping-stop strategy in reducing bus emissions," *Sustainability*, vol. 9, no. 10, p. 1737, 2017.
- [66] Y. Bie, X. Xiong, Y. Yan, and X. Qu, "Dynamic headway control for high-frequency bus line based on speed guidance and intersection signal adjustment," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 1, pp. 4–25, 2020.