IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 9, 2019, accepted July 17, 2019, date of publication July 22, 2019, date of current version August 5, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2930279*

# A Data-Driven Analysis for Operational Vehicle Performance of Public Transport Network

**HUI ZHANG[1], HOUDUN CUI[2], AND BAIYING SHI[1]**
[1]School of Transportation Engineering, Shandong Jianzhu University, Jinan 250101, China
[2]Shandong Hi-speed Group Company Ltd., Jinan 250098, China

Corresponding author: Hui Zhang (zhanghui@sdjzu.edu.cn)

This work was supported in part by the Doctoral funds of Shandong Jianzhu University under Grant XNBS1614, and in part by the Natural Science Foundation of Shandong Province under Grant ZR2018PG002.

**ABSTRACT** The operational stability of public transport is significant for both passengers and operators. Affected by many stochastic factors, such as traffic congestion, traffic signals and passenger demand at stops, the headway always become uneven, which greatly reduces the service quality. This paper used the big global positioning systems (GPS) trajectory data to analyze the headway stability of bus system from the perspective of network. A statistical method is proposed to analyze the operational vehicle performance of bus network. The GPS trajectory data of Jinan is used to test the model. The results show that the average dwell time, actual headway, and headway stability index of stations follow lognormal distributions with obvious right tails. Moreover, the seriously unstable situations do not appear in the peak hours, but in the time periods before peak hours. In addition, the stations with most unstable headway are located in the suburbs and the fringe area of downtown. The outcomes suggest that operators should pay more attention to the suburbs and the fringe area of downtown, and the time periods before peak hours to efficiently improve the service quality.

**INDEX TERMS** Public transport network, stability of headway, GPS trajectory data, data-driven analysis.

## I. INTRODUCTION

Operational stability of bus arrival time at stops plays a crucial role in enhancing the attractiveness and efficiency of public transport. Improving the stability of bus system is conducive to both operators and passengers. However, travel time fluctuates due to many stochastic factors such as road congestion, uneven passenger loads and weather conditions, thereby resulting in irregular arrival time and bus bunching [1], [2]. For operators, understanding the actual performance of bus systems is an import part in rescheduling timetable, avoiding bus bunching etc.

Data-driven methods become a powerful tool to understand the mechanism of public transport system operations and many other systems [3], [4]. In transportation systems, agent behavior plays a significant role in transportation dynamics [5]. Data-driven could exhibit the panorama of system operations that affected by agents and stochastic factors. Recently, exploring the operational characteristic of routes draw many attentions using automatic collected data such as global positioning systems (GPS) point data and smart card data that contain many kinds of useful information like vehicle position, number of passengers and so forth [6], [7]. The GPS point data have been widely used to analyze the operational features of routes such as travel time distribution, variability and reliability [8]–[11]. Besides that, GPS data provide an indispensable source to predict travel time and control vehicles to avoid bus bunching [12]–[14].

Service reliability is a critical indicator of transit performance, which is relevant to traffic condition, demand of stations and other stochastic factors. The irregular headway may result in long waiting time for passengers and low efficiency in transit systems. The extreme case of irregular headway is called bus bunching that the adjacent vehicles are too close on the road. Essentially, bus bunching is caused by minor disturbance that can easily propagate. Fonzone *et al.* [15] pointed out that non-uniform arrival patterns could significantly influence the bus bunching process. Ma *et al.* [16] used AVL data and smart card data to analyze the bus travel time reliability and found congestion, traffic signal and passenger demand at

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

stops were three main important factors. Most studies have focused on passengers' or operators' perspectives on service reliability.

The aforementioned studies mainly highlighted the reliability of a single route, few studies concerned reliability of the entire network. Network performance of public transport describes the operational states of entire network containing all routes. Better understanding network performance is a prerequisite for analyzing the stability, weak points and the reasons for them. Public transport networks are fragile to many uncertainties [17]. This paper proposes a framework to study the operational performance of entire network by analyzing a large-scale empirical arrival-departure time data in Jinan, China.

The rest of paper is organized as follows. Section 2 gives the literature review. Section 3 introduces the performance metrics. Section 4 gives network representation. Data description is introduced in section 5. Section 6 presents the results. Conclusions are given in section 7.

## II. LITERATURE REVIEW

### A. BUS ROUTE PERFORMANCE

It is important for operators to grasp the operational condition for making suitable arrangement of bus operations. Bus performance evaluation has drawn much attention since the automatic data is available. Kathuria *et al.* [18] found that travel time of bus transit varied in different time periods within a day and it was larger in peak hours. Statistical analysis shows that travel time is asymmetric and skewed to right compared with normal distribution [8], [19]. Some researches pointed out Gamma, loglogistic and Weibull distribution could fit the travel time [6], [20].

Travel time consists of running time between stations and dwell time at stations. Studies show that running time between stations is related with traffic condition and traffic signal [18], [21], [22]. Dwell time contains the time spent on opening the door, passengers' boarding and alighting, closing the door, which consume up to 26% of total travel time [23]. In general, the number of passengers who board and alight is the key factor to determine the dwell time [24].

Reliability of bus operation is a crucial indicator that both operators and passengers concern. Sterman and Schofer [25] gave the concept of reliability and qualitative analysis of influential factors. Early studies depend on investigating data and monitoring equipment at the roadside, which has a low accuracy [26], [27]. With the development of advanced automatic data collected devices, more accurate data are available for evaluating the transit performance. Mazloumi *et al.* [28] studied the transit travel time variability by investigating day-to-day variability based on GPS data, and they found travel time variability was higher in the AM peak and lower in the off-peak. In recent decade, a large number of studies focus on the reliability of travel time of route according to the GPS data [29], [30]–[32]. Cats and Gerasimos [33] utilized automatic vehicle location data to provide real-time bus arrival information. Recently, researches start to highlight travel time prediction to overcome the stochastic influence for providing a better service [14], [34], [35].

### B. NETWORK EVALUATION

Bus networks are complex networks with complex structure and functions. It is indispensable to understand the performance of bus network, especially in the place where the service does not match with the demand. Operation synchronization of bus timetable is a key issue to keep transfer efficiency for passengers. Ceder *et al.* [36] developed a model including hold and skip strategies to keep transfer synchronization of bus network. Dou *et al.* [37] proposed a time control point strategy coupled with transfer synchronization to resolve the schedule design problem to improve schedule adherence. Numerous studies focus on the design process of timetable and vehicle scheduling with consideration of transfer synchronization [38]–[41]. Recently, Wang *et al.* [42] built a data-driven model to optimize the bus scheduling, which could largely reduce the waiting time.

Bus network evaluation is a prerequisite to provide better service. Presently, the network evaluation mainly contains network structure and service from survey. Zhang *et al.* [43] constructed a framework to evaluate public transit service with survey data considering convenience, comfort, security, facility etc. For network structure, there is a large number of studies using graph theory and complex network theory to analyze the topology characteristics [44]–[46]. Zhang *et al.* [47] proposed a node failure process to identify hub nodes in bus networks. To better exhibits the network functions, some researches focus on networks with weights such as boarding passenger volume and travel time [48], [49]. Synchronization and robustness of bus network from the perspectives of network topology have been widely studied during the past years [50]–[52]. Recently, Jia et al. used complex network theory to evaluate the urban transit network and proposed a sustainable transit network optimization method considering station and road conditions [53]. Xu and Yang proposed a geographically weighted regression model to determine the correlation between transit accessibility and urban land use characteristics [54]. Park et al. used real-time vehicle location data to study the spatiotemporal patterns of bus operation delays in Columbus and found that the prevailing delays concentrated on certain stops in downtown and core suburban locations [55].

To clearly exhibit the network evaluation of transit network, we summarize the related works in Table 1. The transit network evaluation mainly focuses on topological structure utilizing complex network theory. Transit structure evaluation used route information to construct network and can only reflect the structure characteristics from the perspective of topology structure. In the last two decades, transit networks involving bus and metro have been studied by many works containing evaluation of network, finding important nodes, network robustness. The route information could not reflect the characteristics of economic, society

**TABLE 1. Related studies of transit network evaluation.**

| Authors | Data | Method | Evaluation |
|---|---|---|---|
| Yang et al. (2014) [43] | Route and station | Complex network | Topology structure |
| Chatterjee et al. (2016) [44] | Route and station | Complex network | Topology structure |
| Zhang et al. (2018) [49] | Route and traffic demand | Complex network | Key nodes |
| Xu and Yang (2019) [54] | Multisource data | Geographically regression | Accessibility |
| Park et al. (2019) [55] | Vehicle location data | Spatio-temporal analysis | Reliability |
| Sun et al. (2016) [56] | POI and AFC | Regression analysis | Transit ridership |
| Qi et al. (2019) [57] | POI and smart card | Clustering | Travelers' mobility patterns |
| Quintero-Cano et al. (2014) [59] | Route and station | Graph theory | Network connectivity |
| Ding et al. (2018) [60] | Multisource data | Gravity-based method | Accessibility |

and operations. To overcome the shortage, many studies used multifarious data to evaluate transit networks such as points of interest data (POI) and GPS trajectory data [56]–[58]. POI data could reflect the land use and other information. The GPS trajectory data is always used to measure the operational performance of transit routes. For bus systems, delay is a prevailing, which hinders the development of bus. Furthermore, uneven headway become one of the chief reasons for passengers to dislike to use buses. In the last years, many works emphasized the delay of transit vehicles from the view of routes. However, bus operational vehicle performance from perspective of bus network has seldom studied because of complexity. To fill this gap, this paper intends to study the vehicle performance of bus network based on real GPS trajectory data to provide suggestions for improving the service of bus systems.

## III. PERFORMANCE METRICS

### A. PROCESS OF BUS OPERATION

Each bus run is regarded as a series of events containing arrivals and departures, which has specific arrival time and departure time at each stop along the route. The total travel time can be expressed as

$$T = \sum_{k=1}^{n} DT_k + \sum_{k=1}^{n-1} RT_{k,k+1} \tag{1}$$

where $T$ is the total travel time, $DT_k$ is the dwell time at station $k$, $RT_{k,k+1}$ is the running time between station $k$ and station $k+1$.

### B. HEADWAYS

Headway is defined as time difference between two consecutive vehicles that belongs to a same line. It can be calculated with the formula

$$\Delta H_k^{i,j} = \begin{cases} 0 & j = 1 \\ \left| h_k^{i,j+1} - h_k^{i,j} \right|, & otherwise \end{cases} \tag{2}$$

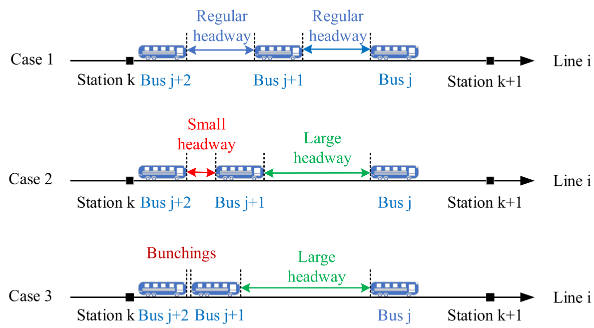


**FIGURE 1. Illustrations of bus operation stability of a line.**

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 0; $h_k^{i,j}$ is the arrival time of vehicle $j$ of line $i$ at station $k$.

The bus runs according to a scheduled headway at the original stations. However, the actual headway fluctuates because of the stochastic factors. For instance, when a bus vehicle suffers from large passengers at stations, it stops a longer time. Then, the following vehicle is close to the vehicle. Under some conditions, the consecutive vehicles arrive at a station simultaneously, which is called bus bunching. Fig. 1 shows the illustrations of bus operation stability of a route. In case 1, the buses have equal headways; in case 2, the bus j + 1 delays and results in small and large headway between it with backward and forward vehicles respectively; in case 3, bus j + 1 and j + 2 bunching, which bring very large headway between bus j and j + 1. The irregular headway will bring longer waiting time for the passengers and the waste of capacity for the bunching vehicles.

### C. HEADWAY STABILITY OF NODES

Firstly, the operation stability of a line at a station is defined as

$$\delta_{k,t}^{i,j} = \frac{\left| \Delta H_{k,t}^{i,j} - H_t^i \right|}{H_t^i} \tag{3}$$

where $i = 1, \cdots, m$ represents the line number, $j = 1, \cdots, n_i$ is the vehicle number of line $i$, $k = 1, \cdots, l_i$ is the station number of line $i$; $t$ is the time period; $H_t^i$ is the plan headway of line $i$ in time period $t$.

The headway varies in a different time, so we calculate the stability of a node for line $i$ in time period $t$ as

$$\delta_{k,t}^{i} = \sum_{j=1}^{n_t} \delta_{k,t}^{i,j} / n_t \tag{4}$$

Then, we define the operation stability of a station as

$$\delta_{k,t}^{-} = \sum_{i=1}^{m'} \delta_{k,t}^{i} / m' \tag{5}$$

where $\delta_{k,t}^{-}$ is the operation stability of station $k$ in time period $t$, $m'$ is the number of lines stop at station $k$.

**TABLE 2. Illustration of GPS trajectory data of route 116.**

| LineNo. | VehicleNo | Date | Time | Latitude | Longitude | Station | Up or down |
|---|---|---|---|---|---|---|---|
| 116 | 1809246 | 2018/12/05 | 5:30:28 | 36.6747 | 117.0501 | 1 | 0 |
| 116 | 1809246 | 2018/12/05 | 5:30:33 | 36.6745 | 117.0500 | 0 | 0 |
| 116 | 1809246 | 2018/12/05 | 5:30:38 | 36.6738 | 117.0495 | 0 | 0 |
| … | … | … | … | … | … | … | 0 |
| 116 | 1809246 | 2018/12/05 | 5:31:59 | 36.6720 | 117.0542 | 2 | 0 |
| … | … | … | … | … | … | … | 0 |

**TABLE 3. Departure-arrival time records achieved from GPS trajectory data.**

| LineNo. | VehicleNo. | Date | Station | Arrival | Departure | SRunTime | StopTime |
|---|---|---|---|---|---|---|---|
| 116 | 1809246 | 2018/12/05 | 1 | 5:30:01 | 5:30:28 | 0:01:08 | 0:00:27 |
| 116 | 1809246 | 2018/12/05 | 2 | 5:31:08 | 5:31:42 | 0:00:40 | 0:00:34 |
| 116 | 1809246 | 2018/12/05 | 3 | 5:32:54 | 5:33:22 | 0:01:12 | 0:00:28 |
| … | … | … | … | … | … | … | … |

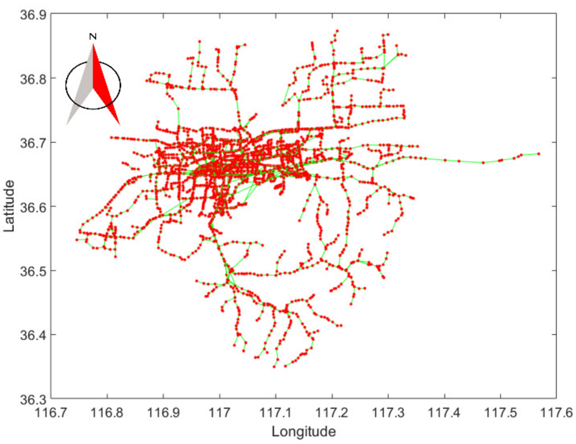


**FIGURE 2.** Layout of Jinan bus network: points represent stations.

## IV. NETWORK REPRESENTATION

In this paper, the bus network is represented as a weighted graph $G = (V, E, W)$, $V$ is the set of nodes and $E$ is the set of edges with weight $W$. $G$ is the $N \times N$ adjacency matrix $\{e_{ij}\}$. If there exists an edges between node $i$ and node $j$, $e_{ij} = 1$; otherwise, $e_{ij} = 0$. $w_{ij}$ is the weight of $e_{ij}$. The weight can represent length of edge, running time of edge, etc.

## V. DATA DESCRIPTIONS

The data was collected from December 1 to December 31, 2018 in Jinan, China, where each bus vehicle was installed with automatic vehicle location equipment. The devices send information such as location, timestamp, velocity and so on every five seconds. There are 280 lines that comprise ten BTR lines and 2167 stations in the main urban area. Fig. 2 shows the layout of Jinan bus network, and red points represent stations.

The data set used in this paper is called departure-arrival data that come from the GPS trajectory point data collected by the devices on the vehicles. The raw information contains line number, the ID of GPS device, time stamp, latitude, longitude, instantaneous velocity, station or not, road name, up or down streams, etc. Table 2 illustrates the GPS trajectory data of route 116. Each raw record is a GPS point data containing the aforementioned information, where the field "Station" is 0 means the point is not in stations and other number means in stations. The field "Up or down" is 0 means the route is upstream while 1 means downstream. The device installed on the buses send a record about 5 seconds.

We extract the departure time and arrival time of all vehicles of routes from the raw data. Table 3 shows the forms of departure-arrival time data. We give a unique ranked number to a station from 1…N, where N is the total number of stations. From the arrival time and departure, section running time (SRuTime) and stopping time at stations (StopTime) can be achieved. The section running time is the difference between the arrival time of a station and the departure time of the former station. The stopping time is the difference between departure time and arrival time of a station. The data will be preprocessed to remove the outliers.

## VI. RESULTS

The evaluation method described in the former section was implemented to study the operational performance of bus system in Jinan, China. There are 1,023,710 raw records used in the studies. We removed 10,537 outliers from the dataset. In this paper, we divided the data into two categories according to weekday and weekend. Results are given separately for the two categories. Each result is corresponding to an average of days' data in the group. We first analyze the dwell time of stations on weekday involving distribution in different time period of a day. Then actual headway performance and operation stability of stations are studied. Finally, we analyze the situation on weekend.

### *A. DWELL TIME OF STATIONS*

Dwell time is a key factor that influence the stability of bus operations, which is subjected to the number of passengers that alight and board, pay modes and the occupancy situation of stations. The data set includes accurate dwell time of each stations. Fig. 3 illustrates the average dwell time (seconds) distributions of stations in different time periods of a day: (a) 8:00–9:00, (b) 12:00–13:00, (c) 18:00–19:00, (d) 20:00–21:00. The time periods can reflect the situations of morning peak time, noon time, evening peak time and

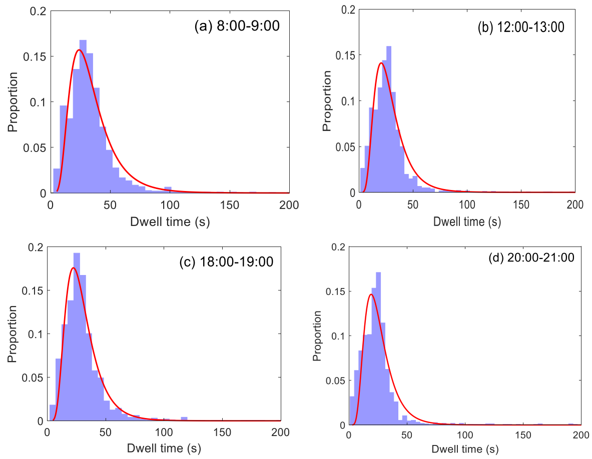


**FIGURE 3.** Average dwell time distribution of stations in different time period of weekday.

**TABLE 4.** Basic statistics of average dwell time of stations: (a) 8:00-9:00; (b)12:00-13:00; (c) 18:00-19:00; (d) 20:00-21:00.

| | Min | Median | Mean | Max | Std. | Skewness |
|---|---|---|---|---|---|---|
| (a) | 2.5 | 29.56 | 34.13 | 189.60 | 30.98 | 7.72 |
| (b) | 1.5 | 25.22 | 28.05 | 178.32 | 25.24 | 7.73 |
| (c) | 2.22 | 27 | 31.3 | 185.21 | 27.38 | 7.54 |
| (d) | 1.6 | 23 | 25.43 | 173.32 | 25.10 | 7.47 |

off-peak time. The value is the average dwell time of each vehicle in the studied time period. As can be seen, most average dwell time of stations concentrate in the range [0, 50]. For instance, the dwell time that is smaller than 50 seconds accounts for 88.2% in 8:00 time section in Fig. 3(a).

The distribution of average dwell time of stations seems normal distribution with right tails. Jarque-Bera test refuses that it follows normal distribution. Table 4 shows the basic statistics of average dwell time of stations. We can see that the mean value is large than median value, which indicates that the distribution is right skewed distribution. The skewness value can reflect the symmetry situation of distribution. The value of zero means it is symmetrical; the positive value means right skewed distribution; the negative value means left skewed distribution. In the section, the skewness values are 7.72, 7.73, 7.54 and 7.47, which validate the distribution is right skewed distribution. In Fig. 3, the red curves are the fit curves of a lognormal distribution, which can fit the distributions of average dwell time of stations.

Fig. 4 exhibits the top 20 stations with large and small values of average dwell time. As can be seen, stations with large values are located in the suburbs, while stations with small values are located in downtown in day time periods: 8:00–9:00, 12:00–13:00 and 18:00–19:00. The reason is that there are few routes in the suburbs and the headway is also large, hence the number of passengers waiting for boarding is large. Fig. 4(d) indicates that the stations with small values are located in the suburbs, while stations with large values are located in both downtown and suburbs. That's because there are some stations in the suburbs with many residential areas,

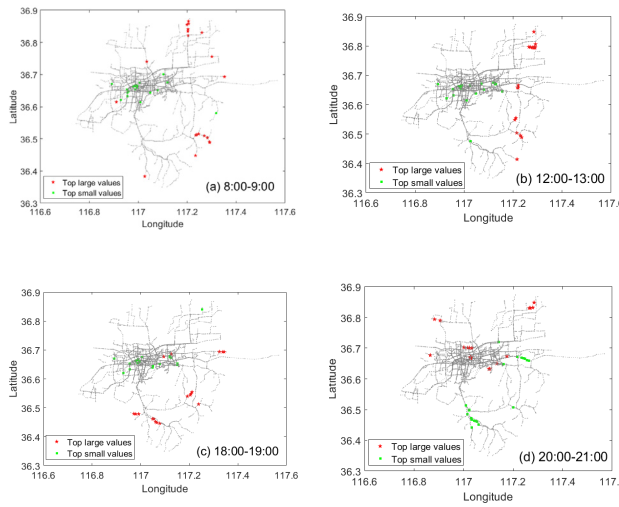


**FIGURE 4.** Stations with top 20 large and small values of average dwell time.

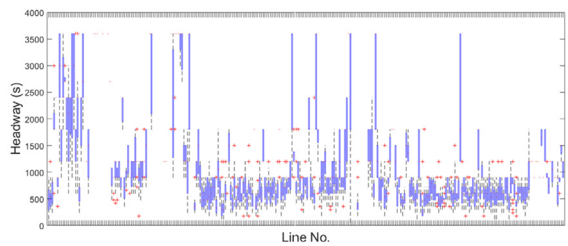


**FIGURE 5.** Plan headway of lines.

where lots of passengers intend to go back home. The results indicate that operators should try their best to improve the service quality for the suburbs.

### *B. ACTUAL HEADWAY PERFORMANCE OF STATIONS*

In transit system, headway reflects the service quality. Passengers prefer small headway to reduce their waiting time, while operators attempt to adopt suitable headway to increase revenue. In the past years, researchers focus on how to make suitable headway to satisfy traffic flow demand and the synchronization of headways. Understanding the actual headway of bus networks will provide a great help for these studies.

In this part, we concern the actual headway of each vehicles at all stations. To better understand the headway of each vehicle, we plot the plan headway first in Fig. 5. We can see that approximate 71.7% headways that are smaller than 1000 seconds. In the figure, there are some lines with large headway and range, which are customized shuttle bus. Take line 313 for instance, the minimum headway is 1200 seconds, while the maximum headway is 3600 seconds. It is noticed that the headway of a line fluctuates, which is in line with the fluctuation of passenger flow demand. In peak hours, small headway is needed, while headway is larger in off-peak hours.

In bus networks, service at stations is an important factor to attract passengers. Understanding the service level of stations provide a deep insight to improve the bus service. To this end, we represent the average headway of "spring square"

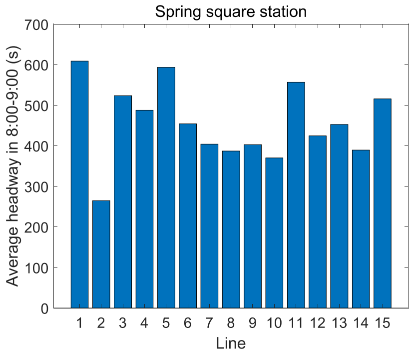


**FIGURE 6.** Average headway of "spring square" station in 8:00-9:00.

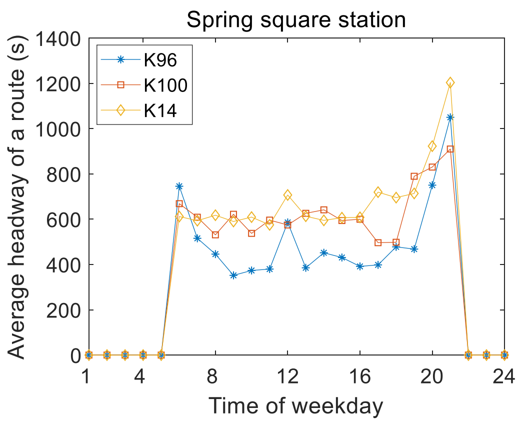


**FIGURE 7.** Average headway during a weekday of three bus lines at spring square station.

station in an hour, which is located in the center of downtown and 15 lines pass by the station. Fig. 6 illustrates the average headway of "spring square" station between 8:00 and 9:00. The range of average headway in 8:00–9:00 is between 265 seconds and 609 seconds.

To better reflect the service of stations, we represent average headway during a weekday of three bus lines at spring square station in Fig. 7. As can be seen, the average headways fluctuate in different hours. Take line "K96" for example, the average headway is 745, 352 and 750 seconds in 6:00–7:00, 9:00–10:00 and 20:00–21:00. Another interesting finding is that the average headway is higher in the beginning and end time periods. That's because passengers are less in the two time periods and the planning headway is larger than other time periods. For a station, the service level is composed of all lines' performance.

Fig. 8 plots the average actual headway distribution of stations in different time period of a day. The value is the average of actual headway of each vehicle that stops at a station in the bus network. We can see that the distribution can also be fitted by lognormal distribution. Most stations' actual headway is less than 1000 seconds, which is in line with the plan headway. Furthermore, the mean values of average actual headway are 784.3, 984.9, 840.0 and 780.4 seconds during the time periods 8:00–9:00, 12:00–13:00,

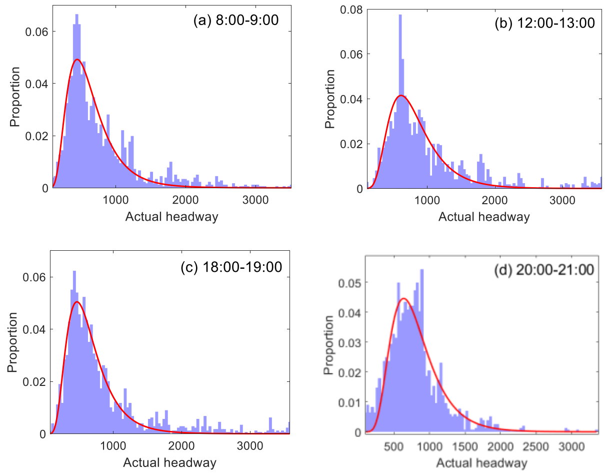


**FIGURE 8.** Average actual headway distribution of stations in four different time periods of weekday.

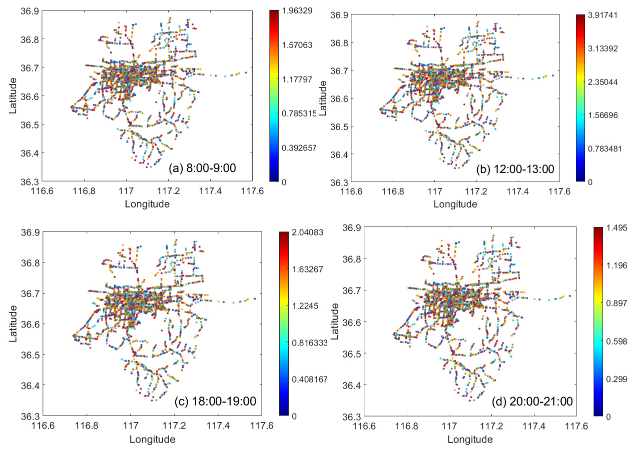


**FIGURE 9.** Average values of headway stability of stations in Jinan bus network.

18:00–19:00 and 20:00–21:00. The results indicate waiting at noon time may spend more time.

### C. STABILITY OF HEADWAY OF STATIONS

Fig.9 presents the average values of average headway stability stations in four different time periods of weekday. The values between 20:00 and 21:00 are larger than other time periods. The outputs illustrate the headway stability is stronger in the off-peak hours when the traffic condition is better than other periods.

We plot the distributions of average headway stability of stations index in Fig. 10 in four different time periods of a day. The distributions are also skewed distribution with right tails and most values are between 0 and 0.5. Specifically, the proportions that the value is smaller than 0.5 are 89.08%, 96.41%, 86.08% and 95.44%. The outputs indicate that the performance in noon time and off-peak hours is superior to peak hours. We notice that the performance in the noon time is good, where 88.67% of the values of stations are smaller than 0.3. However, there are some stations with very high values in this time periods. Approximate 0.8% of stations are

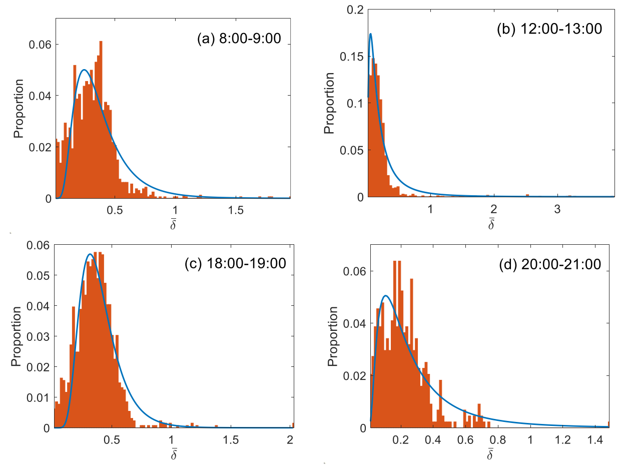


**FIGURE 10. Average headway stability index distributions of stations in four different time periods of weekday.**

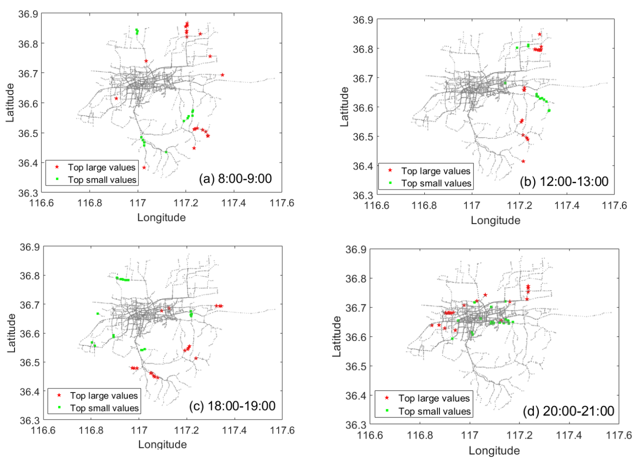


**FIGURE 11. Stations with top 20 large and small values of average headway stability index.**

with values that are larger than 2. The reason is that there are some special lines with irregular headway in the noon time such as customized shuttle bus. These buses always meet big flow, which disturb the stability of headway.

In Fig. 11, we exhibit the stations with top 20 large and small values of average headway stability index. As can be seen that stations with high and small values concentrate on the suburbs during 8:00–9:00, 12:00–13:00 and 18:00–19:00. The results show that the performance of stations in suburbs should be improved in day time. During 20:00–21:00, the stations with large values are located in fringe area of downtown. The reason is that many passengers emerge during this time period, which delay the bus vehicles. Because most lines stop operations after 9:00 in Jinan during winter time, most passengers go to their home in the suburbs before 9:00. Operators should prolong the operation time in winter time to enhance the service quality.

To better understand the average headway stability of stations, Fig.12 plots the median value, mean value and standard deviation of average headway stability index in different

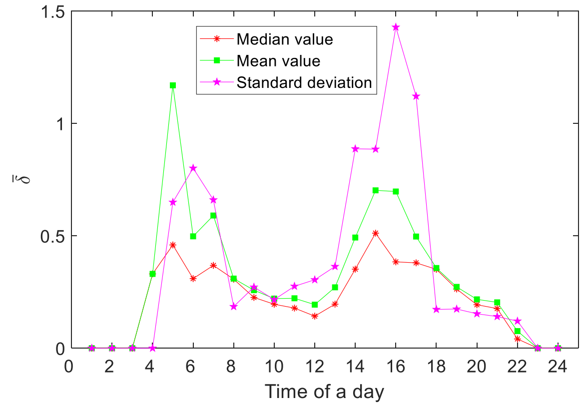


**FIGURE 12. Median, mean and standard deviation values of average headway stability index in different hours of weekday.**

hours of a day. The distributions exhibit two peaks for the three measures. It is noted that the peaks do not appear in the peak hours, but in the time periods before peak hours. The reason is as follows. In peak hours, the plan headway is small, which means there are many vehicles for passengers. In off-peak hours, the traffic demand is small, which cannot bring big fluctuation of vehicle headway. However, the traffic demand fluctuates strongly in the time periods before peak hours.

The headway stability index can only reflect the deviation of actual headway against the plan headway, it cannot reflect the positive or negative deviations. In bus systems, the actual headway does not strictly adhere to the plan headway, because the continuous bus cannot arrive on time. On the other side, the actual headway changes according to the discussed stochastic factors. Understanding the positive and negative deviation is help to even headway by operators.

Similarly, we introduce the concept of advanced headway stability index as follows:

$$\delta_{k,t}^{i,j,*} = \frac{\Delta H_{k,t}^{i,j} - H_t^i}{H_t^i} \tag{6}$$

so, we calculate the advanced headway stability of a node for line $i$ in time period $t$ as

$$\delta_{k,t}^{i,*} = \sum_{j=1}^{n_t} \delta_{k,t}^{i,j,*} / n_t \tag{7}$$

Then, we define the advanced headway stability of a station as

$$\delta_{k,t}^{\bar{*}} = \sum_{i=1}^{m'} \delta_{k,t}^{i,*} / m' \tag{8}$$

Fig.13 illustrates values of advanced headway stability of stations in four time periods of weekday. We observe that the positive values account for small proportions. Concretely, the positive values account for 33.87%, 51.1%, 43.8% and 18% during the time 8:00-9:00, 12:00–13:00, 18:00–19:00 and 20:00–21:00. The results imply that most stations suffer from small headway in reality, which could reduce the

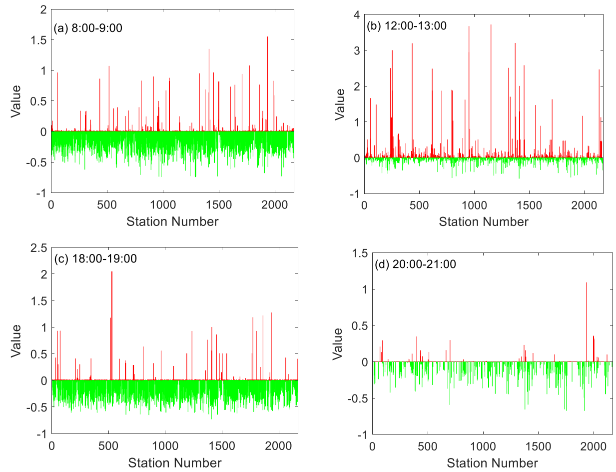


**FIGURE 13.** Average values of advanced headway stability index of stations in four time periods of weekday.

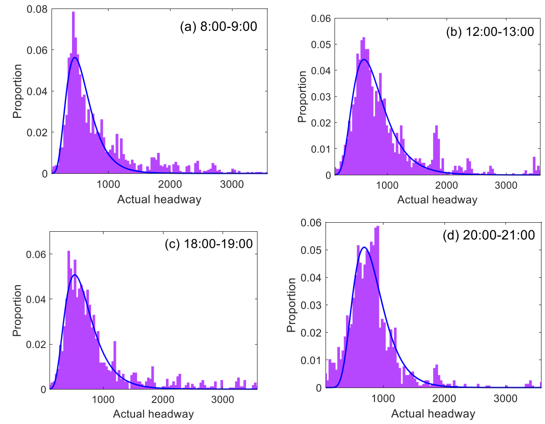


**FIGURE 14.** Average actual headway distribution of stations in four different time periods of weekend.

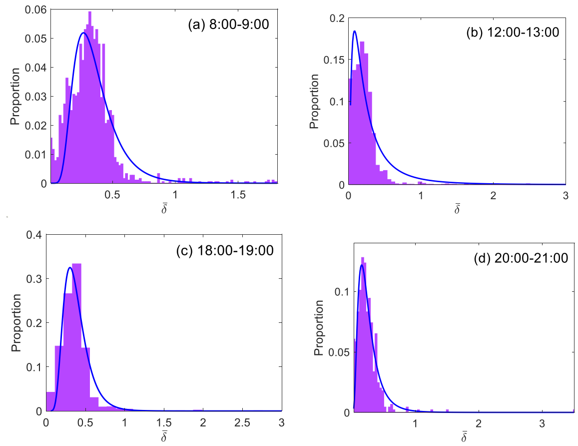


**FIGURE 15.** Average headway stability index distributions of stations in four different time periods on weekend.

service quality. During the noon time, the number of positive values increase sharply, and there are many stations that have very large value. It means that the many vehicles' headway is larger than plan headway and the service quality is worse, which validate the aforementioned results. Moreover, the negative values in the noon time are with small absolute values, which indicate the small headway or bus bunching can bring large headway of the continuous vehicles.

Another reason for the phenomenon that the negative values account for large part is that the actual headway is smaller than plan headway. Operators always to arrange more vehicle to overcome the bus bunching in peak hours. Therefore, the optimization of plan headway is a complicated question that perplex researchers. We also notice that the values of some stations are zeros during 20:00–21:00. That's because many lines stop operations after 20:00 and there is no value of these stations that serve these lines.

### D. PERFORMANCE ON WEEKEED

On weekend, most people do not work and the commuter flow decreases. In this part, we exhibit the performance of bus vehicles on weekend to compare with weekday.

#### 1) HEADWAY PERFORMANCE OF STATIONS

Fig. 14 shows the average actual headway distributions in four time periods of weekend. As can be seen, the distributions can be better fitted by lognormal distribution. Compared with weekday, there is less stations with small value of average headway at noon time. There are more stations on weekend with large average headway between 20:00 and 21:00. Totally, the headway performance of stations on weekend is different from weekday.

#### 2) STABILITY OF HEADWAY OF STATIONS

In order to display the performance of bus stability on weekend, Fig. 15 plots the average headway stability index distributions of stations in four different time periods. The outcomes show that these distributions can also be fitted by lognormal distribution. Compared with weekday, there exists some stations with large values during the time periods 18:00-19:00 and 20:00-21:00.

## VII. CONCLUSIONS

GPS trajectory data from devices on buses could provide many kinds of information to reflect the details of operation process. Traditional studies focused on the travel reliability of single route, while the network performance is lack of exhibition. Indeed, the macroscopic understanding of operation performance of bus networks can give a great help for improving the service quality. This paper proposes a data-driven framework to analyze the headway stability from the perspective of the entire network, which is conducive to identify the key stations and improve the stability of bus operations.

The outcomes show that the average dwell time, average headway and average headway stability index follow lognormal distributions, which have obvious right tails. Specifically, most values of the average headway stability index are lower than 0.5 and a small proportion of stations have the values that are larger than 1. It is worth nothing that large values of average headway stability index of network do not appear in the peak hours, but in the time periods before peak hours. In addition, the stations with most unstable headway are located in the suburbs and the fringe area of downtown. The outcomes suggest that operators should pay more attention to these areas and these time periods to efficiently improve the headway stability.

This study could provide a help for operators to improve the service quality. In the future study, we intend to take into account passenger demand of stations to better evaluate the performance of public transport networks.

## REFERENCES


[1] C. F. Daganzo, "A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons," *Transp. Res. B, Methodol.*, vol. 43, no. 10, pp. 913–921, 2009.

[2] W. Wu, R. Liu, and W. Jin, "Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour," *Transp. Res. B, Methodol.*, vol. 104, pp. 175–197, Oct. 2017.

[3] 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," *J. Transp. Geogr.*, vol. 69, pp. 45–57, May 2018.

[4] J. Jia, C. Liu, and T. Wan, "Planning of the charging station for electric vehicles utilizing cellular signaling data," *Sustainability*, vol. 11, no. 3, Jan. 2019, Art. no. 643.

[5] C. Zhuge and C. Shao, "Agent-based modelling of locating public transport facilities for conventional and electric vehicles," *Netw. Spatial Econ.*, vol. 18, no. 4, pp. 875–908, Dec. 2018. doi: 10.1007/s11067-018-9412-3.

[6] M. M. Rahman, S. C. Wirasinghe, and L. Kattan, "Analysis of bus travel time distributions for varying horizons and real-time applications," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 453–466, Jan. 2018.

[7] J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 292–306, May 2018.

[8] Z. Ma, L. Ferreira, M. Mesbah, and S. Zhu, "Modeling distributions of travel time variability for bus operations," *J. Adv. Transp.*, vol. 50, no. 1, pp. 6–24, Jan. 2016.

[9] Y. Yan, Z. Liu, and Y. Bie, "Performance evaluation of bus routes using automatic vehicle location data," *J. Transp. Eng.*, vol. 142, no. 8, Jan. 2016, Art. no. 04016029.

[10] O. Cats, "Determinants of bus riding time deviations: Relationship between driving patterns and transit performance," *J. Transp. Eng., A, Syst.*, vol. 145, no. 1, Jan. 2019, Art. no. 04018078.

[11] T. Cristóbal, G. Padrón, A. Quesada-Arencibia, F. Alayón, and C. R. García, "Systematic approach to analyze travel time in road-based mass transit systems based on data mining," *IEEE Access*, vol. 6, pp. 32861–32873, 2018.

[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," *Transp. Res. C, Emerg. Technol.*, vol. 87, pp. 197–211, Feb. 2017.

[13] I. I. Sirmatel and N. Geroliminis, "Mixed logical dynamical modeling and hybrid model predictive control of public transport operations," *Transp. Res. B, Methodol.*, vol. 114, pp. 325–345, Aug. 2018.

[14] N. C. Petersen, F. Rodrigues, and F. C. Pereira, "Multi-output bus travel time prediction with convolutional LSTM neural network," *Expert Syst. Appl.*, vol. 120, pp. 426–435, Apr. 2019.

[15] A. Fonzone, J. D. Schmöcker, and R. Liu, "A model of bus bunching under reliability-based passenger arrival patterns," *Transp. Res. C, Emerg. Technol.*, vol. 59, pp. 164–182, Oct. 2015.

[16] Z. L. Ma, L. Ferreira, M. Meshbah, and A. T. Hojati, "Modeling bus travel time reliability with supply and demand data from automatic vehicle location and smart card systems," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2533, pp. 17–27, Nov. 2015.

[17] A. Sumalee, T. Pan, R. Zhong, N. Uno, and N. Indra-Payoong, "Dynamic stochastic journey time estimation and reliability analysis using stochastic cell transmission model: Algorithm and case studies," *Transp. Res. C, Emerg. Technol.*, vol. 35, pp. 263–285, Oct. 2013.

[18] A. Kathuria, M. Parida, and C. S. Chalumuri, "Route performance evaluation of a closed bus rapid transit system using GPS data," *Current Sci.*, vol. 112, no. 8, pp. 1642–1652, Apr. 2017.

[19] 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," *Mod. Phys. Lett. B*, vol. 33, no. 3, Jan. 2019, Art. no. 1950015.

[20] H. Al-Deek and E. B. Emam, "New methodology for estimating reliability in transportation networks with degraded link capacities," *J. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 117–129, 2006.

[21] S. S. Moghaddam, R. Noroozi, J. M. Casello, and B. Hellinga, "Predicting the mean and variance of transit segment and route travel times," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2217, no. 1, pp. 30–37, 2011.

[22] A. Gal, A. Mandelbaum, F. Schnitzler, A. Senderovich, and M. Weidlich, "Traveling time prediction in scheduled transportation with journey segments," *Inf. Syst.*, vol. 64, pp. 266–280, Mar. 2017.

[23] R. Rajbhandari, S. I. Chien, and J. R. Daniel, "Estimation of bus dwell times with automatic passenger counter information," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1841, no. 1, pp. 120–127, 2003.

[24] A. Tirachini, "Bus dwell time: The effect of different fare collection systems, bus floor level and age of passengers," *Transportmetrica A, Transp. Sci.*, vol. 9, no. 1, pp. 28–49, 2013.

[25] B. P. Sterman and J. L. Schofer, "Factors affecting reliability of urban bus services," *Transp. Eng. J.*, vol. 102, pp. 147–159, 1976, Art. no. 11930.

[26] M. Turnquist, "A model for investigating the effects of service frequency and reliability on passenger waiting times," *Transp. Res. Rec.*, vol. 663, pp. 70–73, Aug. 1978.

[27] M. D. Abkowitz and I. Engelstein, "Factors affecting running time on transit routes," *Transp. Res. A, Gen.*, vol. 17, no. 2, pp. 107–113, Mar. 1983.

[28] E. Mazloumi, G. Currie, and G. Rose, "Using GPS data to gain insight into public transport travel time variability," *J. Transp. Eng.*, vol. 136, no. 7, pp. 623–631, 2009.

[29] N. Carrasco, "Quantifying reliability of transit service in Zurich, Switzerland: Case study of bus line 31," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2274, no. 1, pp. 114–125, Dec. 2012.

[30] Y. Huo, J. Zhao, W. Li, and X. Hu, "Measuring bus service reliability: An example of bus rapid transit in Changzhou," *J. Public Transp.*, vol. 17, no. 2, pp. 113–133, Jun. 2014.

[31] A. S. Shalaby and A. Gittens, "Evaluation of bus reliability measures and development of a new composite indicator," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2533, pp. 91–99, Feb. 2015.

[32] 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," *J. Transp. Eng., A, Syst.*, vol. 144, no. 5, May 2018, Art. no. 04018012.

[33] O. Cats and G. Loutos, "Real-time bus arrival information system: An empirical evaluation," *J. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 138–151, 2016.

[34] B. Yu, H. Wang, W. Shan, and B. Yao, "Prediction of bus travel time using random forests based on near neighbors," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 4, pp. 333–350, 2018.

[35] Y. Wang, Y. Bie, and Q. An, "Impacts of winter weather on bus travel time in cold regions: Case study of Harbin, China," *J. Transp. Eng., A, Syst.*, vol. 144, no. 11, Nov. 2018, Art. no. 05018001.

[36] A. Ceder, Y. Hadas, M. Mclvor, and A. Ang, "Transfer synchronization of public transport networks," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2350, no. 1, pp. 9–16, Dec. 2013.

[37] X. Dou, Y. Yan, X. Guo, and X. Gong, "Time control point strategy coupled with transfer coordination in bus schedule design," *J. Adv. Transp.*, vol. 50, no. 7, pp. 1336–1351, Nov. 2016.

[38] P. Fouilhoux, O. J. Ibarra-Rojas, S. Kedad-Sidhoum, and Y. A. Rios-Solis, "Valid inequalities for the synchronization bus timetabling problem," *Eur. J. Oper. Res.*, vol. 251, no. 2, pp. 442–450, Jun. 2016.

[39] Y. Wu, H. Yang, J. Tang, and Y. Yu, "Multi-objective re-synchronizing of bus timetable: Model, complexity and solution," *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 149–168, Jun. 2016.

[40] J. P. Fonseca, E. van der Hurk, R. Roberti, and A. Larsen, "A matheuristic for transfer synchronization through integrated timetabling and vehicle scheduling," *Transp. Res. B, Methodol.*, vol. 109, pp. 128–149, Mar. 2018.

[41] Z. Cao and A. Ceder, "Autonomous shuttle bus service timetabling and vehicle scheduling using skip-stop tactic," *Transp. Res. C, Emerg. Technol.*, vol. 102, pp. 370–395, May 2019.

[42] Y. Wang, D. Zhang, L. Hu, Y. Yang, and L. H. Lee, "A data-driven and optimal bus scheduling model with time-dependent traffic and demand," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2443–2452, Sep. 2017.

[43] C. Zhang, Z. Juan, Q. Luo, and G. Xiao, "Performance evaluation of public transit systems using a combined evaluation method," *Transp. Policy*, vol. 45, pp. 156–167, Jan. 2016.

[44] X.-H. Yang, G. Chen, S.-Y. Chen, W.-L. Wang, and L. Wang, "Study on some bus transport networks in China with considering spatial characteristics," *Transp. Res. A, Policy Pract.*, vol. 69, pp. 1–10, Nov. 2014.

[45] A. Chatterjee, M. Manohar, and G. Ramadurai, "Statistical analysis of bus networks in India," *PLoS ONE*, vol. 11, no. 12, Dec. 2016, Art. no. e0168478.

[46] T. Shanmukhappa, I. W.-H. Ho, and C. H. Tse, "Spatial analysis of bus transport networks using network theory," *Phys. A, Stat. Mech. Appl.*, vol. 502, pp. 295–314, Jul. 2018.

[47] H. Zhang, C. Zhuge, and X. Yu, "Identifying hub stations and important lines of bus networks: A case study in Xiamen, China," *Phys. A, Stat. Mech. Appl.*, vol. 502, pp. 394–402, Jul. 2018.

[48] X. Zhang, G. Chen, Y. Han, and M. Gao, "Modeling and analysis of bus weighted complex network in Qingdao city based on dynamic travel time," *Multimedia Tools Appl.*, vol. 75, no. 24, pp. 17553–17572, Dec. 2016.

[49] X. L. Zhang, Y. B. Ren, B. X. Huang, and Y. Han, "Analysis of time-varying characteristics of bus weighted complex network in Qingdao based on boarding passenger volume," *Phys. A, Stat. Mech. Appl.*, vol. 506, pp. 376–394, Sep. 2018.

[50] X.-L. An, L. Zhang, Y.-Z. Li, and J.-G. Zhang, "Synchronization analysis of complex networks with multi-weights and its application in public traffic network," *Phys. A, Stat. Mech. Appl.*, vol. 412, pp. 149–156, Oct. 2014.

[51] T. Ren, Y.-F. Wang, M.-M. Liu, and Y.-J. Xu, "Analysis of robustness of urban bus network," *Chin. Phys. B*, vol. 25, no. 2, 2016, Art. no. 020101.

[52] H. Pu, Y. Li, C. Ma, and H. Mu, "Analysis of the projective synchronization of the urban public transportation super network," *Adv. Mech. Eng.*, vol. 9, no. 6, Jun. 2017, Art. no. 1687814017702808.

[53] 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, no. 7, Apr. 2019, Art. no. 2007.

[54] W. Xu and L. Yang, "Evaluating the urban land use plan with transit accessibility," *Sustain. Cities Soc.*, vol. 45, pp. 474–485, Art. no. 2019.

[55] 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," *Int. J. Geograph. Inf. Sci.*, Apr. 2019. doi: 10.1080/13658816.2019.1608997.

[56] L.-S. Sun, S.-W. Wang, L.-Y. Yao, J. Rong, and J.-M. Ma, "Estimation of transit ridership based on spatial analysis and precise land use data," *Transp. Lett.*, vol. 8, no. 3, pp. 140–147, 2016.

[57] G. Qi, A. Huang, W. Guan, and L. Fan, "Analysis and prediction of regional mobility patterns of bus travelers using smart card data and points of interest data," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1197–1214, Apr. 2019.

[58] Y. Zhou, G. Jiang, and G. Jiang, "Discover the road sequences of bus lines using bus stop information and historical bus locations," *Int. J. Distrib. Sensor Netw.*, vol. 15, no. 2, Feb. 2019, Art. no. 1550147719830552.

[59] L. Quintero-Cano, M. Wahba, and T. Sayed, "Bus networks as graphs: New connectivity indicators with operational characteristics," *Can. J. Civil Eng.*, vol. 41, no. 9, pp. 788–799, Jul. 2014.

[60] J. Ding, Y. Zhang, and L. Li, "Accessibility measure of bus transit networks," *IET Intell. Transp. Syst.*, vol. 12, no. 7, pp. 682–688, Sep. 2018.


**HUI ZHANG** received the B.S. degree in applied mathematics from Qingdao Agricultural University, in 2009, 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. Since 2016, he has been a Lecturer with the School of Transportation Engineering, Shandong Jianzhu University. His research interests include transportation engineering, big data analysis, complex networks, and engineering management.

**HOUDUN CUI** received the B.S. degree in traffic and transportation engineering from Northeast Forestry University, in 2010, and the M.S. degree in system engineering from Beijing Jiaotong University, in 2012. He is currently an Engineer with Shandong Hi-speed Group Company Ltd. His research interests include traffic engineering and traffic safety.

**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.