

## Research Article

# Identifying Key Bus Stations Based on Complex Network Theory considering the Hybrid Influence and Passenger Flow: A Case Study of Beijing, China

Jianlin Jia <sup>1</sup>, Yanyan Chen <sup>1</sup>, Ning Chen,<sup>1</sup> Hui Yao,<sup>1</sup> Yongxing Li <sup>2</sup>, and Zhuo Liu<sup>1</sup>

<sup>1</sup>Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China

<sup>2</sup>School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

Correspondence should be addressed to Yanyan Chen; [cdyan@bjut.edu.cn](mailto:cdyan@bjut.edu.cn)

Received 16 June 2020; Revised 27 November 2020; Accepted 1 December 2020; Published 17 December 2020

Academic Editor: Lei Wang

Copyright © 2020 Jianlin Jia 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.

In the bus network, key bus station failure can interrupt transfer lines, which leads to the low effectiveness of the whole network, especially during peak hours. Thus, identifying key stations in the bus network before the emergency occurs has a great significance to improve the response speed. In this paper, we proposed a new method considering station hybrid influence and passenger flow to identify key stations in the whole bus network. This method aims to measure the influence of bus stations while combining the topological structure of the bus network and dynamic bus stations passenger flow. The influence of bus stations was calculated based on the local structure of the network, which refines from finding the shortest paths with high computational complexity. To evaluate the performance of the method, we used the efficiency of the network and vehicle average speed at the station to examine the accuracy. The results show that the new method can rank the influence of bus stations more accurately and more efficiently than other complex network methods such as degree,  $H$ -index, and betweenness. On this basis, the key stations of the bus network of Beijing in China are identified out and the distribution characteristics of the key bus stations are analyzed.

## 1. Introduction

A well-developed public transport system can effectively alleviate urban traffic congestion. The bus network is a crucial component of the public transport system and has a great significance to urban traffic. With the rapid development, especially in big cities such as Beijing, China, and New York, USA, the bus system has become too complicated to evaluate. Taking the bus network of Beijing, China, as an example, there are 886 lines and 25424 bus vehicles. Besides, the length of all bus lines is up to 19290 km. The annual passenger volume of the bus has reached 3.36 billion people, and the mean passenger volume is 9.73 million one day. Under the highly complex transit network, how to quickly find influential stations and analyze the topology of the bus network has become a hot topic of current research.

With the small-world network [1] and a scale-free network [2, 3] being proposed, a complex system has made big development. The main function of a complex network is to transform the relationship of each element of the real system into the nodes and edges, emphasizing the topological structure of networks [4]. Therefore, the structural characteristics of the network play an important role in understanding the function of a complex system. For instance, the topological structure of the social network affects the spreading of rumors information [5]. To sum up, complex network theory is a powerful tool in various fields, such as biological networks [6–8], social networks [9–11], and transportation network [12–14].

In terms of transit networks, it is a significant approach to acquire the topological structure using complex network theory. In the past few decades, a lot of scholars have applied complex network theory to carry out related research studies

on urban public transportation networks. Meanwhile, some indices have been proposed to measure the topological characters containing a degree [15], clustering coefficient [16], and betweenness [17, 18]. Sienkiewicz and Hołyst [19] analyzed the topological structure of public transportation networks in 22 cities of Poland and found that all networks have small-world characteristics and the degree distribution of the network obeys power-law distribution or exponential distribution. Ping et al. [20] studied the bus network of 10 cities in China and found that the bus network shows robustness to random attacks and fragility to malicious attacks on a global scale. These studies are useful to find macroscopic properties that are not yet recognized in the real complex system. The scale-free and traffic concentration problem of the network is one of the basic scientific issues in a complex network.

Moreover, once key nodes in the network are damaged, it will quickly spread across the whole network [21, 22]. For example, Jovica proposed that a failure initiated with one critical system could cause cascading effects within the whole system-of-system in a coupled infrastructure system [23]. Thus, there are more and more correlation studies on the efficiency of the whole network when some components of the system failed [24–26]. However, the transportation system is a complex and uncertain system, which is vulnerable when it suffers from bad weather, road construction, and traffic congestion [27, 28]. Bus stations are one of the elementary components of transit networks. Key bus stations play an important role in the operation of the network. The failure of key bus stations could cause large-scale traffic congestion and even paralyze the traffic networks. Therefore, identifying and analyzing the key nodes is critical to the whole transit network, especially during peak hours.

Recently, many types of research focused on identifying key nodes of transit networks. Chen et al. proposed normalized average distance (NAGD), normalized average minimum speed (NAMS), and normalized largest component order (NLCO) algorithms to elevate the vulnerability of networks and help identify the key nodes in networks [29]. Using the supernode graph structure of London city, Shanmukhappa proposed a spatial amalgamation to identify influential nodes in bus transport and metro transport networks [30]. Sun et al. used the  $z$ -score and participation coefficient to analyze the influence of nodes in the Dublin bus network [31]. Zhang et al. studied an approach to extract the hub network from the urban subway transit network, revealing that the hub network is hierarchical. At the same time, transfer stations also played a critical role in key networks [32]. Xu et al. created a weighted network for public transit of Beijing based on key property theory and analyzed the scaling laws and correlations [33, 34]. Moreover, the other approaches for identifying key nodes list as follows: proximity prestige [35], structural holes [36], degree centrality [37],  $K$ -core [38], and betweenness centrality [39].

To best of the authors' knowledge, the studies on identifying key bus stations mainly focus on the topological structure, and the topological evaluation index is treated as unrelated or single, such as degree,  $H$ -index, and  $K$ -core. In reality, the bus network is a typical dynamic complex

network, which not only has the general characteristics of complex network but also has the remarkable feature, such as the travel behaviour of passengers. Generally speaking, passengers tend to travel in areas with active of socioeconomic activities, such as CBD, and transfer hubs. Hence, in addition to considering the connectivity characteristics of physical structure, the travel behaviour of passengers and passenger flow demand of stations should be considered in the bus network. Meanwhile, the passenger flow of bus stations presents the following features [40]: (i) there are two rush hours in the morning and evening on each working day, separately; (ii) the passenger flow of bus stations in the weekend is obviously less than that on a working day. Therefore, in this paper, we select the rush hours in the morning and evening on each working day and weekend as our time periods. Figure 1 shows the passenger flow during different time periods of Si-hui station, Liu-li-qiao station, Beijing west railway station, and Guo-mao station. It is obvious that passenger flow of different time periods has a great difference.

Therefore, in this paper, we firstly explored the heterogeneous relationship between degree and  $H$ -index of the node using complex network theory. Then, based on the relationship, considering the dynamic passenger flow of the bus station, we constructed a new local hybrid influence model (called SHIP) representing the importance of the station in the whole bus network. Compared with the single index of the complex network such as degree,  $H$ -index, and betweenness, we used network efficiency as an evaluation metric to test the effectiveness of the proposed approach. In addition, this paper identified the key station of the Beijing bus network based on this approach and analyzed its spatial distribution characteristics.

The proposed method aims to quickly identify key bus stations before the emergencies really happen, which is essential for improving the response speed, as well as preventing the occurrence of secondary disasters. The main contribution of this research includes mainly two aspects: (i) a new index is proposed based on conventional complex network indices combining the degree and  $H$ -index to better represent the connected relationship of the bus network; (ii) this approach is a local algorithm for mining and identifying the key bus stations based on topological structure and actual travel demand of the station, which avoids calculating the shortest path between any two stations. The computational complexity is greatly reduced.

This paper is divided into four sections. section 2 gives an identification model considering the hybrid influence of degree,  $H$ -index, and passenger flow. Section 3 depicts a case study and verifies the model. Meanwhile, we use the new model to analyze the bus network of Beijing, China. The distribution characteristics of key bus stations are given in Section 3. Section 4 proposes a conclusion and work in the future.

## 2. Identification Model

In the bus network, station failure can cut off transfer lines, which leads to low effectiveness of the whole network. To

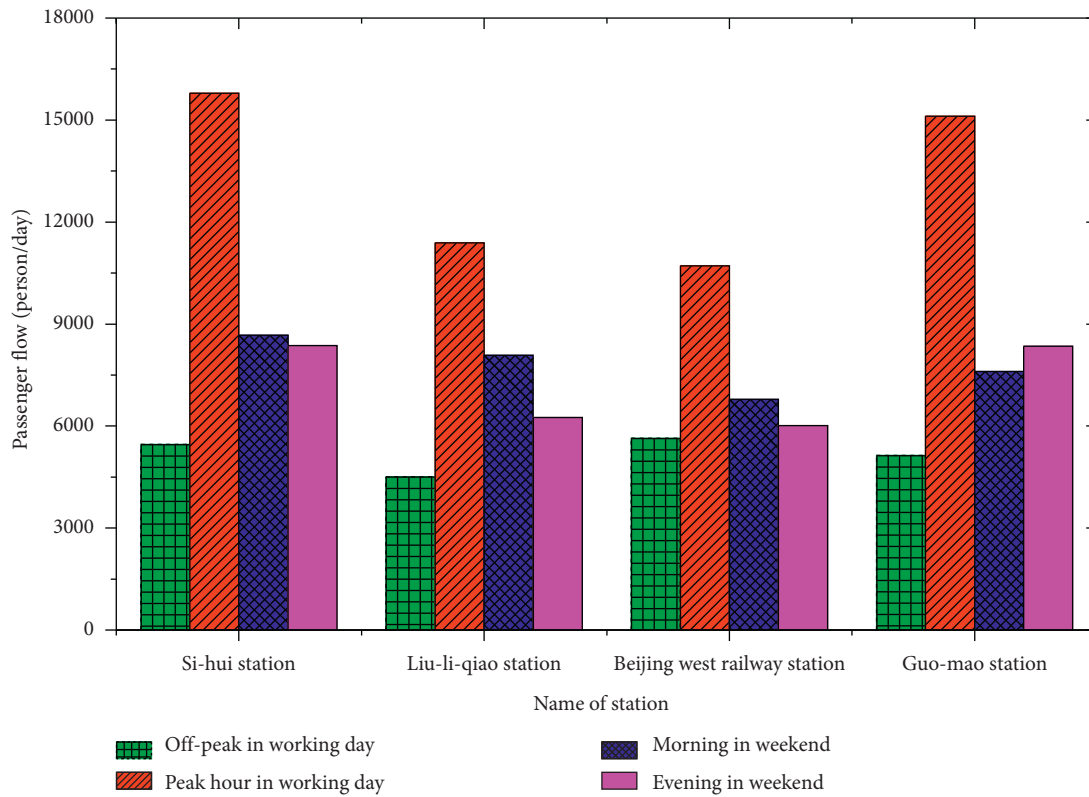


FIGURE 1: The passenger flow in different periods.

enhance the robustness and improve the response speed of the bus network, in this section, we construct a key station identification model. The model includes two stages. Stage one of our model comprises of the bus network construction to calculate the characteristics of the topological structure, while the identification model and evaluation are presented in stage two. The solution framework is shown in Figure 2.

**2.1. Problem Description.** The previous methods mainly focus on topological structure in identifying key nodes in the whole network. The simplest method to measure the importance of a node is to determine its degree, that is, to calculate the number of its linked neighbors. Then, the Hirsch index (i.e.,  $H$ -index) is originally used to measure the citation impact of a scholar or a journal. In this paper, the  $H$ -index of a node is defined to be the maximum value  $h$  such that there is, at least,  $h$  neighbors of degree no less than  $h$  [41–43]. Therefore, the node with a large degree and  $H$ -index play more important roles in the whole network. The large of the degree represents that the influence of the nodes is big. Nevertheless, the  $H$ -index is better than a degree or coreness in quantifying nodes' influence in some cases [44].

Based on that mentioned above, we consider the degree and  $H$ -index in identifying the key bus stations. We give a visual representation for the hybrid of degree and  $H$ -index, as shown in Figure 3. We describe the degree as  $D$  and depict the  $H$ -index as  $H$ . We can see that nodes a1 and c1 have the same degree (namely  $D = 3$ ); however, c1 has a larger  $H$ -index than a1, owing to the larger maximal connected

subgraph. Nevertheless, only considering the  $H$ -index cannot determine the maximal connected subgraph. For example, a1 and b1 have the same  $H$ -index, but the degree is different ( $D = 2$  for a1 and  $D = 4$  for b1). Therefore, b1 has a larger maximal connect subgraph than a1. Hence, degree and  $H$ -index cannot represent the maximal connected subgraph of nodes in networks absolutely. We also find that the product of the degree and  $H$ -index can get a maximal subgraph. For instance,  $D = 4$  and  $H = 2$  for b1,  $D = 3$ , and  $H = 3$  for c1, and the product of b1, c1 is 8, 9; however, the sum is 6. We can find (intuitively see) that c1 has the larger maximal connect subgraph than b1. Therefore, we construct the identification model considering the hybrid influence between the degree and  $H$ -index.

**2.2. Construction of the Network.** The bus network is a typical complex network which consists of bus stations and bus lines. Stations and lines of the bus network can be depicted by nodes and links. Nodes represent stations, and links represent the lines between two stations. Therefore, we construct the bus network using complex network theory. In this section, we describe the construction method of the bus network.

Also, the bus network can be abstracted as a graph. In this paper, we use a graph  $G(V, E)$  to represent the bus network. This  $V$  is the set of nodes, and  $V$  is the station set of the bus network.  $E$  is the set of edges.  $G$  is a  $N \times N$  matrix which consists of 0 or 1.  $N$  is the number of stations. If there is an edge between the station  $i$  and  $j$ ,  $e_{ij} = 1$ ; otherwise

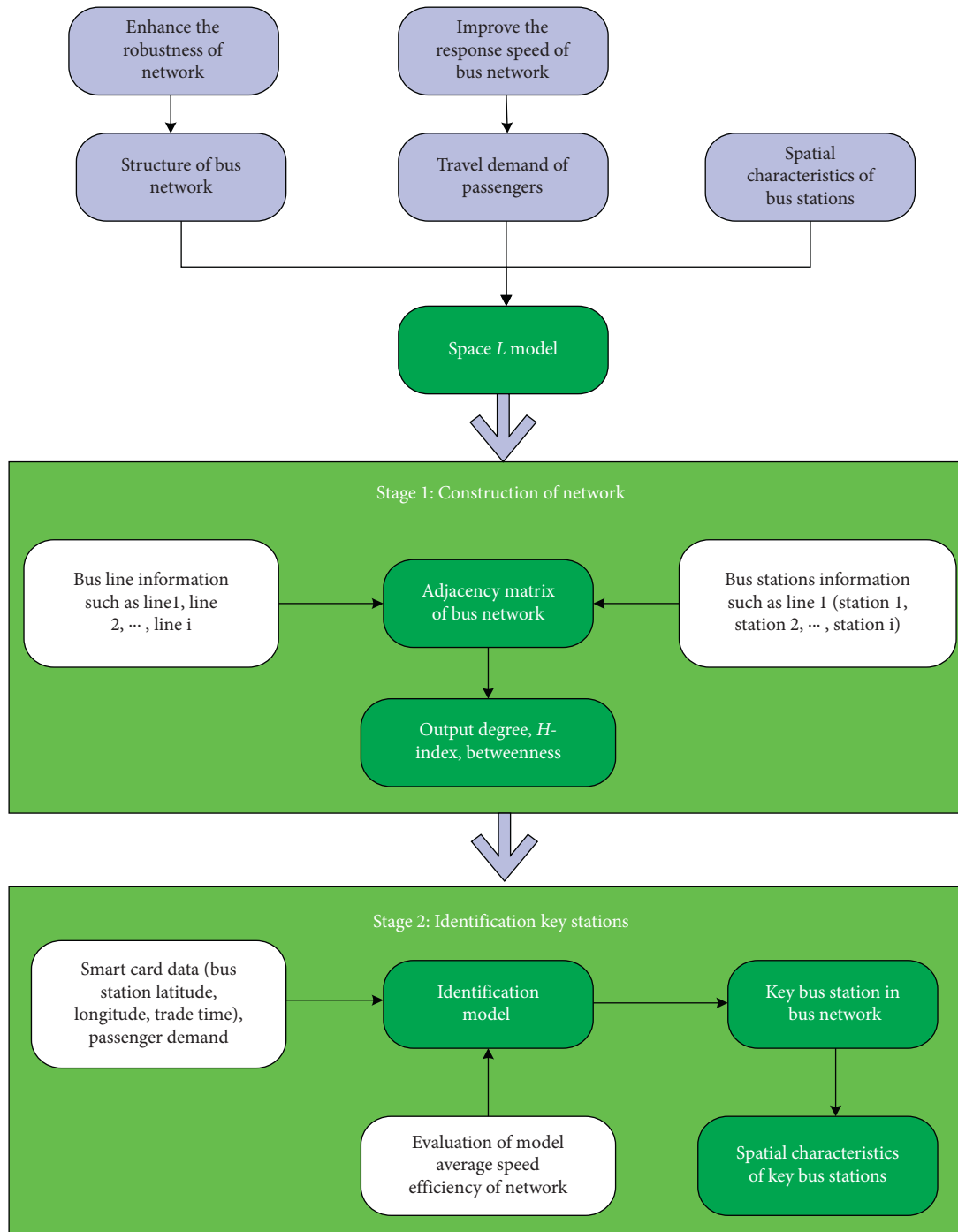


FIGURE 2: The solution framework.

$e_{ij} = 0$ . Stations A and B represent different stations in bus networks. If a bus can be driven from station A to station B, it can also from station B to station A. Therefore, an urban bus network can be regarded as an undirected network.

The description method of the bus network is divided into two types by connection of stations. One is space  $L$ , and the other is space  $P$  [45–47]. Space  $L$  is used to describe the topological properties of the network. Space  $P$  is used to represent the transfer properties of the network. Since the method of space  $L$  reflects the geographical connection

between bus stations, we build the bus network using space  $L$ . The space  $L$  is depicted in Figure 4 by a simple network.

**2.3. Influence of Stations.** The degree and  $H$ -index are the critical characteristics of nodes. The heterogeneity of the degree and  $H$ -index can obtain the maximal connected subgraph. Using hybrid influence as a new metric to measure the importance of stations can better explore the potential influence of stations. The  $H$ -index and degree heterogeneity

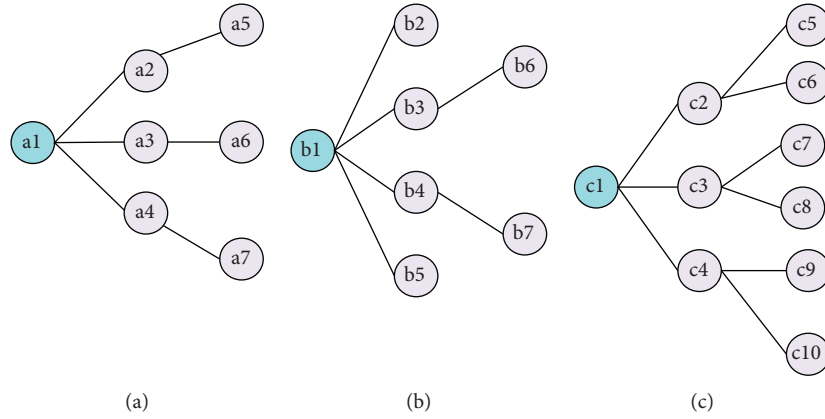


FIGURE 3: The diagram of local hybrid influence of the degree and  $H$ -index. (a)  $D=3$  and  $H=2$  for  $a1$ ; (b)  $D=4$  and  $H=2$  for  $b1$ ; and (c)  $D=3$  and  $H=3$  for  $c1$ .

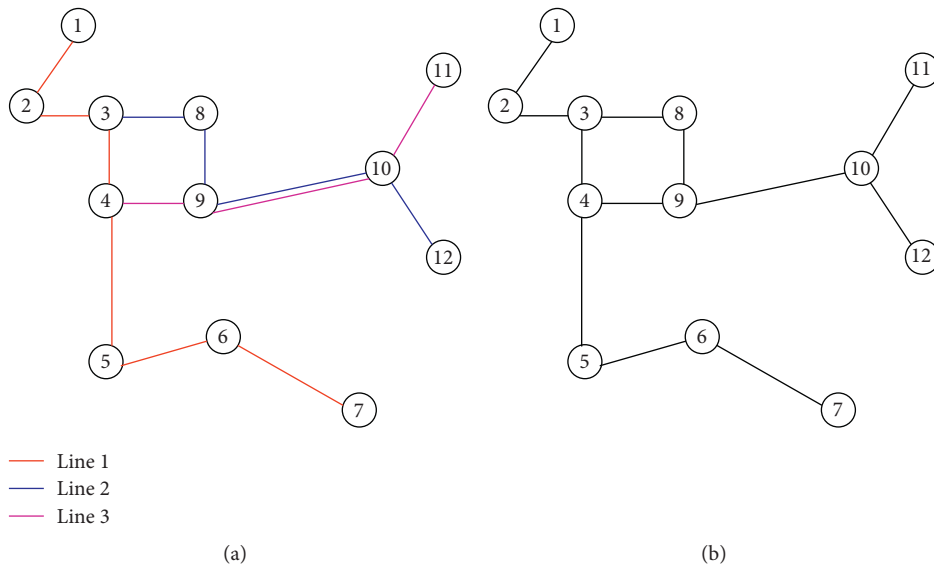


FIGURE 4: The topology of the bus network. (a) A bus network. (b) Bus network under space  $L$ .

are defined as equations (1) and (2). We define the hybrid influence of key bus stations as equation (3).

$$H_{ki} = \frac{Ek_i^2}{E\langle k_i \rangle^2} \quad (1)$$

$$H_{hi} = \frac{Eh_i^2}{E\langle h_i \rangle^2}, \quad (2)$$

$$\vartheta_{SHIP} = \Psi \cdot \frac{\sqrt{k_x^{(1/H_{kx})} \times h_x^{(1/H_{hx})}}}{2|E|}, \quad (3)$$

where  $k_i$  and  $h_i$  represent the degree and  $H$ -index of the station  $i$ .  $H_{ki}$  and  $H_{hi}$  map the heterogeneity of the degree and  $H$ -index.  $|E|$  is the number of links.  $\vartheta_{SHIP}$  is the hybrid influence.  $\Psi$  is the weight parameter.

Passenger flow of stations is a pivotal parameter that can represent the importance of stations to some extent. Meanwhile, passenger flow of stations at different time and locations vary widely. It is clear that the passenger flow of bus stations differs by period from Figure 2. In reality, transfer stations' failure will cause lots of problems when they happen during peak hours. In order to obtain the key station's distribution of the bus network under special time, we select the morning peak (7:30–8:30) and off-peak (14:00–15:00) in a working day and morning peak (7:30–8:30) in the weekend as the research periods. Therefore, we use the normalization of passenger flow as weight called  $\Psi$ . The weight parameter  $\Psi$  can be expressed as

$$\Psi = \frac{x_i}{x_1 + x_2 + \dots + x_i \dots + x_n}, \quad (4)$$

where  $x_i$  represents the passenger flow of the station  $i$ .

### 3. Evaluation Metrics

In this section, we use two metrics to evaluate the identification model: average speed and efficiency of the network.

**3.1. Average Speed.** Key bus stations play an important role in the bus network and have an impact on network topology and vehicle operation. The network connectivity of key bus stations has a good performance, which owns many transfer stations. Meanwhile, there are many points of interest such as catering areas, business district, and medical service around the key bus station, leading the bus lines concentrating, a large number of passenger transfer, and more waiting time around the key bus stations. Therefore, the vehicle average speed of key bus stations is less than that of other stations with a low hybrid influence. In addition, Shanmukhappa et al. pointed out that, with increasing node weight (demand), the maximum speed closer bus station is reduced significantly [48]. In order to easily obtain the average speed of vehicles near the bus stations, we construct a simulation test network by simulation software named VISSIM. In simulation, we set the monitors of speed at 50 meters before the bus station. The average vehicle speed of the bus station  $i$  is depicted as

$$\bar{v}_i = \frac{v_{i,1} + v_{i,2} + \dots + v_{i,n}}{n}, \quad (5)$$

where  $v_{i,j}$  is the speed of the vehicle  $j$  at the station  $i$ .  $n$  is the total number of vehicles at the station  $i$ .

**3.2. Efficiency of the Network.** Meanwhile, we also test the model with traditional algorithms through the effectiveness of the network which can be represented by equations (6) and (7). The efficiency of the network can reflect the connectivity of the network. The higher the connectivity, the higher the efficiency of the network.

$$\eta_{\text{efficiency}} = \frac{1}{N(N-1)} \sum_{v_i \neq v_j} \sigma_{ij}, \quad (6)$$

$$\sigma_{ij} = \frac{1}{d_{ij}}, \quad (7)$$

where  $\eta_{\text{efficiency}}$  is the efficiency of the whole network.  $d_{ij}$  is the shortest distance between nodes  $i$  and  $j$ . When there is no path between nodes  $i$  and  $j$ ,  $d_{ij} = \text{inf}$ .  $N$  is the number of nodes.

### 4. Algorithm Flow

In what follows, we present the detailed procedure of the solution algorithm for calculating the key bus stations. The solution algorithm includes three aspects: construction of the bus network, calculation of hybrid influence, and identification of key bus stations. The flowchart of the solution algorithm is shown in Figure 5.

### 5. Model Verification

We selected the Wang-jing district, Beijing, China, as the test bus network which includes 5 lines and 35 stations to verify the accuracy of identification model results. The distribution of stations and lines is shown in Figure 6.

Based on the bus network of Wang-jing district, Beijing, China, we construct the simulation network on the platform of VISSIM software for obtaining the average speed of vehicles near the bus station. In test bus networks, we set the same traffic flow. We assume that the signal cycle of each intersection is 90 s, and the detailed signal timing is shown in Figure 7.

The verification results of the model are shown in Figure 8 and Table 1. Figure 8(a) expounds on the relationships between average speed and key bus stations. By simulation, we found that, with increasing hybrid influence of stations, the average speed at the bus stations reduces significantly.

The efficiency of the network is the index to measure the connectivity of the network, which is calculated by the shortest distance between two nodes. In this paper, we proposed a station failing process, which is called deliberate attack. When a station is attacked, the edges connecting with the failed station will be removed under space  $L$ . In case one station has been attacked, the efficiency of the bus network will change. Therefore, the stations are attacked in turn according to the importance of the station. The greater the decline in the efficiency of the network, the worse the connectivity of the network, indicating the station is more important in the bus network.

Meanwhile, the results of the comparison with traditional methods, which include the degree,  $H$ -index, and betweenness, are shown in Figure 8(b). From Figure 8(b), it can be seen that the curve of the new index (SHIP) declines fastest, indicating that the new method is more accurate in identifying key stations than other methods. Contemporaneously, the traditional method of computational complexity using the shortest path is  $O(N^2) = 1225$ ; however, the new approach is  $O(N) = 35$ , showing that the computational complexity is greatly reduced.

### 6. Case Study

**6.1. Data Description.** The bus network of Beijing, China, is selected as the case study in our work. We collected the data about the bus network of Beijing, China, from November 1, 2017, to November 7, 2017. The data consist of 5922 stations and 886 lines. The raw data include card ID, card type, trade type, trade time, line ID, bus ID, name of stations and lines, latitude, and longitude of the station. The data format is shown in Table 2.

**6.2. Bus Network Description.** Based on the method of space  $L$ , we construct the bus network, shown in Figure 6. Meanwhile, the characteristics of topology structure in the

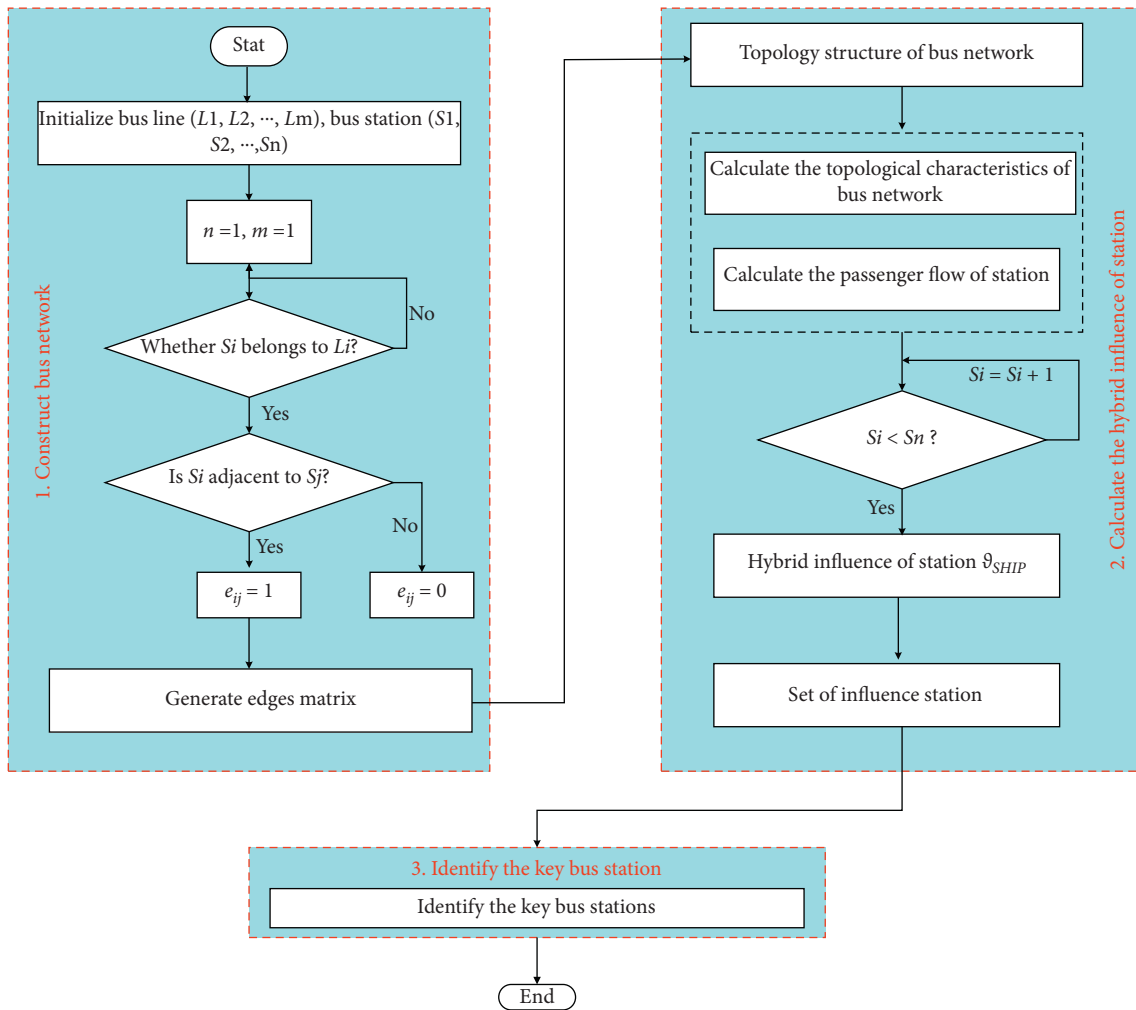


FIGURE 5: The flowchart of the solution algorithm.

network can be described as the measure metrics such as the degree and  $H$ -index. We calculate the metrics, and the results are shown in Figure 9. The abscissa represents the degree and  $H$ -index of the node. The ordinate represents the number of nodes in Figure 10(a). The ordinate represents the frequency of nodes in Figure 10(b). From Figure 10, the degree distribution of nodes is from 1 to 18. The starting or ending stations of each bus line in the bus network have the same degree which is equal to 1. After statistics, the ratio of starting and ending stations is 4.20%. The proportion of ordinary stations in bus networks accounts for the largest, which is 52.01%. The degree of ordinary stations is equal to 2. The value of the average degree in the network is 3.08, which indicates that each bus station has three edges. Similarly, the distribution of nodes is 1 to 8. The nodes'  $H$ -index equal to 2 has the largest proportion, which is 64.35%.

## 7. Application, Analysis, and Discussion

In this section, we use the proposed identification model which has been verified to analyze the distribution characteristics of key bus stations.

**7.1. Regional Distribution.** The results show that the stations with a high value of hybrid influence are most distributed in Dongcheng, Xicheng, Haidian, Chaoyang, Shijingshan, and Fengtai. According to statistics of the top 50 key stations, we count the number of key stations in each region, shown in Figure 11. It is obvious that Chaoyang has the greatest number of key bus stations and Shijingshan owns the least, no matter on working day or weekend. In addition, Haidian has more key bus stations on a working day, indicating that there are more people taking the bus to work.

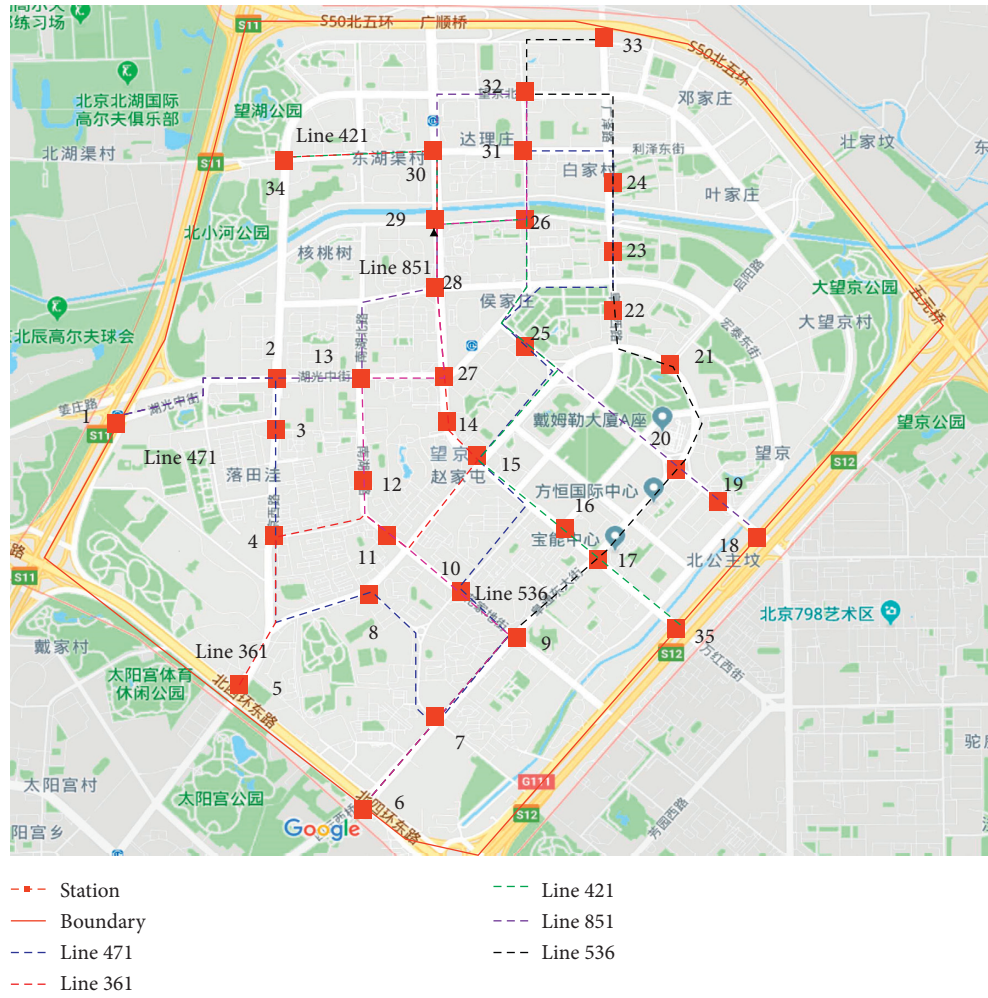


FIGURE 6: Stations and bus lines of Wang-jing district, Beijing, China.



FIGURE 7: The traffic signal timing. Green time, amber time, and red time are equal to 40 s, 5 s, and 45 s, respectively.

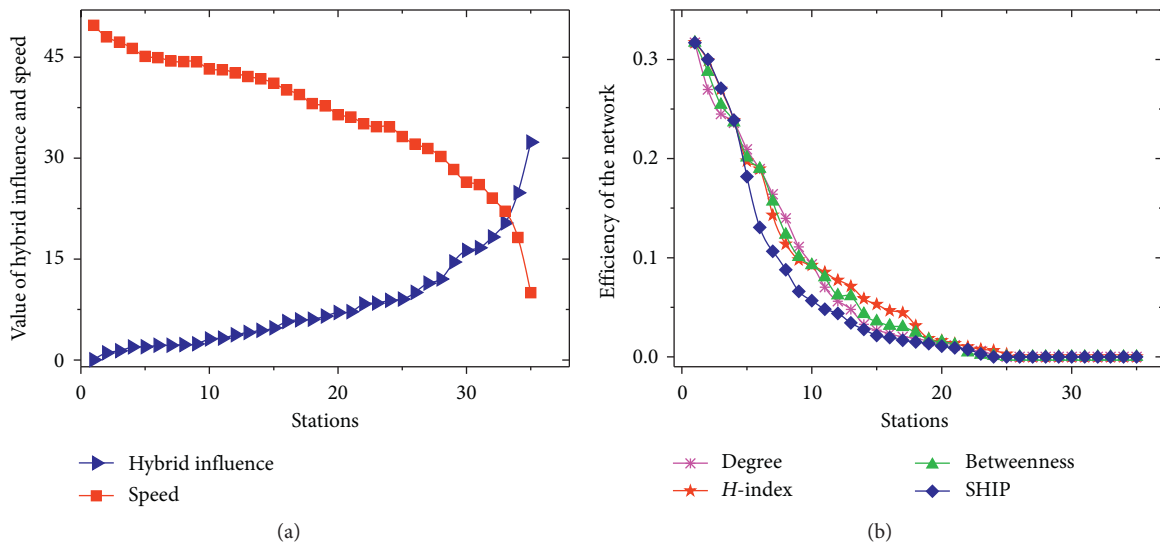


FIGURE 8: The results of model verification.



TABLE 1: The results of the simulation test.

SI	HI ( $\times 10^4$ )	AS (km/h)	SI	HI ( $\times 10^4$ )	AS (km/h)
1	3.240	43.11	2	2.390	44.29
3	1.953	44.78	4	12.026	30.25
5	6.057	38.11	6	4.760	41.13
7	8.316	35.09	8	5.959	39.43
9	24.830	18.21	10	16.284	26.45
11	5.718	40.12	12	1.928	45.34
13	20.361	22.09	14	4.086	42.13
15	18.277	24.07	16	6.506	37.76
17	14.561	28.30	18	1.373	46.11
19	8.838	34.65	20	8.453	34.67
21	3.112	43.24	22	2.274	44.32
23	2.189	44.41	24	0.025	49.71
25	7.179	36.11	26	16.685	26.07
27	9.033	33.21	28	4.402	41.78
29	10.035	32.09	30	3.740	42.65
31	32.360	10.2	32	11.406	31.42
33	1.070	48.01	34	2.178	44.51
35	7.050	36.43			

Note. "SI" represents the station ID. "HI" is the hybrid influence of stations. "AS" maps the average speed of stations.

TABLE 2: The raw data format of the bus network.

Field name	Type of data	Meaning (sample)
Card ID	Varchar 2 (8)	Card number, such as 10007515****4800
Card type	Varchar 2 (2)	Ages (adult, child, senior)
Trade type	Varchar 2 (2)	Type of transaction, such as normal, excess fare
Trade time	Varchar 2 (14)	Get off time
Mark time	Varchar 2 (14)	Boarding time
Line ID	Varchar 2 (5)	Bus lines, such as 171, 486
Bus ID	Varchar 2 (5)	Bus stations, such as Cai-shi-kou station
Station latitude	Varchar 2 (14)	Location of the station (39.940166)
Station longitude	Varchar 2 (14)	Location of the station (116.496255)



FIGURE 9: The topological structure of the bus network under space  $L$ .

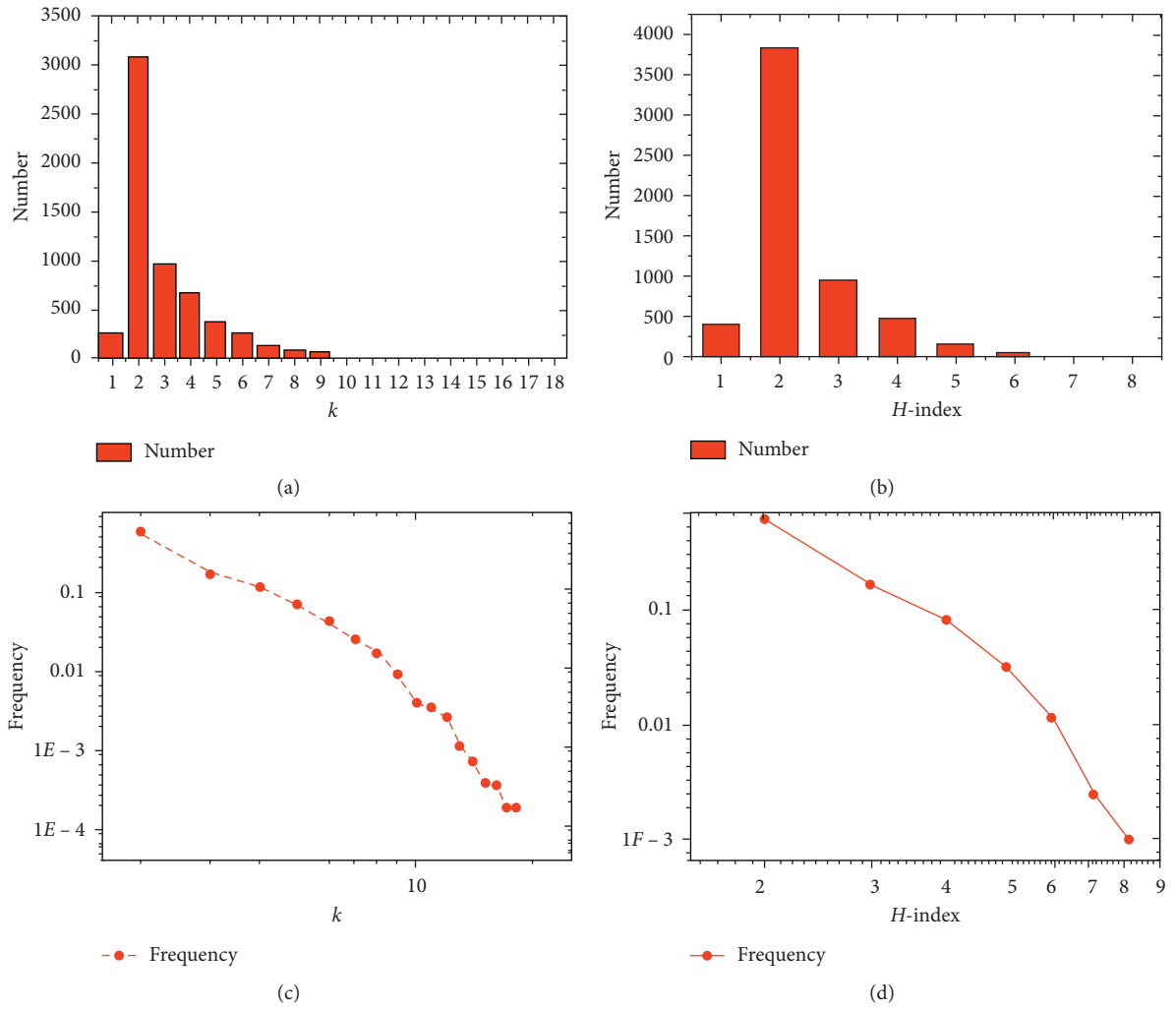


FIGURE 10: The properties of the bus network. (a) The number of stations' degrees. (b) The number of stations'  $H$ -index. (c) The frequency distribution of stations' degree. (d) The frequency distribution of stations'  $H$ -index.

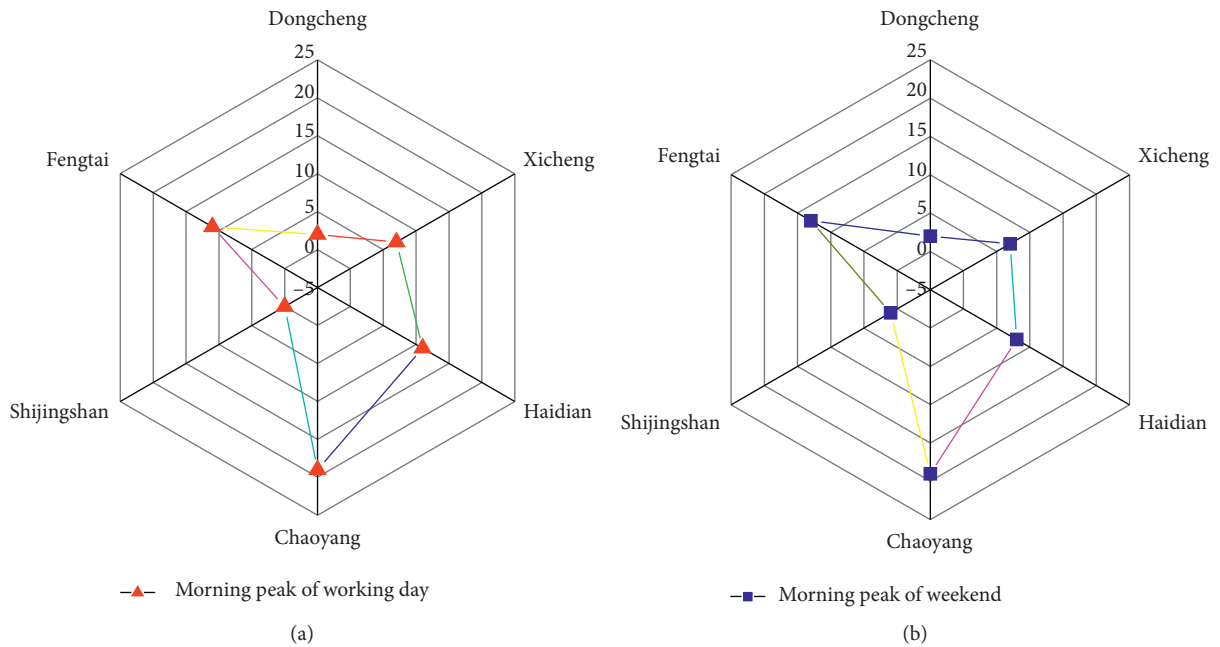


FIGURE 11: Continued.

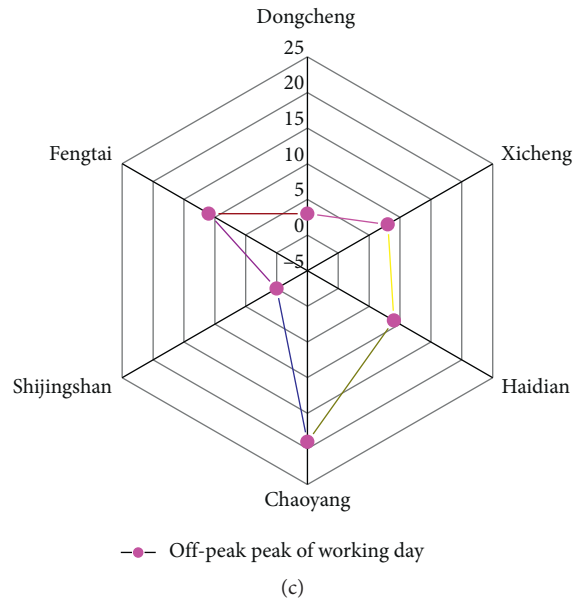


FIGURE 11: The distribution of key bus stations in different regions.

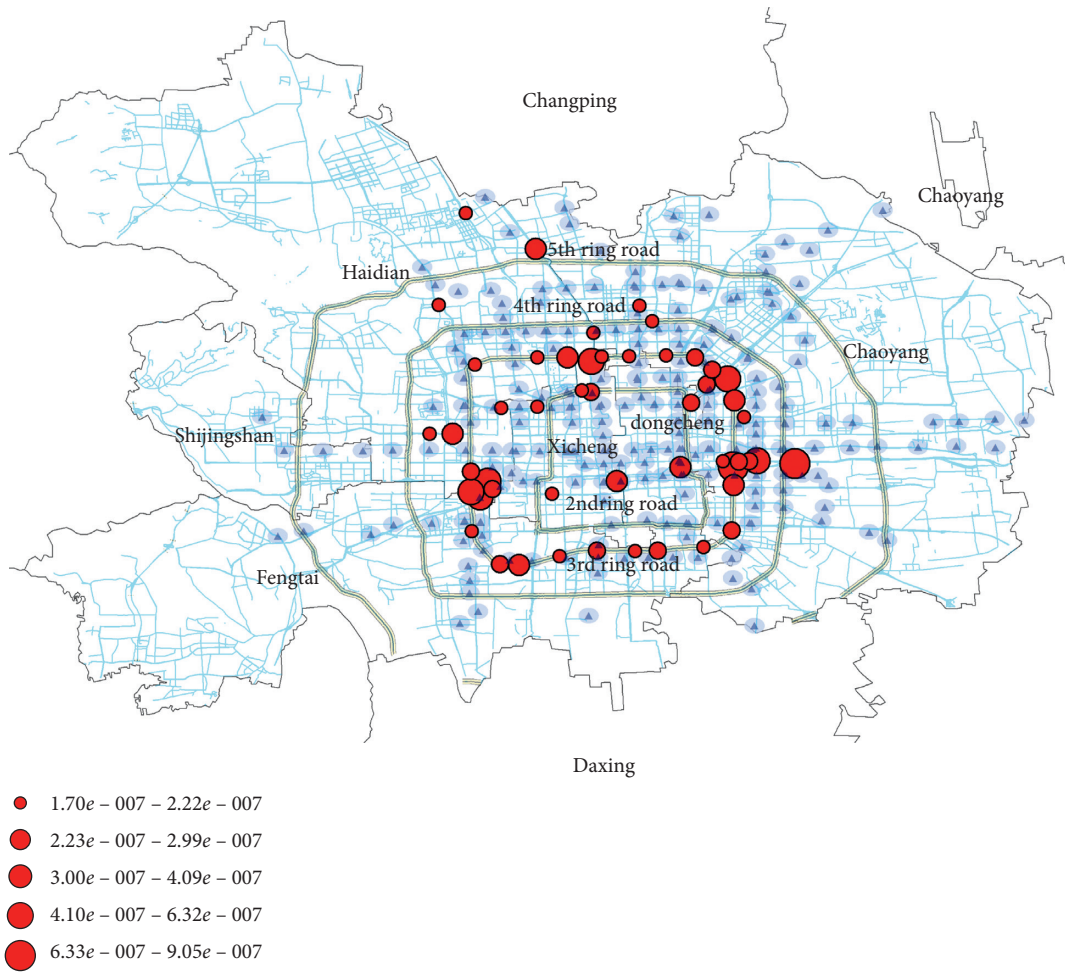


FIGURE 12: Key bus stations occupying the top 50 in the morning peak (working day).

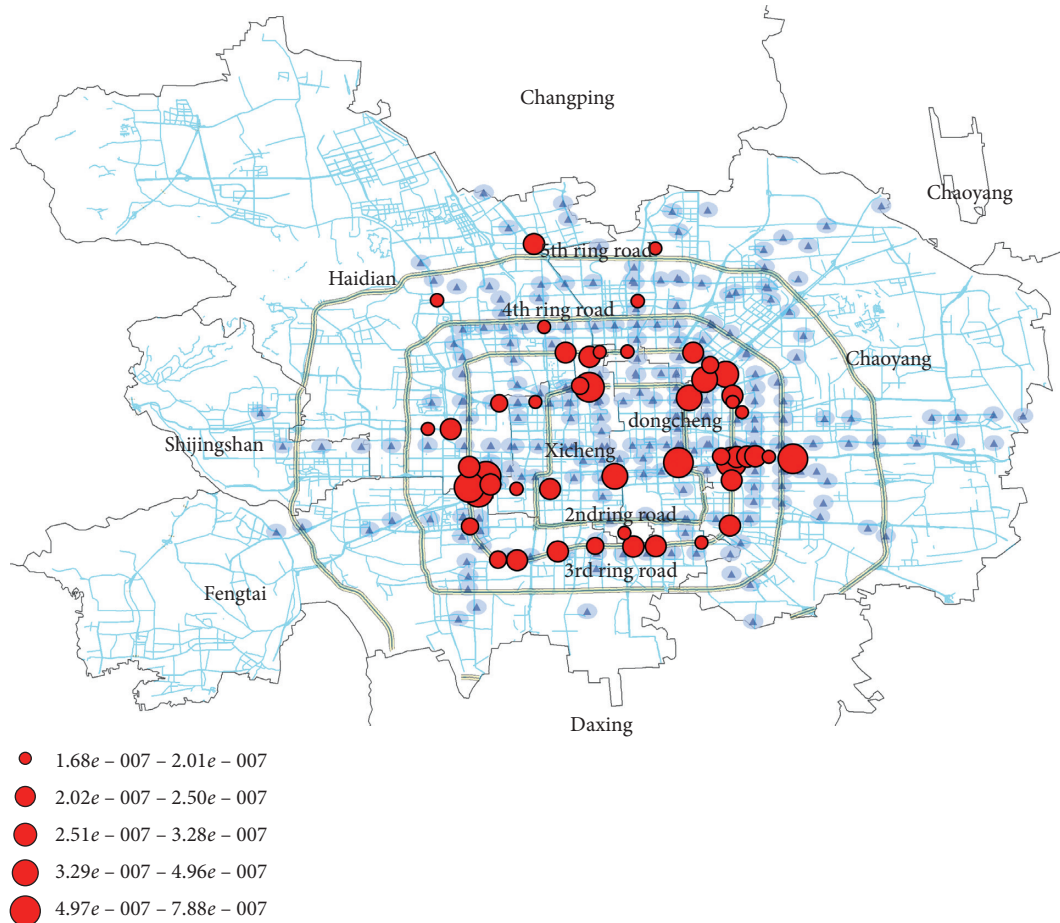


FIGURE 13: Key bus stations occupying the top 50 in off-peak (working day).

**7.2. Spatial Distribution.** The key bus stations occupying the top 50 in different periods are shown in Figures 12–14. We use the red solid circle to represent the measurement of influence. The larger the circle, the greater the influence. According to the spatial distribution of key bus stations, we discuss the underlying reason.

Most of the top 50 key bus stations are located within the 5th ring road. The number of key bus stations on the 3rd ring road is the largest, showing a trend along the loop road. Therefore, the loop roads play a significant role in the bus network of Beijing, China. The distribution of key bus stations at different periods is shown in Table 3.

Meanwhile, in three different periods, the key bus stations are ranged from the passenger transport hub such as Si-hui, Dong-zhi-men, Beijing west railway, Liu-li-qiao station, and Guo-mao. In addition, it can be seen from Figures 12–14 that there are two clustering areas in the top key stations, namely, Liu-li-qiao and Guo-mao. Liu-li-qiao station is adjacent to Beijing west railway station, which is

the main station of Beijing west railway station for passengers to transfer buses. Therefore, the Liu-li-qiao area has attracted a large number of passengers. Guo-mao station is surrounded by commercial areas, including the International Trade Center, SOHO, and many traffic attractions. There are a large number of restaurants, hotels, and business areas, which have a great demand for buses. Hence, the identification results are of great significance to public transportation in reality.

**7.3. Transfer Analysis.** During morning peak and off-peak in a working day, there are 12 key bus stations within 500 m near the subway stations. During the morning peak at weekends, there are 11 key bus stations in the vicinity of the subway stations within 500 m, indicating that the location of key bus stations is close to the subway transfer stations. Therefore, the reasonable setting of bus stations can realize the convenient connection between subway and bus transit.

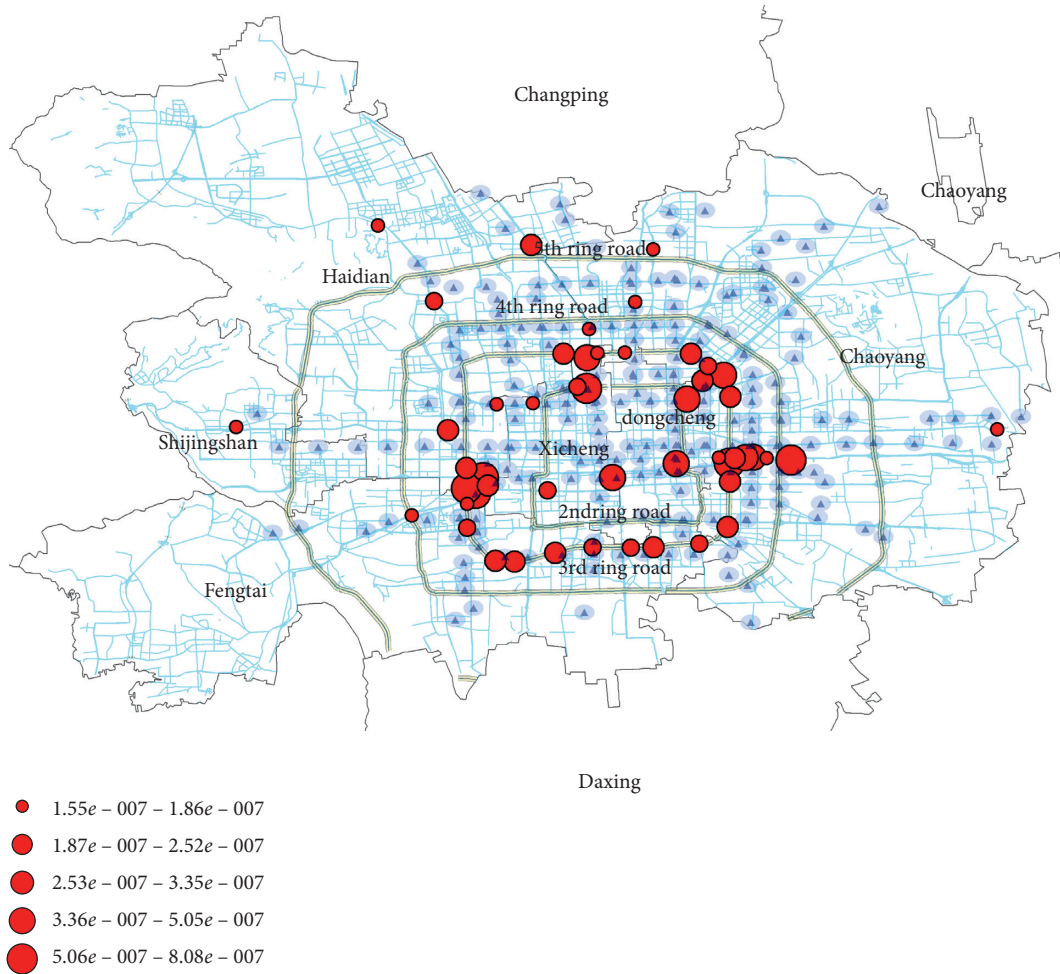


FIGURE 14: Key bus stations occupying top 50 in the morning peak (weekend).

TABLE 3: The number distribution of key bus stations.

Location	Morning peak in the working day	Off-peak in the working day	Morning peak in the weekend
Inner 2nd ring road	3	3	3
2nd ring road	2	2	2
Between 2nd and 3rd ring road	9	11	9
3rd ring road	23	21	21
Between 3rd and 4th ring road	7	8	6
4th ring road	1	0	1
Between 4th and 5th ring road	3	3	3
5th ring road	0	0	0
Between 5th and 6th ring road	2	2	5
6th ring road	0	0	0

### 8. Conclusions

Bus network is a global inefficient network that is difficult to identify key bus stations quickly and accurately with the traditional method. Based on the passenger flow in different periods and the heterogeneous topology structure of the bus network, we used a SHIP model to analyze the spatiotemporal features of key bus stations for the whole network of Beijing, China. The results of this paper can be summarized as follows:

- (i) A new index is proposed based on conventional complex network indices combining the degree and  $H$ -index to better represent the connected relationship of the bus network. From the accuracy of the identification method perspective, we constructed a test network and evaluated the performance of the identification model by two metrics. The results indicate that our method can identify the key stations in the whole bus network more accurately and efficiently when compared to other

methods such as the degree and  $H$ -index. Meanwhile, the computational complexity is greatly reduced compared with the method of betweenness.

- (ii) From the spatial and time perspective, the key bus stations have obvious distribution characteristics. The key bus stations are mostly distributed near the rail transit and passenger terminals, especially in Liu-li-qiao and Guo-mao, which indicates that the key bus stations play an important role in travel convenience. In addition, the distribution of key bus stations has a great difference during the working day and weekend.

This paper provides a useful method to find key bus stations in the bus network, which is helpful to improve the robustness of the whole bus network for planners and managers. However, the method in this study has some limitations which need to be solved. We will consider the transfer time and the number lines of the bus network in the future. Apart from this, the subway and bus network can be seen as a double-layer network, and the interlayer interaction is also a topic worth studying.

## Data Availability

The data are from the official website of Beijing Metro Rail Operation Administration Corporation Limited.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China (2016YFE-0206800 and 2017YFC-0803903).

## References

- [1] F. Nian and S. Yao, "The epidemic spreading on the multi-relationships network," *Applied Mathematics and Computation*, vol. 339, pp. 866–873, 2018.
- [2] J. L. Zhong, F. M. Zhang, and Z. X. Li, "Identification of vital nodes in complex network via belief propagation and node reinsertion," *IEEE Access*, vol. 6, pp. 29200–29210, 2018.
- [3] J. J. Wu, Z. Y. Gao, and H. J. Sun, "Urban transit system as a scale-free network," *Modern Physics Letters B*, vol. 18, pp. 1043–1049, 2004.
- [4] J. Zhang, S. Wang, and X. Wang, "Comparison analysis on vulnerability of metro networks based on complex network," *Physica A*, vol. 496, pp. 72–78, 2018.
- [5] M. Fang, S. Peng, and W. Shang, "Locating the source of asynchronous diffusion process in online social networks," *IEEE Access*, vol. 99, p. 1, 2018.
- [6] D. Bu, Y. Zhao, L. Cai, H. Xue, X. Zhu, and H. Lu, "Topological structure analysis of the protein-protein interaction network in budding yeast," *Nucleic Acids Research*, vol. 31, pp. 2443–2450, 2003.
- [7] C. V. Cannistraci, G. Alanis-Lobato, and T. Ravasi, "From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks," *Scientific Reports*, vol. 3, no. 1, pp. 1613–1616, 2013.
- [8] J. Zhang, H. Yang, H. Song, and Y. Zhang, "An improved archaeology algorithm based on integrated multi-source biological information for yeast protein interaction network," *IEEE Access*, vol. 5, pp. 15893–15900, 2017.
- [9] J. X. Liao, T. C. Liu, and M. L. Liu, "Multi-Context integrated deep neural network model for next location prediction," *IEEE Access*, vol. 99, p. 1, 2018.
- [10] J. Wu, M. Zheng, Z.-K. Zhang, W. Wang, C. Gu, and Z. Liu, "A model of spreading of sudden events on social networks," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 3, Article ID 033113, 2018.
- [11] R. J. Hada, H. Wu, and M. Jin, "Scalable minimum-cost balanced partitioning of large-scale social networks: online and offline solutions," *IEEE Transactions on Parallel & Distributed Systems*, vol. 29, pp. 1636–1649, 2017.
- [12] L. Li, S. Xu, M. Leng, S. Fang, and X. Chen, "Predicting Top- $\$L\%$  missing links: an improved local naïve bayes model," *IEEE Access*, vol. 7, pp. 57868–57880, 2019.
- [13] W. Li and X. Cai, "Empirical analysis of a scale-free railway network in China," *Physica A*, vol. 382, pp. 693–703, 2007.
- [14] R. Mesa-Arango and I. Kumar, "Hierarchical value chains encompassing freight transportation and logistics sectors in the United States: network analysis approach," *Transportation Res. Rec.* vol. 2609, pp. 1–10, 2017.
- [15] D. Gattuso and E. Miriello, "Compared analysis of metro networks supported by graph theory," *Networks and Spatial Economics*, vol. 5, pp. 395–414, 2005.
- [16] P. Y. Zhang, H. P. Yao, and Y. J. Liu, "Virtual network embedding based on the degree and clustering coefficient information," *IEEE Access*, vol. 4, pp. 8572–8580, 2016.
- [17] M. Riondato and E. Upfal, "Abra," *ACM Transactions on Knowledge Discovery from Data*, vol. 12, no. 5, pp. 1–38, 2018.
- [18] Z. J. Sun, B. Wang, J. F. Sheng, and Y. X. Hu, "Identifying influential nodes in complex networks based on weighted formal concept analysis," *IEEE Access*, vol. 5, pp. 3777–3789, 2017.
- [19] J. Sienkiewicz and J. A. Hołyst, "Statistical analysis of 22 public transport networks in Poland," *Physical Review E*, vol. 72, no. 4, Article ID 046127, 11 pages, 2005.
- [20] L. Ping L, X. Xing X, L. Q. Zhong, Q. Y. Gang, and S. Xing, "Topological properties of urban public traffic networks in Chinese top-ten biggest cities," *Chinese Physics Letters*, vol. 23, pp. 3384–3387, 2006.
- [21] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. Watts, "Network robustness and fragility: percolation on random graphs," *Physical Review Letters*, vol. 85, pp. 5468–5471, 2000.
- [22] J. L. Zhong, F. M. Zhang, and Z. X. Li, "Identification of vital nodes in complex network via belief propagation and node reinsertion," *IEEE Access*, vol. 6, pp. 29200–29210, 2018.
- [23] V. Jovica and W. Zhu, "Modeling of interconnected critical infrastructure systems using complex network theory," *IEEE T Smart Grid*, vol. 9, pp. 4637–4648, 2018.
- [24] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, "Identifying, understanding, and analyzing critical infrastructure interdependencies," *IEEE Control Systems*, vol. 21, pp. 11–25, 2001.
- [25] J. Johansson and H. Hassel, "An approach for modelling interdependent infrastructures in the context of vulnerability analysis," *Reliability Engineering & System Safety*, vol. 95, pp. 1335–1344, 2010.
- [26] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network robustness and fragility: percolation on

- random graphs,” *Physical Review Letters*, vol. 85, pp. 5468–5471, 2000.
- [27] B. Ouyang, X. Jin, Y. Xia, and L. Jiang, “Change of network load due to node removal,” *The European Physical Journal B*, vol. 87, no. 3, pp. 40723–40729, 2014.
- [28] L. Dueñas-Osorio and S. M. Vemuru, “Cascading failures in complex infrastructure systems,” *Structural Safety*, vol. 31, pp. 157–167, 2009.
- [29] X. Chen, “System vulnerability assessment and critical nodes identification,” *Expert System Application*, vol. 65, pp. 212–220, 2016.
- [30] T. Shanmukhappa, I. W. H. Ho, and C. K. Tse, “Multi-layer public transport network analysis,” in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, Daegu, South Korea, May. 2018.
- [31] Y. Sun, L. Mburu, and S. Wang, “Analysis of community properties and node properties to understand the structure of the bus transport network,” *Physica A*, vol. 450, pp. 523–530, 2016.
- [32] J. Wang, S. Wang, Z. Zhang, K. Zou, and Z. Shu, “Characteristics on hub networks of urban rail transit networks,” *Physica A*, vol. 447, pp. 502–507, 2016.
- [33] X. Xu, J. Hu, F. Liu, and L. Liu, “Scaling and correlations in three bus-transport networks of China,” *Physica A*, vol. 374, pp. 441–448, 2007.
- [34] B. B. Fu, L. Zhang, S. B. Li, and Y. X. Li, “Survivability of public transit network based on network structure entropy,” *International Journal of Modern Physics C*, vol. 26, pp. 1–11, 2015.
- [35] J. Allen, J. C. Muñoz, and J. Rosell, “Effect of a major network reform on bus transit satisfaction,” *Transportation Research Part A: Policy and Practice*, vol. 124, pp. 310–333, 2019.
- [36] J. Burchard and B. Cornwell, “Structural holes and bridging in two-mode networks,” *Social Networks*, vol. 55, pp. 11–20, 2018.
- [37] Q. Wang, J. D. Ren, Y. Wang, and B. Zhang, “A clustering degree based influential spreader identification algorithm in weighted complex network,” *IEEE Access*, vol. 99, p. 1, 2018.
- [38] M. Eidsaa and E. Almaas, “Investigating the relationship between  $k$ -core and  $s$ -core network decompositions,” *Physica A*, vol. 449, pp. 111–125, 2016.
- [39] A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks,” *Proceedings of the National Academy of Sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.
- [40] Y. Bai, Z. Sun, B. Zeng, J. Deng, and C. Li, “A multi-pattern deep fusion model for short-term bus passenger flow forecasting,” *Applied Soft Computing*, vol. 58, pp. 669–680, 2017.
- [41] J. E. Hirsch, “An index to quantify an individual’s scientific research output,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16569–16572, 2005.
- [42] T. Braun, W. Glänzel, and A. Schubert, “A hirsch-type index for journals,” *Scientometrics*, vol. 69, no. 1, pp. 169–173, 2006.
- [43] J. E. Hirsch, “Does the h index have predictive power?” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 19193–19198, 2007.
- [44] L. Lü, Q.-M. Zhang, and H. E. Stanley, “The  $H$ -index of a network node and its relation to degree and coreness,” *Nature Communications*, vol. 7, no. 1, pp. 10168–10175, 2016.
- [45] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, “Public transport networks: empirical analysis and modeling,” *The European Physical Journal B*, vol. 68, no. 2, pp. 261–275, 2009.
- [46] J. Sienkiewicz and J. A. Hołyst, “Statistical analysis of 22 public transport networks in Poland,” *Physical Review E*, vol. 72, no. 4, pp. 046127–046138, 2005.
- [47] 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.
- [48] T. Shanmukhappa, I. W.-H. Ho, and C. K. Tse, “Spatial analysis of bus transport networks using network theory,” *Physica A: Statistical Mechanics and Its Applications*, vol. 502, pp. 295–314, 2018.