

SHIWEI SUN, M.Sc.<sup>1,2</sup>  
E-mail: 14120881@bjtu.edu.cn  
HAIYING LI, Ph.D.<sup>1</sup>  
E-mail: hyl@bjtu.edu.cn  
XINYUE XU, Ph.D.<sup>1</sup>  
(Corresponding author)  
E-mail: xxy@bjtu.edu.cn

<sup>1</sup> State Key Laboratory of Rail Traffic Control and Safety  
Beijing Jiaotong University, Beijing 100044, China  
<sup>2</sup> School of Traffic and Transportation  
Beijing Jiaotong University, Beijing 100044, China

Traffic Management  
Original Scientific Paper  
Submitted: 12 Apr. 2016  
Accepted: 16 Nov. 2016

# A KEY STATION IDENTIFICATION METHOD FOR URBAN RAIL TRANSIT: A CASE STUDY OF BEIJING SUBWAY

## ABSTRACT

*Congestion occurs and propagates in the stations of urban rail transit, which results in the impendent need to comprehensively evaluate the station performance. Based on complex network theory, a key station identification method is considered. This approach considers both the topology and dynamic operation states of urban rail transit network, such as degree, passenger demand, system capacity and capacity utilization. A case of Beijing urban rail transit is applied to verify the validation of the proposed method. It shows that the method can be helpful to daily passenger flow control and capacity enhancement during peak hours.*

## KEY WORDS

*complex network theory; key station identification; urban rail transit;*

## 1. INTRODUCTION

With the rapid development of urban rail transit network operation technique, the increase of passenger volume leads to congestion in stations [1], causing even safety risk in the network [2]. Besides, the topology of urban rail transit network can also aggravate or relief the congestion situation of subway networks. Therefore, an effective method identified the key station is an imperative to manage the overcrowded urban rail transit.

For the factors of the station performance, many of the current approaches focus on the relationship between passenger flow and facility capacity. Xu et al. [3] combined queuing theory with facility utilization evaluation to conduct the station capacity evaluation, according to the gathering and scattering process of a station. Kepaptsoglou and Karlaftis [4] proposed a model to evaluate passenger evolution on subway station platforms, which usually gathered most passengers. Wang et al. [5] also used the passenger

evacuation capacity to evaluate the station based on the facility utilization. If the service level of the facility is too low, there are big problems in the station.

Besides, passengers' subjective feelings and other factors are also used for evaluation. Matulin et al. [6] proposed a two-level evaluation method to analyze public transport performances from micro and macro levels, using performance indicators and quality of service indicators. Kesten and Öğüt [7] developed a measurement framework of passenger-oriented performance to evaluate rail transportation, using satisfaction and importance ratings obtained from passenger surveys.

Meanwhile, there are many approaches employed in the station evaluation. Some typical methods such as Data Envelopment Analysis (DEA) have played their great role in the station assessment. Sancha et al. [8] used the method to estimate the technical, social and environmental efficiency of transit transfer stations. Pjevčević et al. [9] used DEA method to measure and analyze the efficiencies of ports on the Danube River, which can be borrowed for station evaluation. In addition, simulation is also another method, where relevant simulation or operation data are provided. Zhang et al. [10] proposed a cellular automata-based model to simulate passengers' alighting and boarding behaviours in metro stations, aiming to analyze and evaluate the operation performance of stations. However, most works above do not consider the influence of other stations in the subway system, which results in local rather than global conclusions.

Fortunately, in recent years, a complex network theory has been introduced to the station evaluation from a global viewpoint. Indexes such as degree centrality, betweenness centrality, and closeness centrality have an outstanding ability to identify key nodes from the network [11–14]. Sheikahmadi et al. [12] proposed an improved high degree centrality index to identify the

most influential nodes, which outperformed other indexes on large-scale networks in terms of maximizing the spread of influence with acceptable running time. Yang et al. [13] developed a new weighted composite index to find out the hub stations in a subway network, with the combination of degree and betweenness. Nevertheless, these studies are limited to the topology property, while the property of the operational state is neglected. Moreover, the lack of analyses on operation conditions would considerably influence the station evaluation, which would be raised in the current study.

The aim of this paper is to present a new method to identify key stations of urban rail transit networks. Not only the network topology, such as degree, betweenness, and closeness centrality, but also dynamic operational parameters, such as passenger volume and load factor are considered in this method for the first time. With strong influence on nearby stations, high passenger flow loads, and congestion propagation along the network, these key stations should take measures to keep the subway system safety.

The rest of the paper is organized as follows. Section 2 introduces three centralization indexes of the complex network. Section 3 proposes indexes of key station identification and gives the evaluation method and data sources. In Section 4, a case of Beijing urban rail transit illustrates the application of the proposed approach. Section 5 gives the conclusions.

## 2. COMPLEX NETWORK

This section introduces the background and several common centralization indexes of the complex network. It is the base of the method of key stations identification mentioned in Section 3.

### 2.1 Background

Most social, biological, and technological networks display substantial non-trivial topological features, which makes complex network theory arise to discover the character of the network. A complex network can be modelled as a directed or undirected network (or graph) consisting of nodes and links. Not all nodes (stations) are equivalent in the network [13], which are ranked based on the topology of the network and the location of each node. Especially, many centrality indexes have been proposed to rank nodes in the networks. Those with the highest/lowest indexes are taken as influential nodes of the network.

There are well-known indexes that mostly deal with the location of nodes in the network, which have strong ability to find the key nodes in the network. A simple one is the degree centrality, considering only very limited information. Another group of indexes considering the global information gives better ranking

results, such as betweenness centrality and closeness centrality [11].

### 2.2 Centralization indexes

#### Degree centrality

Degree centrality describes the direct influence of a node by the number of connections with other nodes.

For a directed network, the degree centrality of node  $i$  ( $C_d(i)$ ) is defined as:

$$C_d(i) = d(i)^{in} + d(i)^{out} \tag{1}$$

where  $d(i)^{in}$  is in-degree denoted by the number of connections from other nodes to node  $i$ ;  $d(i)^{out}$  is out-degree denoted by the number of connections from node  $i$  to other nodes.

A node with a larger degree is likely to have higher influence than a node with a smaller degree. This index is a straightforward and efficient metric; however, in some cases, it fails to identify important nodes since it considers only very limited information.

#### Betweenness centrality

Betweenness centrality, an index of the influence of a node over the information spread through the network, is defined as the fraction of shortest paths between node pairs that pass through the node of interest.

In a network, the betweenness centrality of node  $i$  ( $C_b(i)$ ) is formulated as follows:

$$C_b(i) = \frac{\sum_{j \neq k} g_{jk}(i)}{\sum_{j \neq k} g_{jk}} \tag{2}$$

where  $g_{jk}$  is the number of shortest paths from node  $j$  to node  $k$ ;  $g_{jk}(i)$  is the number of shortest paths, which pass through node  $i$  from node  $j$  to node  $k$ .

According to this indicator, influential nodes are those that are visited by the largest number of shortest paths from all nodes to all others within the network.

#### Closeness centrality

Closeness centrality, an index indicating how long it will take to spread the information from a node to all other nodes, is calculated by the average distance from the evaluation node to all other nodes in the networks.

In a network with  $n$  nodes, the closeness centrality of node  $i$  is defined as:

$$C_c(i) = \frac{\sum_{j=1, i \neq j}^n d_{ij}}{n-1} \tag{3}$$

where  $d_{ij}$  is the shortest distance from node  $i$  to node  $j$ .

A node is considered as an influential (key) node by the closeness strategy if its total distance to all other nodes is the lowest.

### 3. KEY STATION IDENTIFICATION METHOD

This section combines the above-mentioned centralization indexes with the dynamic operational state to identify key stations in urban rail transit network. In addition, the evaluation procedure and data sources are given in details.

#### 3.1 Notation

In this study, stations are regarded as nodes, and sections are denoted by links, similarly described in [15]. Notations of the urban rail network are as follows:

$V=\{i | i=1,2,\dots,N\}$  - the set of nodes, where  $N$  refers to the number of nodes in the network;

$E=\{e_{ij} | i,j=1,2,\dots,N, i \neq j\}$  - the set of edges, where  $e_{ij}=(i,j)$  refers to the edge from node  $i$  to its adjacent node  $j$ ;

$O$  - the set of origin nodes,  $D$  is the set of destination nodes, and  $O, D \subseteq V$ ;

$R$  - the set of all feasible paths in the network;

$SR$  - the set of all the shortest paths in the network;

$R_{od}$  - the set of feasible paths from node  $o, o \in O$  to node  $d, d \in O$ , and  $R = \cup R_{od}$ ;

$SR_{od}$  - the set of the shortest paths from node  $o$  to node  $d$ ,  $|SR_{od}|$  is the number of elements in the set, and  $SR = \cup SR_{od}$ ;

$r_{od}$  - a feasible path from node  $o$  to node  $d$ ;

$sr_{od}$  - one shortest path of  $SR_{od}$ .

#### 3.2 Key station identification index

##### Throughput Degree (TD)

The degree centrality of the station only reflects its importance in the network structure, but it does not consider the passenger volume. To overcome the above shortcoming, a new index TD, which integrates the passenger volume and degree centrality, is proposed. The TD of station  $i$  is defined as follows:

$$D_i = \sum_j (V_{ij} + V_{ji}) \quad (4)$$

where  $V_{ij}$  is the passenger volume of section  $e_{ij}$ , and  $V_{ji}$  is the passenger volume of section  $e_{ji}$ . These parameters can be obtained by the Automatic Fare Collection (AFC) data.

Not only connecting sections from the complex network perspective, but also the served passengers from the operation view, are considered in the TD. It is used to identify stations that have a strong local influence on nearby stations.

##### Capacity Utilization Betweenness (CUB)

According to the betweenness concept in Section 2, the CUB of station  $i$  is denoted as follows:

$$B_i = \frac{\sum_{(o,d) \in \{O,D\}} \sum_{sr_{od} \in SR_{od}} [u(sr_{od}) \cdot \varphi_i(sr_{od})]}{\sum_{(o,d) \in \{O,D\}} \sum_{sr_{od} \in SR_{od}} u(sr_{od})} \quad (5)$$

where a certain shortest path  $sr_{od}$  goes through station  $i$ , then  $\varphi_i(sr_{od})=1$ , otherwise  $\varphi_i(sr_{od})=0$ ;  $u(sr_{od})$  is the capacity function of  $sr_{od}$ , which will be defined by the following Formula 6.

Generally, the path with least sections is considered as the shortest path for an O-D pair. However, in subways, highly utilized, with large load factors should be also considered in the shortest paths. So, a new identification method of the shortest paths is proposed:

$$SR_{od} = \left\{ r_{od} \mid \max_{e \in r_{od}} \frac{\max(\alpha_e \cdot \beta_e)}{|r_{od}|} \right\},$$

$$u(sr_{od}) = \frac{\max_{e \in sr_{od}} (\alpha_e \cdot \beta_e)}{|sr_{od}|} \quad (6)$$

where  $e$  is a section in path  $sr_{od}$ ,  $\alpha_e$  is the real number of trains through the section,  $\beta_e$  is average section load factor in a certain period,  $|r_{od}|$ ,  $|sr_{od}|$  is the actual topology length of path  $r_{od}$ ,  $sr_{od}$ , respectively, that can be calculated by the section length,  $\max_{e \in r_{od}} (\alpha_e \cdot \beta_e)$  is the maximum capacity utilization of this path.

CUB is designed for identification of stations that have a broader range of influence. Note that if a station has large CUB value in an emergent situation, a huge passenger flow will widely propagate in the network. To reduce the range of crowded stations effectively, these stations with high values of CUB should be considered first.

##### Connection Closeness (CC)

According to the Closeness concept in Section 2, the CC of station  $i$  is denoted as follows:

$$C_i = \frac{\sum_{d \in D, d \neq i} \sum_{sr_{id} \in SR_{id}} c(sr_{id})}{n-1} \quad (7)$$

where  $c(sr_{id})$  is the spare capacity function of the shortest path  $sr_{id}$ , and its value will be defined by Formula 8.

Analogously, not only the distance between them but also the spare capacity of the path is considered in the process of finding the shortest path. Unlike the similar set mentioned in the previous subsection, the set of the shortest paths and their function length for CC is given as follows:

$$SR_{od} = \left\{ r_{od} \mid \max_{e \in r_{od}} \frac{\min[\alpha_e \cdot (\beta_{\max} - \beta_e)]}{|r_{od}|} \right\},$$

$$c(sr_{id}) = \frac{\min_{e \in sr_{id}} [\alpha_e \cdot (\beta_{\max} - \beta_e)]}{|sr_{id}|} \quad (8)$$

where  $\beta_{\max}$  is the largest load factor of all sections in the network for the evaluation period,

$\min_{e \in r_{od}} [\alpha_e \cdot (\beta_{max} - \beta_e)]$  is the remaining capacity of path  $r_{od}$ .

CC is used to identify stations that have a strong spreading ability. If CC value is larger in one station, passengers of this station will reach other stations more easily, which means that congestion at this station will spread faster.

### 3.3 Evaluation production

The evaluation is conducted by the following steps: *Step 1: Data preparation.* Network topology, the set of feasible paths, the set of the shortest paths, section length, section load factor, section passenger volume, and actual train diagram are included. The relationship between data and indexes is shown in *Figure 1*.

*Step 2: Index calculation.* TD, CUB, and CC are calculated by the above-mentioned approaches in section 3.2, respectively.

*Step 3: Key station identification.* The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is applied to make a comprehensive evaluation [16]. If a station ranks high by this method, the station will be the key station.

## 4. CASE STUDY

### 4.1 System description

The Beijing subway has a total of 18 lines, 279 stations, 41 transfer stations and a mileage of more than 544 km by 2014 (*Figure 2*). The number of stations with routine passenger flow control amounted to 65, including 58 stations in the morning and 19 stations in the evening peak hours. So, it is urgent to identify

the key stations to improve the operation quality of the Beijing subway.

The actual operational data of March 3, 2014, are collected to conduct this evaluation, especially the data from between 7:00 and 9:00. The details of data sources are shown in *Table 1*:

Table 1 – Sources of data

Data	Source
Section load factor Section passenger volume	Simulation/ Data statistics platforms
Network topology Section length Actual train diagram The set of feasible paths	Operation data
The set of the shortest paths	Calculation platforms

### 4.2 Key station identification

First, the results of each station are calculated respectively by each index, and then a comprehensive evaluation is made. Finally, the set of key stations is achieved.

#### Throughput Degree (TD)

The TD result is shown in *Figure 2*. Note that the top 20 stations are marked in black in the descending order, while stations ranked from 21<sup>st</sup> to 40 are marked in dark gray. The following facts can be seen from this result:

- 1) Most large TD stations are located in the circular region of Line 10, including 38 in the top 40 stations with large TD. Both many commuters and the

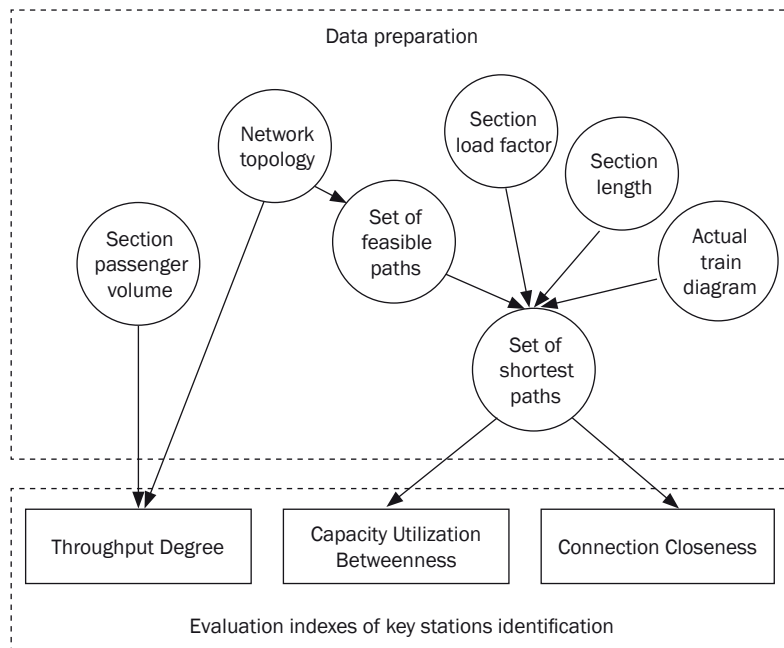


Figure 1 – Relationship between data and indexes

complex topology result in large TD values of these stations in the north and east of this area.

- 2) The top 20 stations with large TD are transfer stations except for Yong'anli. Furthermore, non-transfer stations with large TD are usually located near these transfer stations. These mean that the high connection of transfer stations are usually along huge passenger flows [17], and have strong influences on nearby stations.

**Capacity Utilization Betweenness (CUB) and Connection Closeness (CC)**

The evaluation results of CUB and CC are shown in Figure 3, where the following facts can be concluded:

- 1) There are 15 transfer stations in the top 20 stations with high CUB as shown in Figure 3a. The non-transfer stations between two of these transfer stations also have a similar CUB because they have

the same opportunity to be passed through by the shortest paths with these transfer stations.

- 2) The top 20 stations with high CC values are located in the circular region of Line 2, which is the central part of the Beijing subway. Further, two groups of stations are formed as shown in the circle areas of Figure 3b.

**Comprehensive evaluation**

According to these three indexes, the key stations are located in the central part of the network although the station set, the rank of stations, and the range of distribution are not the same. To balance these different indexes, an integrated result is achieved by TOPSIS, denoted as the Comprehensive importance of stations (CI-S). The top 10 stations with high CI-S are shown in Figure 4.

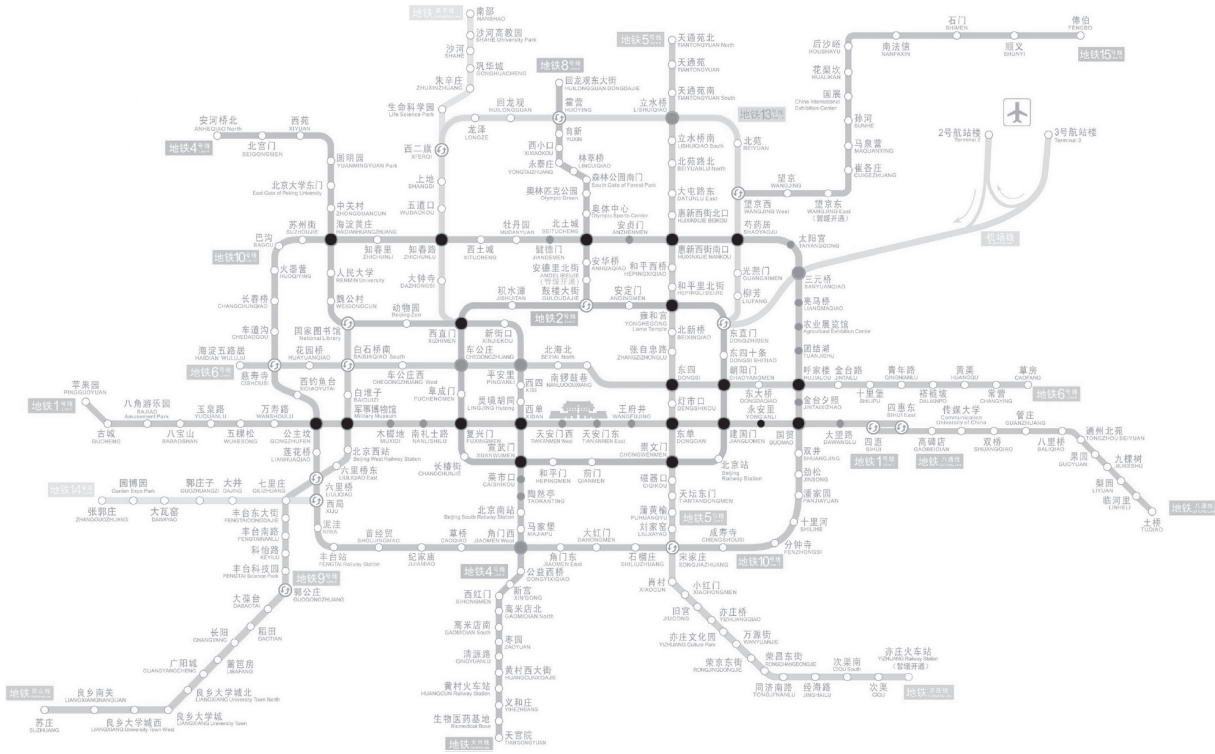
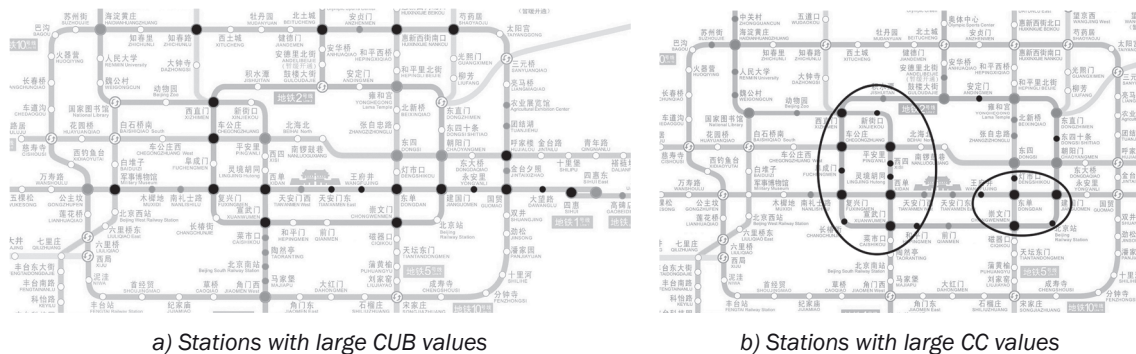


Figure 2 – Stations with large TD values



a) Stations with large CUB values

b) Stations with large CC values

Figure 3 – Stations with large CUB and CC values

Table 2 – Comparison of top 10 stations with CTI-N

Rank	CTI-N ([18])	CI-S	TD	CUB	CC
1	Guomao	Guomao	Guomao	Guomao	Guomao
2	Xizhimen	Jianguomen	Jianguomen	Jianguomen	Xizhimen
3	Huixinxijie Nankou	Dongdan	Hujialou	Dongdan	Dongdan
4	Hujialou	Xidan	Huixinxijie Nankou	Xidan	Fuxingmen
5	Zhichunlu	Fuxingmen	Dongdan	Fuxingmen	Xinjiekou
6	Xidan	Xizhimen	Xidan	Xuanwumen	Xidan
7	Shaoyaoju	Xuanwumen	Xizhimen	Huixinxijie Nankou	Fuchengmen
8	Jianguomen	Huixinxijie Nankou	Fuxingmen	Xizhimen	Changchunjie
9	Dongdan	Hujialou	Zhichunlu	Hujialou	Xuanwumen
10	Fuxingmen	Zhichunlu	Xuanwumen	Yong'anli	Jianguomen

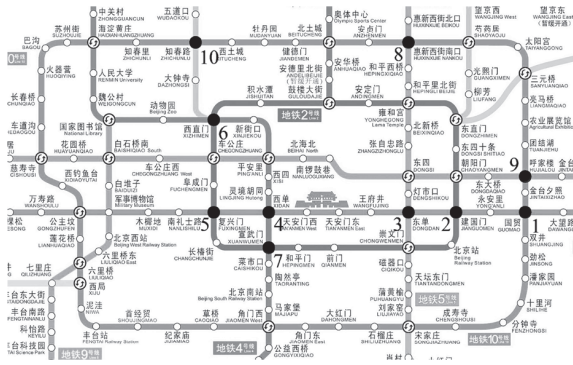


Figure 4 – Top 10 stations of CI-S

### 4.3 Comparison with existing research

One previous work of the key station identification made by Xiao [18] has been applied to compare the proposed method. In that work, structure degree, structure betweenness, and flow degree of nodes are chosen to construct the index comprehensive transportation importance of nodes (CTI-N). Comparison of top 10 stations is shown in Table 2, and some conclusions are drawn as follows:

- 1) Most key stations identified by CI-S are similar to those identified by CTI-N, which shows the accuracy of the proposed indexes.
- 2) CTI-N and CI-S can only rank stations from a holistic perspective, but they cannot tell which aspects are specifically in trouble. TD, CUB, and CC can identify stations in a specific view, offering a targeted guidance to adjustment in the vital part of networks. Moreover, some additional non-transfer

stations are considered as the key stations, which is consistent with the reality.

## 5. CONCLUSION

The main conclusions of this paper are given as follows:

- 1) Three evaluation indexes are constructed to identify the key stations from the urban rail transit, which include properties of both the network topology and operational state. Moreover, an evaluation method is proposed to easily find out the key stations. A case study of Beijing subway illustrates the identification of the key stations by the proposed method.
- 2) The key station identification method for the urban rail transit developed in this study provides a theoretical account and empirical basis to evaluate the network topology and operation state, and the method can also be used by an independent aspect for operation purposes. In addition, the identification result would be helpful to control the key station for improving the performance. Especially, which and how many passengers of stations should be determined during the morning peak hours, which is the hot topic of these days in the field of urban rail transit [3],[19-20].

## ACKNOWLEDGEMENT

This work has been financially supported by the State Key Lab of Rail Traffic Control and Safety, China (RCS2014ZTY5, RCS2015ZZ002), the Beijing Natural Science Foundation (No.9164033), the Beijing Municipal Science and Technology Project (Z141106004414106), and the National Natural Science Foundation of China (No.71601018). The authors acknowledge with gratitude the suggestions by Professor Liu, Ph.D. Candidate Qin, and anonymous

referees that helped to improve the presentation of the paper.

孙世炜, 硕士研究生<sup>1,2</sup>

邮箱: 14120881@bjtu.edu.cn

李海鹰, 教授, 博士<sup>1</sup>

邮箱: hyl@bjtu.edu.cn

许心越, 讲师, 博士<sup>1</sup>

(通讯作者)

邮箱: xxy@bjtu.edu.cn

<sup>1</sup> 北京交通大学 轨道交通控制与安全国家重点实验室

北京, 中国 100044

<sup>2</sup> 北京交通大学 交通运输学院

北京, 中国 100044

城市轨道交通关键车站识别方法—以北京地铁为例

摘要

目前, 城市轨道交通枢纽客流拥挤情况非常严重, 并在路网中传播, 这使得对车站评价的需求十分紧迫。本文提出了一种基于复杂网络理论的关键车站识别方法, 综合考虑了路网静态拓扑结构和轨道交通动态运营属性等不同因素, 形成了包括复杂网络度等中心化和客流需求、路网能力和能力利用率等动态运营评价指标的车站识别指标; 并以北京市城市轨道交通为例验证本文提出方法的有效性。结果表明: 该方法为高峰小时的客流控制和能力提高指明了方向, 具有很强的应用价值。

关键词

复杂网络理论; 关键车站识别; 城市轨道交通

## REFERENCES

- [1] Zhang M, Du S. Transfer coordination optimization for network operation of urban rail transit based on hierarchical preference. *Journal of the China Railway Society*. 2009;31(6):9-14.
- [2] Lam WHK, Cheung C-Y, Lam CF. A study of crowding effects at the Hong Kong light rail transit stations. *Transportation Research Part A: Policy and Practice*. 1999;33(5):401-15.
- [3] Xu X, Liu J, Li H, Hu J-Q. Analysis of subway station capacity with the use of queueing theory. *Transportation Research Part C: Emerging Technologies*. 2014;38:28-43.
- [4] Kepaptsoglou K, Karlaftis MG. A model for analyzing metro station platform conditions following a service disruption. In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC; 2010*. p. 1789-94.
- [5] Wang Z, Chen F, Li X. Comparative analysis and pedestrian simulation evaluation on emergency evacuation test methods for urban rail transit stations. *Promet-Traffic & Transportation*. 2012;24(6):535-42.
- [6] Matulin M, Mrvelj Š, Jelušić N. Two-level Evaluation of Public Transport Performances. *Promet – Traffic & Transportation*. 2011;23(5):329-39.
- [7] Kesten AS, Ögüt KS. A New Passenger-Oriented Performance Measurement Framework for Public Rail Transportation Systems. *Promet – Traffic & Transportation*. 2014;26(4):299-311.
- [8] Román-De la Sancha A, Mayoral JM, Román LI. Modeling Urban Transfer Stations Efficiency. *Procedia Computer Science*. 2016;83:18-25.
- [9] Pjevčević D, Radonjić A, Hrle Z, Čolić V. DEA Window Analysis for Measuring Port Efficiencies in Serbia. *Promet – Traffic & Transportation*. 2012;24(1):63-72.
- [10] Zhang Q, Han B, Li D. Modeling and simulation of passenger alighting and boarding movement in Beijing metro stations. *Transportation Research Part C: Emerging Technologies*. 2008;16(5):635-49.
- [11] Chen D, Lu L, Shang M-S, Zhang Y-C, Zhou T. Identifying influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*. 2012;391(4):1777-87.
- [12] Sheikhahmadi A, Nematbakhsh MA, Shokrollahi A. Improving detection of influential nodes in complex networks. *Physica A: Statistical Mechanics and its Applications*. 2015;436:833-45.
- [13] Yang Y, Liu Y, Zhou M, Li F, Sun C. Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing Subway. *Safety Science*. 2015;79:149-62.
- [14] Murali P, Ordóñez F, Dessouky MM. Modeling strategies for effectively routing freight trains through complex networks. *Transportation Research Part C: Emerging Technologies*. 2016;70:197-213.
- [15] Miler M, Medak D, Odobašić D. The shortest path algorithm performance comparison in graph and relational database on a transportation network. *Promet – Traffic & Transportation*. 2014;26(1):75-82.
- [16] Zhang H, Gu C, Gu L, Zhang Y. The evaluation of tourism destination competitiveness by TOPSIS & information entropy - A case in the Yangtze River Delta of China. *Tourism Management*. 2011;32(2):443-51.
- [17] Jacura M, Týfa L. Utilisation of Decision Tables for Proposal of Transfer Node Conception. *Promet – Traffic & Transportation*. 2012;24(5):425-31.
- [18] Xiao X. Risk and safety assessment of urban rail transit networks operation – a topology based approach. Beijing Jiaotong University; 2014.
- [19] Xu X, Liu J, Li H, Jiang M. Capacity-oriented passenger flow control under uncertain demand: Algorithm development and real-world case study. *Transportation Research Part E: Logistics and Transportation Review*. 2016;87:130-48.
- [20] Liu J, Zhou X. Capacitated transit service network design with boundedly rational agents. *Transportation Research Part B: Methodological*. 2016;93:225-50.