Ruiyong TONG, Ph.D. candidate[1]
E-mail: 20114047@bjtu.edu.cn
Qi XU, Ph.D.[1,2]
(Corresponding author)
E-mail: xuqi@bjtu.edu.cn
Runbin WEI, Ph.D. candidate[1]
E-mail: 19114056@bjtu.edu.cn
Junsheng HUANG, Ph.D. candidate[1]
E-mail: 19114037@bjtu.edu.cn
Zhongsheng XIAO, Ph.D. candidate[1]
E-mail: 18114047@bjtu.edu.cn
[1] School of Traffic and Transportation
Beijing Jiaotong University
No.3 ShangYuanCun, Haidian District,
Beijing 100044, China
[2] Key Laboratory of Transport Industry of Big Data
Application Technologies for Comprehensive Transport
Beijing Jiaotong University
No.3 ShangYuanCun, Haidian District,
Beijing 100044, China

Information and Communication Technology
Original Scientific Paper
Submitted: 29 Aug. 2021
Accepted: 21 Feb. 2022

WEIGHTED COMPLEX NETWORK ANALYSIS OF THE DIFFERENCE BETWEEN NODAL CENTRALITIES OF THE BEIJING SUBWAY SYSTEM

ABSTRACT

The centrality of stations is one of the most important issues in urban transit systems. The central stations of such networks have often been identified using network topological centrality measures. In real networks, passenger flows arise from an interplay between the dynamics of the individual person movements and the underlying physical structure. In this paper, we apply a two-layered model to identify the most central stations in the Beijing Subway System, in which the lower layer is the physical infrastructure and the upper layer represents the passenger flows. We compare various centrality indicators such as degree, strength and betweenness centrality for the two-layered model. To represent the influence of exogenous factors of stations on the subway system, we reference the alpha centrality. The results show that the central stations in the geographic system in terms of the betweenness are not consistent with the central stations in the network of the flows in terms of the alpha centrality. We clarify this difference by comparing the two centrality measures with the real load, indicating that the alpha centrality approximates the real load better than the betweenness, as it can capture the direction and volume of the flows along links and the flows into and out of the systems. The empirical findings can give us some useful insights into the node centrality of subway systems.

KEYWORDS

node centrality; betweenness; alpha centrality; subway system; passenger flow.

1. INTRODUCTION

Network characteristics of transportation present great opportunities, which can be analysed from the viewpoint of network science [1–4]. Meanwhile, infrastructure networks have been extensively studied and many valuable insights have also been obtained in recent years. Graphs derived from such networks were analysed on the examples of road networks [5–6], inter-urban transit networks [7–8], airline networks [9–10] and urban transit networks [11–12]. Nevertheless, subway systems present smaller-sized networks compared to most studied transportation networks, and they exhibit a rich variety of fascinating network properties, which include scale-free structures [13–14], small-world properties [15–16], network efficiency [17–18], network vulnerability [19–20] and network growth patterns [21–22]. In this paper, we focus on one of the most important features of subway systems: node centrality.

The identification of the most central nodes in the complex systems is an important issue in network characterisation [23–25]. Some scholars have also conducted research on the network centrality of urban transit networks. Derrible [26] applied the notion of betweenness centrality to 28 worldwide subway systems based on a simplified network representation and found that betweenness centrality becomes more evenly distributed with network size, which allows the systems to distribute the flow of passengers more evenly. Tang [27] constructed the subway and bus networks by L-space and P-space methods, respectively, and proposed an identification method in Multiplex Network based on Dempster–Shafer evidence theory (MNDS). On the other hand, the use of network centrality is mostly served as a basis to study matters of vulnerability [19–20], to measure the traffic flows as a load estimator [28–29], to study the relationship between urban structures and land use [30], and to explore the match between the subway system and the functional zoning of the city [31].

However, the traditional centrality measures, such as degree centrality [32] and betweenness centrality [33], based exclusively on depicting configurations of their physical infrastructure networks have strong limitations, as they do not take into account the real-life traffic flow patterns in the network. This may provide an incomplete view of network functionalities because transportation of passengers or goods is the ultimate goal of every transportation system. In fact, the discussion about whether traffic flows can be approximated by network properties in urban street networks has lasted for decades among urban planning researchers [34–35]. Therefore, some studies on urban networks have focused on the “gap” between the centrality measure and the actual flow [36–37]. The most important nodes and links from a topological point of view may not necessarily carry the most traffic flows [38]. In other words, the geographical patterns formed by traffic flows and betweenness differ substantially in the transit networks [28].

With the rapid development of information technology, large amounts of data over time and space are available to model urban dynamics. In particular, monitoring, recognising and analysing human mobility patterns is becoming a hot issue in transport and urban planning [39–40]. Urban traffic arises from an interplay between the dynamics of the individual movements and its underlying structure. For subway systems, passenger flow can be seen as person-trips aggregately distributed in the networks, which has motivated various studies based on trip data for inferring statistical properties of passenger flows in the Metropolitan Seoul Subway [41], and analysing the travel routes [38] and the spatiotemporal density of passengers [42] for the Singapore Mass Rail Transit to revelling the urban spatial mobility patterns [43] and hierarchical organisation of urban polycentricity [44] for the London Underground. According to some studies, dynamic traffic flow in a network really matters in determining the relative importance of a station within the network, leading to the results that show that stations that experience more large passenger flows are more important or central in a subway system [38, 43, 44]. Therefore, the framework to analyse physical topologies and traffic patterns together should bridge the gap between static infrastructure network centrality and dynamic traffic flow process derived from human mobility in real transportation networks.

In addition to the above, in the case of a subway network, the passenger flow at each station is not only related to the transmission of passenger flow from adjacent stations, but also to the flow into and out of the station. The flow into the station from outside and the flow out of the network from the station are important exogenous factors, which are related to the population density, land use and urban functions around that stations. Therefore, we introduce the alpha centrality [45] and consider the exogenous factors of stations, in order to highlight that station centrality should consider the communication between the station and the outside.

In this paper, we introduce a two-layered model to analyse subway networks, where the networks of traffic flows and physical infrastructures interact with each other, and identify nodal importance and centrality based on trip data obtained from the Beijing Subway System. These results suggest that commonly used topological centrality measures such as degree centrality and betweenness fail to determine nodal prominence in a weighted flow network as well as to approximate the real load, but the alpha centrality with reference to traffic flow patterns works better in our dataset. Our findings confirm that layered view of transportation systems can be very helpful to capture the fundamental differences between these coexisting topologies as a part of such networks, and understand the reasons standing behind the differences. The overloading of lines and stations is a real problem in many subway systems around the world. Our research results on critical nodes identification can help subway systems take measures to distribute flows and prevent bottlenecks and dangerous crowding levels at particular nodes.

This paper is organised as follows. In Section 2, we describe the studied subway network and the dataset used in this study, and we introduce the layered model to facilitate the studied system. In Section 3, we present the results from the nodal centrality analysis and compare them with each other. In Section 4, we discuss the relationship between the node centrality and the real load. Finally, Section 5 contains our conclusions and future work.

2. METHODOLOGY

2.1 Study network and data description

We consider the Beijing Subway System (BSS), which serves as the major transportation mode in the urban and suburban districts of Beijing municipality, China. The network in 2014 consisted of 16 lines, 236 unique stations and 265 sections between them, connecting all major districts across the Beijing city. The network has two operators, the state-owned Beijing Subway Operation Corporation, and the Beijing MTR Corporation, a public-private joint venture with the Hong Kong MTR Corporation. Here, we study the importance of stations in a subsystem, which is managed by the Beijing Subway Operation Corporation. Note that the subnetwork includes 14 lines, 203 unique stations and 227 sections.

Our data analysis is based on a dataset describing the passenger trip data on a single weekday, collected from the transaction data of the smart Yikatong card on April 15, 2014. This contactless smart card is currently the main payment method employed by the Beijing public transportation system, which is similar to the Singapore's EZ-Link and Oyster card used in London. In the smart card-based electronic ticketing system, we are able to record individual person movements within the BSS network and to capture specific travel information such as the origin-destination stations as well as the corresponding travel time of the trip. Our database that is provided by the Beijing Subway Operation Corporation captures a total of 6,161,646 passenger individuals arriving at stations and leaving stations on a single weekday. These trip data serve as a basis to generate information on load profile, flow characteristics and spatiotemporal variation. The whole experiment is conducted by MATLAB 2016, a programming and numeric computing platform.

Subway passenger flows can be seen as individual trips aggregately distributed in the subway network. To analyse the passenger trip data of the BSS network, from this database, we construct the passenger flow matrix $\mathbf{B}_1$ whose elements are defined to be the number of passengers taking trips between a pair of two adjacent stations over a given period of time. We also build another two flow matrices, $\mathbf{B}_2$ and $\mathbf{B}_3$, which describe the number of passengers arriving at stations and leaving stations over a given period of time, respectively. It is important to note that the data of the day (from 5 a.m. to 24 p.m.) serve as a basis for the nodal centrality analysis presented in the next section. We also divide the entire trip data into half-hourly pieces, so 30 min is the time period for gathering the aggregated number of passenger movements. The analysis of these flow matrices will explicitly probe spatiotemporal patterns of nodal centrality discussed in Section 4.

2.2 Two-layered model of the subway system

Weighted networks provide a good description and explanation for the rich dynamics observed in the real-life subway systems [4]. However, in the presence of passenger flows through a subway network, one-layer weighted physical graph is not sufficient because the physical network is only a part of a larger complex system, where coexisting topologies such as the network of traffic flows and physical infrastructures interact with and depend on each other. Accordingly, borrowing from the general multilayer model [28, 36], we refer to the flows through a system as a two-layered model. In our layered model, the lower-layer topology represents the physical infrastructure network of a subway system and the upper-layer topology abstracts the passenger movements within the system. *Figure 1* presents a simplified example of the two-layered model and we explain it in detail below.

In *Figure 1*, the graph G^d is the physical structure of the subway infrastructure, and the graph G^f is the corresponding network of the passenger flows. The weights of directed links in the graph G^f represent the numbers of passenger flows.

For the lower layer, we represent the topology of the subway infrastructure based on L-space representation [4, 17, 36] that depicts the original configuration of real transportation networks. In L-space, subway stations are nodes, and two stations are connected only if they are physically directly connect-

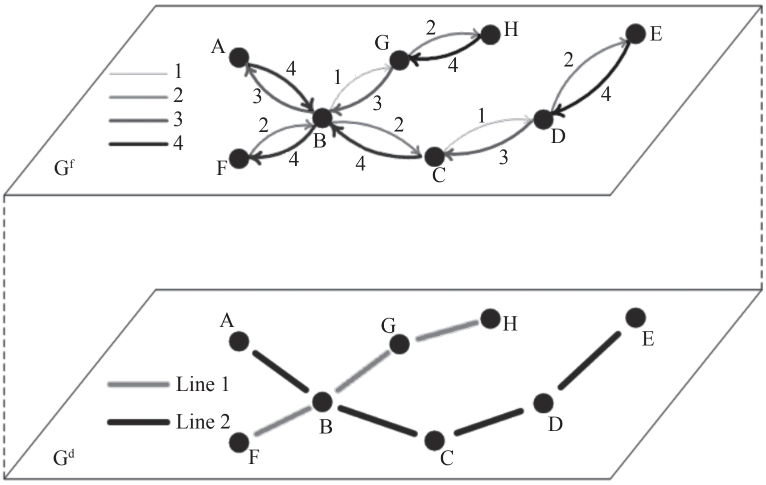


Figure 1 – Schematic representation of the two-layered model

ed. Let the subway infrastructure topology be represented as the weighted physical graph G^d with N^d nodes and L^d links, there are an associated adjacency matrix $\mathbf{A}=\{a_{ij}\}$ and a weight matrix $\mathbf{W}^d=\{w^d_{ij}\}$ representing the physical distance between two adjacent nodes n^d_i and n^d_j.

For the upper layer, we construct a directed weighted network of the passenger flows from topology of the lower layer by incorporating passenger trip data. This directed weighted network can be obtained by considering the directions and volumes of the flows between two adjacent stations. Let the network of passenger flows be represented as the weighted flow graph G^f with N^f nodes and L^f links. The sets of nodes at both two layers are apparently identical in transportation systems, i.e. $N^d=N^f$, but without loss of generality, we will keep the superscripts d and f to make the description unambiguous. Each logical flow link l^f is mapped on the physical graph G^d as the physical paths connecting the node n^f_i and n^f_j. These paths can be given explicitly by converting an un-directed physical link into two directed flow links since in reality, a pair of adjacent stations is often joined by a double-track section where the train services run on the up and down directions, respectively (i.e. typical travel is bidirectional).

We find that, after introducing directed weighted links [41, 43] to present the directionality of flows, it is easy to derive the weight matrix $\mathbf{W}^f=\{w^f_{ij}\}$ of the flow graph G^f from the flow matrix $\mathbf{B}_1$. An element w^f_{ij} denotes the flow from the node n^f_i to n^f_j.

The above subscripts i and j appear as destination-source station, e.g. w^f_{ij} denotes the flow of departure (outflow) of node n^f_i, moving in the $i \rightarrow j$ direction. It is noted that the subscripts particularly refer to a pair of two adjacent stations unless otherwise mentioned.

In addition to the flows between a pair of stations, a weighted flow graph G^f has a flow moving into node n^f_i from outside the network and a flow moving out of the network from the same node. The flows from outside to node n^f_i correspond to passengers entering into the station from nearby areas or other transportation means. The flows from node n^f_i to the outside correspond to passengers leaving the station to end their trip. We introduce the compartment labelled 0 (zero) to present the network exchanging with the source of exogenous flows. Let the flow into node n^f_i from outside be denoted by v_{0i}, and the flow out of the network for node n^f_i be denoted by u_{i0}.

It is noted that the exogenous flows, v_{0i} and u_{i0}, can be derived from the flow matrices $\mathbf{B}_2$ and $\mathbf{B}_3$, respectively. Accordingly, building a weighted flow graph G^f from trip data guarantees that w^f_{ij} always represents the passengers taking trips between adjacent stations, while v_{0i}/u_{i0} always represents these passengers entering into/leaving one station. Therefore, the flow characteristics as network properties can be calculated at the node level.

3. RESULTS

3.1 Degree and strength

The degree of a node can be seen as the simplest of the node centrality measures [46]. In the case of a directed network, the degree is classified into indegree and outdegree respectively; while in the case of a weighted network, the degree is extended to the sum of weights on the links incident upon a node and labelled node strength [46–47]. For a given node i in the graph G^d, node degree is the number of links a node has, $k_i = \sum_{J}^{N} a_{ij}$. In the graph G^f, node degree is the sum of indegree and outdegree, $k_i = k_i^{in} + k_i^{out} = \sum_{J}^{N} a_{ij}$, then the average degree in both graphs can be denoted by $\langle k \rangle$:

$$\langle k \rangle = \frac{1}{N}\sum_{i}^{N} k_i = \frac{1}{N}\sum_{i}^{N}\sum_{j}^{N} a_{ij} \tag{1}$$

Strength is the natural generalisation of node degree in weighted networks. Subsequently, a node's strength in the graphs G^d and G^f is the sum of the weights on the links incident upon a node, $s_i = \sum_{J}^{N} a_{ij} \cdot w_{ij}$, and the average strength can be denoted by $\langle s \rangle$. It is important to note that in the case of the flow graph G^f, the node strength accounts for the total traffic handled by a station. As a load estimator, this quantity is a natural measure of centrality in weighted flow networks [47].

$$\langle s \rangle = \frac{1}{N}\sum_{i}^{N} s_i = \frac{1}{N}\sum_{i}^{N}\sum_{j}^{N} a_{ij} \cdot w_{ij} \tag{2}$$

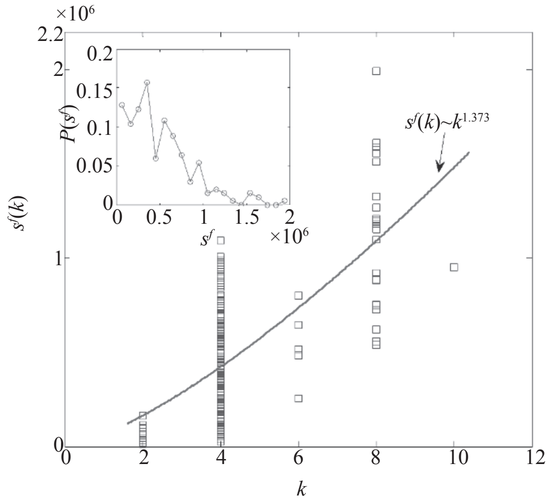
The flow graph G^f possesses a low average node degree, $\langle k \rangle = 4.5320$, which indicates that the BSS network has a few hub nodes that allow transfer passengers to change from one route to the other route. To study the relationship between topological centrality of nodes and the geographical patterns of flows, we obtain the strength, $s^f(k)$, as a function of node degree and its corresponding spatial distribution of the subway system, as shown in *Figure 2*. It can be seen that the probability distribution $P(s^f)$ can be fitted by the Weibull distribution with parameters a=54439 and b=1.3772 rather than a broad law such as the power-law distribution suggested in [47]. However, we find that the strength $s^f(k)$ of nodes with node degree k follows a power-law behaviour $s^f(k) \sim k^\beta$ with β=1.3373. This value denotes strong correlation between the traffic handled by a station and the number of its connections. In fact, the strength $s^f(k)$ grows exponentially with degree k. In other words, the allometric growth of node strength indicates that the more connected station tends to have more traffic it handles, and the traffic will grow much faster than the number of connections, as shown in *Figure 2a*.

The inset in *Figure 2a* is the distribution of $P(s^f)$, which can be approximately fitted by the Weibull distribution. In *Figure 2b*, the nodes in the network are placed into eight quartile groups, and the colour of nodes is proportional to the measured value.

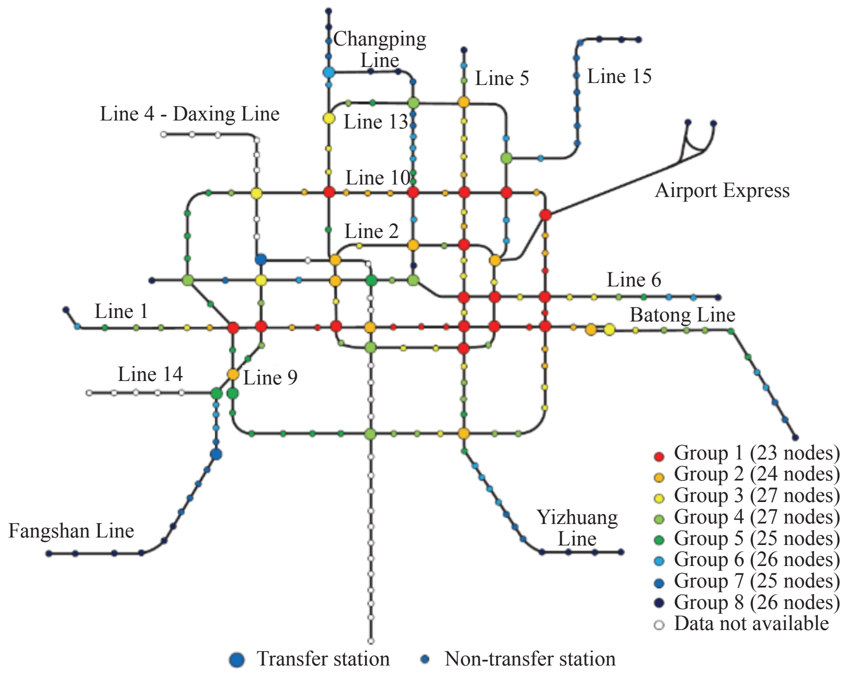
The fact is clear that the most connected nodes with a topological measure of centrality may not necessarily carry the most traffic in transportation networks. We can also see from *Figure 2a* that although many stations share the same connections, the traffic flows handled by each station differ significantly in terms of volume. For example, hub nodes or transfer stations Guomao and Cishousi both possess the degree k=8 but handle significantly different volumes of traffic, s=19992366 for the former station and s=537754 for the latter one, respectively. To shed more light on the relationship between the nodes' strength and degree, we investigate the spatial distribution of node strength in the BSS network, as illustrated in *Figure 2b*. All stations are divided into 8 groups using a hierarchical classification method and are represented by different colours. It is clear that the dependence of s_i^f on k_i may not always follow the power-law distribution. For instance, Xizhimen station that possesses the highest degree (k=10) in the network has the strength s=950148, whereas Guomao station handles the higher volume of traffic, significantly greater (2.09 times) than that of Xizhimen station. The reason for this violation is that node strength only takes into consideration a node's total link magnitudes in the network, but fails to take into account the characteristics (direction and volume) by which the flows move through a network. Therefore, our findings confirm the statement that the commonly used topological measures such as node degree and strength fail to approximate the most central stations in transportation networks.

3.2 Node centrality

In addition to knowing the degree and strength of nodes, we would like to further investigate the identification of the most central nodes in both the physical infrastructure network (graph G^d) and weighted flow network (graph G^f), and the difference between them. Degree and strength are the simplest of the node centrality measures by focusing solely on the local structure and information around nodes. However, more connected or loaded stations may not be necessarily more central in a subway network. There are other complex measures of the



a) Strength $s^f(k)$ as a function of the degree k of nodes of the BSS network



b) The corresponding spatial distribution of the node strength s^f

Figure 2 – Function and distribution related to the strength $s^f(k)$

node centrality, such as *betweenness* and *eigenvector-like centrality* to identify which nodes are more central than others in a network.

In this study, we use two complex measures to identify the centrality of nodes in the graph G^d and graph G^f, respectively. For the physical graph G^d, we use the betweenness centrality measure, which highlights the importance of a node acting as a "bridge" for connecting different regions of the transportation network [26, 47]. This *transfer* characteristic is clearly of paramount importance in subway systems and thus can help design the infrastructure network to distribute the flow of passengers more evenly [26]. Let b_i be the fraction of shortest paths that pass through node i and then the betweenness centrality of a node b_i can be represented as:

$$b_i = \sum_{i \neq j \neq k \in N} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{3}$$

where σ_{jk} is the number of physical distance shortest paths between two stations and $\sigma_{jk}(i)$ is the number of paths that go through node i. Note that betweenness relies on the identification of the shortest paths. For transportation infrastructure networks, it is more appropriate to use physical distance rather than the network distance in identifying the shortest paths.

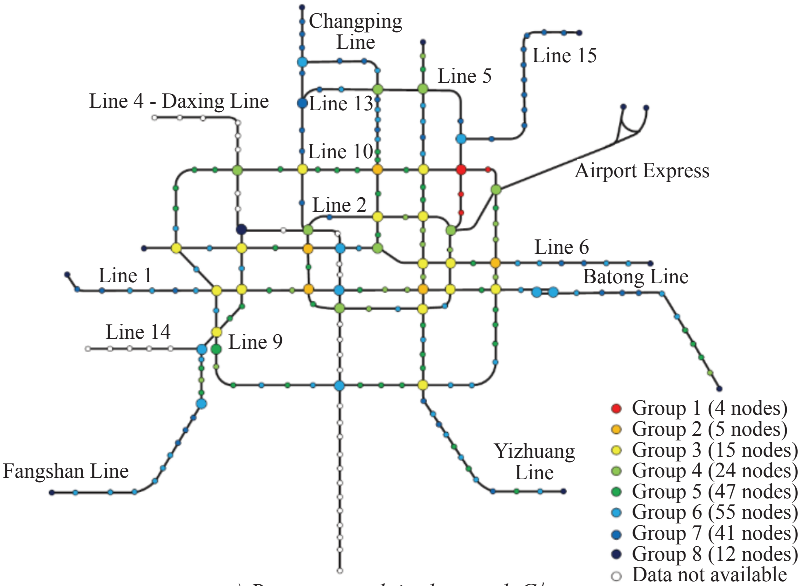
For weighted flow network of transportation systems, eigenvector centrality was used to measure the importance or centrality of nodes [38]. The concept of underlying eigenvector centrality is that those stations that are themselves receiving much traffic from others will be more central in transportation networks. However, this measure of centrality does not take into account stations' external status characteristics such as passengers entering into the station and leaving the station. The flow into the station from outside and the flow out of the network from the station are important exogenous factors, and are related to the population density, land use, and urban functions around that station. Therefore, for the flow graph G^f, we use the alpha centrality measure [45], which allows stations to have external sources of influence. Besides, the influence of adjacent nodes on nodes in complex networks is considered [48]. Let e_i be the amount of exogenous flows that node i receives, namely $e_i=v_{0i}+u_{i0}$, and then the alpha centrality of a node c_i can be defined as:

$$c_i = \alpha \sum_{j=1}^{N} a_{ji} w_{ji}^{f} c_j + e_i \tag{4}$$

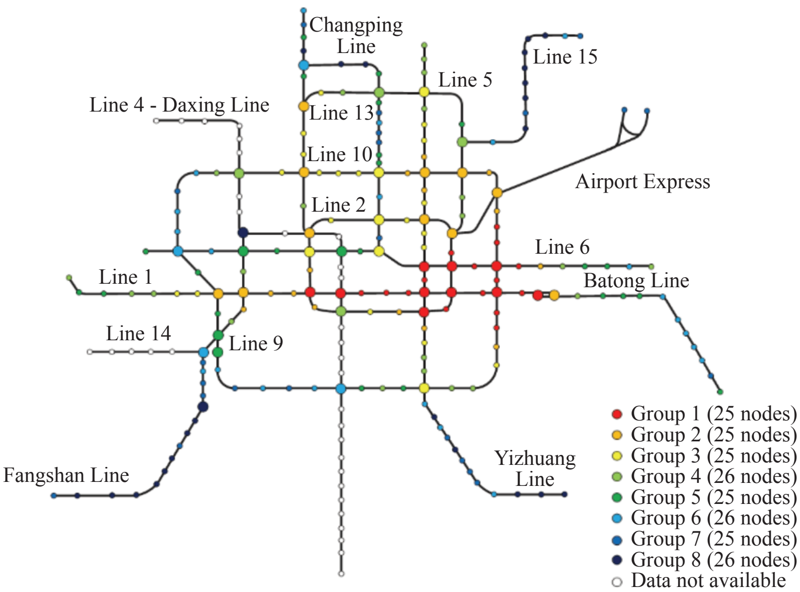
where α is a parameter that trades off the importance of exogenous factors against the importance of connectivity, which is influenced by the network topology. When α=0, only the consideration of the exogenous factors matters. When α is very large then only the connectivity matters, i.e. it reduces to the eigenvector centrality case. Generally, the parameter α has to be $1/\lambda$, where λ is the largest eigenvalue of the weighted matrix **W** [45].

Figure 3 shows the spatial distributions of the betweenness in the graph G^d and the alpha centrality in the graph G^f. *Figure 4* shows their corresponding probability distributions. To make the node centrality scores comparable across graphs, we normalise them by the total value of centrality of each network. In *Figure 3a*, we measure the importance of stations in the physical infrastructure network, corresponding to the nodes' betweenness score ranging from 0 to 0.095. The shape of the betweenness distribution is asymmetrical, being skewed to the left hand side and can be fitted by the Weibull distribution with parameters a=0.0025 and b=0.4309, as shown in *Figure 4a*. In fact, more than 80% of stations have relatively low and similar centrality scores. This is not surprising since the BSS network has a large fraction, 80% of stations that do not offer transferring and thus necessarily help to the possible shortest path. From a topological point of view, these stations appear nearly equal in terms of importance. On the other hand, the stations with higher centrality scores (Groups 1-3) almost belong in the type of transfer stations. These most central stations host the Line 2 and 10, which are both rectangular loop lines to offer a majority of transferring in the network. From this analysis, we conclude that betweenness centrality can be viewed as an appropriate measure to determine which stations are topologically more central in the network, since it captures the importance of a station as transfer point to join pairs of nodes. Furthermore, our finding shows the evenly distributed betweenness in the network, which is similar with other worldwide subway systems [26]. Compared to other researches on the centrality of the Beijing subway network, our results are also similar. The top 10 stations with higher degree and betweenness in reference [33] are all within the top three groups in *Figure 3a*.

Compared with *Figures 3a and 3b* provides much of the same spatial pattern, but shows it in a way that allows one to determine the importance of a station in the system with reference to how passenger flows are distributed through the network. It is evident that the identifications of the most central stations are roughly similar to these results presented in *Figure 3a*. In *Figure 4b*, the alpha centrality score of stations ranges from 0.001 to 0.076 and its distribution can be well fitted by the Weibull distribution with parameters a=0.0016 and b=0.2921. This skewed distribution indicates that passenger flows are concentrated on a few centres and hub stations with the high alpha centrality, whereas many stations of little importance have less passengers flowing into the station and out of the station. The result is compatible with our understanding of subway networks, namely the number of trips or traffic between adjacent stations is scale-free whose features are common to many subway systems. We also observed that a station that has a high centrality score is one that is adjacent to stations that are themselves high scorers. As such, the centrality of a station is expected to depend on the centrality of its neighbouring stations. This tendency accounts for the hierarchical organisation of the passengers flows



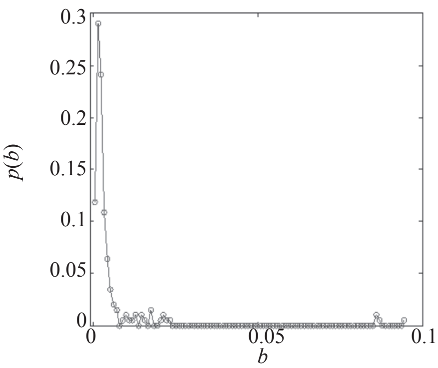
a) Betweenness b in the graph G^d



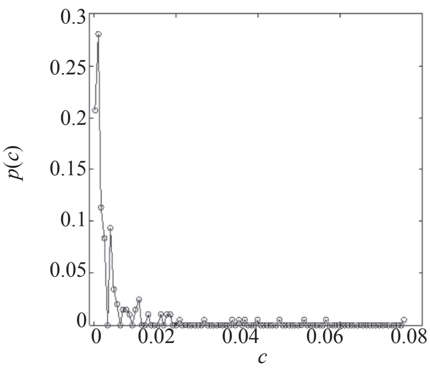
b) Alpha centrality c in the graph G^f

Transfer station Non-transfer station

Figure 3 – The spatial distributions of the betweenness and the alpha centrality



a) Betweenness b



b) Alpha centrality c

Figure 4 – The probability distributions

through the network, which can be empirically observed at an intra-urban level, in agreement with the studies for London Underground [44].

The results show that the spatial distribution of the alpha centrality is almost similar to that of the betweenness, but differences can be found. For instance, Guomao station, which is ranked the 10th by betweenness centrality, is found to receive far more traffic than expected. Indeed, it is ranked the highest score by the alpha centrality. The reason is that Guomao is situated in the financial centre of the Beijing city, the Beijing Central Business District (Beijing CBD), even if it does not lie in the centre of the network geographically. Conversely, Shaoyaoju station possesses the highest score of betweenness centrality, but handles less flow of passengers. In fact, it is only ranked 31st by the alpha centrality. From the comparison, we conclude that in a subway network the stations that are highly central in the geodesic-based processes such as the betweenness need not be highly central in the dynamic processes such as the passenger flow along links. This can serve as a basis for the better understanding of the node centrality in urban transit systems and may strongly affect the results of the network performance analysis based solely on topological measures.

4. DISCUSSION

For comparison, we have presented the two centrality measures based on the layered network framework. The first metric is betweenness centrality b based exclusively on the physical network. The betweenness centrality quantifies the fraction of a station and acts as a "bridge" along the shortest path between two other stations. The betweenness of a node thus captures the amount of flows passing through the node. Indeed, existing studies have already taken betweenness as a measure of the load estimator either directly [28–29] or with slight modifications [49]. The second metric is the alpha centrality c under the consideration of the manner in which flows are distributed through the network. It seems that the stations with high centrality score carry more flows than the less central stations. In fact, node centrality turns out to be equivalent to the load in many networks, so there are two quantities to keep track of in our study, which we refer to as betweenness and alpha centrality.

Although we can easily compute the centrality of nodes with *Equations 3 and 4*, there is still a question that whether the two load estimators can mimic the load pattern in a real-life transportation network. To study this question, we first define the real load of nodes in the weighted flow network. Borrowing from ecology [36], we refer to the sum of the flow weights of all links and the exogenous flows whose both paths traverse a node i as its real load, l_i:

$$l_i = v_{0i} + u_{i0} + \sum_j w_{ij}^f + \sum_j w_{ji}^f \tag{5}$$

where in a subway network l_i can be computed as the total amount of passengers moving into/out of the station and passengers arriving at/leaving the same station.

In *Figure 5*, we present the spatial distribution of the real load for the BSS network, which reveals a roughly similar pattern comparable to the patterns of the two node centrality measures. The nodes are classified into eight quartile groups. The probability distribution of $P(l)$ is fitted by the Weibull distribution of parameters a=579241 and b=1.2650. The difference can be observed in the geographical patterns formed by the distributions of the real load, the betweenness and the alpha centrality, by comparing *Figures 3a, 3b, and 5*. To quantify these differences, we present the scatter plots of the two node centrality measures versus the real load l in *Figure 6*. In the top left corner of every plot, we give the value of the corresponding Pearson's correlation coefficient. Unsurprisingly, it is found that in our data the alpha centrality approximates the real load, which is better than the betweenness, corresponding to their Pearson's coefficients, 0.6945 and 0.3045, respectively. These results can be understood by the layered view of the subway system, which is described as below.

The physical graph G^d, which is mapped on the physical infrastructure network of a subway system, is undirected, weighted, with the weighted distribution of physical distance being skewed to the left hand side. The betweenness is a physical topological property of such physical networks, which is evenly distributed with large size for our studied network and other worldwide subway systems [26]. Indeed, betweenness can be viewed as the node load estimator in transportation networks and any efforts in predicting traffic would therefore assume that the traffic flows through the network are completely affected by the geographical constraints of the network. Nevertheless, many authors took betweenness as a measure of load [28, 29, 49]. It is very far from being satisfactory because of the fact that not all passengers use geodesic paths to travel and that the distribution of passengers in a subway network is not evenly distributed. The exponential distribution of flows arises

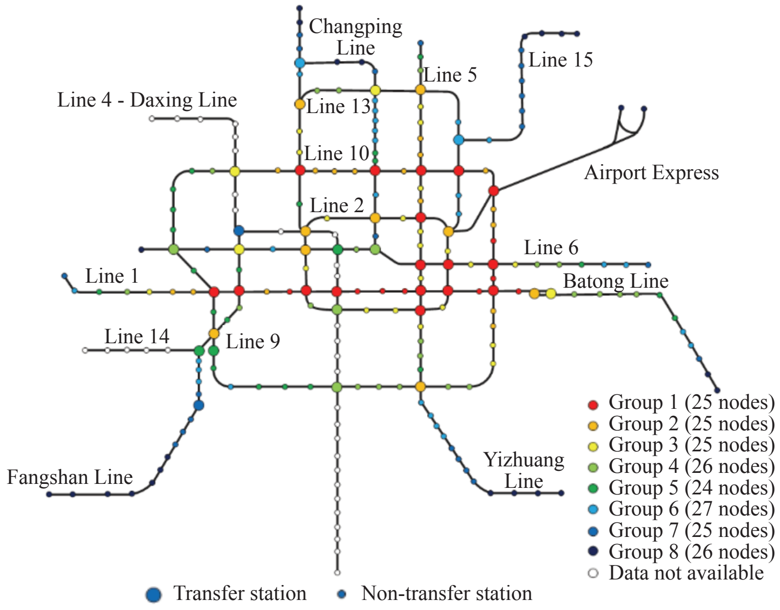
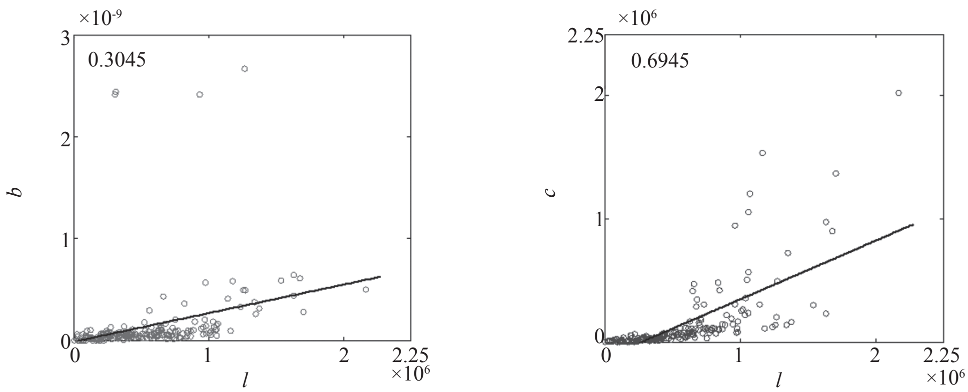


Figure 5 – Spatial distribution of the real load l



a) Betweenness b

b) Alpha centrality c

Figure 6 – Scatter plots of the betweenness and the alpha centrality versus the real load l

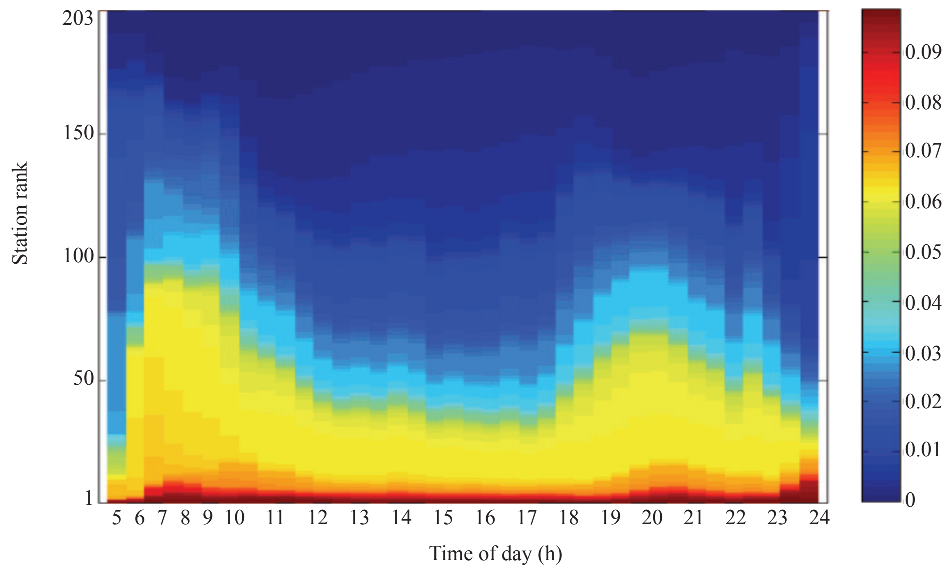


Figure 7 – The Beijing Subway System: node centrality and spatiotemporal pattern

from the interplay between the dynamics of traffic and the underlying structure. The concept of betweenness largely neglects the dynamic flow processes that unfold along the links of a network, and thus fails to approximate the real load in subway systems. We also expect to observe similar results in other transportation networks such as inter-urban rail network [28] and urban road network [6].

In contrast, the alpha centrality identifies the importance of nodes based on the weighted flow network. Such a logical network, which is derived naturally from the passenger movements within a subway network, is directed, weighted, with the flow weight distribution decaying exponentially. In fact, the characteristics of the flow process affect which stations will receive flows from other stations frequently and abundantly, and finally determine the importance of a station in the system. The empirical results, for the most part, support our claim that the alpha centrality is better able to determine important stations than betweenness because it more closely models the spread of flow on the subway system, which occurs via "broadcasts" from one station to its neighbourhood stations, and thus is more easily perceived and correlated with real traffic patterns such as the heterogeneity and hierarchical organisation of the flows.

The time evolution of passengers travelling between a pair of stations in a subway system can be modelled as a dynamic evolving system [41]. From a dynamic perspective, the alpha centrality can be viewed as a time-dependent measure for assessing nodal prominence and its value, possibly is stochastic due to the uncertainty of travel demand patterns of the network. To uncover the time evolution of the nodal centrality, we present the spatiotemporal pattern of the stations' alpha centrality of the BSS network, from 5 a.m. to 24 p.m. in intervals of 30 mins, as shown in *Figure 7*. The colours indicate the normalised alpha centrality of stations at a certain time and location. Intuitively, the stations' centrality has a distinct morning peak around 7:00. Likewise, the evening peak can be observed around 19:00. Such characteristics are very similar to those easily observed in the travel demand patterns of a subway system. This spatiotemporal pattern provides a basic understanding of what stations in the BSS network are more central in terms of passengers handled over a typical weekday. Better understanding of the evolution of nodal centrality therefore has clear benefits such as effective management of crowds and early detection of service breakdowns.

Conversely, stations' betweenness centrality is static and deterministic unless the topological structure of the system is changed, such as the opening of new stations or lines. Accordingly, one advantage of the alpha centrality is that we can easily identify the nodes' importance or centrality by taking consideration of the temporal patterns of traffic demand. Such dynamic characteristics can be helpful to define a more effective response strategy regarding the operation of BSS services. Furthermore, our analysis shows that the dynamic properties of a system may differ significantly with its topological properties, suggesting a change in thinking from static topology analysis to a dynamic (weighted) perspective. This changed in view strongly affects the results of the network performance based exclusively on the physical topology of a network. For example, the evolution of cascading failure or robustness in the past crucially depended on the topological measures [19, 20] and now it relies on dynamic flow processes or dynamic flow redistribution [50].

5. CONCLUSIONS AND FUTURE RESEARCH

Network methods are useful for studying transportation systems. In this paper, we distinguish a subway system on two coexisting topologies depending on each other, in which the lower layer is the physical infrastructure and the upper layer represents the passenger flows within the network. As an example of its application, we applied this layered approach to the identification of which stations are more central and important than other stations within the Beijing Subway System. We examined strength as a function of node degree in the weighted flow network and suggested that using only node degree and strength in a weighted flow network is unsuitable for identifying the nodal centrality. We clarified that it only computes a node's total link magnitude, but ignores the direction and volume of the flows along links. We further investigated the node centrality in the two coexisting layers of the system by using the corresponding appropriate measures (betweenness and alpha centrality). By comparing with the two centrality indicators, we found that stations that are more central in terms of betweenness in the weighted physical network need not be highly central in the weighted flow network quantified by the alpha centrality. We discussed this difference by comparing the centrality indicators with the real load.

The empirical findings can give us some useful insights into the node centrality of subway systems. As communities grow, urban public transportation is thriving. However, current subway systems are not able to accommodate the increasing traffic demand and undergoing congestion. The overloading of lines was a real problem in Beijing before COVID-19 and it is still the same now. Crowding is a threat to the subway network, especially for central stations. Therefore, adopting a complex network approach to study the node centrality for transport can be beneficial. Based on the analysis of the network and critical nodes, it becomes feasible to design systems to distribute flows and prevent bottlenecks and dangerous crowding levels at particular nodes.

On the other hand, subway systems are becoming a part of the growing metropolitan multi-modal traveling. Our modelling approach needs to be extended to cover other urban transit modes (bus, tram, trolley, etc.), which will enrich the analysis of network performance from a new point of view.

ACKNOWLEDGEMENT

This work was supported by the Fundamental Research Funds for the Central Universities (2021YJS090) and the National Natural Science Foundation of China (71971021).

童瑞咏，博士研究生[1]
电子邮箱：20114047@bjtu.edu.cn
许奇，博士[1,2]（通讯作者）
电子邮箱：xuqi@bjtu.edu.cn
魏润斌，博士研究生[1]
电子邮箱：19114056@bjtu.edu.cn
黄俊生，博士研究生[1]
电子邮箱：19114037@bjtu.edu.cn
肖中圣，博士研究生[1]
电子邮箱：18114047@bjtu.edu.cn
[1] 北京交通大学交通运输学院，
中国北京市海淀区上园村3号，100044
[2] 北京交通大学综合交通运输大数据应用技术交通运输行业重点实验室，
中国北京市海淀区上园村3号，100044

加权复杂网络中的节点中心度差异分析——以北京地铁系统为例

摘要

车站的中心性是城市交通系统中最重要的问题之一。此类网络的中心车站通常使用网络拓扑中心性进行识别。但在真实的网络中，客流是由个体运动的动力学和底层物理结构之间的相互作用产生的。在本文中，我们应用一个两层模型来识别北京地铁系统中的大多数中心车站，其中下层代表物理基础设施，上层代表客流。我们比较了两层模型的几种中心性指标，如度、强度和介数中心性。为了表示车站外部因素对地铁系统的影响，我们引入了*Alpha*中心性。结果表明，物理基础网络中的中心车站的介数中间性与客流网络中的中心车站的*Alpha*中心性并不一致。我们将两个中心性结果与实际客流负载进行对比来解释这一差异，表明*Alpha*中心性比介数中心性更接近实际，因为它考虑了轨道上的客流的方向和大小，以及进入和离开地铁系统的客流。这些实证结果可以为我们深入了解地铁系统的节点中心性提供一些有用的信息。

关键词

节点中心性；介数中心性；*Alpha*中心性；
地铁系统；客流

REFERENCES

[1] Newman MEJ. The structure and function of complex networks. *SIAM Review.* 2003;45(2): 167-256. doi: 10.1137/S003614450342480.

[2] Boccaletti S, et al. Complex networks: Structure and dynamics. *Physics Reports*. 2006;424(4): 175-308. doi: 10.1016/j.physrep.2005.10.009.

[3] Derrible S, Kennedy C. Applications of graph theory and network science to transit network design. *Transport Reviews*. 2011;31(4): 495-519. doi: 10.1080/01441647.2010.543709.

[4] Lin J, Ban Y. Complex network topology of transportation systems. *Transport Reviews*. 2013;33(6): 658-685. doi: 10.1080/01441647.2013.848955.

[5] Xu M, et al. Discovery of critical nodes in road networks through mining from vehicle trajectories. *IEEE Transactions on Intelligent Transportation Systems*. 2018;20(2): 583-593. doi: 10.1109/TITS.2018.2817282.

[6] Kocur-Bera K. Scale-free network theory in studying the structure of the road network in poland. *Promet – Traffic&Transportation*. 2014;26(3): 235-242. doi: 10.7307/ptt.v26i3.1316.

[7] Calzada-Infante L, Adenso-Díaz B, Carbajal SG. Analysis of the European international railway network and passenger transfers. *Chaos, Solitons & Fractals*. 2020;141: 110357. doi: 10.1016/j.chaos.2020.110357.

[8] Wang W, et al. Analysis of the Chinese railway system as a complex network. *Chaos, Solitons & Fractals*. 2020;130: 109408. doi: 10.1016/j.chaos.2019.109408.

[9] Bombelli A, Santos BF, Tavasszy L. Analysis of the air cargo transport network using a complex network theory perspective. *Transportation Research Part E: Logistics and Transportation Review.* 2020;138: 101959. doi: 10.1016/j.tre.2020.101959.

[10] Verma T, Araújo NAM, Herrmann HJ. Revealing the structure of the world airline network. *Scientific Reports*. 2014;4: 5638. doi: 10.1038/srep05638.

[11] Gallotti R, Porter MA, Barthelemy M. Lost in transportation: Information measures and cognitive limits in multilayer navigation. *Science Advances*. 2016;2(2):

e1500445. doi: 10.1126/sciadv.1500445.

[12] Yuan G, Kong DW, Sun LS, Luo W. Connectivity contribution to urban hub network based on super network theory–case study of Beijing. *Promet – Traffic&Transportation*. 2021;33(1): 35-47. doi: 10.7307/ptt.v33i1.3536.

[13] Wu J, Gao Z, Sun H, Huang H. Urban transit system as a scale-free network. *Modern Physics Letters B*. 2004;18(19n20): 1043-1049. doi: 10.1142/S021798490400758X.

[14] Louf R, Roth C, Barthelemy M. Scaling in transportation networks. *PLoS One*. 2014;9(7): e102007. doi: 10.1371/journal.pone.0102007.

[15] Latora V, Marchiori M. Is the Boston subway a small-world network?. *Physica A: Statistical Mechanics and its Applications*. 2002;314(1-4): 109-113. doi: 10.1016/S0378-4371(02)01089-0.

[16] Derrible S, Kennedy C. The complexity and robustness of metro networks. *Physica A: Statistical Mechanics and its Applications*. 2010;389(17): 3678-3691. doi: 10.1016/j.physa.2010.04.008.

[17] Latora V, Marchiori M. Efficient behavior of small-world networks. *Physical Review Letters*. 2001;87(19): 198701. doi: 10.1103/PhysRevLett.87.198701.

[18] Ek B, VerSchneider C, Narayan DA. Efficiency of star-like graphs and the Atlanta subway network. *Physica A: Statistical Mechanics and its Applications*. 2013;392(21): 5481-5489. doi: 10.1016/j.physa.2013.06.055.

[19] Latora V, Marchiori M. Vulnerability and protection of infrastructure networks. *Physical Review E*. 2005;71(1): 015103. doi: 10.1103/PhysRevE.71.015103.

[20] Yang Y, et al. Robustness assessment of urban rail transit based on complex network theory: A case study of the Beijing Subway. *Safety Science*. 2015;79: 149-162. doi: 10.1016/j.ssci.2015.06.006.

[21] Roth C, Kang SM, Batty M, Barthelemy M. A long-time limit for world subway networks. *Journal of The Royal Society Interface*. 2012;9(75): 2540-2550. doi: 10.1098/rsif.2012.0259.

[22] Leng B, Zhao X, Xiong Z. Evaluating the evolution of subway networks: Evidence from Beijing subway network. *Europhysics Letters*. 2014;105(5): 58004. doi: 10.1209/0295-5075/105/58004.

[23] Freeman LC. Centrality in social networks conceptual clarification. *Social Networks*. 1978;1(3): 215-239. doi: 10.1016/0378-8733(78)90021-7.

[24] Wang J, Li C, Xia C. Improved centrality indicators to characterize the nodal spreading capability in complex networks. *Applied Mathematics and Computation*. 2018;334: 388-400. doi: 10.1016/j.amc.2018.04.028.

[25] Newman MEJ. A measure of betweenness centrality based on random walks. *Social Networks*. 2005;27(1): 39-54. doi: 10.1016/j.socnet.2004.11.009.

[26] Derrible S. Network centrality of metro systems. *PLoS ONE*. 2012;7(7): e40575. doi: 10.1371/journal.pone.0040575.

[27] Tang J, Li Z, Gao F, Zong F. Identifying critical metro stations in multiplex network based on DS D–S evidence theory. *Physica A: Statistical Mechanics and its Applications*. 2021;574: 126018. doi: 10.1016/j.physa.2021.126018.

[28] Kurant M, Thiran P. Layered complex networks. *Physical Review Letters*. 2006;96(13): 138701. doi: 10.1103/PhysRevLett.96.138701.

[29] Ramli MA, Monterola CP, Khoon GLK, Guang THG. A method to ascertain rapid transit systems' throughput distribution using network analysis. *Procedia Computer Science*. 2014;29: 1621-1630. doi: 10.1016/j.procs.2014.05.147.

[30] Crucitti P, Latora V, Porta S. Centrality measures in spatial networks of urban streets. *Physical Review E*. 2006;73(3): 036125. doi: 10.1103/PhysRevE.73.036125.

[31] Liao C, Dai T, Zhao P, Ding T. Weighted centrality and retail store locations in Beijing, China: A temporal perspective from dynamic public transport flow networks. *Applied Sciences*. 2021;11(19): 9069. doi: 10.3390/app11199069.

[32] Wang Z, Li J, Huang L, Yang Z. Discovering the evolution of Beijing Rail Network in fifty years. *Modern Physics Letters B*. 2020;34(21): 2050212. doi: 10.1142/S0217984920502127.

[33] Xiao X, Jia L, Wang Y, Zhang C. Topological characteristics of metro networks based on transfer constraint. *Physica A: Statistical Mechanics and its Applications*. 2019;532: 121811. doi: 10.1016/j.physa.2019.121811.

[34] Cats O, Jenelius E. Dynamic vulnerability analysis of public transport networks: Mitigation effects of real-time information. *Networks and Spatial Economics*. 2014;14(3): 435-463. doi: 10.1007/s11067-014-9237-7.

[35] Luo D, Cats O, Lint HV. Can passenger flow distribution be estimated solely based on network properties in public transport systems?. *Transportation*. 2020;47(6): 2757-2776. doi: 10.1007/s11116-019-09990-w.

[36] Gao S, Wang Y, Gao Y, Liu Y. Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality. *Environment and Planning B: Planning and Design*. 2013;40(1): 135-153. doi: 10.1068/b38141.

[37] Feng J, et al. Weighted complex network analysis of the Beijing subway system: Train and passenger flows. *Physica A: Statistical Mechanics and its Applications*. 2017;474: 213-223. doi: 10.1016/j.physa.2017.01.085.

[38] Soh H, et al. Weighted complex network analysis of travel routes on the Singapore public transportation system. *Physica A: Statistical Mechanics and its Applications*. 2010;389(24): 5852-5863. doi: 10.1016/j.physa.2010.08.015.

[39] Gonzalez MC, Hidalgo CA, Barabasi AL. Understanding individual human mobility patterns. *Nature*. 2008;453(7196): 779-782. doi: 10.1038/nature06958.

[40] Sun SW, Li HY, Xu XY. A key station identification method for urban rail transit: A case study of Beijing subway. *Promet – Traffic&Transportation*. 2017;29(3): 267-273. doi: 10.7307/ptt.v29i3.2133.

[41] Lee K, Jung WS, Park JS, Choi MY. Statistical analysis of the Metropolitan Seoul Subway system: Network structure and passenger flows. *Physica A: Statistical Mechanics and its Applications*. 2008;387(24): 6231-6234. doi: 10.1016/j.physa.2008.06.035.

[42] Sun L, Lee DH, Erath A, Huang X. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 12 August 2012*.

2012. p. 142-148.

[43] Hasan S, Schneider CM, Ukkusuri SV, Gonzalez MC. Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics*. 2013;151(1-2): 304-318. doi: 10.1007/s10955-012-0645-0.

[44] Roth C, Kang SM, Batty M, Barthelemy M. Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLoS ONE*. 2011;6(1): e15923. doi: 10.1371/journal.pone.0015923.

[45] Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*. 2001;23(3): 191-201. doi: 10.1016/S0378-8733(01)00038-7.

[46] Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*. 2010;32(3): 245-251. doi: 10.1016/j.socnet.2010.03.006.

[47] Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*. 2004;101(11): 3747-3752. doi: 10.1073/pnas.0400087101.

[48] Li C, Wang L, Sun S, Xia CY. Identification of influential spreaders based on classified neighbors in real-world complex networks. *Applied Mathematics and Computation*. 2018;320: 512-523. doi: 10.1016/j.amc.2017.10.001.

[49] Albert R, Albert I, Nakarado GL. Structural vulnerability of the North American power grid. *Physical Review E*. 2004; 69(2): 025103. doi: 10.1103/PhysRevE.69.025103.

[50] Huang A, et al. Cascading failures in weighted complex networks of transit systems based on coupled map lattices. *Mathematical Problems in Engineering*. 2015;940795. doi: 10.1155/2015/940795.