Identification of Important Nodes on Urban
Based on CRITIC Method

Shuang ZHANG
School of Highway, Chang’an University
Middle-section of S. 2nd Rd, Xi'an, Shanxi,China
514439627@qq.com

Kuan-min CHEN
School of Highway, Chang’an University
Middle-section of S. 2nd Rd, Xi'an, Shanxi,China

Quan CHEN
School of Highway, Chang’an University
Middle-section of S. 2nd Rd, Xi'an, Shanxi,China
917166531@qq.com

ABSTRACT

Urban rail transit stations play an important role in passenger distribution and connectivity intervals in the network. If a node fails, it will lead to the embarrassment of large-area orbital networks. How to effectively evaluate its influence in the passenger flow network is the focus of research on network structure optimization and reducing operational risks. Based on the complex network theory, the station is taken as the research object. By establishing the network passenger flow distribution model and combining the rail transit smart card data, the CRITIC method is used to determine the objective comprehensive weight and determine the important nodes of the urban rail transit network. The research method was applied to the Xi'an subway network, and the key stations of Xi'an subway were identified and the current situation of Xi'an subway early peak passenger flow was systematically analyzed, and opinions on subway network operation were put forward.

Keyword: CRITIC method, Urban rail transit, important node evaluation, AFC data

1 INTRODUCTION

The subway network is an important part of the urban transportation network. (Traversi D 2014) Due to the heterogeneity of the network structure, the importance of each node in the network is different. Effective identification of important nodes in the subway network is of great significance for rational allocation of limited network node resources, improvement of network efficiency and reliability, and prevention of node failure caused by random factors such as passenger flow fluctuations and equipment trouble. As a public transportation, urban rail transit has the advantages of large volume, fast speed, safety and reliability, and on-time comfort. It has gradually become the preferred means of transportation for residents, and played a vital role in solving urban traffic congestion problems. At present, urban rail transit has gradually started to operate in the network, passenger traffic continues to increase, and the passenger traffic of some stations and lines has exceeded its carrying capacity. It is imperative to ensure the safety, reliability and comfort of the urban rail transit network.
Complex network theory is to reveal the internal laws of complex systems, and to abstract the complex system using a graphical way. It originated from Euler's research on the "seven bridge problems" and applied to power engineering, computer networks, transportation, etc. Field (Lili, 2015). In 1964, Garrison and Marble began to use the graph theory knowledge to quantitatively analyze the topology of the traffic network, and defined the α, β, and γ indicators (Garrison W, 1964). Ball proposed to judge the importance of an edge by calculating the change of network shortest path after removing an edge in the network (Ball, 1989). Gattuso and Miriello first used graph theory to study the network structure of the subway, and assigned weights to the network nodes to analyze the station's transfer ability and passenger flow attraction (Gattuso D, 2005). Zhao Yizheng proposed a kind of correlation between nodes by defining the node importance contribution matrix. The communication network node importance evaluation method considers the influence of the connection relationship between different nodes in the network on the importance of the node (Zhao Yizheng, 2009). However, the urban rail transit network is a complex dynamic system. The existing research on the importance of the station is mainly based on the network topology, mainly qualitative analysis, and only considers the station's position information. There is a lack of research on the network function structure and the influence of passenger flow on the station.

Based on the theory of complex network, based on the model of the actual urban rail transit network, with road traffic flow as weights to build directed weighted network model, such as station weighted betweenness and weighted vulnerability to evaluate traffic impact on the importance of the station, considering the influence of the passenger flow to the station at the same time, through the CRITIC method to calculate the importance ranking, to provide decision support for managers, the control for the main station is beneficial to improve the level of urban rail transit operation management and services.

2 BUILDING NETWORK MODEL

2.1 Construction of Network Model with Sectional Passenger Flow as Weight

The operation of the city rail transit line is divided into two different directions, and the passenger flow situation seen in the upper and lower sections is different. Therefore, the urban rail transit network is abstracted into a directed weighted network map, the nodes represent stations, and the sides of the map represent adjacent stations. The directed segments between the points and the edges form a network structure, denoted by $G = (V,E)$. Where $V = \{v_1, v_2, \ldots, v_n\}$ represents the set of all stations, and $E$ represents the set of directed sections $l_{ij}$ between all adjacent stations (the directed section $l_{ij}$ represents the direction from station $i$ to station $j$ road section). The adjacency matrix $A = \{a_{ij}\}$ of the network graph is expressed as:

$$a_{ij} = \begin{cases} 0 & i, j \text{ no connection} \\ 1 & i, j \text{ connection} \end{cases}$$  \hspace{1cm} (1)

The definition of an element of the adjacency matrix $A'$ in a weighted directed graph can be expressed as:

$$w_{ij} = \begin{cases} w_{ij} & i, j \text{ connection} \\ 0 & i, j \text{ no connection} \end{cases}$$  \hspace{1cm} (2)

It can construct a weighted network of urban rail transit.
2.2 Passenger flow weighted network node importance assessment index

In order to better identify the key nodes of the urban rail transit network, it is first necessary to select the indicators that can properly reflect the topological characteristics and passenger flow characteristics of the urban rail transit network, and highlight the difference between the key nodes and other nodes. Therefore, the following indicators are selected to evaluate the degree of node importance.

(1) Station degree $D_i$. It is divided into $D_i^{in}$ and $D_i^{out}$. The degree of $D_i^{in}$ is the number of connected sides of the station directly connected to station $i$, and the degree of $D_i^{out}$ is the number of connected sides of station directly from the station $i$. In the network topology, the larger the station degree value is, the more important the station is and the more important contribution to its adjacent stations is.

$$D_i = D_i^{in} + D_i^{out} \tag{3}$$

In the formula, $D_i$ represents the degree of station $i$, $D_i^{in}$ is the degree of entry of station $i$, and $D_i^{out}$ is the degree of exit of station $i$.

(2) Node number $B_i$: The ratio of the number of paths passing through the node in all shortest paths in the network to the total number of shortest paths.

$$B_i = \frac{d_{st}}{\sum_{s \neq t} d_{st}} \tag{4}$$

Where $d_{st}^i$ represents the number of shortest paths from the station $s$ to the station $t$ and through the station $i$, $\sum_{s \neq t} d_{st}$ the total number of shortest paths from the station $s$ to the station $t$.

(3) Node efficiency $E_i$: The average of the reciprocal distances between station $i$ and all stations on the shortest path in the network.

$$E_i = \frac{1}{n} \sum_{j=1}^{N} \frac{1}{d_{ij}} \tag{5}$$

Where, for the number of stations in the network, $d_{ij}$ is the distance on the shortest path from station $i$ to station $j$.

(4) Node connection strength $C_i$: In the real rail vehicle operation, due to the differences in the real transport passenger flow on the adjacent sides, the key degree of the station with the same degree value is also different, so an index is needed to reflect its important position in the actual operation and network structure at the same time.

$$C_i = w_{ix} + w_{xi} \tag{6}$$

$w_{ix}$, $w_{xi}$ represents the cross-section passenger flow between node $i$ and its neighbor node $x$.

(5) Weighted node median: The weighted node median is the ratio of the total passenger flows on all sections of the shortest path through node $i$ to the total passenger flows on all sections of the shortest path in the network. It contains the information of the number of nodes passing through node $i$ in all shortest paths in the whole network, and gives different weights to each shortest path. That is, the sum of passenger flows in sections on the path, so it can more truly reflect the capacity of the station to load passenger flows.

$$C_C(i) = \frac{\sum_{o,d \in V} \sum_{i \neq o,d (\sum_{e \in R_{od}} w_e) \varphi_i(o,d)}}{\sum_{o,d \in V} \sum_{i \neq o,d} \sum_{e \in R_{od}} w_e} \tag{7}$$

Where $R_{od}$ is the shortest path in OD, $e$ is an interval of $R_{od}$, $w_e$ is the up-down cross-section flow of interval $e$, and if $i \in R_{od}$ then $\varphi_i(o,d) = 1$, otherwise $\varphi_i(o,d) = 0$. 
(6) Weighted node vulnerability $V_i$: The weighted node vulnerability is the ratio of the sum of the whole network OD quantity and the shortest circuit OD quantity passing through node $i$ to the sum of the whole network OD quantity. The passenger flow loss caused by the damage of the node can truly reflect the influence of the station on the passenger flow.

$$V_i = \frac{\sum_{e \in R_{od}} w_e \cdot \sum_{e \in R_{od}} w_e}{\sum_{e \in R_{od}} w_e}$$ (8)

### 2.3 Network Passenger Flow Allocation Model

The model is built to predict the route selection of passengers in the rail transit network, which is the basis of analyzing the node importance after passenger flow loading, and also the key technology of the subway separation system. This paper considers passenger travel time, comfort and transfer times to establish a generalized road resistance function model to simulate passenger path selection resistance:

$$L_{od}^w = \sum_{t \in R_{od}} d_t + \sum_{t \in R_{od}} l_t + \varepsilon_{R_{od}}$$ (9)

$R_{od}^w$ indicates that the $w$th effective path is selected in the OD pair, $L_{od}^w$ is the path resistance of the selected path, and $d_t$ indicates the time when one line is transferred to another line in the path, and $l_t$ is the ride on the path. Time, $\varepsilon_{R_{od}}$ is a random error value.

According to the generalized road resistance function model, the dijkstra algorithm search is performed by matlab software programming, and the shortest path between each OD pair can be obtained. Finally, the single path has all the passenger flow allocation.

### 3 COMPREHENSIVE EVALUATION OF NODE IMPORTANCE

#### 3.1 Establish an Evaluation Matrix

Establish an evaluation matrix. According to the calculation formula of each index, each index value is obtained and the evaluation matrix is obtained.

$$X_{N \times M} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$ (10)

In the matrix $X_{N \times M}$, $x_{ij}$ represents the value of the $j$-th indicator of the $i$-th node, $N$ represents a total of $n$ nodes in the network, and $M$ represents an index defining $m$ evaluation importance. The ideal solution in the evaluation of node importance is the most important urban rail transit node in the expectation. The index values of the nodes are better than the corresponding indicators in other locations in the urban rail transit network (take the maximum value of each indicator).

#### 3.2 Gini Coefficient

The Gini coefficient is an important indicator defined by the lorenz curve to measure the difference in income distribution (that is, the distribution of wealth among members of a society).

(ZHANG, 2015; WANG Lei, 2017) The larger the value, the more uneven the income distribution of members of society, the greater the gap between the rich and the poor in society; On the contrary, if the value is smaller, it indicates that the income distribution of social members is more even, that is, the gap between the rich and the poor is smaller, so the Gini coefficient of the $m$-th indicator is defined as:
\[ \xi_m = \frac{\sum_{i=1}^{N} \sum_{k=1}^{N} |x_{im} - x_{km}|}{2 \sum_{i=1}^{N} \sum_{k=1}^{N} x_{km}} = \frac{\sum_{i=1}^{N} \sum_{k=1}^{N} |x_{im} - x_{km}|}{2N \sum_{i=1}^{N} x_{im}} \]  

(11)

The value ranges from \([0, 1]\). 1 indicates that the value of the indicator is completely dissimilar, and 0 indicates that the value of the indicator is completely similar. In the problem of node importance evaluation of urban rail transit network, the six indicators used to evaluate the importance of nodes can be regarded as members of society in the calculation of Gini index, and the attribute values of each index can be regarded as members of society in the calculation of Gini index. The wealth value (ie income), so the Gini index of each indicator in the urban rail transit network node importance evaluation can be solved by the above formula.

3.3 Kendall Coefficient

The Kendall coefficient is a correlation coefficient used to measure the degree of correlation of multi-column level variables (DAS S, 2017). For two columns of variables with \(N\) elements \(Z_{m}^C = [x_{1m}, x_{2m}, \ldots, x_{Nm}]^T\) and \(Z_{t}^C = [x_{1t}, x_{2t}, \ldots, x_{Nt}]^T\) the ith variable The values are \(x_{im}, x_{it}(1 \leq i \leq N)\). Suppose \(X_{R,m}\) and \(X_{R,t}\) are the sorted values of \(x_{im}\) and \(x_{it}\) in \(Z_{m}^C\) and \(Z_{t}^C\), respectively, then \(X_{R,m}\) and \(X_{R,t}\) correspond to the sort value variable. With the set \(X_R\) that makes up the variable pair, the ith variable pair of the variable pair is \((X_{R,m}, X_{R,t})\). Therefore, the Kendall correlation coefficient between the \(m\)th and \(t\)th indicators can be defined as:

\[ \eta_{mt} = \frac{N_{cc} - N_{dc}}{\sqrt{(D - \sum_{i=1}^{N} \sum_{j=1}^{N} (N_{Tj}^2(N_{Tj} - 1))) / 2}} \]  

(12)

Where: \(D = \frac{N(N-1)}{2}\) \(N_{cc}\) and \(N_{dc}\) respectively represent the variable \(X_R\) of the variable pair, the variable order value is equal, the number and the variable value are not equal, and the variable pair number and the variable value are sorted. The number of unequal variables; \(N_{Ti}\) and \(N_{Tk}\) represent the number of variables having the same variable value in the variables \(Z_{m}^C\) and \(Z_{t}^C\), respectively.

Therefore, the overall Kendall correlation coefficient of the \(m\)th indicator and other indicators can be defined as:

\[ \eta_m = \sum_{t=1}^{M} \eta_{mt} / M \]  

(13)

When the Kendall coefficient of the index \(m\) is 1, it indicates that the index has a consistent rank correlation with other indicators; and when the Kendall coefficient is 0, it indicates that the indicator is independent of other indicators.

In summary, the Gini coefficient and the Kendall coefficient can be used to measure the conflict between the contrast strength of the evaluation index and the evaluation index. Therefore, the integrated Gini coefficient and Kendall coefficient can be used to determine the objective weight of each indicator, that is, the indicator. The objective weight of \(m\) can be expressed as

\[ \omega_m = \frac{\xi_m(1-\eta_m)}{\sum_{i=1}^{M} \xi_i(1-\eta_i)} \]  

(14)

4 ANALYSIS OF KEY POINTS IDENTIFICATION APPLICATION OF URBAN RAIL TRANSIT NETWORK

By the end of 2017, Xi’an City has opened three rail transit lines with a total operating mileage of
91.35km, forming the city's main skeleton transportation network, connecting the east, west, north, southeast and southwest of Xi'an. There are 63 stations and 3 transfer stations for rail transit 1, 2 and 3. The overview of the operation of each line during peak hours is shown in Table 1:

<table>
<thead>
<tr>
<th>Line</th>
<th>Length/km</th>
<th>Maximum number</th>
<th>TIME (7:30-9:30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rated load capacity(10,000 person/h)</td>
</tr>
<tr>
<td>NO.1</td>
<td>25.4</td>
<td>1880</td>
<td>3.78</td>
</tr>
<tr>
<td>NO.2</td>
<td>26.8</td>
<td>1880</td>
<td>2.62</td>
</tr>
<tr>
<td>NO.3</td>
<td>39.15</td>
<td>1880</td>
<td>3</td>
</tr>
</tbody>
</table>

According to the operation of the peak period of each line of Xi'an rail transit, the two-way cross-section passenger flow of each section of the morning peak on October 26, 2017 is selected as the weight of the road segment, and the load weight is constructed. The directed weighted network topology is shown in Figure 1.

![Figure 1: Xi'an Urban Rail Transit Network](image)

First, calculate the value of each index according to the definition of each index, so as to form the evaluation matrix of node importance degree of rail transit network, and then carry out dimensionless processing on it; Finally, according to the Gini coefficient, Kendall coefficient and objective comprehensive weight calculation formula, the results are shown in table 2.

Kendall coefficient is a kind of correlation coefficient used to measure the correlation degree of multi-rank variables. Weighted node vulnerability measures the importance of a node by the degree of passenger flow reduction after the node is damaged. Other indicators mainly measure the importance of nodes from the perspective of topological structure and passenger flow concentration. The correlation with the other 5 indicators is not significant. This is consistent with the calculated Kendall correlation coefficient of the index.

After the comprehensive gini coefficient and Kendall coefficient, the comprehensive weight values of the six indexes obtained are 0.2118, 0.0643, 0.0480, 0.2294, 0.1910 and 0.2556.
The practical urban rail transit network shown in figure 1 is taken as an example to illustrate the guiding significance of the proposed key node identification method for practical work. In fact, each node of the urban rail transit map shown in figure 1 represents a station in the actual line network. According to the analysis of subway passenger flow data, large passenger flow will have a huge impact on the safety and normal of the station, which may cause the occurrence of accidents. In the actual work, due to the limited resources, it needs to be maintained according to the degree of importance. Therefore, the practical significance of the node identification method of urban rail transit network studied in this paper lies in identifying important nodes, which is conducive to guiding and strengthening daily maintenance and operation management, formulating fault emergency plan, and ensuring the normal operation of rail transit network. As can be seen from the results in table 2, Beidajie station is at the most important position and is the key station for daily maintenance. In case of failure, the station should be given priority for maintenance.

A large number of passenger flows in the peak period is the main factor that significantly increases the vulnerability of the station, and it will also need to be paid close attention to in the operation and management. The simple structure topological index can only reflect the role of the station in the structure, and can not reflect the state of the passenger flow in rail transit agglomeration, and has practical significance. In urban rail transit network, how to quantitatively describe the degree of difference between stations or sections in the network is an important aspect to study network heterogeneity.
At the same time, this paper uses the Etaroby index, which describes the degree of agglomeration in the market. (Adamic I, 2002) Here, the vehicle passenger flow aggregation index is expressed over a period of time, which is defined as the passenger flow through a station in a period of time. The ratio is given by:

$$E.I = \sum_{i=1}^{N} \left( \frac{P_i}{P} \right)^2$$  \hspace{1cm} (14)

$P_i$ represents the passenger flow through node $i$ during the defined time period, and $P$ represents the passenger flow in the entire network in the defined time period.

The passenger flow concentration of the Xi’an urban rail transit network is as shown in Fig3:

![Fig 3 Passenger flow agglomeration effect](image)

It can be found from Fig. 2 that there are a large number of stations in the online network where there are a large number of passengers, including North Street Station, Xiaozhai Station, Tonghuamen Station, and Zhonglou Station. The actual situation is also basically consistent with the assessment of the results from important nodes and passenger flow. The effect of the node relative ratio, the node importance is consistent, that is, the more obvious the passenger flow agglomeration phenomenon of the network node, the more important the node is in the network. Therefore, the evaluation results show that the identified key urban traffic sections are accurate and effective. It shows that the passenger flow has a serious agglomeration phenomenon in the network communication, the
difference between the stations is obvious, and the line with long operation time is particularly obvious. Stations with a large number of passenger flow clusters should give greater weight in the importance assessment, so the importance of the station will be more sensitive to passenger flow, and it can more accurately reflect the difference in the importance of the stations in the network caused by the change of passenger flow, and To some extent, it reflects the influence of the station on the passenger flow network. It can be seen that the station with higher importance is basically a transfer station, a station adjacent to the transfer station, and a station with large passenger flow such as tourist passenger flow and external traffic flow. These stations are concentrated in the central area of the network, highlighting The contribution of station location and station function to station importance is consistent with the actual situation. Focus on these stations to ensure the reliability of the overall network of urban rail transit operations.

5 CONCLUSION

In the evaluation of the importance of complex network nodes, the evaluation indicators and evaluation methods used in decision-making when judging node importance are not uniquely determined, so the evaluation results are different. In the important evaluation price of urban rail transit network, not only the topology of the network itself but also the network passenger flow is also a key factor affecting the importance of the network, and the importance of each node is evaluated in turn. The conclusions of the analysis of the example are as follows:

1) The larger the Gini coefficient of the indicator, the greater the difference in the value of each node on the indicator, the greater the proportion of the indicator in the comprehensive importance assessment of the node; conversely, the smaller the proportion of the indicator.

2) The smaller the Kendall coefficient of the indicator, the smaller the correlation between the indicator and other indicators, the greater the proportion of the indicator in the comprehensive importance evaluation of the node; conversely, the smaller the proportion of the indicator.

3) The proposed method can better identify the key nodes of the urban rail transit network, avoid the subjectivity of artificially determining the weight, and the evaluation results are closer to the actual situation of the urban rail transit network operation, which can be better and more accurate. Identifying the importance of urban rail transit network nodes, combined with the characteristics of the topology structure and the characteristics of passenger flow agglomeration, the station can be maintained in advance, which is conducive to improving the efficiency of operation management and maintaining important stations.

In this paper, the passenger flow distribution model, AFC data processing flow and topological network indicators are applied to the Xi'an subway network. It is found that the phenomenon of network passenger flow agglomeration in Xi'an subway peak period is significant, mostly concentrated in the transfer station. This paper analyzes the importance ranking of urban rail transit stations. The link cost is used as the weight of the road segment, which can directly reflect the time cost and congestion cost between the road segments. The importance evaluation matrix is used as the evaluation method to obtain the station importance ranking. It is also proved from the results that the station with higher importance is basically a transfer station, a station adjacent to the transfer station, and a station with a large passenger flow such as tourist passenger flow and external traffic flow, which is also located at the center of the network and bears important importance. Features. For stations with higher importance, daily maintenance and operation management should be strengthened,
and emergency response plans should be formulated to ensure the normal operation of the rail transit network. The identification of important stations can provide a basis for the decision-making of urban rail transit operators, and has strong practical value.

**REFERENCES**


Zhao Yizhen, Wang Zulin, Zheng Jing, Guo Xujing. Using the importance contribution matrix to determine the most important node in communication network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2009(9): 1076-1079


