Research on evaluation algorithm of key nodes in urban road traffic network based on complex network

Canshi Zhu¹, Xiaoyang Wang²

¹Information Engineering College, Xijing University, Xi’an 710123, China
²Xi’an Satellite Control Center, Xi’an 710043, China

Corresponding author and e-mail: Canshi Zhu: 3399241158@qq.com

Abstract. Urban road traffic planning is a complicated project. If this project is abstracted as a complex network, the smoothness of important nodes in the network will have a significant impact on the smooth flow of the entire city. How to evaluate the importance of urban road traffic network nodes optimizing key nodes as much as possible is the key to transportation planning. This paper proposes a complex network node "structural hole" impact matrix to evaluate key nodes, constructs key node evaluation criteria, and simulates and analyses smaller-scale networks. The results show that this method can accurately and effectively evaluate the criticality of nodes. The use of its research results can provide a decision basis for alleviating the current urban traffic congestion.

1. Introduction

With the rapid development of social economy, the scale of cities continues to expand, the urban population and vehicles increase sharply, urban road traffic is facing huge challenges, and the relative smoothness of the road traffic network has attracted more and more attention from scholars. Traffic congestion and smoothness is one of the major concerns of the government and citizens. Urban road traffic planning is a complex project. If this project is abstracted as a complex network, then the criticality of urban road traffic nodes will be evaluated from the perspective of the complex network, and then corresponding measures will be taken at traffic nodes of higher importance (establishing traffic etc.) Is very important to solve the phenomenon of urban vehicle congestion and improve the timeliness of urban traffic.

At present, different methods are needed to evaluate the key nodes of networks of different sizes. The criticality of complex network nodes can be evaluated from four aspects of the network, namely local attributes, global attributes, network location and random walk. The local attributes of the network mainly consider the information of the node itself and its neighbors. Due to the simple calculation and low time complexity, it can generally be used in large networks. For example, Wang Jianwei believed that the importance of nodes in the network is related to the degree of the node itself and the degree of the node's neighbors. A simple method for measuring the importance of nodes is also proposed. Taras Agryzkov et al. Use PageRank algorithm to establish the classification of nodes. Combined with the idea of random walk intermediary centrality, a centrality measurement method is proposed for urban network nodes Sort. Kitsak et al. First proposed the idea that node importance depends on its position in the entire network in 2010, and K-core decomposition was used to obtain a node importance ranking index (k-shell), which has low time complexity. It is suitable for large
networks, and can more accurately identify the most influential nodes in the spread of disease than degrees and intermediaries.

The above methods only evaluate the node importance or rank the nodes from a certain aspect. In fact, the importance of network nodes is not only related to the local attributes of the nodes, but also closely related to their location in the network and the degree of interdependence between the nodes. Related8. It is necessary to use multiple importance indexes of nodes to carry out comprehensive evaluation from different angles. Therefore, to evaluate the importance of a node, it is necessary to consider not only the attributes of the node itself, but also the global attributes of the node. Literature [10] put forward the concept of importance contribution matrix on the basis of considering node local importance and combining node location information (median) to characterize the interdependence between nodes and use it to evaluate the importance of nodes.

In fact, the key nodes of a complex network should have the following characteristics: nodes that act as a "bridge" in each local area network. Therefore, the study of the importance of nodes in complex networks should be based on the local centrality measurement method of the nodes and their neighbor structure holes, comprehensively considering the number of neighbors of the node and the topology between the neighbors, while reflecting the degree attribute and "bridge" attribute. Literature [12] also believes that the focus of the key node sequencing problem cannot be limited to the core nodes in the network, nor can it ignore the nodes at the position of structural holes. Reference 13 pointed out that the node tightness can better reflect the influence of the node on other nodes, and also reflect the difference of the node in the network topology.

This paper proposes a complex network node "structure hole" impact matrix to evaluate key nodes and construct key node evaluation criteria. This method takes into account the local information (constraint coefficients) and global information (closeness) of the nodes. The "hole" impact matrix can more accurately and efficiently evaluate key nodes in the network. And on this basis, an evaluation algorithm (Cc-Burt) based on the "structure hole" influence matrix is designed, and then on the "Karate network", the algorithm is respectively compared with degree centrality and intermediary centrality. ) Compared with the N-Burt algorithm 11, and finally uses the SIR model to simulate the propagation process, which shows the effectiveness and reliability of the algorithm.

2. Key node evaluation theory

2.1. Theoretical basis

Suppose the undirected unauthorized network is represented by graph \( G = (V, L) \), where: graph \( G \) contains \( n \) nodes, \( m \) edges, \( V = \{v_1, v_2, ..., v_n\} \) represents the set of all nodes in the network, and \( L = \{L_1, L_2, ..., L_m\} \subseteq V \times V \) represents the set of edges in the network. From set \( L \), construct adjacency matrix \( A = [a_{ij}]_{n \times n} \) of network \( G \), where:

\[
a_{ij} = \begin{cases} 
1, & \text{if } i \text{ and } j \text{ are connected} \\
0, & \text{if } i \text{ and } j \text{ are not connected}
\end{cases}
\]  

(1)

Definition 1: Closeness centrality

Let \( N \) denote the number of nodes in a complex network, and \( d_{ij} \) denote the shortest distance from node \( i \) to node \( j \).

Then: the tightness is expressed as the reciprocal of the average of the sum of the distances from node \( i \) to all other nodes in the network.

\[
C_i = (N - 1) / \sum_{j=1}^{N} d_{ij}
\]  

(2)
From a definition point of view, tightness is used to measure the ability of nodes in the network to exert influence on other nodes through the network. The greater the node's compactness, the greater the node's position in the center of the network, and the more important it is accordingly. The tightness depends on the topology of the network. For a network similar to a star structure, it can accurately find the central node.

Definition 2 structural holes

If there is no direct connection between two individuals or two groups in the network, and there is no indirect redundant relationship between them, the obstacle between the two is the structural hole. Burt proposed to calculate the network constraint coefficients of structural holes to measure the network closure and structural holes.

$$p_{ij} = \frac{a_{ij}}{\sum_{j \in \Gamma(i)} a_{ij}}$$ (3)

Among them, $p_{ij}$ represents the proportion of the energy invested by the node to maintain the neighbor relationship with node $j$ in the total energy, and $\Gamma(i)$ represents the set of neighbor nodes of node $i$. Then the calculation expression of the network constraint coefficient of node C5 is:

$$C_i = \sum_{j \in \Gamma(i)} (p_{ij} + \sum_{q} p_{iq}p_{qj})^2 \quad q \neq i, j$$ (4)

In the formula, $q$ is an indirect node connecting node $i$ and node $j$, and $p_{iq}$ and $p_{qj}$ are the proportion of the energy invested by nodes $i$, $j$ and the common neighbor $q$ in maintaining the relationship. The smaller the network constraint coefficient $C_i$, the greater the degree of structural holes, the greater the influence of nodes in information dissemination, and the more important the position of nodes.

2.2. "Structure hole" influence matrix construction

A complex network is a unified whole composed of nodes and their edges. Any one of the nodes does not exist in isolation, but is constrained and affected by other nodes in the network. That is to say, in a network where no isolated nodes exist, any node can influence neighbor nodes by connecting edges. The adjacency matrix of the network reflects the directly connected relationship between nodes, and the most direct importance relationship between nodes exists in the interaction between adjacent nodes. This mutual influence relationship between nodes can be described in the form of importance contribution matrix. The compactness index of the centrality theory of complex networks measures the importance of nodes from the perspective of their influence on other nodes in the network. At the same time, literature believes that the importance of network nodes is weakly related to the degree of nodes. Based on this, drawing on the idea of the node importance contribution matrix, using the adjacency matrix relationship and combining the node tightness, a node influence factor matrix $H_A$ is established, which is defined as follows:

$$H_A = \begin{bmatrix}
1 & a_{12}C_c(2) & \cdots & a_{1n}C_c(n) \\
 a_{21}C_c(1) & 1 & \cdots & a_{2n}C_c(n) \\
 \vdots & \vdots & \ddots & \vdots \\
 a_{n1}C_c(1) & a_{n2}C_c(2) & \cdots & 1
\end{bmatrix}$$ (5)

Among them, $H_A(i,j) = a_{ij}C_c(j)$ represents the influence factor of node $j$ on node $i$, and 1 on the diagonal of the matrix means that the influence factor of node on itself is 100%. It can be seen that the node impact factor matrix reflects the degree of influence of any node on other nodes in the
network. However, the importance of a node depends on two factors: the location information of the node (global importance) and the neighboring information of the node (local importance). The structural hole theory not only reflects the mutual influence relationship between nodes, but also reflects the topology structure between neighboring nodes. From the perspective of calculating the constraint coefficient of the measured node, it also reflects the degree attribute and "bridge" attribute of the node. Therefore, in this paper, the node constraint coefficients are used to construct the importance influence between nodes. As the adjacent information of the nodes, combined with the node influence factor matrix $H_A$ (global importance), the "structure hole" influence matrix $H_C$ can be obtained as:

$$H_C = \begin{bmatrix}
C_i^{-1} & a_{i1}C_i(2)C_2^{-1} & \cdots & a_{in}C_i(n)C_n^{-1} \\
a_{i1}C_i(1)C_1^{-1} & C_1^{-1} & \cdots & a_{in}C_i(n)C_n^{-1} \\
\vdots & \vdots & \ddots & \vdots \\
a_{i1}C_i(1)C_1^{-1} & a_{i2}C_i(2)C_2^{-1} & \cdots & C_n^{-1}
\end{bmatrix}$$

(6)

Among them, $C_i^{-1}$ represents the reciprocal of the node's constraint coefficient, and $H_A(i, j) = a_{ij}C_i(j)C_j^{-1}$ represents the influence value of node $j$ on node $i$. It can be seen that the influence value of a node on its neighbors is negatively correlated with its constraint coefficient, and positively correlated with its tightness. Effect, in turn, the greater the impact it has on neighboring nodes. Here, the importance of node $i$ is defined as $M_i$:

$$M_i = \frac{1}{n} \left( C_i^{-1} + \sum_{j=1,j\neq i}^{n} a_{ij}C_i(j)C_j^{-1} \right)$$

(7)

Among them: $M_i$ reflects the average of the sum of the influence values of the importance of all nodes adjacent to node $i$ and the reciprocal of the constraint coefficient of node $i$ itself. From equation (7), it can be seen that the importance of a node depends on Due to its own constraint coefficient, the closeness of adjacent nodes and the size of the constraint coefficient, this method can more accurately evaluate the importance of nodes.

2.3. Evaluation algorithm (Cc-Burt) based on "structure hole" influence matrix

The simplest and most direct form of the dependency relationship between interconnected nodes is between adjacent nodes. This algorithm comprehensively considers the node's tightness (global information) and constraint coefficient (local information), and proposes an evaluation algorithm based on the "structure hole" influence matrix, which can obtain more accurate evaluation results.

The algorithm steps are as follows:

Input: Complex network adjacency matrix $A = [a_{ij}]_{n \times n}$

Output: importance $M_i$ of node $i$

Step 1 According to definition 1, calculate the tightness of the node and write the matrix of node influence factors (5);

Step 2 According to definition 2, calculate the constraint coefficient of the node and write into the "structure hole" to affect the matrix formula (6);

Step 3 According to equation (7), calculate the importance $M_i$ of all nodes in the network

The importance of the nodes obtained in step 3 is sorted in order from large to small, and the node in front is the relatively important node.

It can be seen from the above algorithm steps that the time complexity of the entire algorithm depends on the calculation of closeness, because the shortest path between all pairs of nodes needs to
be calculated in closeness. This paper uses the shortest path faster algorithm (SPFA, shortest path faster algorithm) applied in [8] to calculate the node tightness. The complexity of the entire algorithm can reach \( O(n^2) \), which has great advantages over the classic Floyd algorithm.

3. (Cc-Burt) algorithm validity verification and analysis

3.1. Transportation network abstraction and data extraction

First, select the verification network in to verify the feasibility of the algorithm. As shown in Figure 1, although the structure of the traffic network graph is simple, it contains "bridge" nodes and the nature of the community. Use this graph to verify the network can explain the effectiveness of the algorithm to a certain extent.

![Figure 1. Schematic diagram of the verification traffic network.](image)

Table 1. The top 5 importance ranking results of the verification network.

| Algorithm     | Rank 1   | Rank 2   | Rank 3   | Rank 4   | Rank 5   |
|---------------|----------|----------|----------|----------|----------|
| Cc-Burt       | i(1.06)  | G(0.94)  | A(0.54)  | F(0.51)  | E(0.43)  |
| K             | i(6)     | A(4)     | G(4)     | B(3)     | F(3)     |
| Betweenness   | F(0.51)  | G(0.49)  | i(0.47)  | E(0.18)  | A(0.15)  |

3.2. Network node analysis and algorithm validity verification

From Figure 1, we can see that the verified transportation network has two transportation hubs, namely nodes B1 and G, and node F is similar to a "bridge" node to connect the two hubs, so from the perspective of the network topology, Intuitive feeling is that the three nodes i, G, F must occupy a certain degree of traffic importance. Compared with the degree-centrality method, the degree-centrality method cannot evaluate the importance of bridge connection points, and cannot distinguish the propagation differences of nodes with the same degree. The Cc-Burt algorithm can distinguish the importance of nodes G and A, F and B, and E as a bridge node of K is more important than B. Compared with the Betweenness algorithm, although F has a "bridging" effect, it is not a hub center. Betweenness ranks F at the forefront. Obviously, it does not consider the local information of nodes, and node i is not only the hub center, but also the structural hole of node i. The constraint is smaller than node F, so it is more reasonable for C-Burt algorithm to rank node i in front of F. The N-Burt algorithm only considers the local information of the nodes and does not consider the global information. Although nodes i and G both serve as the centers of their respective communities, node i is more compact than G, indicating that i has greater control over other nodes. And therefore of greater importance than G. In summary, the C-Burt algorithm in this paper not only considers the local topology but also integrates the global information, making the ranking results better and more reliable.

In order to further analyze the difference of the evaluation results of each algorithm, the nodes that appear in the top 10% of each algorithm at different times are selected and analyzed based on the SIR
propagation model. These top 10% nodes in each algorithm reflect the emphasis of different algorithms.

The curve is shown in Figure 2:

![Figure 2. The top 10% of nodes in the algorithm spread the influence.](image)

It can be seen from Figure 2 that the difference nodes mined in the Cc-Burt algorithm in this paper and other algorithms have stronger propagation capabilities than the nodes mined by the degree centrality method, Betweenness algorithm and N-Burt algorithm. It further illustrates the accuracy and precision of the algorithm in this paper.

4. Conclusions
The importance evaluation of complex network nodes is the basis for mining key nodes. Practice has proved that a small number of nodes in the network have strong control and influence on the network. Once the key nodes are found, effective measures can be taken on the network. However, the importance of nodes in the network cannot be measured by only one indicator, and the importance of nodes needs to be comprehensively evaluated locally and globally. This paper presents a method for evaluating the key nodes of the "structure hole" influence matrix of complex network nodes. This method builds a "structure hole" influence matrix to evaluate the importance of nodes based on the comprehensive consideration of the local and global information of the nodes. Then the effectiveness of the algorithm is verified, and further comparison with degree centrality, Betweenness and N-Burt algorithms is performed on a small-scale "Karate network". The ranking results show that the results of the Cc-Burt algorithm in this paper have The higher the accuracy, the higher the accuracy of the assessment of importance. Finally, the SIR model is used to simulate the propagation process, and compared with other methods. It shows that the algorithm in this paper has better advantages in the ability to mine key nodes, and further shows the effectiveness and accuracy of the algorithm, which can be more accurate and efficient. Critical nodes in the evaluation network.

References
[1] Xiao Weidong, Tan Wentang, Ge Bin, Li Fangfang. A fast evaluation method of network node importance [J]. System Engineering Theory and Practice. 2013, 33 (7): 1898-1904.
[2] Xiao-gang JIN, Yong MIN. Modeling dual-scale epidemic dynamics on complex networks with reaction diffusion processes [J]. Journal of Zhejiang University-SCIENCE C (Computers & Electronics). 2014 15 (4): 265-274.
[3] Liu Jianguo, Ren Zhuoming, Guo Qiang, Wang Binhong. Research progress of node importance ranking in complex networks [J]. Journal of Physics. 2013, 62 (17): 1-10.
[4] Ren Zhuoming, Shao Feng, Liu Jianguo, Guo Qiang, Wang Binhong. Research on network node importance measurement method based on degree and agglomeration coefficient [J]. Journal of Physics. 2013, 62 (12): 1-5.
[5] Xu Lixiong, Liu Junyong, Liu Yang, Liu Youbo, Gou Jing, Bazargan Masoud. Comprehensive evaluation of the classification of node importance [J]. Chinese
Journal of Electrical Engineering. 2014, 34 (10): 1609-1617.

[6] Lu L Y, Zhang Y C, Yeung C H, Zhou T. Leaders in social networks, the Delicious case. PLoS One. 2011,6: e21202.

[7] Kleinberg J M. Authoritative sources in a hyperlinked environment [C] // Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’98, Society for Industrial and Applied Mathematics. Philadelphia: [sn], 1998: 668-677.

[8] Fan Wenli, Liu Zhigang. Sorting method of node importance in complex network based on transmission efficiency matrix [J]. Journal of Southwest Jiaotong University. 2014,49 (2): 338-342.

[9] Yu Hui, Liu Zun, Li Yongjun. Comprehensive evaluation method of node importance in complex networks based on multi-attribute decision-making [J]. Acta Phys. Sin. 2013, 62 (2): 1-9.

[10] Zhao Yihuan, Wang Zulin, Zheng Jing, Guo Xujing. Using the importance contribution matrix to determine the most important nodes in the communication network [J]. Journal of Beijing University of Aeronautics and Astronautics. 2009,35 (9): 1076-1079.

[11] Su Xiaoping, Song Yurong. Using neighborhood "structural holes" to find the most influential nodes in social networks [J]. Journal of Physics. 2015,64 (2): 020101.

[12] Han Zhongming, Wu Yang, Tan Xusheng, Duan Dagao, Yang Weijie. Sorting of key nodes in a complex network oriented to structural holes [J]. Journal of Physics. 2015, 64 (5): 1-9.

[13] Wang Haoxiang, Zeng Shan, Liu Huanyang. Analysis of node importance in virtual social networks [J]. Journal of Shanghai Jiaotong University. 2013, 47 (7): 1055-1059.

[14] Burt R. S. Structural holes. Cambridge: Cambridge University Press. 1992.

[15] Li Fang, Hu Bin, Di Peng. Dynamic invulnerability optimization of scale-free network based on limited resource model [J]. System Engineering and Electronic Technology. 2012,34 (1): 175-178.