# Research on edge importance measure in double-layer small-word networks

**Luo Zhangkai, Pei Zhongmin, Wang Xinmin**

Science and Technology on Complex Electronic System Simulation Laboratory, Space Engineering University, Beijing, China, 101416

luo_zhangkai@126.com, peizhongmin@tsinghua.org.cn, 13391709986@163.com

**Keywords:** two-layer network; edge importance; Newman algorithm; edge betweenness

**Abstract:** to overcome the problem that edge sorting methods based on the centrality index cannot accurately sort the edge according to the impact on information transmission, this paper puts forward a method that combines the community structure excavation algorithm and the edge betweenness index to measure the importance of each edge. In the proposed method, the community structure is firstly excavated based on the Newman algorithm, which realize the redefinition of the number of community. Then, the edge betweenness of each edge is calculated. Based on the community structure and the betweenness of each edge, the edge sorting method in two-layer network is designed. Finally, the network efficiency is used as the evaluation index to evaluate the performance of the proposed method, and the maximum degree method , the maximum edge betweenness method and random method are used as contrast methods. Simulation results demonstrate that the proposed method performs better than these three methods and can better locate the important edges that have a large impact on the information transmission in the two-layer network.

## 1. Introduction

With the rapid development of computer technology and communication technology, the process of network is accelerating, and the network can be seen everywhere in human society[1], Such as social networks QQ, wechat, Weibo, TikTok and other platforms, greatly accelerated the efficiency of real-time information transmission, and to a large extent changed the organizational structure of human society. At the same time, common systems in nature and human society, such as biological, social, electricity, transportation systems and so on. In addition, there are many abstract systems, such as the macroscopic motion of a large collection of particles in statistical mechanics, all of which can be described through networks. With the advent of the 5G era, the internet of everything has become the mainstream of development, and more and more devices are connected to the network,which indicates that human activities are increasingly dependent on the network[2]. Since the discovery of the small-world networks and scale-free networks[3] features, network science and technology have developed vigorously, and more and more scholars use the theories and methods of complex networks to study the problems in various networks. All kinds of entities in the network could be abstracted as the nodes in the network, and the relationship between the entities could be abstracted as the connected edges in the network. The strength of the relationship is measured by the weight of the edges. Thus, the entity network could be totally abstracted as the network represented by the nodes and the edges. Therefore, nodes and connecting edges are the crucial constituent units in the network, and their properties will reflect the properties of the network to some extent. For example, in QQ instant messaging network, the server node has the largest degree value, when compared with the mobile terminal, it has a greater impact on the communication than the mobile terminal. For the edge between the community, which has a greater impact on the communication and network connectivity between the community.

In real networks, except for uniform networks, most networks have heterogeneity, that is, different connected edges or different nodes in the network have different effects on the various properties of the network, such as connectivity, damage resistance, and information transmission efficiency. In most cases, only a small number of nodes and connections play a key role and have a great influence

  DOI: 10.25236/iwmecs.2023.014

on the structure and function of the network; while many other nodes and connections have little effect on the function and structure of the network. The mining of the important nodes and links in the network is an effective method to change the nature of the network, for example, the cascade faults in the power grid[4], diseases spreading in biological networks and innovation and the spread of rumors on social networks[5]. By identifying the key nodes and edges, we can quickly control the propagation behavior in the network and minimize the loss. In addition, by strengthening the protection of important edges and nodes, the resistance of network network can be effectively be enhanced[6].

In the field of network science, there are many methods to identify key nodes, such as node importance evaluation method designed based on the nature of nodes themselves, such as degree centrality method based on degree value, medium number centrality method based on node medium number, and *k*-shell decomposition method[7], the proximity centrality method et al. In addition, node importance can be judged combined with network structure, such as the eigenvector centrality method based on the eigenvectors of the adjacency matrix[8-9], combining with information entropy in information theory[10], in which the "influence" of the nodes on the information transmission is analyzed, thus the node sorting is realized. However, when the network is large, it is difficult to measure the importance of nodes in the network by relying only on a single node index. To over the problem, in [10], the degree of the node, the negative effect of the node clustering coefficient and the positive effect of the second-order neighbor node clustering coefficient are integrated to design the measurement index. In [11], the node degree, node location information and the node importance model of network community structure are comprehensively considered to realize the evaluation of node importance.

Compared with the studies of the evaluation method of key nodes in the network, there are relatively few studies on how to measure the importance of connecting edges. However, it is also important to mine important links in the network. For example, when controlling information spreading in the internet, it is unrealistic to delete nodes like the server, for there is a large probability that leads to normal communication interruption, and often some important links can be cut off to block information spreading.

In order to realize the control of information spreading, this paper focuses on the important edge evaluation method in two-layer networks. In the proposed method, we considering the large network size condition, in which there will inevitably be community structures in the network. Therefore, at first, we use the community structure mining algorithm to mining the community structure in the network. Then we define the connection between the clubs as Separate community structure. Secondly, we calculate the number of edge betweenness for each edge in the whole network, and rank the connected edges in each independent community network. Finally, we remove the edges in parallel for each community to test the performance of the proposed method in controlling the information spreading. When there are a lot of edges to remove, the edge with the same rank index number in each community is identified as the same importance, When the number of the edge to remove is small, we treat the edge in the community with more nodes are more important. For example, there are 5 communities and it needs to remove 6 edges. First of all, delete the edges with the same rank index for each community. Then, for the sixth edge, we remove remove the second most important edge in the community with the largest number of edge connections in the network. Finally, the performance of the proposed method is verified through simulations. The simulation results prove that the method of the centrality method and the random edge deletion method can more accurately locate the key edges in the network and effectively interfere with information dissemination.

## 2. Traditional methods

The edge betweenness method, this method characterizes the importance of the edges by the number of the shortest paths passing through the edges $e\ (u, v)$. The more the shortest paths between node pairs pass through edges $e\ (u, v)$, the more important the edge $e\ (u, v)$ is. The specific definition of the edge interface number is as follows:

$$EBC_e = \sum_{i \neq j} \frac{E_{v_i v_i}(e)}{E_{v_i v_i}} \tag{1}$$

$E_{v_i v_j}$ denotes the shortest paths between node $v_i$ and node $v_j$, $E_{v_i v_i}(e)$ denote the edges that pass through the edge $e$. $EBC_e$ denotes the betweenness of the edge $e$. Obviously, the larger the value $EBC_e$, the more the shortest paths passing through the edge, the more important the edge is. In real network, in order to ensure the lowest cost of information transmission or data packet transmission, the edge betweenness is often used to find the least jump number (*i. e.* the shortest path). In addition, it should be noted that the connection edge with the largest betweenness changes in real time. When the edge with the largest betweenness is deleted, the betweenness of the edge of the whole network will change. Therefore, when the network scale is large, a large calculation amount is often needed to continuously discover the edge with the largest betweenness. In addition, deleting the connection edges with the largest betweenness will affect the efficiency of network on information spreading, but have little impact on the connectivity of the network. This feature is useful for controlling the information spreading in the network while ensures the network connectivity.

Maximum degree method, that is, ranked the edge by the product of the degree of two edge end points.

*k*-path centrality is to simulate the random walk process of information spreading, thus describing whether an edge is important by the number of times the information passes through that edge. The *k*-path centrality is defined as:

$$KP_e^k = \sum_{u \in V} \frac{\sigma_u^k \quad e}{\sigma_u^k} \tag{2}$$

It is assumed that information passes at most *k* steps, *u* is the starting node of information spreading, $\sigma_u^k$ representing the number of all paths propagating from node *u*, $\sigma_u^k \quad e$ denotes the number of paths propagating *k* steps from node *u* through edge *e*.

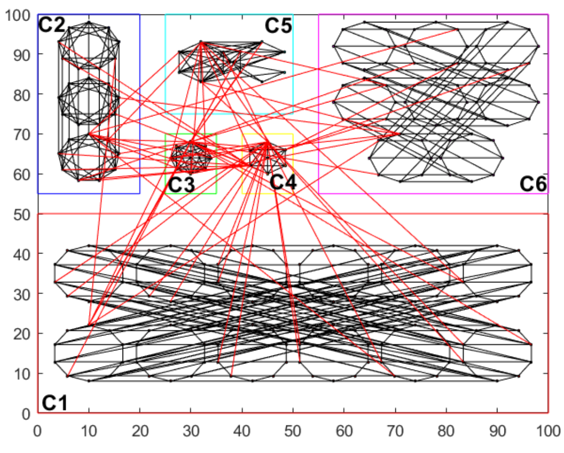


Fig. 1. A Network with a six-community structure

In addition, there are Jaccard coefficient, accessibility index, second-order field index and other indicators to measure the importance of the edge. Its common feature is to directly calculate edge properties in the network. However, these methods have some limitations, such as the network with six community structures shown in Fig. 1 [12-13], where a close connection within communities and sparse connections between communities. When the network scale is larger, there will be a large number of communities in the network, and the community structure will make the traditional edge importance evaluation index useless, such as the edge betweenness. It is obviously that the largest edge betweenness will appear in community C1 or the connected edges with C1 community, and the edge betweenness in other communities or edge between the communities will be small, and if the information spreading source in communities other than the C1 community, disconnecting the edge with the largest betweenness has no impact on the information spreading. Similarly, the same situation occurs for measuring the importance of edges by the size of the product of the edge's endpoint degree value .

As we know, there is obvious community structure in most large networks. In order to achieve effectively control of information spreading by deleting important connections, the community structure of the two-layer network should be taken into account when designing the evaluation index of the importance of network connections. In addition, for a large network, the number of connected edges in the network is far greater than the number of nodes. By deleting an important node, the connect edges are deleted together, that is, deleting a node is equivalent to deleting more edges. Therefore, through delete edges one by one to realize the control of information spreading requires more accurate judgment of the importance of connecting edges.The mainly goal of the proposed method is to reduce the efficiency of network information spreading in the case of deleting fewer edges.

## 3. The principle of the proposed method

In view of the above considerations, the edge importance evaluation method proposed in this paper is mainly divided into three steps:

1) through the community mining of the single-layer network and the double-layer network, we analyze the network structure to obtain the number of communities and the set of connection edges between communities in the double-layer network;

2) Then, redefine the community structure that we treat the edges between communities as a separate community. Thus the number of the community becomes large to $n(n+1)/2$;

3) Calculating the edge betweenness of all edges in the network, and then rank the edges according to their betweenness for each community redefined in the second step.

By dividing the community structure, the network structure can be more accurately understood, and the important links within each community can be located quickly and equally, which can prevent information transmission within all communities and between communities at the same time, which can also avoid the influence of random distribution of information spreading sources in the network.

First of all, in order to mine the community structure of the two-layer network more accurately, in the proposed method, we take the advantage of Newman community structure mining algorithm[13] and assume each node is a community and constantly merging the communities that has edge connected. At the same time, a valuable called modularity represented by $Q$ is used to help us determine the optimal mining of communities, so as to obtain the results of community division. The main steps of the proposed method are as the following six steps:

1) Initialization: Initialize the network as $N$ communities, that is, each node is one community. Make the modularity $Q = 0$, the elements of matrix F and auxiliary vector $a$ are initialized as

$$f_{ij} = \begin{cases} 1/(M), & \text{edge connect node } v_i \text{ and } v_j \\ 0 & ,\text{other} \end{cases} \qquad (3)$$
$$a_i = k_i/(M)$$

where $k_i$ denote the degree of node $v_i$ and $M$ is the total number of edges in the network.

2) Mergers: merge the two communities connected by edges in turn $G_i$ and $G_j$. And calculate the modularity increment as

$$\Delta q_{ij} = f_{ij} + f_{ji} - 2a_i a_j = 2\left(\frac{1}{M} - \frac{k_i k_j}{(M)^2}\right) \qquad (4)$$

Select the merge method that increases the most or reduces the least $Q$ to merge communities (computational complexity is $O(M)$), then, add the rows and columns of the two communities related to the merge and update the elements of the matrix $\boldsymbol{F}$ and auxiliary vector $a$ (computational complexity is $O(N)$), and update $Q$=$Q$ + max $\{\triangle q_{ij}\}$.

3) Termination: repeat step 2 to merge the community until the entire network is merged into one community, which cost up to $N$-1 merger.

Note: the algorithm selects nodes with small degree but have connected edges and merges them into a community. The computational complexity of each cycle is $O(N+M)$, with at most $N$-1 cycles, so, taking the sparse network as an example, the total computational complexity is $O(N^2)$.

4) Find the maximum $Q$ and the corresponding community merges. Let's assume the total network is divided into $n$ communities. After redefining the community, there are $n(n+1)/2$ communities, for example, in the figure below, we assume the whole network is divided into two communities after processed by Newman algorithm, the solid lines represent internal edge and dotted lines represent the edges between communities. According to the principle of the proposed method, the dotted line is divided into a community. In this case, there are three communities in the network.

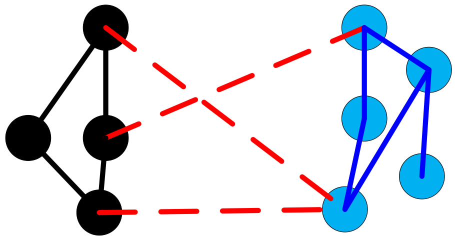

Fig. 2. Society discovery diagram

5) Calculating the betweenness of all edges, and then for each community of the $n(n+1)/2$ communities, the edges are ranked according to the betweenness and marked with the sorted number. In addition, according to the size of the number of edges in the community, the communities are sorted and numbered from large to small.

6) Delete edges: all communities are treated equal when delete edges, which means the importance of the edge with the same sorted number are the same. For example, if one percent of the edges in the network need to be deleted, then for each community, 1% edges are deleted (at least one edge) according to the sorted number in parallel. Notably, when the number of deleted edges is small, that is, the number of deleted edges is less than the number of community, one of the edges is deleted in turn according to the sorted number of community.

## 4. Simulations

To verify the performance of the algorithm, the network efficiency index is used to verify the performance of the algorithm, and the network efficiency is

$$\varepsilon = \sum_{i \neq j} \frac{1}{d_{ij}} \tag{5}$$

where $d_i$ is the distance between $v_i$ and $v_i$. It is obviously that the bigger the $\varepsilon$ is, the faster the information spreads.

Firstly, we build a two-layer small-world network, which is consist of A network and B network, as shown in Fig. 3.

Initially, the nearest coupled network is set, where each node is connected to its left and right $k$ nodes. Network A is set to have 500 nodes, and the edges connect to each node will randomly connect to non-neighbor nodes in the network with probability $p_A$=0.2, $k_A$=2. For network B, the node number is 400, its neighbor edges will randomly connect to non-neighbor nodes in the network with the probability is $p_B$=0.3, $k_B$=2. The number of connected edges between communities is half of the number of nodes on network A and the edges are randomly connected.

If we try to find the most important edges in the two-layer network, according to the proposed method, the Newman algorithm is firstly used to mining the community structure.

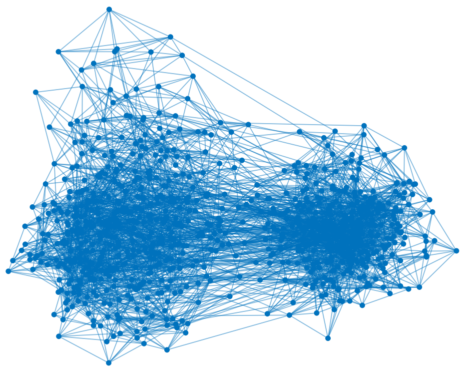

Fig. 3. Two-layer small-word

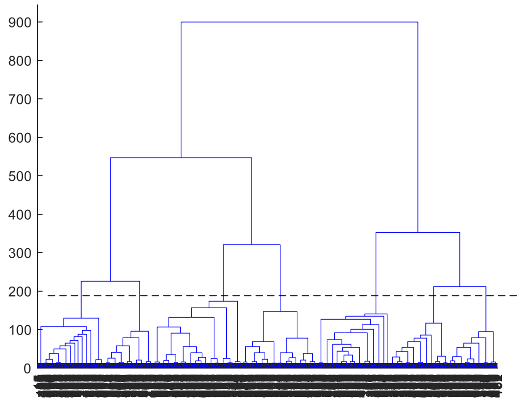


Fig. 4. Tree diagram of community structure division

Fig. 4 shows the tree diagram obtained of the community structure, which is mined by the Newman algorithm. The horizontal coordinate is the node label, which cannot be shown in detail because there are too many nodes. According to the principle of the proposed method, the optimal community division is broken at the dotted line, that is, the connected nodes below the dotted line are divided into a community, and the entire network can be divided into 7 communities. Notably, When the scale of the whole network is large, the number of communities cannot be judged manually. In this case, the Q value can be used to assist the judgment. In the above two-layer network, the Q value curve obtained by community merger at each step is shown in Fig. 5.

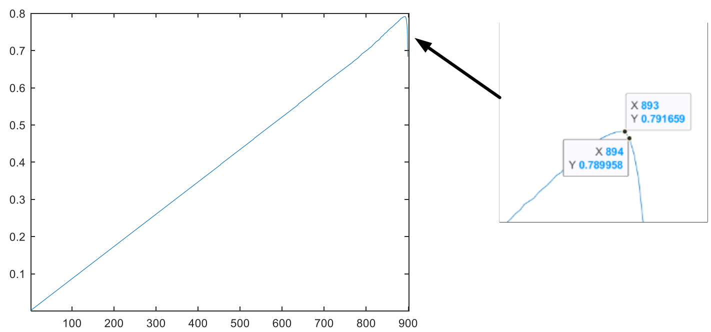


Fig. 5. Q value changes during merging

As can be seen in the figure, after the 893rd merger, the Q value of the whole network will decrease, that is, the connections within the communities will gradually decrease, while the connections between the communities will increase. According to the index, the six mergers from 894th to 899th are not helpful for the division of the community structure, so it is reasonable to divide the association into seven associations. It is more accurate to judge the number of community division by $Q$ value.

Further, according to the process of the proposed method, the connected edges between the divided communities are also regarded as communities, with a total of 28 communities. By calculating the number of edge betweenness, the inner edges of each community are sorted, and the network efficiency is obtained after deleting the edges in proportion.

Here, we set three algorithms for comparison:

1) Max degree method, that is, ranking edges according to the degree value product of the end points of the edge from large to small order;

2) Max betweenness method, that is, ranking edges according to the edge betweenness value of the whole network from large to small;

3) Random method, that is, randomly select the same persentage of connected edges as other methods for deletion.

The simulation results are shown in Fig. 6. Compared with the other three methods, the method proposed has a faster network efficiency decline when deleting the same persentage of nodes and can effectively control information propagation. The main reason is that after the network is divided into multiple communities, important links in each community can be deleted, which not only controls the information spreading within the community, but also inhibits the information spreading between the communities. The network efficiency curves of the maxi degree and the max betweenness will decrease with the increase percentage of deleted edges. Although the overall trend of the network efficiency curve of randomly selected connected edges is downward, the curve is not smooth, because the importance of each randomly deleted connected edge is different, and the network efficiency decline rate is different.

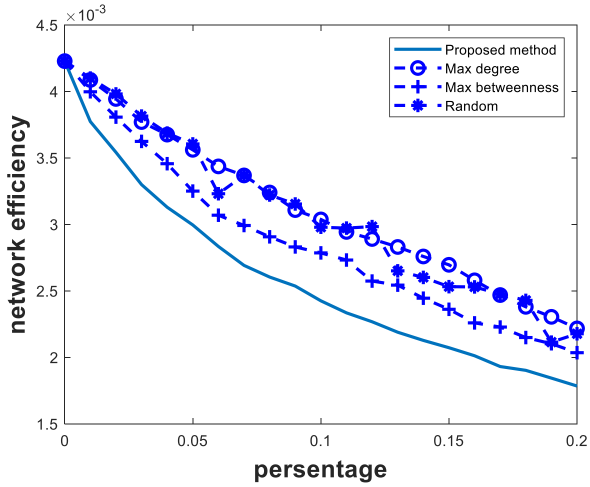


Fig. 6. Network efficiency VS different edge delete percentage

## 5. Summary

This paper designed the two-layer small word network edge importance measurement method. Through the community structure mining process, we realize the redefinition of network community. Furthermore, the edges are ranked separately according to its betweenness in its community. In this way, we can locate all importance edges that has largest impact on the information spreading. In addition, the proposed method also can avoid uneven information transmission source distribution, which would lead to the deletion of invalid edges. Compared with the max degree and max betweenness methods, the information spreading efficiency can be lower by deleting the same persentage of connected edges, which proves the advantages of the proposed method in locating the important connected edges of the network.


## Acknowledgment

Fund Project: Fund of Complex System Simulation Laboratory (6142401004032106)

## References

[1] Yu Ying. Analysis of the dynamics of information propagation and control immunization strategy in Ying complex network [D]. Yanshan University, 2022

[2] Guo Bin, Liu Sicong, Liu Yan, etc. Intelligent Internet of Things: concept, architecture and key technologies [J]. Journal of Computer Science: 2023, 1-20.

[3] Reka Albert, Albert Laszlo Barabasi. Statistical mechanics of complex networks [J]. Reviews of Modern Physics, 2002, 26(1): xii.

[4] Li Baojie, Liu Yan, Li Hongjie, etc. To see the security problem of power information system from the power failure accident in Ukraine [J]. China Electric Power, 2017, 50 (05): 71-77.

[5] Gu Ziyu. Modeling and analysis of the propagation dynamics of complex networks [D]. Shaanxi University of Science and Technology, 2023

[6] Guilherme Ramos, Daniel Silvestre, Carlos Silvestre. Node and network resistance to bribery in multi-agent systems[J]. Systems & Control Letters, 2021, 147: 104842.

[7] Li Gang, Wang Yuda, Cui Rong. KiC: a key node identification algorithm of social network that combines the constraint value of "structural hole" and K shell decomposition [J]. Modern Intelligence, 2020, 40 (12): 9.

[8] Li Peng, Wang Shilin, Chen Guangwu, etc. Based on the improved local structure entropy complex network important node mining [J]. Computer Applications, 2023, 43 (04): 1109-1114.

[9] Kamal, Berahmand, Asgarali, et al. A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks – Science Direct [J].Chaos, Solitons & Fractals, 2018, 110: 41-54.

[10] Liguo Fei, Yong Deng. A new method to identify influential nodes based on relative entropy [J].Chaos, Solitons and Fractals, 2017, 104: 257–267.

[11] Liu Meng, Pei Zhongmin, Luo Zhangkai to evaluate the importance of spatial information network nodes based on triangular structure [J]. Ordnance Automation, 2022, 41 (11): 27-31.

[12] Ren Lintao, Pei Zhongmin, Xiong Wei, etc. Research on the failure of multi-community network cascade considering the node load [J]. Computer Engineering and Application, 2022, 58 (06): 95-102.

[13] Fu Changlei. An improved community division algorithm based on Newman fast algorithm [J]. Computer Technology and Development, 2018, 28 (01): 33-35, 40.