

Research on the Recognition Algorithm of Important Nodes in Complex Network

Yue Su

Guangzhou Huali Science and Technology Vocational College, Guangzhou 511325, China

E-mail: suyue_ys@yeah.net

Student paper

Keywords: complex network, important node, K-shell algorithm, node degree, zachary network

Received: December 26, 2019

It is crucial to identify the important nodes in the network. On the basis of the K-shell algorithm, this study researched the recognition of important nodes in complex networks. The study introduces concepts of edge weight and influence coefficient, the design of an IKS algorithm, and an analysis of its recognition effect in Zachary network and real micro blog network. The results showed that the partition results of the K-shell algorithm were coarse, while the partition results of the IKS algorithm were refined; the IKS algorithm could sort the important nodes accurately on the basis of the K-shell algorithm, and its rationality was higher than that of the closeness centralization and PaperRank algorithm. The partition results in the microblog network also verified the effectiveness of the improved method. The experimental results show that the IKS algorithm is reliable in the important node identification, which is beneficial to the recognition of important nodes in complex network.

Povzetek: V študentskem prispevku je opisan študij algoritma za določitev pomembnih vozlišč v kompleksnih omrežjih.

1 Introduction

Because of constant societal developments, people's daily activities are progressively closely linked. Most of these activities can be regarded as a kind of complex network. In the network, some areas are closely connected with each other, while the connection with other areas is very weak. These areas can be called the clustering property of the network. The network is composed of different communities, such as social network and information network, traffic network, etc. The number of members in complex networks is large, and individual behaviors and mutual relations are very complex [1]. With the development of technology, it is feasible to mine information in complex networks. The important nodes in the network community are the nodes that have an important impact on the network, which has very important practical significance for controlling rumor spread, the spread of infectious diseases [2], traffic conditions [3], etc. This is why the recognition algorithm of important nodes has been widely researched. Gu et al. [4] designed an important node identification algorithm based on LeadeRank algorithm and node similarity and verified the reliability of the algorithm by simulation on SIR model and Superman model. Based on the shortest path, Zheng et al. [5] combined proximity and centrality to find important nodes and found through performance analysis that the method was more effective in finding important nodes. Hu et al. [6] designed a multi-index method which integrated the centrality of feature vectors and degree centrality to recognize important nodes and found through simulation experiments that the algorithm was more reasonable and accurate. Wen et al. [7] designed

a "No Return" method for the importance evaluation of aviation network nodes and found that this method could effectively find potential important nodes with high accuracy after testing it on the aviation networks of China and the United States. In this study, the classical K-shell algorithm was improved and an IKS algorithm was designed for identifying the important nodes. Experiments on the Zachary and the Weibo networks verified the reliability of the method, which is conducive to effectively measuring the important nodes in complex networks and can be used in controlling the spread of infectious diseases and rumors.

2 Complex network

The complex network is usually described by the network graph. According to graph theory, the network is composed of node V and edge E , denoted as $G = (V, E)$, and the related concept includes:

- (1) degree. The number of edges of node v is called degree, represented by k . It is divided into out-degree k_i^{out} (number of edges which points at other nodes) and in-degree k_i^{in} (number of edges with other nodes pointing at the target node) according to the directions of edges.
- (2) average path length. Path with the least edges connecting node i and j is the shortest path, and the number of edges on the shortest path is the distance between the two nodes, d_{ij} . If the number of nodes is n , then average path length L can be expressed as:

$$L = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i \geq j} d_{ij}$$

(3) clustering coefficient. For node i , k_i nodes connect with it, then there are at most $\frac{1}{2}k_i(k_i - 1)$ edges between them. If the number of edges is E_i , then the clustering coefficient can be expressed as $C_i = \frac{2E_i}{k_i(k_i - 1)}$, and the

average clustering coefficient is $\bar{C} = \frac{1}{N} \sum_{i=1}^N C_i$.

(4) diameter. The maximum value of distance between nodes in the network is $D = \max_{i,j} d_{ij}$.

3 Improved K-shell based important node recognition algorithm

3.1 Classic K-shell algorithm

K-shell algorithm is a method proposed by Kitsak et al. [8]. The algorithm has low time complexity, a good practicability in a large complex network, and high recognition accuracy. The recognition steps are as follows:

- (1) the degree of all nodes in the network is calculated;
- (2) all nodes with degree of 1 are searched, and the nodes and their connecting edges are deleted. In time, nodes with a degree of 1 may reappear, and they are deleted as well until there is no node with a degree of 1 in the network. The deleted node forms 1-shell.
- (3) all nodes with degree of 2 are searched, and the deletion repeats until there is no node with degree of 2 in the network. The deleted nodes form 2-shell.
- (4) the above steps repeat until all nodes in the network have corresponding shell values.

Take Figure 1 as an example, (1) represents the original picture, (2) represents 1-shell, (3) represents 2-shell, and (4) represents 3-shell.

In K-shell algorithm, the larger the K-shell value of a node, the greater the influence on the network, and the lower the computational complexity of K-shell algorithm, which has a good division of the hierarchy. However, K-shell cannot show the differences of nodes in the same layer of network, and the results are coarse and not refined enough.

3.2 Improved K-shell algorithm

In order to recognize the importance of nodes better, the K-shell algorithm is improved in this study. Firstly, concepts of weight of edge w_{ij} and influence coefficient e_{ij} are introduced:

(1) $w_{ij} = k_i + k_j$, where k_i and k_j represent the degree of node, and the weight is the sum of the degree of node. If two nodes have many neighbor nodes, it means that

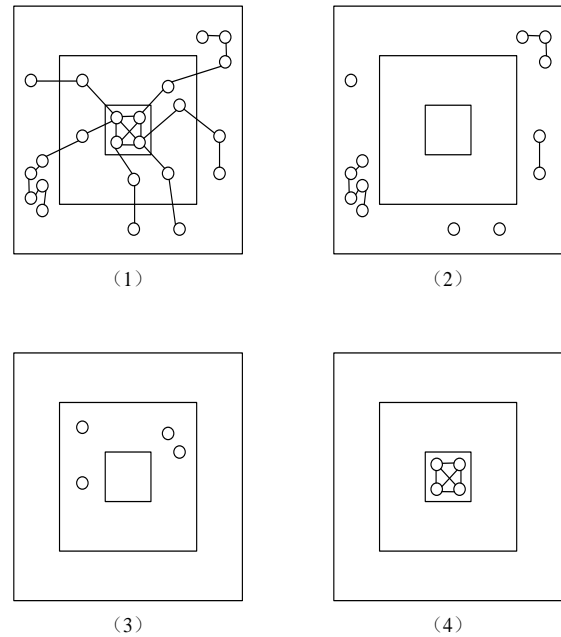


Figure 1: Examples of K-shell.

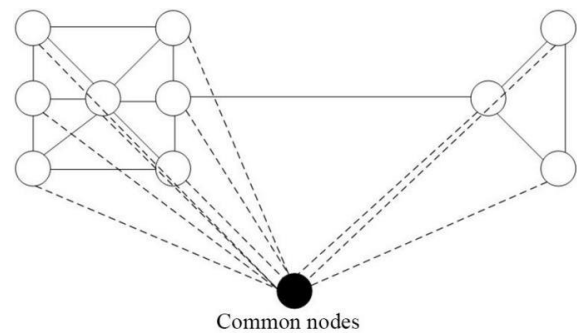


Figure 2: An example of common nodes.

more nodes will be involved if the information propagates in the two nodes.

$$(2) e_{ij} = \frac{|N_i \cap N_j|}{|N_i \cup N_j|},$$

where N_i and N_j represent the sets

of neighbor nodes, and the influence coefficient is the ratio of the number of common neighbors of two nodes to the total number of neighbors.

For the influence coefficient, if two nodes connect to two network communities but have no common friends, then e_{ij} of them is zero. In order to avoid this situation, common node is introduced (Figure 2), which increases the number of common neighbors and the total number of neighbors by 1 and avoids the influence coefficient to be zero.

Based on the above two concepts, weighted degree W_k of nodes is proposed:

$$W_{k_i} = \sum_{j \in N_i} e_{ij} k_{s_j} + \sum_{j \in N_i} w_{ij},$$

where k_{s_j} stands for the K-shell value of node j .

The improved K-shell algorithm is denoted as IKS algorithm, and its specific steps are shown in Figure 3.

In the IKS algorithm, the larger the IKS value, the larger the importance of the node, that is, the important node.

4 Example analysis

4.1 Zachary network analysis

Zachary network is a common data set in network analysis, and has 34 nodes, including two communities with node 1 as the core and node 34 as the core, as shown in Figure 4.

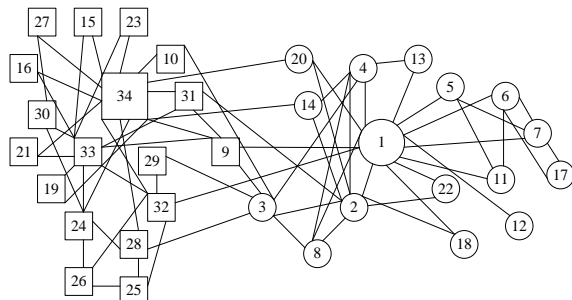


Figure 4: Zachary network.

The classical K-shell algorithm was used to identify the important nodes, and the results are shown in Table 1.

1-shell	12
2-shell	10, 13, 15, 16, 17, 18, 19, 21, 22, 23, 27
3-shell	5, 6, 7, 11, 20, 24, 25, 26, 28, 29, 30, 32
4-shell	1, 2, 3, 4, 8, 9, 14, 31, 33, 34

Table 1: Recognition results of the classic K-shell.

It was found from Table 1 that K-shell divides the whole network into four layers, among which 10 nodes in 4-shell could be regarded as important nodes, but the importance of these 10 nodes was further distinguished. It was found from Figure 4 that the importance of these 10 nodes was different, for example, node 1 was obviously more important than node 8.

In order to verify the reliability of the method, the Zachary network was identified by using closeness centralization [9], PaperRank [10] and IKS algorithm designed in this study. The top 10 nodes were taken as important nodes, and the results are shown in Table 2.

Rank	Closeness centralization	PaperRank	IKS
1	1	34	1
2	3	1	34
3	34	33	33
4	32	3	2
5	9	2	3
6	33	32	4
7	14	4	14
8	20	24	8
9	2	9	9
10	4	14	32

Table 2: Comparison of recognition results.

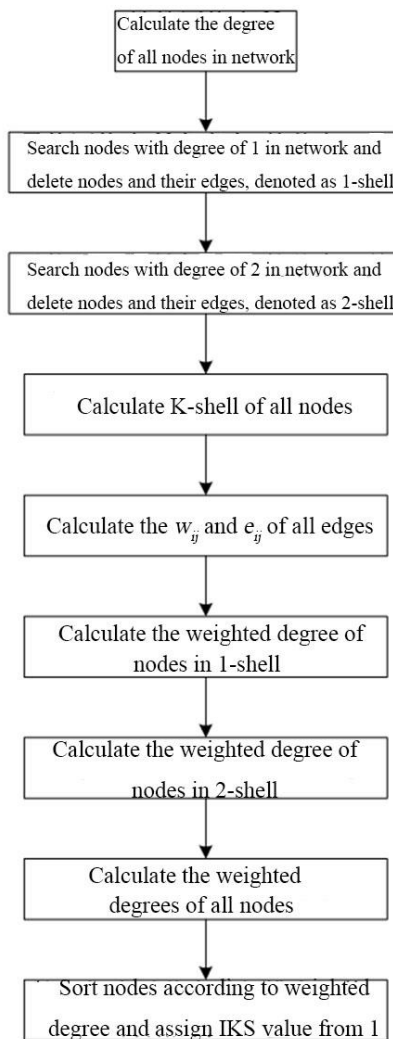


Figure 3: Steps of IKs algorithm.

It was found from Table 2 that the important nodes obtained by the three algorithms were basically similar, but there were some differences in the specific sorting. In proximity centrality, the importance of nodes was measured by the average distance from nodes to other nodes, and node 3 ranked the second place, while node 3 did not appear in the top three in PaperRank and IKS. Comparing node 3 with node 34, although node 3 was in the center of the network, node 34, as the core of a community, was significantly more important than node 3.

The sorting results of PaperRank and IKS were highly similar. Next, the differences were analyzed:

(1) node 34 and node 1. In the 4-shell, nodes 2, 3, 4, 8, 9 and 14 nodes connected with node 1, and nodes 9, 14, 31 and 33 connected with node 34. In the aspect of the number of neighbor nodes, node 1 was more important than node 34, which showed that the results of IKS algorithm were more reasonable.

(2) node 2 and node 3. It was found from Figure 2 that node 2 was more closely related to the community with node 1 as the core, while node 3 was related to both communities, but less closely than node 2, indicating that node 2 was more important than node 3.

(3) node 24 and node 8. Node 8 is not classified as the important node by PaperRank, and node 24 was also not classified as the important node by IKS. The K-shell division results showed that node 8 belonged to 4-shell and node 24 belonged to - 3 shell, indicating that node 8 was more important than node 24, and IKS division results were more accurate.

4.2 Microblog network analysis

Collecting micro blog data with web crawler obtained a data set with 4833 nodes, including 14256 edges, 13 network diameter, 1 minimum degree and 127 maximum degree, which is a complex network. Firstly, it was decomposed using K-shell algorithm, and the results are shown in Figure 5.

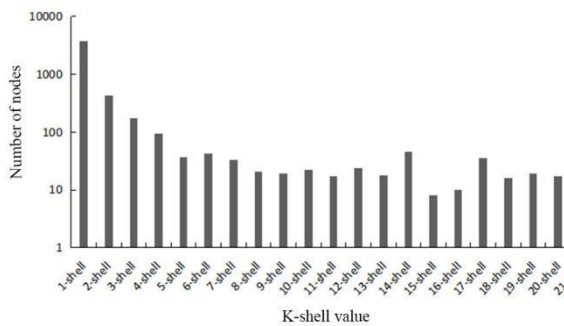


Table 3: Decomposition results of K-shell.

It was found from Figure 5 that the classic K-shell divided the whole network into 21 layers, of which the number of 1-shell nodes reached 3678 and the number of 21-shell nodes was 73. These 73 nodes could be regarded as important nodes, but the importance was not further distinguished. Therefore, the network was subdivided by the IKS algorithm, and the division results are shown in Table 3.

It was found from Table 3 that the IKS algorithm finally divided the network into 77 layers, the number of

IKS value	Number of nodes
1	1627
2	952
3	658
4	364
5	246
6	183
7	96
...	...
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1

Table 4: Division results of IKS.

nodes with the minimum IKS value was 1627, which were the nodes with the smallest influence in the network, and the number of nodes with the maximum IKS value was 1, which was the most important node in the whole network. In the comparison of the division results between the classical K-shell algorithm and IKS algorithm, the number of nodes with low importance was large, while the number of nodes with high importance was small, which was consistent with the actual situation of microblog network. Compared with K-shell algorithm, IKS had a more precise recognition of important nodes, and the number of the top 10 important nodes was 1, which showed that IKS algorithm had strong recognition ability. It showed that the IKS algorithm made up for the defect of the rough division of the classic K-shell algorithm and could effectively sort the important nodes.

5 Discussion

Complex networks are pervasive in many fields, including biology [11], physics, computer science and others. Presently, the research encompasses important node identification [12], community discovery [13], link prediction [14], etc. Important node identification is a key problem in complex networks [15]. In any network, there are differences in the importance of nodes, where the important nodes play a key role, and it is of great significance to identify them [16]. For example, in the network of criminal gangs, through the analysis of the relationship between people, their leader can be localized and the police force can be centralized for control; in the network of infectious diseases, the source of infectious diseases can be found through the analysis of the network, so as to effectively isolate the source of diseases and slow down the spread of diseases; in the network of rumor propagation, the key figures can be mined; in the power network, the important nodes can be recognized and protected to effectively avoid large-scale failure [17]. Therefore, the identification of important nodes has a high practical value, and it is also of great significance to promote the development of related fields.

Currently, the commonly used recognition algorithms for important nodes are degree centrality (the larger the node degree, the more important it is), betweenness centrality (the more information the node propagates, the more important it is), closeness centralization (the more central the node is in the network, the more important it is), PaperRank algorithm (judging the importance of nodes according to the number and quality of other nodes pointing to the target node), etc. Based on the K-shell algorithm, an improved K-shell (IKS) algorithm was proposed in this study. On the basis of K-shell division, the importance of nodes was further sorted, so as to get more precise results. Additional experiments were carried out in Zachary network and a real micro blog network. It was found that the classical K-shell algorithm could identify important nodes, but the important nodes were not distinguished carefully, and the IKS algorithm could effectively improve the results of the K-shell algorithm and accurately identified the important nodes in the network. In the comparison with closeness centralization

and PaperRank algorithms, it was also found that the IKS algorithm designed in this study had higher reliability in the recognition of important nodes.

Although some achievements were made in the research of important nodes identification, there are still some shortcomings due to the limited time and ability:

(1) only the static network was analyzed, but the actual network has dynamic changes; the effectiveness of the algorithm on the dynamic network needs to be studied;

(2) whether the algorithm is equally applicable to the important node identification of weighted networks needs to be studied.

6 Conclusion

The recognition of important nodes in complex networks was studied in this work, the IKS algorithm was obtained by improving the classical K-shell algorithm, and experiments were carried out on the Zachary network and a micro blog network. The results showed that:

(1) the classical K-shell algorithm could divide important nodes, but it cannot sort them in details;

(2) compared with closeness centralization and PaperRank algorithm, the results of IKS were more reasonable;

(3) the IKS algorithm could effectively improve the coarse division results of the K-shell algorithm and realize the accurate identification of important nodes.

7 References

- [1] Tong C, Lian Y, Niu J, Xie Z, Zhang Y (2016). A novel green algorithm for sampling complex networks. *Journal of Network and Computer Applications*, pp. 55-62.
<https://doi.org/10.1016/j.jnca.2015.05.021>
- [2] Jahanpour E, Chen X (2013). Analysis of complex network performance and heuristic node removal strategies. *Communications in Nonlinear Science and Numerical Simulation*, pp. 3458-3468.
<https://doi.org/10.1016/j.cnsns.2013.04.030>
- [3] Wang HY, Wen RY, Zhao YF (2015). Empirical Research on Topological Characteristics of Air Traffic Situation Network. *Applied Mechanics and Materials*, pp. 1975-1979.
<https://doi.org/10.4028/www.scientific.net/AMM.744-746.1975>
- [4] Gu YR, Zhu ZY (2017). Node Ranking in Complex Networks Based on LeaderRank and Modes Similarity. *Journal of the University of Electronic Science and Technology of China*, pp. 441-448.
<https://doi.org/10.3969/j.issn.1001-0548.2017.02.020>
- [5] Zhang Z, Zhang ZY, Song MM (2013). Important node searching algorithm based on shortest-path betweenness. *Computer Engineering & Applications*, pp. 98-97.
- [6] Hu F, Liu Y (2015). Multi-index algorithm of identifying important nodes in complex networks based on linear discriminant analysis. *Modern Physics Letters B*, pp. 1450268.
<https://doi.org/10.1142/s0217984914502686>
- [7] Wen X, Tu C, Wu M (2018). Node importance evaluation in aviation network based on “No Return” node deletion method. *Physica A: Statistical Mechanics and its Applications*, pp. S0378437118301997.
<https://doi.org/10.1016/j.physa.2018.02.109>
- [8] Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E (2007). From the Cover: A model of Internet topology using k-shell decomposition. *Proceedings of the National Academy of Science*, pp. 11150-11154.
<https://doi.org/10.1073/pnas.0701175104>
- [9] Basu S, Maulik U, Chatterjee O (2016). Stability of Consensus Node Orderings Under Imperfect Network Data. *IEEE Transactions on Computational Social Systems*, pp. 1-12.
<https://doi.org/10.1109/TCSS.2016.2596041>
- [10] Shuang X, Wang P, Zhang CX, Lu J (2018). Spectral Learning Algorithm Reveals Propagation Capability of Complex Networks. *IEEE Transactions on Cybernetics*, pp. 1-9.
<https://doi.org/10.1109/TCYB.2018.2861568>
- [11] Saucède T, Laffont R, Labrière C, Jebrane A, François E, Eble GJ, David B (2015). Empirical and theoretical study of atelostomate (Echinoidea, Echinodermata) plate architecture: using graph analysis to reveal structural constraints. *Paleobiology*, pp. 436-459.
<https://doi.org/10.1017/pab.2015.7>
- [12] Han ZM, Wu Y, Tan XS, Duan DG, Yang WJ (2015). Ranking key nodes in complex networks by considering structural holes. *Acta Physica Sinica*, pp. 58902-058902.
<https://doi.org/10.7498/aps.64.058902>
- [13] Deng X, Wen Y, Chen Y (2016). Highly efficient epidemic spreading model based LPA threshold community detection method. *Neurocomputing*, pp. 3-12.
<https://doi.org/10.1016/j.neucom.2015.10.142>
- [14] Li L, Qian L, Wang X, Luo S, Chen X (2015). Accurate similarity index based on activity and connectivity of node for link prediction. *International Journal of Modern Physics B*, pp. 1550108.
<https://doi.org/10.1142/S0217979215501088>
- [15] Zhong LF, Shang MS, Chen XL, Cai SM (2018). Identifying the influential nodes via eigen-centrality from the differences and similarities of structure. *Physica A: Statistical Mechanics and its Applications*, pp. 77-82.
<https://doi.org/10.1016/j.physa.2018.06.115>
- [16] Xu S, Wang P (2017). Identifying important nodes by adaptive LeaderRank. *Physica A: Statistical Mechanics and its Applications*, pp. 654-664.
<https://doi.org/10.1016/j.physa.2016.11.034>
- [17] Li Y, Zhu W, Huang C, Xiong N (2017). Research on power heterogeneous communications network stability with SOC. *Power System Protection and Control*, pp. 118-122.
<https://doi.org/10.7667/PSPC160342>

