Journal of Computers Vol. 29 No. 3, 2018, pp. 179-187
doi:10.3966/199115992018062903016

An Improved Weighted K-shell Decomposition Method for Complex Networks

Zhenjiang Zhang[1*], Xiaoyu Tan[2], Shiyuan Tong[2]

[1] Department of Software Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
zhjzhang1@bjtu.edu.cn

[2] Department of Electronic and Information Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
{15125042, 17111010}@bjtu.edu.cn

Received 13 September 2017; Revised 13 October 2017; Accepted 9 November 2017

Abstract. The identification of influential nodes in complex networks has attracted much attention due to their significant theoretical significance and wide applicability. When designing an identification method for an unweighted network, existing methods also consider edges. In this paper, we propose a new improved K-Shell algorithm based on weight for complex networks, short for CNW-IKS algorithm. We use edge load-bearing and edge influence factor measure the local features of nodes, turning the problem of unweighted network to a weighted network. With the help of SIR information dissemination model to verify the validity and accuracy of CNW-IKS algorithm, the real social network simulation results show that the CNW-IKS algorithm is more accurate for the division granularity to influence the size of the nodes. This method can provide theoretical support for the application of public opinion control and advertising marketing in complex networks.

Keywords: influence subject identification, information dissemination, K-Shell, nodes ranking

1 Introduction

The importance assessment of nodes has always been a fundamental issue in the study of complex networks [1]. Many mechanisms are highly influenced by a small number of influential nodes, such as propagation, concatenation, synchronization and controllability [2]. As we all know, degree, intimacy and intermediate degree [3] are the first three measures developed to distinguish which nodes are more important than others [4]. With the rapid development of investigating the importance of nodes, many central measures have been put forward in recent years, such as PageRank [5], LeaderRank [6], semi-local centrality [7] and so on. However, it is still an open question to design an effective way to identify the importance of nodes. Pei et al. searched for influential communicators by tracking the actual propagation dynamics in many online social networks and find that the extent of widespread use and PageRank cannot rank the impact of users [8]. The influential communicators which proposed by Borge Holthoefer et al. may limit the spread of the rumor in rumors of dynamics [9]. Moreover, finding multiple influential spread spectrum devices in complex networks is a very arduous task [10]. More recent overviews of identifying the node importance and some related applications are mentioned in [11]. Recently, Kitsak et al. proposed an interesting measure, k-shell decomposition (Ks), classifying nodes into core nodes and edge nodes [12]. They show that the most influential nodes are those which located in the core of the network. After this original work was completed, a lot of work will be down to improve the k-shell decomposition. In order to overcome the original k-shell decomposition of many nodes

* Corresponding Author

distributed in the same k-shell, Zeng et al. proposed the mixed degree decomposition (MDD) method, taking into account the excess and depletion levels [13]; Basaras et al. put forward the power community index (PCI) for balancing the core and intermediate center principles [14]; Bae and Kim proposed a novel measure recently, called coreness centrality (Cnc), which makes use of the k-shell index of its neighbor to estimate the propagation impact of nodes in the network [15]. The k-shell decomposition extends to weighted complex networks [16].

In this paper, we continue to improve the K-shell algorithm in complex networks. In most unweighted networks, edges are treated equally, but the importance of each edge in network structure and function may not be the same [17]. Therefore, when we design centrality measures for unweighted networks, it is important to consider the potential importance of edges. Here, we first propose edge bearers based on the joining degree of the two nodes.

And then, we consider the edge influence factor by using the total number of friends divide the number of common friends. Besides, we consider the nodes' own features by weighted degree. Last but not the least, we use the improved K-Shell algorithm based on weight for complex networks (CNW-IKS Algorithm). The rank results show that our method gives wider rank list and overcomes the original k-shell decomposition assigning many nodes in an identical k-shell. To evaluate the effectiveness of the method proposed in this paper, we use the SIR model to study epidemiological propagation processes. By measuring the time which the top-10 nodes as the initial source of infection infect the entire network cost, it shows that our method can rank the spreading ability of nodes more accurately than the k-shell algorithm.

The rest of this paper is organized as follows: In Section 2, we first briefly introduced the k-shell algorithm and proposed our algorithm. In Section 3, we apply the SIR models to evaluate the effectiveness of the proposed algorithm in real complex networks. Finally, we make a summary of the obtained results in the last section.

2 Materials and Methods

In order to more accurately measure the potential influence of the edge, the Edge Load-bearing and the Edge Influence Factor are proposed in the case of taking full account of the influence of the neighbor's influence. The Edge Load-bearing will be used to characterize the influence of the adjacent nodes on the edge of the relationship between them. The influence factor will complement the difficulty of describing the mutual influence between adjacent nodes.

For a given unweighted complex network $G=\left(V,E\right)$, N is the number of nodes, and M is the number of edges. $e_{uv} \in E$ means The node u is connected to the node v.

2.1 Edge Load-bearing

The Edge Load-bearing of e_{uv} is defined as ω_{uv}:

$$\omega_{uv} = C_d(u) + C_d(v) \tag{1}$$

where $C_d\left(u\right)$ and $C_d\left(v\right)$ are respectively means the degree of node u and node v.The meaning of the formula (1) is to characterize the influence of the neighbor nodes involved in the interaction between the two adjacent nodes on the load of the edge.

2.2 Edge Influence Factor

The Edge Load-bearing of e_{uv} is defined as φ_{uv}:

$$\varphi_{uv} = \frac{\left|\Gamma_u \cap \Gamma_v\right|}{\left|\Gamma_u \cup \Gamma_v\right|} \tag{2}$$

where Γ_u and Γ_v are respectively means the neighbor node set of node u and node v. The meaning of the formula (2) is to characterize the degree of difficulty between adjacent two nodes of influencing each other.

The calculation formula of φ_{uv} as follows:

$$\varphi_{uv}=\frac{|\Gamma_u \cap \Gamma_v|}{|\Gamma_u \cup \Gamma_v|}\ \varphi_{uv}=\frac{|\Gamma_u \cap \Gamma_v|}{|\Gamma_u \cup \Gamma_v|} \tag{3}$$

where $C_d(u)$ and $C_d(v)$ are respectively means the degree of node u and node v, g_{uv} means the total number of friends with node u and node v.

2.3 Weighted Degree

The degree of centrality is the most direct measure of the centrality of nodes in network analysis. The K-Shell algorithm measures the node's own attributes as a measure of the node's degree and gives the K-Shell value of the node. In order to fully consider the potential impact of the edge to effectively improve the K-Shell algorithm, while maintaining the node's own attributes are not ignored, we introduce the concept of Weighted Degree.

The Weighted Degree of node v is defined as $C'_d(v)$:

$$\mathrm{C}_d'(v)=\sum_{u\in\Gamma_v}\omega_{uv}+C_d(v)\sum_{u\in\Gamma_v}\varphi_{uv} \tag{4}$$

where ω_{uv} means the Edge Load-bearing of e_{uv}, φ_{uv} means the Edge Load-bearing of e_{uv}, Γ_v means the neighbor node set of node v. Since the Weighted Degree may no longer be an integer, we round the Weighted Degree down to the nearest integer.

2.4 Calculation Steps

After these preparing, we generalize the k-shell algorithm for the structured weighted networks, which is called CNW-IKS algorithm.

The calculation steps of weighted k-shell algorithm is designed as followed:

CNW-IKS Algorithm for Complex Networks

For a given complex network $G=(V,E)$

$N=\|V\|$;
$C'_d(G)=[]$
for i = 1 : N
  calculate the neighbor set U_i of node v_i;
  $M_i=\|U_i\|$
  for j=1 : M_i
  calculate the $\varphi v_i u_j$ and $\omega v_i u_j$;
  end for
end for
while N > 0
 for i = 1 : N
   calculate the $C_d(V_i)$;
  end for
 $v_{\min}=\arg\min\ C_d(V_i)$;
 $key=\|v_{\min}\|$;
 for i = 1 : key
   calculate the $C'_d(v_{\min}(i))$;
   put the $C'_d(v_{\min}(i))$ into $C'_d(G)$;
   delete the node $v_{\min}(i)$ in G;

---

end for
$N = \| V \|$;
end while
sort the set $C'_d(G)$ from smallest to biggest;
sort the order of $C'_d(G)$ and assign the node CNW-IKS value from 1.

2.5 Case Analysis

This non-destructive network $G = (15,19)$ consists of 15 nodes and 19 edges. The node distribution is shown in Fig. 1.

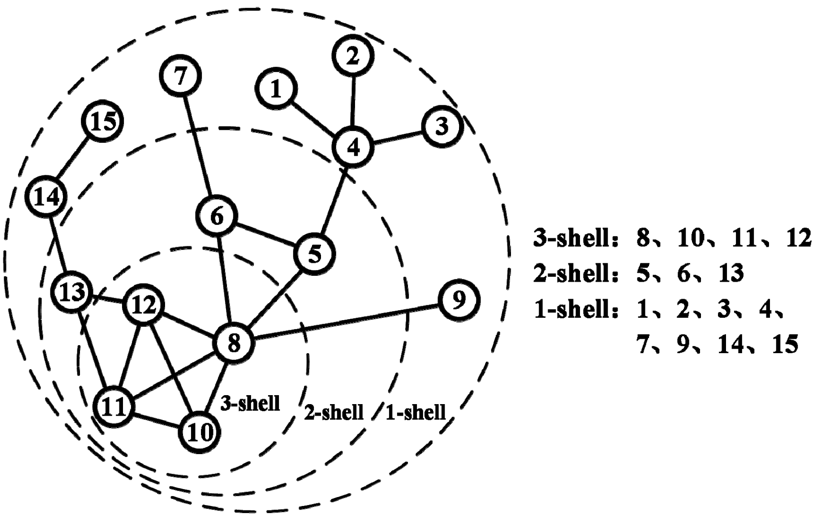

Fig. 1. Non-destructive network after breaking down by K-Shell

The decomposition analysis is applied to this network by the k-shell algorithm. Although k-shell decomposition efficiently detects a group of influential nodes, the nodes in the same shell are not distinguishable by the k-shell decomposition.

The non-destructive network $G = (15,19)$ consists of 15 nodes and 19 edges. The Edge Load-bearing and the Edge Influence Factor after breaking down by CNW-IKS algorithm is shown in Fig. 2 and Table 1.

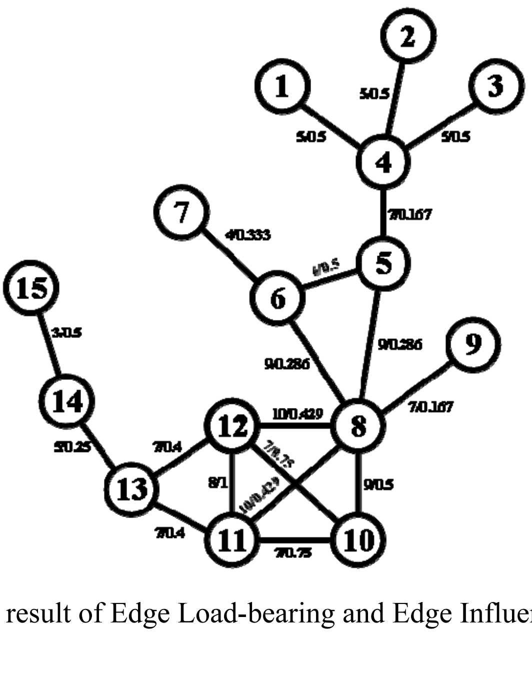

Fig. 2. The result of Edge Load-bearing and Edge Influence Factor

Table 1. The CNW-IKS result of network $G=(15,19)$

Node number	$C_d(v)$ value	$C'_d(v)$ value	$C'_d(v)$ (Rounded down)	K-Shell value	CNW-IKS value
15	1	3.5	3	1-shell	W1-shell
7	1	4.333	4	1-shell	W2-shell
14	2	5.25	5	1-shell	W3-shell
1	1	5.5	5	1-shell	W3-shell
2	1	5.5	5	1-shell	W3-shell
3	1	5.5	5	1-shell	W3-shell
4	4	7.167	7	1-shell	W4-shell
9	1	7.167	7	1-shell	W4-shell
13	3	15.6	15	2-shell	W5-shell
5	3	16.571	16	2-shell	W6-shell
6	3	16.571	16	2-shell	W6-shell
10	3	29	29	3-shell	W7-shell
11	4	31.536	31	3-shell	W8-shell
12	4	31.536	31	3-shell	W8-shell
8	6	33.071	33	3-shell	W9-shell

Through the above example, we find that after the CNW-IKS algorithm is decomposed by the CNW-IKS algorithm, the network is divided into 9 sets of nodes. The CNW-IKS values of all nodes in each layer are the same, but the values are not necessarily the same 4 and node 9 belong to the W4-shell, but the value is different, the node 4 degrees value is 4, the node 9 degree value is 1. In addition, the two nodes with the same value do not necessarily belong to the same layer, such as node 5 and node 13 degrees are 3, but node 5 belongs to W6-shell, node 13 belongs to W5-shell.

CNW-IKS algorithm that the larger the CNW-IKS value is closer to the core of the network, that is, node 8 is at the core of the undirected network, the smaller the CNW-IKS value is, the closer the node is to the edge of the network, The edge of the network.

By analyzing the CNW-IKS decomposition results of the undirected network, we find that the CNW-IKS is more fine than the K-Shell algorithm. The K-Shell algorithm divides the undirected network into three layers, and the result is too coarse and the CNW-IKS The algorithm divides the network into nine layers. In addition, K-Shell algorithm that nodes 8, 10, 11, 12 the same influence, belong to the same 3-shell, but the CNW-IKS algorithm can further compare the node 8, 10, 11, 12 between the influence of the size of the final calculation The influence of node 8 is the largest, and the influence of node 10 is relatively small, which is the advantage of CNW-IKS algorithm.

3 Simulation

3.1 Data Set

In order to verify the effectiveness of the proposed method, we carry out it in the following real-world networks.

Simulation of the data used by Sina microblogging. By parsing the JSON object to retrieve the data returned by the request, and then the data is cleaned and denoised, and finally a data set containing 5906 user information and its attention is obtained. Topological features of Weibo data set is shown in Table 2.

Table 2. Topological features of Weibo data set

Data name	Data sources	Number of nodes	Number of edges	Network diameter	Clustering coefficient
Microblogging data set	Sina Weibo	5906	13191	11	0.0772

Degree distribution curve of Weibo data set is shown in Fig. 3. By observing the network distribution curve of Sina microblogging data set, it can be seen that most nodes in the network have very small degree, only a few nodes have large degree (with more friends), and meet the scaleless nature of complex networks.

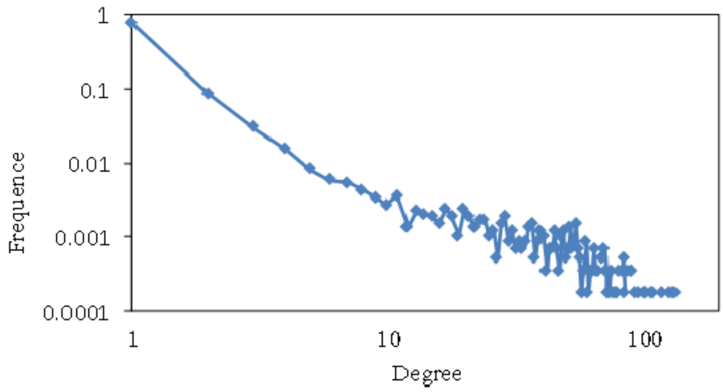

Fig. 3. Degree distribution curve of Weibo data set

3.2 Simulation Results

The CNW-IKS algorithm is implemented by the Python language program, and the CNW-IKS decomposition of the Sina microblogging data set is carried out. After the CNW-IKS decomposition of the data set network, each node is given the corresponding CNW-IKS value, because the data set network contains a large number of nodes, Here only show the number of nodes included in the different CNW-IKS values, the specific node number will be ignored.

The number of node to CNW-IKS value as Table 3 shows. By observing the simulation results of Table 3, it can be seen that this contains 5906 user information and its interest in Sina microblogging data, after the CNW-IKS algorithm decomposition, the decomposition of the maximum CNW-IKS value of 78, the minimum CNW-IKS value 1, the whole network is divided into 78 layers by CNW-IKS algorithm, the number of nodes with the largest CNW-IKS value is 1, the number of nodes with the smallest CNW-IKS value is 1947, that is, the CNW-IKS algorithm decomposes and thinks that the data set network is the most The core layer is located on the 78th floor, which has only one node and belongs to the core of the network, and the node is the most influential node in all nodes of the whole network. The edge layer of the data set network is in the first layer, The layer has 1947 nodes, and these nodes for the entire network of all nodes in the least influential batch of nodes.

Table 3. The number of node to CNW-IKS value

CNW-IKS value	The number of node	CNW-IKS value	The number of node	CNW-IKS value	The number of node
1	1947	27	7	53	3
2	921	28	4	54	1
3	743	29	10	55	2
4	459	30	5	56	1
5	347	31	6	57	1
6	276	32	7	58	1
7	177	33	5	59	2
8	192	34	4	60	1
9	127	35	2	61	1
10	82	36	1	62	1
11	96	37	3	63	1
12	77	38	2	64	1
13	61	39	4	65	2
14	42	40	1	66	1
15	55	41	3	67	1
16	47	42	1	68	1
17	28	43	1	69	1
18	19	44	2	70	1
19	17	45	4	71	1
20	23	46	1	72	1
21	16	47	3	73	1
22	12	48	2	74	1
23	9	49	1	75	1
24	6	50	2	76	1
25	8	51	1	77	1
26	5	52	1	78	1

CNW-IKS decomposition result distribution curve as Fig. 4 shows. By observing the distribution curve of Fig. 4, the number of nodes with small CNW-IKS value is large, the number of nodes with large CNW-IKS value is small, the number of nodes with small CNW-IKS value occupies a large part of the total number of users in the whole network, The number of nodes with a large CNW-IKS value occupies only a small fraction of the total number of users in the entire network. For example, the node with the CNW-IKS value of 1 accounts for 33.0% of the total number of nodes, and the CNW-IKS value of 78 accounts for 0.02% of the total number of nodes.

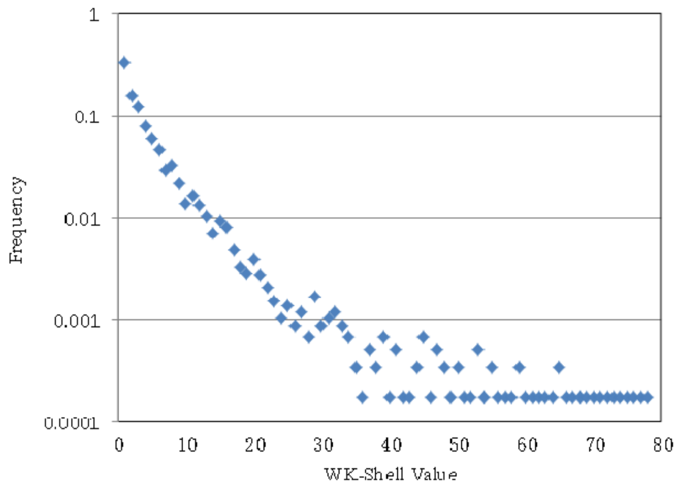

Fig. 4. CNW-IKS decomposition result distribution curve

3.3 Performance on the SIR Model

In order to demonstrate the accuracy of the CNW-IKS algorithm for the ranking of nodes, the SIR model is set up and the CNW-IKS algorithm is selected as the initial infection node, the infection model, the simulation process, and the whole network At the end of the infection, the time unit is deduced and the average of the 500 experiments is repeated. The simulation results of the nodes are compared with the simulation results of the different nodes as the initial infection nodes. The accuracy of the CNW-IKS algorithm is analyzed.

SIR model infection probability $\lambda = 0.7$, immune probability $\mu = 0.0001$, The time unit consumed is $F(t)$. Table 4 shows the CNW-IKS algorithm to calculate the impact of the top 10 nodes as the SIR model of the initial infection nodes infected with the entire network of the average time spent, as well as the value of each node.

Table 4. CNW-IKS Algorithm ranking the first 10 nodes related indicators

The number of node	Shell value	CNW-IKS value	F(t)	$C_d(v)$
1	23	78	127.41	102
2	23	77	127.57	120
3	23	76	127.67	93
4	23	75	127.73	103
5	23	74	127.88	108
6	23	73	127.94	133
7	23	72	128.01	85
8	23	71	128.17	127
9	23	70	128.23	97
10	23	69	128.34	111

By observing the simulation results in Table 4, we found that the node with the influence of the first node as the initial infection node infected the entire network to spend the shortest average time. On the whole, the WJ algorithm has a negative correlation with the average infection time, that is, the node with the greatest influence, the shorter the average time it takes for the initial infection node to infect the whole network, and the influence The longer the average time spent by the small node as the initial infection node, the CNW-IKS algorithm is used as an improved algorithm to calculate the influence of nodes, and the result is accurate.

4 Conclusion

In this paper, we proposed a new improved K-Shell algorithm based on weight for complex networks. We modified the original k-shell decomposition method to identify the importance of the node. The method proposed in this paper is to consider both the degree of the node and the links' weights. The decomposition results show that the CNW-IKS centrality produces more monotonic ranking than the original k-shell method. In order to evaluate the effectiveness of the algorithm proposed in this paper, we compare the ranking of CNW-IKS centrality to the infected population ranking in the SIR model.

CNW-IKS algorithm takes into account the node's own attributes, location attributes and local attributes, the network according to different CNW-IKS values are divided into different levels, CNW-IKS value of the larger nodes, the core of the network location, And the greater the influence, the smaller the CNW-IKS value of the node, its location in the network where the edge, and the smaller the influence.

Simulation results show that CNW-IKS algorithm is more accurate than the K-Shell algorithm, and the granularity of the network is more detailed. CNW-IKS algorithm as a new social network influence on the main body recognition algorithm for the social network influence on the main body of the theoretical research and practical application provides some useful exploration.

Acknowledgements

This research is supported by the Fundamental Research Funds for the Central Universities, No. 2017JBZ107.

References

[1] F. Bauer, J.T. Lizier, Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: a walk counting approach, arXiv:1203.0502, 2012.

[2] Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks, Nature 473(2011) 167-173.

[3] T. Opsahl, F. Agneessens, J. Skvoretz, Node centrality in weighted networks: generalizing degree and shortest paths, Social Networks 32(2010) 245-251.

[4] H. Wang, J.M. Hernandez, P. Van Mieghem, Betweenness centrality in a weighted network, Phys. Rev. E 77(2008) 046105.

[5] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30(1998) 107-117.

[6] Q. Li, T. Zhou, L. Lü, D. Chen, Identifying influential spreaders by weighted leaderrank, Physica A 404(2014) 47-55.

[7] S. Gao, J. Ma, Z. Chen, G. Wang, C. Xing, Ranking the spreading ability of nodes in complex networks based on local structure, Physica A 403(2014) 130-147.

[8] S. Pei, L. Muchnik, J.S. Andrade, Z. Zheng, H.A. Makse, Searching for superspreaders of information in real-world social media, Sci. Rep. 4(2014) 5547.

[9] J. Borge-Holthoefer, Y. Moreno, Absence of influential spreaders in rumor dynamics, Phys. Rev. E 85(2012) 026116.

[10] Z.-L. Hu, J.-G. Liu, G.-Y. Yang, Z.-M. Ren, Effects of the distance among multiple spreaders on the spreading, Europhys. Lett. 106(2014) 18002.

[11] Q. Jin, L. Wang, C.-Y. Xia, Z. Wang, Spontaneous symmetry breaking in interdependent networked game, Sci. Rep. 4(2014). <https://www.nature.com/articles/srep04095>, 2014.

[12] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, Identification of influential spreaders in complex networks, Nat. Phys. 6(2010) 888-893.

[13] A. Zeng, C.-J. Zhang, Ranking spreaders by decomposing complex networks, Phys. Lett. A 377(2013) 1031-1035.

[14] P. Basaras, D. Katsaros, L. Tassiulas, Detecting influential spreaders in complex, dynamic networks, Computer 46(2013) 24-29.

[15] J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, Physica A 395(2014) 549-559.

[16] M. Eidsaa, E. Almaas, s-core network decomposition: a generalization of k-core analysis to weighted networks, Phys. Rev. E 88(2013) 062819.

[17] C.-J. Zhang, A. Zeng, Network skeleton for synchronization: identifying redundant connections, Physica A 402(2014) 180-185.