

Postprint: Decomposition Algorithms for Central Node Identification in Weighted Directed Networks

Authors: Liu Chen, Li Dandan, Han Lin, An Yongxue

Date: 2018-05-20T00:00:00+00:00

Abstract

Currently, node importance identification algorithms for complex networks primarily focus on unweighted and undirected networks, which cannot comprehensively describe real-world complex network scenarios. For example, most centrality metrics only consider a single indicator, ignoring the differences between node out-degree and in-degree, and neglecting the importance of weights. Based on directed weighted complex networks, and comprehensively considering the differences between node out-degree and in-degree as well as the actual importance of weights in real networks, we propose a central node identification algorithm based on out-degree, in-degree, and weights—the cw-shell decomposition algorithm. To verify the effectiveness of the algorithm, virus propagation simulation experiments are conducted on real complex networks using the W-SIR propagation model. The results demonstrate that the cw-shell decomposition method can effectively rank nodes hierarchically and identify nodes with high diffusion capability.

Full Text

Preamble

Decomposition Algorithms for Identifying Central Nodes in Directed-Weighted Networks

Liu Chen, Li Dandan, Han Lin, An Yongxue

(Business School, University of Shanghai for Science & Technology, Shanghai 200093, China)

Abstract: Existing evaluation methods for node importance in complex networks primarily focus on undirected and unweighted networks, which cannot fully characterize real-world network structures. Most centrality measures, for

instance, consider only a single metric, ignoring differences between node out-degree and in-degree while neglecting the significance of weights. Based on directed-weighted complex networks, this paper proposes a central node recognition algorithm—the cw-shell decomposition method—that incorporates out-degree, in-degree, and weight considerations, accounting for both the distinction between out-degree and in-degree and the practical importance of weights in real networks. To verify the effectiveness of this algorithm, we conduct viral spreading simulations on real-world complex networks using the W-SIR propagation model. Results demonstrate that the cw-shell decomposition method can efficiently rank nodes hierarchically and identify those with high diffusion capacity.

Keywords: cross-degree; c-shell; k-shell; node importance; transmission dynamics

0 Introduction

With the rapid development of information technology, complex networks have permeated every aspect of human life. Whether examining virus propagation or the spread of opinions and rumors, a critical question emerges: “Which nodes are most important in the network, and how can information diffuse effectively?” This has become a focal point for researchers. Goel et al. [1] observed data from seven online platforms and found that most cascades are small and shallow, describable through similar tree structures. They noted that even when large cascades occur, they typically result from media effects, such as fans’ enthusiastic attention to favorite artists or products. Consequently, the authors suggest that to expand network propagation scope, online marketers and researchers should reduce sharing costs, identify central node groups within populations, and focus on enhancing the propagation potential of important nodes.

Numerous studies have addressed how to identify important nodes. For example, classic algorithms for ranking node importance in undirected networks include degree, betweenness, and k-shell values. For directed networks, the two classic algorithms—HITS and PageRank—originate from web page ranking on the WWW [2–5]. Newman et al. [6] studied email networks and found that removing nodes with maximum degree renders networks extremely fragile. The degree metric is computationally simple with low time complexity, making it suitable for large networks. However, degree only reflects local node characteristics without describing the connection status of neighboring nodes. Freeman [7] defined node importance using betweenness—the number of shortest paths passing through a node—which characterizes a node’s control over information transmission along shortest paths between node pairs. However, betweenness suffers from high computational complexity due to shortest path calculations, making it unsuitable for large complex networks.

Kitsak et al. [8] proposed that a node’s position in the network is the pri-

mary factor influencing its importance. With a single propagation source, the most influential nodes are not those with maximum degree or betweenness, but those with maximum k-shell values. The k-shell metric has low time complexity, works well for large networks, and more accurately identifies the most influential nodes in disease propagation than degree or betweenness metrics. However, k-shell decomposition fails for tree-like networks and BA networks because all nodes in such networks share identical ks values, preventing the ks metric from measuring node importance. Bryan et al. [9] proposed the PageRank algorithm, a classic method for search engine page ranking applicable to directed networks. PageRank' s fundamental principle states that a webpage' s importance depends on the quantity and quality of other pages linking to it. However, when networks contain isolated nodes or communities, ranking non-uniqueness issues arise. Kleinberg [10] proposed the HITS algorithm in 1998, based on the idea that each webpage' s importance has two characterization metrics—authority and hub. HITS is widely used in academia due to requiring fewer iterations and faster convergence, reducing time complexity. However, HITS cannot identify abnormal webpage citations, causing deviations between calculated results and reality.

Zhou Lixin et al. [11] proposed the c-shell decomposition method for directed networks based on k-shell decomposition, demonstrating its good discriminative power in identifying important nodes in directed networks. However, c-shell decomposition cannot be directly applied to weighted networks. Most centrality methods, including k-shell decomposition, are designed for unweighted networks. Yet real networks are weighted, and their weights describe performance attributes of underlying systems. For example, in economic trade networks, weights represent measurable attributes such as trade flows and capital flows. In weighted networks, nodes can be described by at least two attributes: degree and weight. However, weight is a property of network links, so a node' s weight can be defined as the sum of weights of all links passing through it [12]. Networks contain both high-degree low-weight nodes and low-degree high-weight nodes. Previous research has employed two main approaches: completely ignoring weights, or considering only links with weights exceeding arbitrarily chosen thresholds while filtering the remainder. These methods' drawbacks include the need to select appropriate weight thresholds, potentially deleting high-degree low-weight nodes (below threshold). Additionally, ignoring links below the threshold may disconnect some nodes, making networks increasingly sparse [12].

Addressing these issues, this paper proposes a node importance evaluation index—the cross-point weight (cw) index—based on out-degree, in-degree, and weight, along with a general method for cw-shell decomposition in weighted directed networks. Simulations on US airport and adolescent health networks demonstrate that this index effectively measures node importance. This paper is organized as follows: First, we introduce shell decomposition-based central node identification algorithms and propose the cw-shell decomposition method. Then, we simulate viral propagation on real complex networks using the W-SIR propaga-

tion model to verify the method's effectiveness. Finally, we present conclusions.

1 Shell-Based Central Node Identification Algorithms

1.1 k-Shell Decomposition Algorithm

Kitsak et al. [8] first proposed in 2010 that a node's position in the network is the primary factor influencing its importance. With a single propagation source, based on SIR and SIS models and real network data, they proved that the most influential nodes are not those with maximum degree or betweenness, but those with maximum k-shell values. However, some nodes with small k-cores also exert tremendous influence on network propagation dynamics, which traditional k-core decomposition cannot measure [13]. Additionally, k-shell decomposition is unsuitable for tree-like networks and BA networks because all nodes in such networks share identical ks values, preventing the ks metric from measuring node importance.

The k-shell decomposition method recursively removes all nodes with degree less than or equal to K from the network, revealing the network's shell structure properties and hierarchical characteristics [13]. Consider network $G=(E,V)$ consisting of V nodes and E edges. The k-shell is defined as: the subnetwork $H=(C,E/C)$ derived from set C , where for any node V in C , its degree is greater than k . The maximum subnetwork satisfying this condition is called the k-core. The portion belonging to the k-core but not belonging to $(k+1)$ -core is called the k-shell, abbreviated as ks .

Since nodes in the minimum k-core share identical ks values, and according to k-core decomposition principles, nodes with degree 1 and most with betweenness 0 belong to the minimum k-core, relying solely on nodes' own k-core, degree, or betweenness information cannot effectively distinguish these nodes' propagation capabilities [13].

1.2 c-Shell Decomposition Algorithm Based on Cross-Degree

Zhou Lixin et al. [11] proposed the c-shell decomposition method for directed networks based on k-shell decomposition, proving its good discriminative power in identifying important nodes in directed networks. However, c-shell decomposition cannot be directly applied to weighted networks. This method transforms directed networks into undirected networks before performing classic k-shell decomposition. During important node identification in directed networks, the degree metric simply adds in-degree and out-degree, failing to effectively distinguish their differences [12].

The physical meaning of node cross-degree is the number of times connections between a node's neighbors pass through that node. Node cross-degree is denoted as kc :

λ

's different values affect nodes' cross-degree values, thereby distinguishing the importance of out-degree versus in-degree. In equation (1), the authors add an extra node to the network and bidirectionally connect it to all existing nodes, obtaining an $n+1$ network. This allows measuring the importance of nodes with zero out-degree or in-degree, making the new network strongly connected.

Based on [Figure 1: see original paper], we demonstrate the c -shell decomposition of the network and argue for the cross-degree metric. The c -shell decomposition method transforms directed networks into undirected networks before performing classic k -shell decomposition. This method suits the hierarchical process of k -shell decomposition but uses cross-degree as an alternative metric. Calculation results are shown in .

Zhou Lixin et al. [11] conducted simulation experiments using the SIR model to study the average propagation range $M(kc, k)$ of nodes with identical cross-degree and degree values in networks, comparing c -shell decomposition and degree methods for accurately identifying important nodes. Results show that nodes with larger kc values have larger average propagation ranges $M(kc, k)$ in networks. Nodes with similar kc values have similar propagation capabilities, and $M(kc, k)$ values correlate weakly with node degree values. Nodes with identical degree values exhibit significant differences in propagation capability, indicating that networks contain nodes with large degree but limited propagation capacity. Similarly, this proves that the c -shell decomposition algorithm more accurately measures node propagation capability and importance than betweenness centrality.

However, the c -shell decomposition method cannot be directly applied to weighted networks. Therefore, this paper proposes the concept of cross-point weight.

2 cw -Shell Decomposition Algorithm

2.1 Cross-Point Weight

This method considers node out-degree and in-degree, as well as weights of links to neighboring nodes. Since a node' s weight can be expressed as the sum of weights of all links passing through it, we assign cross-point weight to each node. Node i ' s cross-point weight is defined as:

where N is the set of node i ' s neighbors,

$$\sum_{j \in N} w_{ij}$$

is the sum of weights of all links starting from node i , and

$$\sum_{j \in N} w_{ji}$$

is the sum of weights of all links ending at node i . Parameter

$$\lambda$$

has the same meaning as above—to distinguish the importance of out-degree versus in-degree. This paper discusses only the case

$$\lambda = 1/2$$

, considering in-degree and out-degree equally important.

In unweighted networks where $w_{ij}=1$, a node's cross-point weight equals its cross-degree, allowing c-shell method to continue layering the same network.

The cw-shell decomposition method transforms weighted directed networks into unweighted undirected networks before performing classic k-shell decomposition. This method suits the hierarchical process of k-shell decomposition but uses cross-point weight as an alternative metric. The cw-shell decomposition algorithm recursively removes all nodes with cw values less than or equal to cw_min (the smallest cw value in the network), thereby dividing the network into layers where all nodes in each shell share equal cw values.

The specific algorithm implementation proceeds as follows: a) Calculate all nodes' cw values in the network according to equation (3); b) Sequentially remove all nodes with cw values less than or equal to cw_min , and remove links between removed nodes and their neighbors. At this point, some new nodes with cw values less than or equal to cw_min may appear in the remaining network; remove these nodes and their links to neighbors. Repeat this operation until no nodes with cw values less than or equal to cw_min remain in the network. The set of removed nodes is called the network's cw-shell. The remaining node set is called the network's cw-core; c) Continue shell-peeling operations, repeating steps a) and b) until all nodes in the network are removed. Larger cw values indicate nodes closer to the innermost layer of cw-shell decomposition, with greater influence. The cw-shell decomposition algorithm can determine all nodes' hierarchical positions in the network, provide node importance ranking, and identify central nodes.

2.2 Example

To highlight the defects of unweighted methods, assume the network is weighted. Based on [Figure 1: see original paper], assume $W_{1,10}=2.5$, $W_{5,12}=3.5$, $W_{1,6}=0.5$, $W_{3,7}=0.5$, $W_{12,17}=2$, $W_{4,3}=2$, with all others equal to 1. Decompose the network in [Figure 1: see original paper] using the cw-shell decomposition method. Results are shown in .

According to k-shell decomposition, node 10 would be placed at the network's periphery ($k_s=1$), even though it is strongly connected to core nodes. In real networks, such strong connections indicate that the particular node is more important [8]. However, this cannot be depicted through the layered structure calculated by k-shell decomposition. Through cw-shell decomposition, we calculate $cw_{10}=0.8708 > cw_{11}=0.4142$, making node 10's importance higher than node 11's, thereby highlighting its actual importance.

3 cw-Shell Decomposition Algorithm Effectiveness Analysis

This section introduces the application of the cw-shell algorithm in real complex networks. First, we obtain datasets from the Koblenz website and describe the data. Then, we use the cw-shell algorithm for network decomposition to obtain the network's cw-shell structure and determine nodes' centralities. Finally, we simulate viral propagation using the W-SIR model to examine the spreading scope and importance of nodes with different centrality levels, thereby verifying the cw-shell decomposition algorithm's effectiveness.

3.1 Datasets

We obtain the following datasets from the Koblenz website, as shown in : a) **Adolescent health:** A 1994 survey on adolescent friendship relationships. Nodes represent students, directed edges indicate student A choosing student B as a friend, and weights represent intimacy levels (higher values indicate closer relationships). No edge between nodes indicates no interaction. b) **US airport:** A 2010 US airport operations network. Nodes represent airports, directed edges indicate flights from airport A to airport B, and weights represent the number of flights in that direction.

provides detailed statistical properties of the networks. D+ indicates weighted directed networks, NN is total nodes, NE is edges, k is average degree, C is average clustering coefficient, and d is diameter.

3.2 W-SIR Model

In recent years, the SIR model has been used in network research to explore disease propagation, economic crisis spread, and information/rumor diffusion [14-18]. Since we focus on weighted networks, we introduce weight-related concepts into the SIR model, as shown in equation (4). This model was originally designed to simulate economic crisis propagation [19], assuming node i infects node j with probability P_{ij} , where each link has different infection probability:

where w_{ij} is the total weight of all links from node i to node j , and w_j is the total weight of all links ending at node j . The ratio w_{ij}/w_j serves as a factor for node j 's infection likelihood by node i . Parameter m is a free parameter

that can amplify factors such as crisis severity, virus contagiousness, or rumor importance. This model is called the Weighted SIR model (W-SIR).

The W-SIR modeling process proceeds as follows. First, define all nodes as susceptible (S). Then select a node i as the infected source (I). This node infects its neighbors with probability P_{ij} in the first time step, causing all infected nodes to transition from state S to I, while the initiating node transitions to state R and can no longer infect others or be infected. Repeat these steps, with all infected nodes attempting to infect their neighboring S nodes. This process continues until no more nodes can be infected or all nodes are infected.

3.3 Simulation Analysis

For a given m value, we conduct 100 experiments for each individual node on the W-SIR model, with each node's infection capability being the average across these 100 experiments. We then analyze each shell's spreading capability according to the shell structure obtained by cw-shell decomposition.

3.3.1 Precision of cw-Shell Decomposition Method First, we decompose the network based on node degree, betweenness, and cw-shell decomposition method to obtain the network's shell structure, and rank shells according to nodes' distance from the core. Then we calculate the average infection range of all nodes in each shell to estimate the shell's propagation potential, as shown in [Figure 2: see original paper].

[Figure 2: see original paper] shows that central nodes obtained through cw-shell decomposition can trigger infection outbreaks, with stronger infection capability closer to the core. This means cw-shell decomposition can locate nodes with higher average diffusion capacity and place them near the core.

For further quantitative comparison, we measure these metrics' accuracy in identifying node importance by calculating the imprecision function [20]:

The smaller the (p) value, the higher the metric's precision, stronger propagation capability, and more effective identification of central nodes in the network.

[Figure 3: see original paper] shows that as the selected proportion p increases, the accuracy of each metric continuously improves. For the US airport network ($m=1$), when p is less than 1.5%, $_cw(p) < _k(p)$ and $_cw(p) < _CB(p)$, meaning cw-shell decomposition's error is smaller than both degree and betweenness metrics, indicating greater accuracy. For the Adolescent health network ($m=4$), $_cw(p) < _CB(p)$ always holds, showing cw-shell decomposition is more accurate than betweenness. Compared with the other two methods, betweenness-based identification of the most effective spreaders is the least accurate. Simulation results demonstrate that cw-shell decomposition more reasonably and effectively ranks nodes hierarchically, placing high-diffusion-capacity nodes near the core.

3.3.2 Homogeneity of Each Shell To verify cw-shell decomposition' s accuracy, we test each shell' s homogeneity—all nodes in the same shell should have similar infection capabilities. For a given m value, we layer the network according to cw values, with r representing the number of layers and n representing total nodes. This verification becomes a one-way ANOVA testing whether infection capabilities differ significantly across shells at different r-shell levels.

To test whether network layering structure significantly affects node infection capability, we use SPSS software for One-way ANOVA analysis ($\alpha=0.05$). Results are shown in and .

For $\alpha=0.05$, the significance probability $P<0.05$ indicates significant differences between groups. That is, the shell structure obtained using cw-shell decomposition can effectively rank and layer the network, producing shells with strong homogeneity. Additionally, pairwise comparisons can further verify each shell' s homogeneity. In summary, cw-shell decomposition effectively decomposes networks. For a given cw-shell value, nodes in the same cw layer have similar infection capabilities, and the most effective spreaders are located in the innermost layer of cw-shell decomposition with maximum cw values.

4 Conclusion

This paper proposes a general method—cw-shell decomposition—for identifying central nodes in weighted directed networks. This method comprehensively considers differences between node out-degree and in-degree, as well as the practical importance of weights in real networks. Simulation studies on US airport and adolescent health networks demonstrate that the cw index can effectively rank network node importance, produce shell structures with strong homogeneity, and identify nodes with high diffusion capacity. When comparing the cw index with degree and betweenness metrics, the results are not ideal because degree and betweenness metrics ignore link direction and weight factors. Additionally, this paper discusses only $\alpha=1/2$, assuming equal importance for out-degree and in-degree. However, for different networks, out-degree and in-degree have different influences. Therefore, further discussion of α values is needed according to specific networks.

References

- [1] Goel S, Watts D J, Goldstein D G. The structure of online diffusion networks [C]// Proc of the 13th ACM Conference on Electronic Commerce. New York: ACM Press, 2012: 623-638.
- [2] Liu Jianguo, Ren Zhuoming, Guo Qiang, et al. Research progress on node importance ranking in complex networks [J]. Acta Physica Sinica, 2013, 62(17):

178901-178901.

- [3] Li Xiang, Liu Zonghua, Wang Binghong. Network propagation dynamics [J]. *Complex Systems and Complexity Science*, 2010, 7(Z1): 33-37.
- [4] Pei S, Muchnik L, Jr A J, et al. Searching for super spreaders of information in real-world social media [J]. *Scientific Reports*, 2014, 4: 5547-5547.
- [5] Ren Zhuoming, Shao Feng, Liu Jianguo, et al. A measure of network node importance based on degree and clustering coefficient [J]. *Acta Physica Sinica*, 2013(12): 128901-128901.
- [6] Newman M E J, Forrest S, Balthrop J. Email networks and the spread of computer viruses [J]. *Physical Review E*, 2002, 66(3 Pt 2A): 035101.
- [7] Freeman L C. A set of measures of centrality based on betweenness [J]. *Sociometry*, 1977, 40(1): 35-41.
- [8] Kitsak M, Gallos L K, Havlin S, et al. Identifying influential spreaders in complex networks [J]. *Nature Physics*, 2010, 6(11): 888-893.
- [9] Bryan K, Leise T. The \$25,000,000,000 Eigenvector: The linear algebra behind Google [J]. *Siam Review*, 2006, 48(3): 569-581.
- [10] Kleinberg J M. Authoritative sources in a hyperlinked environment [J]. *Journal of the ACM*, 1999, 46(5): 604-632.
- [11] Zhou Lixin, Liu Chen, Huo Liang' an, et al. Research on central node identification algorithm for directed-weighted networks based on cross-degree [J]. *Computer Application Research*, 2015, 33(11): 1-1.
- [12] Garas A, Schweitzer F, Havlin S. A k-shell decomposition method for weighted networks [J]. *New Journal of Physics*, 2012, 14(8): 083030.
- [13] Ren Zhuoming, Liu Jianguo, Shao Feng, et al. Analysis of propagation capability of minimum k-core nodes in complex networks [J]. *Acta Physica Sinica*, 2013(10): 108902.
- [14] Wang Changchun, Chen Chao, Dong Zhiqiang. The impact of network structure on rumor propagation [J]. *Journal of System Simulation*, 2013, 25(1): 127-131.
- [15] Moreno Y, Nekovee M, Pacheco A F. Dynamics of rumor spreading in complex networks [J]. *Physical Review E*, 2004, 69(6 Pt 2): 066130.
- [16] Ge Xin, Zhao Hai, Zhang Jun. Simulation research on network virus propagation with variable correlation coefficients [J]. *Journal of System Simulation*, 2012, 24(8): 1723-1727.
- [17] Borge-Holthoefer J, Moreno Y. Absence of influential spreaders in rumor dynamics [J]. *Physical Review E*, 2011, 85(2 Pt 2): 026116.
- [18] Klemm K, Serrano M Á, Eguíluz V M, et al. A measure of individual role in collective dynamics [J]. *Scientific Reports*, 2010, 2(2): 292.

- [19] Garas A, Argyrakis P, Rozenblat C, et al. Worldwide spreading of economic crisis [J]. *New Journal of Physics*, 2010, 12(2): 185-188.
- [20] Wang Xiaofan, Li Xiang, Chen Guanrong. *Network Science: An Introduction* [M]. Beijing: Higher Education Press, 2012.
- [21] Shi Yimin, Xu Wei, Qin Chaoying, et al. *Mathematical Statistics* [M]. 3rd ed. Beijing: Science Press, 2009.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.