Applied Mathematics and Computation 334 (2018) 388–400

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Improved centrality indicators to characterize the nodal spreading capability in complex networks

Juan Wang[a,b], Chao Li[c,d], Chengyi Xia[c,d,*]

[a] Tianjin Key Laboratory for Control Theory and Complicated Industry Systems,Tianjin University of Technology, Tianjin 300384, PR China
[b] School of Electrical and Electronic Engineering, Tianjin University of Technology, Tianjin 300384, PR China
[c] Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, PR China
[d] Key Laboratory of Computer Vision and System (Ministry of Education), Tianjin University of Technology, Tianjin 300384, PR China

A R T I C L E I N F O

Keywords:
Centrality indicator
Degree centrality
k-shell value
Mixed gravitational centrality
Influential spreaders
Complex networks

A B S T R A C T

In this paper, we deeply investigate the identification of influential spreaders in complex networks based on various centrality indices. At first, we introduce several frequently used centrality indices to characterize the node influence. Then, based on the standard SIR model, we integrate various centrality indicators into the characterization of the nodal spreading capability, and then starting from the gravitational centrality formula, we systematically compare the ranking similarity and monotonicity under various centrality algorithms over 6 real-world networks and Barabási-Albert model networks. The extensive simulations indicate that the mixed measure of gravitational centrality combining the $k-$shell value and degree will display the best performance as far as the ranking results are concerned, in which the focal node used the $k-$shell value as his mass while his neighboring nodes viewed the degree value as their masses. The current results are beneficial for us to develop the effective methods to discover and protect the significant nodes within many networked systems.

1. Introduction

As is well known, the contagion phenomena, which include the information and rumor diffusion, epidemic spreading and computer virus transmission, are widespread within the increasingly networked society and human life [1–10]. Deeply understanding the properties of transmission dynamics within the population is significant for us to fully utilize the spreading characteristics to improve the system performance [11,12], as an example, how to effectively identify the influential spreaders within the network or population is an important step to monitor the public opinion [13,14], control the epidemic outbreak [15,16] and promote the new product, technology or innovation idea [17,18], and thus it is of high importance in the theory and practice. At present, various centrality indicators have been proposed to judge whether an individual node is influential [19,20], and even to characterize the nodal spreading capability such as degree centrality [21], betweenness centrality [22,23], eigenvector centrality [24], closeness centrality [25], page rank centrality [26], leader rank centrality and core centrality [27] and so on. All these indices assume that there exists some relationship between the topology position of an

* Corresponding author at: Tianjin Key Laboratory of Intelligence Computing and Novel Software Technology, Tianjin University of Technology, Tianjin 300384, PR China.
E-mail address: xialooking@163.com (C. Xia).

https://doi.org/10.1016/j.amc.2018.04.028
0096-3003/

individual node in the network and its influence or capacity on the basis of the spreading dynamics [28]. In particular, Kitsak et al. [29] put forward a k-shell decomposition method to divide various nodes into different ranks, within which each node has been thought to hold the same influence, and meanwhile the higher the k-shell value that a node has, the more influential the node is. Large quantities of evidences demonstrate that k-shell value is a better measure for the epidemic spreading process when compared to the degree centrality, especially for the initial outbreaks.

However, k-shell value is usually not very large within many real-world networks and k-shell is still a coarse-grained index, and thus the differentiation for the impact of nodes is not sufficient since many nodes of different degrees may be contained within the same ranking slot. Some recent works have also shown that the nodes may have distinct influences even if they own the same k-shell value, which may lose its effect when confronting some networks without core-like structure, e.g., Barabási-Albert (BA) network [30]. After this pioneering work, several works are devoted to further improving the performance of k-shell decomposition algorithm. For instance, Zeng et al. [31] integrated the residual degree and exhausted degree of each node into the decomposition process so that a mixed degree decomposition process was proposed. Liu et al. [32] showed that the k-shell method is invalid when a number of core-like groups exist within the whole network, and then they designed a method to improve the accuracy of k-shell method by removing the redundant links inside networks. Lin et al. [33] incorporated the shortest path distance between a target node and the node set with the highest k-core value into the ranking process, and then devised an enhanced ranking method to accurately locate the influential spreaders. Bae et al. [34] proposed a new measure centrality index to characterize the spreading capability of focal node by summing all its neighbors's k-shell values of its neighbors. Recently, Li et al. [35] proposed a novel classified neighbors algorithm to quantify the nodal spreading capability and further to differentiate the influence of various nodes, and then to classify the neighbors of the focal node according to the removal order of the neighbor during the process of k-shell decomposition.

In reality, in order to accurately characterize a node's influence, we may not only consider the focal node, but also take the nearest neighbors and the interaction between them into account. As an example, Ma et al. [36] borrowed the idea of universal gravity formula of Isaac Newton to propose a new centrality index, termed as the gravity centrality indicator, in which each node's $k-$shell value can be regarded as its mass and the shortest path length between two nodes can be viewed as their distance. Then, they summed all the gravity of the focal node's neighbors within limited radius as its influence. On the basis of susceptible-infected-recovered (SIR) model [37,38], the experimental results on several real-world data sets have proved that this index is a little more superior to those traditional ones, such as degree centrality, betweenness centrality, closeness centrality and k-shell method and so on. Nevertheless, as mentioned above, k-shell value is a relatively coarse-grained index to measure the node influence, and we further refine the gravity centrality index by replacing the k-shell value with the degree centrality for the neighboring nodes so that the redundant information about the k-shell value can be ignored in the evaluation of individual influence. Extensive experiments have been performed within 6 real-world networks and 4 BA model networks, and the results indicate that the modified gravity centrality index can better measure the node influence when compared to other centrality indices.

The rest of this paper is structured as follows. In Section 2, we summarize several centrality indices and propose our gravity centrality index in detail. After that, we briefly describe the SIR epidemic model and the corresponding real-world network data sets adopted in this paper, and then we will perform extensive numerical simulations upon 6 real-world networks and 4 model networks, and we also compare the present method and other typical algorithms in Section 3. Finally, we conclude the paper with some concluding remarks in Section 4.

2. Methods and models

At present, the complex network has provided a powerful framework to gauge the nodal importance or ranking in the whole population, which is beyond the well-mixed topology and can better characterize the interaction between nodes. Among them, an undirected and unweighted network has often been described as a simple graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ is the set of nodes and $\mathbf{E}$ means the set of links between nodes, correspondingly, the number of all nodes and links are assumed to be N and M, respectively. Meanwhile, $\mathbf{A} = (a_{ij})_{N\times N}$ denotes the adjacent matrix of the graph, where $a_{ij} = 1$ if node i is connected to node j, or else a_{ij} is set to be 0. In this section, starting from the basic definitions mentioned above, we firstly introduce several frequently used centrality indices and their potential applications, and then propose our modified gravitational indicator to further enhance the ranking performance.

2.1. Frequently used centrality indices

Centrality-based methods are often used to rank nodes, and it is supposed that the higher the ranking value is, the more influential the node is. Thus, the top ranked nodes are considered to hold the largest spreading capabilities, and we will here introduce several typical centrality indices, which include degree, betweenness, closeness, eigenvector, k-shell, mixed degree decomposition, gravitational centrality and so on.

Degree centrality (DC) [21] may be the simplest and most intuitive indicator to measure the nodal importance. The degree of a node can be calculated as the number of nearest neighbors, and the higher degree node is viewed as the influential node since it can spread his impact into more other nodes. Taking the online social networks as an example, a person with a huge number of fans may be popular, who can attract a great deal of concern from the fans and transmit the products