Physica A 512 (2018) 379–391

Contents lists available at ScienceDirect

Physica A

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

Leveraging local h-index to identify and rank influential spreaders in networks

Qiang Liu [a,*], Yu-Xiao Zhu [b], Yan Jia [a,c], Lu Deng [a], Bin Zhou [a,c], Jun-Xing Zhu [a], Peng Zou [a]

[a] College of Computer, National University of Defense Technology, Changsha 410073, China
[b] School of Management, Guangdong University of Technology, Guangzhou, 510520, China
[c] State Key Laboratory of High Performance Computing, National University of Defense Technology, Changsha 410073, China

HIGHLIGHTS

- We propose a local h-index centrality (LH-index) method for identifying and ranking influential spreaders in networks.
- The LH-index can resolve the resolution limit problem of the H-index.
- The LH-index does not need the global topological information of the network.
- The LH-index value of a node is not sensitive to the small variation of the degree information of the node itself and its neighbors.
- The performance of the LH-index shows its superiority in both real world and simulated networks.

ARTICLE INFO

Article history:
Received 20 January 2017
Received in revised form 14 April 2018
Available online xxxx

Keywords:
Complex networks
Influential nodes
Local h-index centrality
SIR model

ABSTRACT

Identifying influential nodes in complex networks has received increasing attention for its great theoretical and practical applications in many fields. Some classical methods, such as degree centrality, betweenness centrality, closeness centrality, and coreness centrality, were reported to have some limitations in detecting influential nodes. Recently, the famous h-index was introduced to the network world to evaluate the spreading ability of the nodes. However, this method always assigns too many nodes with the same value, which leads to a resolution limit problem in distinguishing the real influences of these nodes. In this paper, we propose a local h-index centrality (LH-index) method to identify and rank influential nodes in networks. The LH-index method simultaneously takes into account of h-index values of the node itself and its neighbors, which is based on the idea that a node connecting to more influential nodes will also be influential. Experimental analysis on stochastic Susceptible–Infected–Recovered (SIR) model and several networks demonstrates the effectivity of the LH-index method in identifying influential nodes in networks.

1. Introduction

Identifying influential nodes in networks has become a hot topic in recent years for its wide applications in many fields, such as social network analysis [1–3], viral marketing [4], epidemic spreading and controlling [5,6]. By finding influential nodes, we can have a better understanding of the characteristics of the network structure and function so as to fulfill various applications being closely related to human lives. Meanwhile, some further applications based on influential nodes

* Corresponding author.
E-mail address: liuqiang1981@nudt.edu.cn (Q. Liu).

https://doi.org/10.1016/j.physa.2018.08.053
0378-4371/

identification also receive increasing attention from scholars. For instance, some have utilized influence propagation method to find densely connected parts in networks [7,8], while some others try to find laws of information transmission in networks by utilizing influential nodes [9,10]. Eventually, all these methods contribute greatly to the requirements of new ways for identifying influential nodes in networks. Therefore, how to effectively identify influential nodes in networks has become an urgent problem.

Traditional centrality measures, such as degree centrality [11], betweenness centrality [12], closeness centrality [13] and coreness centrality [14], have been adopted to evaluate the influences of the nodes in networks. Given one node, the degree centrality [11] measures its influence by counting the number of its connected neighbors. However, the degree centrality neglects the location and structure information of a node, which may not fairly reflect the real influence of the node. For example, one high degree node may be located in the periphery of the network, then it can only affect very few other nodes. Both the betweenness centrality [12] and closeness centrality [13] are global centrality methods, and they measure the influence by computing the number of shortest paths in the whole network, so these two methods are not fit for the large-scale networks due to their high computational complexity. The coreness centrality [14] measures the influence by k-core decomposition, the node with higher coreness value means that its location is more central in the network. Note that, the computing of coreness needs the global topological information of the whole network.

In recent years, some new centrality measures have been put forward. Lü et al. [15] devised an adaptive and parameter-free LeaderRank algorithm to identify influential nodes in social networks, which achieves good performance for directed networks but is weak for undirected networks. Chen et al. [16] proposed a semi-local centrality measure which considers both the nearest and the next nearest neighbors' degree information of a node. However, the degree information of a node sometimes cannot reflect its real influence fairly, because the influences of its different neighbors always vary greatly. Whereafter, some core-based methods were proposed. Bae et al. [17] studied the neighbor's coreness centrality of a node for identifying influential nodes in networks. Ma et al. [18] proposed the neighbor's gravity centrality to detect influential nodes in networks. Liu et al. [19] generalized the neighbors' centrality methods [16,17] in networks, and made detailed analyses of the neighbor's degree centrality and coreness centrality. Malliaros et al. [20] capitalized on the properties of the K-truss decomposition to locate influential nodes, which can filter out less important information and detect influential nodes. Basu et al. [21] proposed a group density method based on core analysis to identify influential nodes in weighted networks, this method was inspired by the idea of game theoretic concept of voting mechanism. However, the core-based method always needs global topological information of the network, thus methods of this kind always own high computational complexity, which hinders their practical applications in large networks. Besides, recent research [22] pointed out that the node with the highest coreness value may not be the most influential spreader due to the existence of the core-like groups. Later, Liu et al. [23] found that the core-like groups can be attributed to the existence of redundant links, then they proposed one new method to improve the accuracy of the K-core decomposition by removing these redundant links. Nonetheless, all the centrality measures that mentioned above are sensitive to the small variation of nodes' degree information. For instance, if the collected data set misses some connection information of few nodes, then the influences of these nodes, which are measured by degree centrality or coreness centrality, will be affected directly.

The h-index, proposed by Hirsch in 2005 [24], is a classical metric to measure both the productivity and citation impact of the publications of a scholar. Recently, some scholars introduced the h-index to measure a node's influence in networks. For example, Korn et al. [25] illustrated that the h-index can make a well-balanced mix of traditional centrality measures. Lü et al. [26,27] also demonstrated the effectivity of the h-index in evaluating a node's influence in many real world networks. However, we find that the h-index method always assigns the same value to many nodes, which makes it difficult to distinguish the real influences of these nodes. So it is necessary to design one better ranking method which can improve its resolution power in identifying influential nodes.

Inspired by the classical h-index, we put forward one novel method to better qualify the spreading ability of the nodes in this paper. The new proposed local h-index (LH-index) method simultaneously considers two factors: the h-index value of the node itself and the h-index values of its neighbors. On one hand, the h-index of one node indicates the direct influences exerted by its nearest influential neighbors. On the other hand, the h-index values of its neighbors indicate the two-hop indirect influences exerted by further influential neighbors. That is to say, the node which is surrounded by much more high h-index neighbors is more influential. We adopted the epidemic spreading process to evaluate the performance of the LH-index method in identifying influential nodes, The experimental results demonstrate that the LH-index outperforms many traditional methods in many networks.

The rest of this paper is organized as follows. In Section 2, we will briefly review some centrality measures and present the definition of the new LH-index. In Section 3, we will introduce the network data sets, the spreading model and the evaluation method. In Section 4, we will present our experimental results. Finally, conclusions will be given in Section 5.

2. Methods

In this paper, we first consider four classical centrality measures, which are degree centrality, betweenness centrality, closeness centrality and coreness centrality. Consider an undirected network $G = (V, E)$, where V is the vertex set, and E is the edge set. The degree centrality (DC) [11] of node i is defined as the number of i's nearest neighbors. The betweenness centrality (BC) [12] of node i is defined as the fraction of the number of shortest paths between node pairs that travel through