Physica A 451 (2016) 205–212

Contents lists available at ScienceDirect

Physica A

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

Identifying influential spreaders in complex networks based on gravity formula

Ling-ling Ma [a], Chuang Ma [a], Hai-Feng Zhang [a,b,c,*], Bing-Hong Wang [d]

[a] School of Mathematical Science, Anhui University, Hefei 230601, PR China

[b] Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 211189, PR China

[c] Department of Communication Engineering, North University of China, Taiyuan, Shan'xi 030051, PR China

[d] Department of Modern Physics, University of Science and Technology of China, Hefei 230026, PR China

HIGHLIGHTS

- Each node's k-shell value is considered as its mass and the shortest path distance between two nodes is viewed as their distance.
- A new method based on gravity formula is proposed to identify the influential nodes in complex networks.
- Our method yields better performance of identifying the influential nodes than many previous methods.
- The method can be further generalized.

ARTICLE INFO

Article history:
Received 11 May 2015
Received in revised form 6 November 2015
Available online 29 January 2016

Keywords:
Complex networks
Influential spreader
Gravity formula

ABSTRACT

How to identify the influential spreaders in social networks is crucial for accelerating/hindering information diffusion, increasing product exposure, controlling diseases and rumors, and so on. In this paper, by viewing the k-shell value of each node as its mass and the shortest path distance between two nodes as their distance, then inspired by the idea of the gravity formula, we propose a gravity centrality index to identify the influential spreaders in complex networks. The comparison between the gravity centrality index and some well-known centralities, such as degree centrality, betweenness centrality, closeness centrality, and k-shell centrality, and so forth, indicates that our method can effectively identify the influential spreaders in real networks as well as synthetic networks. We also use the classical Susceptible–Infected–Recovered (SIR) epidemic model to verify the good performance of our method.

1. Introduction

To effectively identify influential spreaders in social networks is of theoretical and practical significance [1–11], since it is crucial for developing efficient strategies to control epidemic spreading, accelerate information diffusion, promote new products, and so on. In view of this, many centrality indices have been proposed to address this problem, including degree centrality [12], betweenness centrality [13], neighborhood centrality [14] and closeness centrality [15], etc. In particular, Kitsak et al. proposed a *k*-shell decomposition method to identify the most influential spreaders based on the assumption that nodes in the same shell have similar influence and nodes in higher shells are likely to infect more nodes. *k*-shell

* Corresponding author at: School of Mathematical Science, Anhui University, Hefei 230601, PR China.
E-mail address: haifengzhang1978@gmail.com (H.-F. Zhang).

http://dx.doi.org/10.1016/j.physa.2015.12.162

method is found to be better than the degree centrality index in many real networks [1]. However, recent researches have demonstrated that the nodes within the same shell often have distinct influences, and this method may fail in some networks without core-like structure, e.g., Barasási–Albert network [16]. Thus, after this, some methods were proposed to further improve the performance of the k-shell method. For example, Zeng et al. proposed a mixed degree decomposition method by incorporating the residual degree and the exhausted degree [17]; Liu et al. have demonstrated that the existence of the core-like groups can result in the invalidation of k-shell method [18], and then they showed that the accuracy of k-shell method can be improved once the redundant links in networks are removed [19]. Chen et al. designed a semi-local index by considering the next nearest neighborhood [20]; Lin et al. presented an improved ranking method by taking into account the shortest path distance between a target node and the node set with the highest k-core value [21]; Recently, Bae et al. defined a novel measure-coreness centrality index, which is given by summing all neighbors' k-shell values [22].

In general, a node's influence is not only dependent on its nearest neighbors but also on the nodes who are not the nearest neighbors [23,24], meanwhile, their interaction influence commonly decreases with their shortest path distance. If the k-shell value of each node is viewed as its mass, and the shortest path distance between two nodes is defined as their distance, then we can use the idea of gravity formula proposed by Isaac Newton to measure the influence of nodes. Inspired by these factors, in the work, we propose a new centrality index to measure the influence of nodes, which is called gravity centrality index. We apply the susceptible–infectious–recovered (SIR) spreading dynamics to evaluate the effectiveness of our proposed method, the experimental results indicate that gravity centrality index can better evaluate the influence of nodes than the ones generated by degree centrality, betweenness centrality, k-shell centrality, closeness centrality, and so on.

The layout of the paper is as follows: In Section 2, we first briefly review several typical centrality indices which are used to compare in this work, and the description of our method is presented. Then the experimental results are presented in Section 3. Finally, Conclusions and discussions are summarized in Section 4.

2. Method

An undirected network is represented by $G = (N, M)$ with N nodes and M edges, and its structure can be described by an adjacent matrix $A = (a_{ij})_{N\times N}$ where $a_{ij} = 1$ if node i is connected to node j, and $a_{ij} = 0$ otherwise.

Here we briefly review the definitions of several centrality indices that will be discussed in this work.

The degree centrality (DC) of a node is defined as the number of nearest neighbors. The betweenness centrality (BC) of a node is defined as the fraction of all shortest paths travel through the node. The closeness centrality (CC) of a node is defined as the reciprocal of the sum of the lengths of the geodesic distance to every other node. The k-shell decomposition method (ks) is implemented by the following steps: Firstly, remove all nodes with degree one, and keep deleting the existing nodes until all nodes' degrees are larger than one. All of these removed nodes are assigned 1-shell. Then *recursively* remove the nodes with degree no larger than two (i.e., remove all nodes with degree two, and keep deleting the existing nodes until all nodes' degrees are larger than two.) and include them to 2-shell. This procedure continues until all nodes have been assigned to one of the shells [17].

To improve the exactness of k-shell method, the mixed degree decomposition (MDD) method was proposed by Zeng et al. [17]. The mixed degree $k_m(i)$ for a node i is defined by considering the residual degree $k_r(i)$ and the exhausted degree $k_e(i)$ simultaneously, which is written as:

$$k_m(i) = k_r(i) + \lambda * k_e(i). \tag{1}$$

At each step of the MDD procedure, the nodes are removed according to the mixed degree, and the mixed degrees of remaining nodes are also updated. Where λ is a tunable parameter between 0 and 1. As in Ref. [17], we take $\lambda = 0.7$ in this work.

Recently, Baus et al. designed a ranking method-neighborhood coreness C_{nc} by considering the degree and the coreness of a node simultaneously, the $C_{nc}(i)$ for a node i is defined as [22]

$$C_{nc}(i) = \sum_{j\in\Lambda_i} ks(j), \tag{2}$$

where Λ_i is the neighbor node set of node i. They further developed an extended neighborhood coreness C_{nc+}, which is described as:

$$C_{nc+}(i) = \sum_{j\in\Lambda_i} C_{nc}(j). \tag{3}$$

Chen et al. proposed a semi-local centrality measure as a tradeoff between low-relevant degree centrality and other time-consuming measures (labeled as SL index). It considers both the nearest and the next nearest neighbors. The semi-local centrality $SL(i)$ of node i is defined as [20]

$$Q(s) = \sum_{j\in\Lambda_s} N(j), \tag{4}$$

$$SL(i) = \sum_{s\in\Lambda_i} Q(s), \tag{5}$$

where Λ_i is the neighbor node set of node i. $N(j)$ is the number of the nearest and the next nearest neighbors of node j.