Chaos, Solitons and Fractals 86 (2016) 1–7

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Identifying influential spreaders by weight degree centrality in complex networks

Yang Liu[a], Bo Wei[b], Yuxian Du[a], Fuyuan Xiao[a], Yong Deng[a,c,*]

[a] School of Computer and Information Science, Southwest University, Chongqing 400715, China
[b] Institute of Intelligent Control and systems, Harbin Institute of Technology, Harbin 150080, China
[c] School of Electronics and Information, Northwestern Polytechnical University, Xian, Shaanxi, 710072, China

ARTICLE INFO

Article history:
Received 9 January 2016
Accepted 27 January 2016
Available online 1 March 2016

Keywords:
Complex networks
Weight degree centrality method
Influential spreaders
Assortativity

ABSTRACT

The problem of identifying influential spreaders in complex networks has attracted much attention because of its great theoretical significance and wide application. In this paper, we propose a successful ranking method for identifying the influential spreaders. The proposed method measures the spreading ability of nodes based on their degree and their ability of spreading out. We also use a tuning weight parameter, which is always associated with the property of the networks such as the assortativity, to regulate the weight between the degree and the ability of spreading out. To test the effectiveness of the proposed method, we conduct the experiments on several synthetic networks and real-world networks. The results show that the proposed method outperforms the existing well-known ranking methods.

1. Introduction

Spreading processes are ubiquitous in nature [1], such as the disease spreading [2,3], reaction diffusion processes [4], pandemics [5], cascading failures [6] and Game theory [7–10]. The problem of identifying influential spreaders in complex networks has gained much attention [11,12] in the recent years because of its theoretical significance and remarkable practical value in controlling rumor spreading [13,14], disease spreading [15] and creating new marketing tools [16–19]. Some new and advanced math tools, such as fuzzy sets and evidence theory [20–24], are widely used to model and combine uncertain information, which make them successfully applied to identify influential spreaders [25,26]. In addition, many bio-inspired methods, such as genetic algorithm [27], particle swarm optimization [28], physarum polycephalum algorithm [29], shows great advantages and promising aspect in this field.

Degree centrality is a straightforward and efficient local metric, however, it is less relevant since it neglects the global structure of the network: a node with a few high influential neighbors may have much higher influence than a node with a larger number of less influential neighbors [11]. Betweenness centrality [30] and closeness centrality [31] are well-known global metrics to give better results, but the both are incapable to be applied in large-scale networks because of their high computational complexity. Recently, Kitsak et al. [1] presented a new method, the k-shell decomposition, which restructured the network with different levels. Although the most influential spreaders were located in the core of the network, there are difficult to distinguish the influence of each other. To overcome the shortcoming of monotonicity, mixed degree decomposition [32], θ method [33] and coreness centrality [16] were developed. Pei et al. showed that k-shell may be more efficient in complete real-world structure network than other methods [34]. Furthermore, they showed the sum of the

* Corresponding author at: School of Computer and Information Science, Southwest University, Chongqing 400715, China. Tel.: +86 2368254555.
E-mail address: ydeng@swu.edu.cn, prof.deng@hotmail.com (Y. Deng).

http://dx.doi.org/10.1016/j.chaos.2016.01.030

nearest neighbors' degree is a reliable local proxy for user's influence when the complete global network structure is unavailable.

In this paper, motivated by authors in Refs. [16,35–37], we proposed a novel ranking method for identifying the influential spreaders of the complex networks, which is more efficient than the accepted ranking methods. The main idea of the proposed method is to quantify the spreading capability of a node according to the degree and the ability of outspreading of it. To test the performance of our proposed strategy, we conduct the experiments on Erdős–Rényi (ER) networks [38], scale-free (SF) networks [39] and also several real-world networks. The results show that the proposed method outperforms the betweenness centrality method, the degree centrality method, the k-shell decomposition method, the neighborhood coreness method and the extended neighborhood coreness method.

The rest of the paper is organized as follows: in Section 2, the proposed strategy is described; in Section 3, the experimental setup is introduced; in Section 4, related experiments are conducted and experimental results are evaluated and conclusions are presented in Section 5.

2. Method

Consider a network $G(N, E)$ where N and E are the set of nodes and edges, respectively. In [16], Bae and Kim proposed a coreness centrality method to measure the ability of spreaders using the neighborhood coreness:

$$C_{nc}(v) = \sum_{u \in \Gamma(v)} ks(u) \tag{1}$$

where $C_{nc}(v)$ is the neighborhood coreness of node v, $\Gamma(v)$ is the set of the neighbors adjacent to node v and $ks(u)$ is the k-shell index of its neighbor node u. The extended neighborhood coreness C_{nc+} is defined as follows:

$$C_{nc+}(v) = \sum_{u \in \Gamma(v)} C_{nc}(u) \tag{2}$$

C_{nc} and C_{nc+} are highly effective, but it requires global information about the network and may be infeasible in some networks such as scale-free (SF) networks [39], which makes it impractical in real case. In [36,37], Wang et al. used the power-law function of degree w_{ij} to measure the weight of an edge ij:

$$w_{ij} = (d_i d_j)^{\theta} \tag{3}$$

in which d_i the degree of node i and θ is a tuning parameter. They found that $\theta = 1$ leads to the strongest robustness on various networks.

Inspired by authors in Refs. [16,36,37], we propose a weight degree centrality (W_{dc}) method which consists of the degree of a node and the degree of its neighborhoods and holds the advantage of local information and wider application compared to C_{nc}:

$$D_i = \sum_{j \in \Gamma(i)} d_j \tag{4}$$

$$W_{dc}(i) = (D_i - d_i) \cdot d_i^{\alpha} \tag{5}$$

where D_i is the sum of the degree of the neighborhoods of node i and α is a tuning parameter which regulates the weight of d_i in W_{dc}. $(D_i - d_i)$ denotes the ability of outspreading of node i. Recursively, the extended weight degree centrality method (EW_dc) of node i is defined as follows:

$$ED_i = \sum_{j \in \Gamma(i)} W_{dc}(j) \tag{6}$$

$$EW_{dc}(i) = (ED_i - W_{dc}(i)) \cdot W_{dc}(i)^{\alpha} \tag{7}$$

In Ref. [35], the author classifies the networks into assortative networks (where high degree nodes tend to connect to other high degree nodes), disassortative networks (high degree nodes mostly have neighbors with a small number of connections.) and neutral networks by the assortativity coefficient r. In this paper, we try to use $\alpha = |r|$ as r to divide the networks into two types: neighborhood-relevant networks (assortative and disassortative networks) and neighborhood-irrelevant networks (neutral networks). Eqs. (5) and(7) are rewritten as follows:

$$W_{dc}(i) = (D_i - d_i) \cdot d_i^{|r|} \tag{8}$$

$$EW_{dc}(i) = (ED_i - W_{dc}(i)) \cdot W_{dc}(i)^{|r|} \tag{9}$$

Note that α, as a tuning parameter, can be assigned other values.

3. Experimental setup

To evaluate the performance of the proposed method, the SIR model [11,40–42] is used to examine the spreading influence of the ranked nodes. In the SIR model, each node belongs to one state of the susceptible state, the infected state and the recovered state. At the initial stage, we set a node that we are interested in to be infected to investigate the influence of this node, and the others to be susceptible. Then, at each time step, the infected nodes infect its susceptible neighbors with infection probability λ, and they recover with probability η. The recovered nodes are removed from the network. This process is repeated until there is no infected node in the network. At the end of an epidemic process, the number of recovered nodes (removed nodes) is approximate to estimate the influence of the initially infected node. Notice that in order to measure the spreading ability of each node more efficiently, we set the recovered probability $\eta = 1$.

Kendall's tau as a rank correlation coefficient [16,43,44] is usually used to quantify the correctness of the ranking methods. Let (x_i, x_j) and $(y_i, y_j)\{i, j \in V\}$ be a set of joint variables from two ranking lists, X and Y, respectively. Any pair of observations (x_i, y_i) and (x_j, y_j) is said to be concordant if the ranks for both elements agree: that is, if $x_i > x_j$ and $y_i > y_j$ or if $x_i < x_j$ and $y_i < y_j$. They are said to be discordant, if $x_i > x_j$ and $y_i < y_j$ or if $x_i < x_j$ and $y_i > y_j$. If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant. Then, the Kendall's tau τ is defined as:

$$\tau(X, Y) = \frac{n_c - n_d}{0.5n(n-1)} \tag{10}$$