Chaos, Solitons and Fractals 104 (2017) 485–494

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

Influential nodes ranking in complex networks: An entropy-based approach

Ahmad Zareie, Amir Sheikhahmadi*, Adel Fatemi

Department of Computer Engineering, Islamic Azad University, Sanandaj Branch, Sanandaj, Iran

ARTICLE INFO

Article history:
Received 1 February 2017
Revised 30 July 2017
Accepted 7 September 2017

Keywords:
Social networks
Influential nodes
Entropy
Node centrality
Node ranking

ABSTRACT

Measurement of the spreading capability of nodes has been one of the most attractive challenges in the field of social networks. Because of the huge number of nodes in a network, it has appealed to many researchers to find an accurate measure which can potentially detect the spreading capability and rankings of nodes. Most of the available methods determine the spreading capability of nodes based on their topological locations. In this paper, however, we have proposed a new measure based on the basic notions in information theory to detect the spreading capability of nodes in networks on the basis of their topological information. The simulation and experimental results of investigating real-world and artificial networks show that the proposed measure is more accurate and efficient than the similar ones.

1. Introduction

Extension and development of social networks has offered a new opportunity to spread and propagate ideas, news, messages and advertisement. Through online social networks, a large part of the society can be apprised of special news and ideas in the easiest way and shortest time. Therefore, it is considered as an important issue in these networks to detect influential nodes to maximize the influence of a message. Influence maximization can be defined as identification of a small subset of nodes in order to maximize the spread of influence; in other words, it involves maximization of the number of nodes which are influenced, under a specific diffusion model [1]. Identifying the spreading capability of nodes in networks is a key parameter in successful treatment of this problem [1–4].

A variety of measures and approaches have been proposed so far by many scholars in various branches of science. A significant number of these approaches are based on the structural measures which determine the spreading capability of nodes using their topological locations. The above measures can be divided into centrality and link topological measures [5]. In the former category, the centrality of a node in the network is a determinative factor in its spreading capability. The centrality of nodes can be calculated with different measures such as degree centrality [6], closeness [7], betweenness [8] and eigenvector [9]. These measures may be in conflict with each other; for example, a node may have high centrality in a measure but lower centrality in another [10 ch3]. Despite these potential differences, the common ground for all approaches of this type is their emphasis on the importance of having more and closer neighbors as a key factor in determination of the spreading capability of a node. [10 ch3].

HITS (Hyperlink-Induced Topic Search) [11], PageRank [12], Leader Rank [13] and Weighted Leader Rank [14] can be referred to as some measures of the second type. Appreciating the critical value of neighboring nodes constitutes the backbone of these measures; moreover, acknowledging the fact that it is more valuable to be a neighbor of important nodes than of the others is generally thought of as another key factor in a thorough understanding of the measures of the second type [5].

Kitsak, 2010, proposed a new centrality measure called “k-shell decomposition” [15]. This measure, which determines the centrality of nodes based on their locations in the network, considers nodes topographically located in the core of the network as influential nodes. It assigns a k-shell index to every node, so that higher k-shells are assigned to nodes located close to the core of the network. Compared to older approaches, k-shell decomposition has shown a significant improvement in specification of node influence. Different measures have been proposed so far based on this approach, and some of them are mixed degree decompositions [16], shortest distance to highest k-shell value node [17], minimum k-shell method [18], neighborhood coreness centrality measure [19], k-shell iteration factor [20] and mixed core, degree and entropy [21]. H-index [22] is another centrality measure, which identifies the spreading capability of nodes according to the concept of h-index and the degrees of the neighbors.

* Corresponding author.
E-mail address: sheikhahmadi@eng.ui.ac.ir (A. Sheikhahmadi).

http://dx.doi.org/10.1016/j.chaos.2017.09.010
0960-0779/

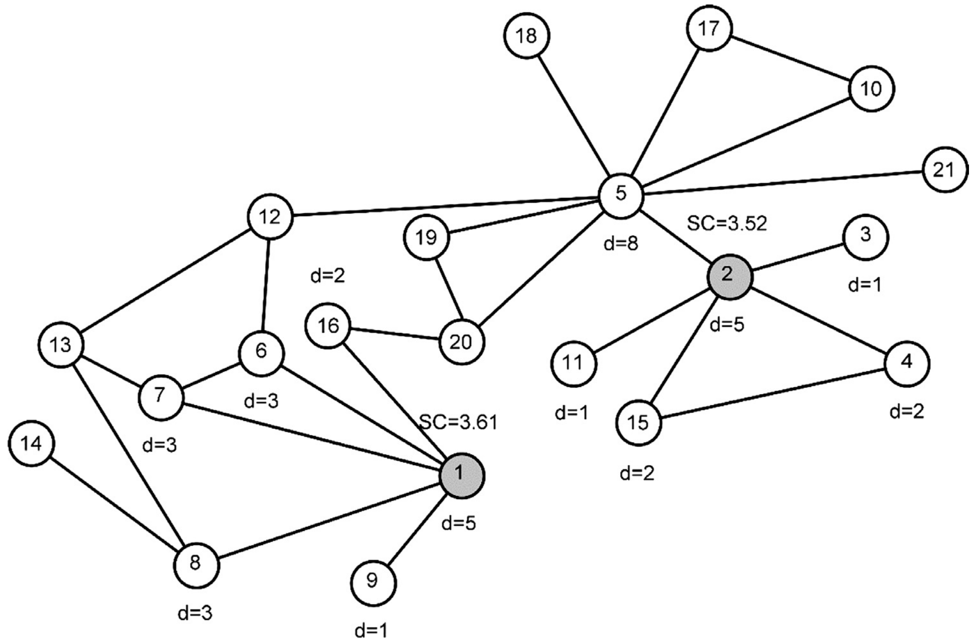

Fig. 1. Schematic network.

In this study, we will propose a measure for specification of the spreading capability of nodes in networks through the concept of entropy. Susceptible-Infected-Recovered (SIR) model [23–26] has been used to evaluate the proposed measure, called Entropy-Based Ranking Measure (ERM) hereafter. The experimental results of investigating real-world and artificial networks show that ERM outperforms the state-of-the-art approaches in terms of monotonicity, correctness of the ranking list and efficiency.

The reminder of the paper is structured along the following sections. Our motivation and the details of the proposed approach are described in Section 2. The simulation strategies and exhaustive experimental results are reported in Section 3. Section 4 is devoted to the discussions and concluding remarks.

2. Motivation and proposed method

One of the various approaches that have been proposed so far for identifying the spreading capability of nodes in complex networks is using node degrees [6]. However, as shown in [19,20], it is not sufficient for specification of the spreading capability of nodes to employ their degrees, since it is just the number of neighboring nodes that is considered. It can improve the accuracy of the approach for measuring the spreading capability of nodes to take into account the degrees of the neighbors of the node; a node with neighbors that have high degrees has greater spreading capability. It is worth pointing out, however, that the high degree of a node, or its neighbors, is not sufficient for specification of its spreading capability. For example, consider nodes 1 and 2 in Fig. 1. In this figure, d_i is the degree of node i, and SC_i (Spreading Influence) refers to the spreading capability of node i, which is the average number of nodes infected by node i in 1000 repetitions of the SIR model. The degree of node 1, d_1, is 5, and that of node 2, d_2, is 5. Moreover, the total degrees of the neighbors of nodes 1 and 2 are 12 and 14, respectively. A relevant point that merits mentioning here is that the presence of node 5 in the neighborhood of node 2 has increased the sum of its neighbors' degrees. Although the sum of the neighbors' degrees is larger in node 2, the *SC* of node 1 is proven to be larger than that of node 2. This is the case since any probable problem in spread of the message into node 5 may potentially result in dramatic decrease in the spreading capability of node 2. A possible hypothesis at this point to account for this fact is "a node has high spreading capability if its neighbors' degrees are both high and uniform." The concept of entropy, details about which will be discussed in the next section, is used for evaluating the above hypothesis.

2.1. Information entropy

Shannon, an American scholar, founded information entropy in 1948. Entropy is a concept of information theory which determines the amount of information in any event. Based on Shannon entropy, if X is a set of possible events $x_1, x_2, \ldots, x_n$, and p_i is the probability of x_i, the entropy of X can be calculated by Eq. (1).

$$E(X) = -\sum_{i=1}^{n} p_i \times log(p_i) \tag{1}$$

On the one hand, if the probabilities have uniform distribution over the domain "X" the maximum value of entropy will be obtained [27]. On the other hand, as the value of n increases, so does the entropy value. Therefore, employing entropy can be useful for detection of nodes with high-degree, more uniform neighbors.

2.2. Proposed approach

The Social Network graph is shown by $G=(V, E)$, where nodes $V=\{v_{1,\ldots,} v_{|V|}\}$ denote the network users and edges $E=\{e_{1,\ldots,} e_{|E|}\}$ show the relations between the users. The numbers of nodes and edges are represented by $|V|$ and $|E|$ respectively. Nodes v_i and v_j are called neighbors if there is an edges between them. The degree of v_i is denoted by d_i, and N_i shows its neighbor set. The total degree of the neighbors of v_i is shown by $d_i{}^1$, and is calculated as $d_i{}^1 = \sum_{v_j \in N_i} d_j$; the total degree of the neighbors of v_i's neighbors, which will be named second-order neighbors hereafter, is shown by $d_i{}^2$, and is calculated as $d_i{}^2 = \sum_{v_j \in N_i} d_j^1$.

In the proposed ERM, the entropy of the degrees of the neighbors and second-order neighbors of node $d_i{}^1$ are calculated by Eqs.