Physica A 465 (2017) 312–330

Contents lists available at ScienceDirect

Physica A

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

Identifying and ranking influential spreaders in complex networks with consideration of spreading probability

Qian Ma, Jun Ma *

School of Computer Science and Technology, Shandong University, Jinan, 250101, China

HIGHLIGHTS

- The common centralities are sensitive to the spreading probability.
- A hybrid degree centrality is proposed to evaluate nodes' spreading ability.
- Our method integrates degree and local centrality with spreading probability.
- Our method is not sensitive to the spreading probability.
- Our method outperforms other centralities in both real and artificial networks.

ARTICLE INFO

Article history:
Received 4 March 2016
Received in revised form 21 June 2016
Available online 23 August 2016

Keywords:
Complex network
Ranking
Spreading
Centrality measures

ABSTRACT

Identifying the influential spreaders in complex network has great theoretical and practical significance. In order to evaluate the spreading ability of the nodes, some centrality measures are usually computed, which include degree centrality (DC), betweenness centrality (BC), closeness centrality (CC), k-shell centrality (KS) and local centrality (LC). However, we observe that the performance of different centrality measures may change when these measures are used in a real network with different spreading probabilities. Specifically, DC performs well for small spreading probabilities and LC is more suitable for larger ones. To alleviate the sensitivity of these centrality measures to the spreading probability, we modify LC and then integrate it with DC by considering the spreading probability. We call the proposed measure hybrid degree centrality (HC). HC can take the advantages of DC or LC depending on the given spreading probability. We use SIR model to evaluate the performance of HC in both real networks and artificial networks. Experimental results show that HC performs robustly under different spreading probabilities. Compared with these known centrality measures such as DC, LC, BC, CC and KS, HC can evaluate the spreading ability of the nodes more accurately on most range of spreading probabilities. Furthermore, we show that our method can better distinguish the spreading ability of nodes.

1. Introduction

The adoption of new ideas, products or innovations often depends on a few early adopters. Influential adopters may cause a large cascade of followers through the influence spread [1]. Therefore, identifying the influential nodes in complex networks has gained much attention in recent years [2–4]. It has great theoretical and practical value in rumor controlling [5], viral marketing [6,7], opinion leader detection [8], epidemic controlling [9,10], etc. The key to the influential nodes identification problem is how to measure the spreading ability of the nodes.

* Corresponding author.
E-mail addresses: maqiansdu@126.com (Q. Ma), majun@sdu.edu.cn (J. Ma).

http://dx.doi.org/10.1016/j.physa.2016.08.041
0378-4371/

Various centrality measures, such as degree [11], betweenness [12], closeness [13], and eigenvector centralities [14–16], have been used to evaluate the influence of the nodes. Degree centrality is a simple and widely used measure in ranking user's influence and it is efficient in many situations [17]. However as degree centrality neglects the global topological structure of the network, not all the nodes with high degree are influential spreaders. Betweenness and closeness are well known global centrality measures. Both of them are proposed based on the assumption that the influence spreads through the shortest paths [17]. However betweenness and closeness need to compute the shortest paths between any two nodes in the network, so the computation complexity of them is high. Furthermore it is hard to get the complete network structure for most large-scale networks nowadays. Kitsak et al. [18] argued that the most efficient spreaders are those located in the core of the network which are identified by k-shell decomposition. Compared with degree, the nodes in the core must be the high degree nodes while the high degree nodes may locate in the periphery of the network. Sen et al. [4,17] found that k-shell index of a node is a better predictor of spreading influence than degree. However the k-shell method usually assigns too many nodes with the identical shell value and it cannot distinguish the nodes' spreading ability well. For this reason a series of methods [18,2,19,20] have been proposed to improve the distinguish ability of k-shell measure. Chen et al. [11] proposed local centrality, which considers the nearest and the next nearest neighbors of a node. Local centrality is a tradeoff between the low-relevant degree centrality and other time-consuming global centralities. Local centrality is likely to be more effective to identify influential nodes than degree centrality as it utilizes more information while it has much lower computational complexity than betweenness and closeness centrality. On this basis, Gao et al. [12] also considered the topological connections among the neighbors and proposed local structural centrality (LSC). However we find that LC and LSC perform better than other centralities only when the spreading probability is large. When the spreading probability is small, LC and LSC are not effective indicators. Moreover, there are also many researches that work on identifying multiple spreaders [10,7,6]. The collective influence (CI) defined by F. Morone et al. [10] is an effective quantity to find the influential nodes which can keep the whole network connected.

However, all of these centralities only evaluate the node's influence by considering the topological structure of the network. They do not take the effect of spreading probability on the nodes into account. Nevertheless, it is obvious that the spreading probabilities of the networks describing diseases, opinions, rumors should be different. It raises a problem naturally whether the performance of these centralities is not sensitive for different spreading probabilities. Our experiments show the answer is no. There are also some references which point out the problem. For example, degree performs well at small spreading probabilities and global centralities are more suitable for larger ones [12,21]. So the methods of identifying influential spreaders should be robust to different spreading probabilities. In other words, the methods should give different influence ranking lists under different spreading probabilities. Until now most works only use spreading probability in spreading simulations and few works consider it in evaluating the influence of the node, which means they give the same ranking result under different spreading probabilities. Min et al. [21] proposed ranking influential spreaders by extension of degree to alleviate this issue, but this method is time-consuming as it needs to calculate the proper extension level index for each certain spreading probability.

In this paper, we propose a simple and efficient method to alleviate the sensitivity of single centrality to the spreading probability and discover the influential spreaders. Firstly we validate that the performance of commonly used centrality measures on evaluating the nodes' spreading ability is sensitive to the spreading probability. Then we divide the influence into near-source influence and distal influence. Degree centrality and modified local centrality are used as near-source influence and distal influence respectively. Degree is an efficient metric for small spreading probabilities and modified local centrality can achieve better results for larger spreading probabilities as they consider topological structure with different ranges. Based on this we propose an ensemble method with consideration of spreading probability to evaluate the influence of the node. To evaluate the performance of the method, we simulate the epidemic spreading process in real networks and artificial networks with standard Susceptible–Infected–Recovered (SIR) model and standard Susceptible–Infected (SI) model [9,22]. Experimental results show that our method is robust to the spreading probability and our method evaluates the nodes' spreading ability more accurately than other single centralities on most range of spreading probabilities. HC can also distinguish the spreading ability of nodes better than other centralities in different datasets.

The remainder of this paper is organized as follows. We briefly review the definition of centrality measures for comparison in Section 2 and introduce our method in Section 3. In Section 4 we present the details of the datasets, the spreading model and the evaluation measure. We present the experimental results in Section 5. Finally in Section 6, we expose the conclusions of the work.

2. Centrality measures

For an undirected network $G = (V, E)$, where V and E denote the set of nodes and edges in the network respectively. n and m are used to denote the number of nodes and edges. The graph G can be presented as an adjacency matrix $A = \{a_{uv}\} \in \{0, 1\}^{n\times n}$. If two nodes u and v are adjacent, then $a_{uv} = 1$, otherwise, $a_{uv} = 0$.

Degree centrality (DC) [23] of node v is defined as the number of v's neighbors. It reflects the influence of the node to other nodes directly.

$$C_D(v) = \sum_{u=1}^{n} a_{vu}.$$