Physica A 403 (2014) 130–147

Contents lists available at ScienceDirect

Physica A

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

Ranking the spreading ability of nodes in complex networks based on local structure

Shuai Gao[a], Jun Ma[a], Zhumin Chen[a,*], Guanghui Wang[b], Changming Xing[a,c]

[a] *School of Computer Science & Technology, Shandong University, Jinan, 250101, China*
[b] *School of Mathematics, Shandong University, Jinan, 250101, China*
[c] *School of Continuing Education, Shandong University of Finance and Economics, Jinan, 250014, China*

HIGHLIGHTS

- The structure of the neighbors of a node can affect its spreading ability.
- A local structural centrality method for ranking node's spreading ability is proposed.
- The proposed method considers both the number and structure of node's neighbors.
- The proposed method outperforms other measures on both real and artificial networks.
- The proposed method is robust to different network sizes and community structure.

ARTICLE INFO

Article history:
Received 3 October 2013
Received in revised form 26 December 2013
Available online 25 February 2014

Keywords:
Complex networks
Centrality measures
Local structure
Spreading

ABSTRACT

Ranking nodes by their spreading ability in complex networks is a fundamental problem which relates to wide applications. Local metric like degree centrality is simple but less effective. Global metrics such as *betweenness* and *closeness centrality* perform well in ranking nodes, but are of high computational complexity. Recently, to rank nodes effectively and efficiently, a semi-local centrality measure has been proposed as a tradeoff between local and global metrics. However, in semi-local centrality, only the number of the nearest and the next nearest neighbors of a node is taken into account, while the topological connections among the neighbors are neglected. In this paper, we propose a local structural centrality measure which considers both the number and the topological connections of the neighbors of a node. To evaluate the performance of our method, we use the *Susceptible–Infected–Recovered* (SIR) model to simulate the epidemic spreading process on both artificial and real networks. By measuring the rank correlation between the ranked list generated by simulation results and the ones generated by centrality measures, we show that our method can rank the spreading ability of nodes more accurately than centrality measures such as degree, *k*-shell, betweenness, closeness and local centrality. Further, we show that our method can better distinguish the spreading ability of nodes.

1. Introduction

The study of the spreading process on complex networks has drawn much attention recently because of its great theoretical significance and remarkable practical value in many areas including epidemic controlling [1–5], information

---

* Corresponding author.
E-mail addresses: grey_shine@hotmail.com (S. Gao), chenzhumin@sdu.edu.cn (Z. Chen).

http://dx.doi.org/10.1016/j.physa.2014.02.032
0378-4371/

dissemination [6,7] and viral marketing [8,9] etc. One of the fundamental problems in understanding and controlling spreading process is evaluating the *spreading ability* for each node in the network, i.e. how many nodes will finally be covered when the spreading origins from this single node [10–16]. The knowledge of node's spreading ability shows new insights for applications such as finding social leaders [17], ranking reputation of scientists, publications [18] and designing efficient methods to either hinder epidemic spreading or accelerate information dissemination.

Over the recent years, various centrality measures such as degree, betweenness [19], closeness [20] and eigenvector [21] centralities have been proposed to rank nodes in the network. *Degree centrality* is a simple and efficient local metric, but it is less relevant since it neglects the global structure of the network. Some well-known global metrics such as *betweenness centrality* and *closeness centrality* can give better results. However due to their high computational complexity, they are incapable to be applied in large-scale networks. Recently, Kitsak et al. found that the most efficient spreaders are those located within the core of the network as identified by the *k*-shell decomposition analysis [10]. After this, some modified network decomposition algorithms have been introduced to further improve the ranking performance [16,22]. In directed networks, several iterative process based ranking methods such as PageRank [23], HITS [24] and LeaderRank [17] have been proposed to rank nodes.

Since the scale of online social systems keep growing, they can have millions or even billions of user, e.g. the total number of monthly active Facebook users is 1.1 billion till June 2013.[1] Thus the ranking algorithms which are based on global information of the network will be very time-consuming and incapable to be applied. Hence, in order to rank nodes effectively and efficiently, it is better to design the ranking algorithms based on the local information of the network. For example, a *semi-local centrality* measure which considers both the nearest and the next nearest neighbors of a node has been proposed in Ref. [11]. This centrality measure has been shown to well rank the spreading ability of nodes and achieves a good tradeoff between low-relevant degree centrality and other time-consuming measures.

However, when the local centrality is used to rank nodes, only the number of the nearest and the next nearest neighbors of a node is considered, while the topological connections among the neighbors are completely ignored. Actually, the topological connections among the neighbors are also very important. For nodes with the same local centrality, the one with denser connected neighbors is supposed to have stronger spreading ability since denser connected neighbors get more chance to influence each other. Inspired by this idea, we propose a *local structural centrality* measure which considers both the number and the topological connections of node's neighbors, where the local clustering coefficient of a node is used to measure the topological connections among its neighbors. We use the *susceptible–infected–recovered* (SIR) model [25] to simulate the epidemic spreading process on both artificial and real networks. By measuring the rank correlation between the ranked list generated by simulation results and the ones generated by centrality measures, we show that our method can rank the spreading ability of nodes more accurately than centrality measures such as degree, *k*-shell, betweenness, closeness and local centrality. Through experiments on artificial networks generated by Barabási–Albert (BA) [26,5] network model and Lancichinetti–Fortunato–Radicchi (LFR) network model [27], we show that our method can outperform other centrality measures in scale-free networks with different sizes and different community structure. Further, we show that our proposed method can better rank the most influential nodes than other measures considered. Moreover, we use the *susceptible–infected* (SI) model [25] to simulate the epidemic spreading process and show that our proposed method can better rank the spreading ability of nodes under the SI model. Finally, we examine the ability of different methods to distinguish the spreading ability of the nodes and show that our proposed method performs better.

Following parts are organized as follows. We briefly review the definition of centrality measures used for comparison in Section 2 and introduce our local structural centrality measure in Section 3. In Section 4, we present the data, the spreading model and the evaluation measure that are used to evaluate the performance of our method. The experimental results are presented in Section 5. We conclude our paper and give a discussion in Section 6.

2. Centrality measures

Consider an unweighted and undirected simple network $G = (V, E)$ with $n = |V|$ nodes and $m = |E|$ links. G could be described by an adjacent matrix $A = \{a_{uv}\} \in R^{n,n}$, where $a_{uv} = 1$ if node u is connected with node v and $a_{uv} = 0$ otherwise. We use $\Gamma_h(v)$ to denote the set of neighbors within h-hops from node v.

The *degree centrality* (DC), $C_D(v)$, of node v can be calculated as

$$C_D(v) = \sum_{u=1}^{n} a_{uv} = |\Gamma_1(v)|. \tag{1}$$

The computational complexity for degree centrality is $O(n)$.

The *k-shell centrality* (KS) [10], C_{KS}, is obtained by employing *k*-shell decomposition algorithm on the network. The algorithm can be performed iteratively. First, we remove all nodes with degree equals to 1 and assign their $C_{KS} = 1$. After removing all these nodes, some nodes may be left with one link, so we keep pruning the network until there is no node with degree equals to 1 left in the network. The removed nodes, together with their corresponding links, form a *k* shell with

---

[1] http://www.statisticbrain.com/facebook-statistics/.