Physica A 510 (2018) 77–82

Contents lists available at ScienceDirect

Physica A

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

Identifying the influential nodes via eigen-centrality from the differences and similarities of structure

Lin-Feng Zhong [a,b,*], Ming-Sheng Shang [d], Xiao-Long Chen [a,b,c], Shi-Ming Cai [a,b]

[a] Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, PR China
[b] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 610054, PR China
[c] Center for Polymer Studies and Department of Physics, Boston, MA 02215, USA
[d] Chongqing Key Laboratory of Big Data and Intelligent Computing, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, PR China

HIGHLIGHTS

- Differences and similarities of the structure can be used to identify the influential nodes.
- We present the ECDS centrality which determined by the Jaccard similarity of the random two nodes.
- We use a parameter to adjust the relationship between the differences and similarities in the ECDS centrality.
- The proposed centrality could identify the influential nodes more accurately than the K-shell, degree, and closeness centralities.

ARTICLE INFO

Article history:
Received 2 April 2018
Received in revised form 11 June 2018
Available online xxxx

Keywords:
Complex network
Influential node
Eigen-centrality
SIR
Kendall

ABSTRACT

One of the most important problems in complex network is the identification of the influential nodes. For this purpose, the use of differences and similarities of structure to enrich the centrality method in complex networks is proposed. The centrality method called ECDS centrality used is the eigen-centrality which is based on the Jaccard similarities between the two random nodes. This can be described by an eigenvalues problem. Here, we use a tunable parameter α to adjust the influence of the differences and similarities. Comparing with the results of the Susceptible Infected Recovered (SIR) model for four real networks, the ECDS centrality could identify influential nodes more accurately than the tradition centralities such as the k-shell, degree and closeness centralities. Especially, in the Erdös network, the Kendall's tau could be reached to 0.93 when the spreading rate is 0.12. In the US airline network, the Kendall's tau could be reached to 0.95 when the spreading rate is 0.06.

1. Introduction

Spreading is a ubiquitous phenomenon in nature. A lot of activities can be seen as spreading in society [1–5]. In the past few years, the spreading in complex networks is concerned more and more with its great theoretical significance and remarkably practical value, that is, epidemic controlling [6–8], information dissemination [9–11] and viral marketing [12,13]. One of the fundamental problems is to identify the influential nodes in the networks. The knowledge of the node's spreading ability shows new insights for applications such as identifying influential nodes [14–19], designing efficient methods to either hinder epidemic spreading or accelerate information dissemination.

* Corresponding author.
E-mail address: googlezlf@163.com (L.-F. Zhong).

https://doi.org/10.1016/j.physa.2018.06.115
0378-4371/

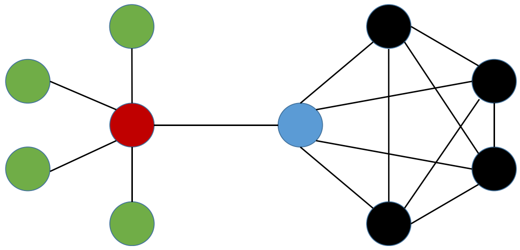

Fig. 1. In the illustrated network, there exists similarity between the blue node and black node, because they share neighbors between them. The red node and blue node are different because they do not share neighbors between them. The blue node reaches the green nodes only through the red node and therefor its contribution to the centrality of blue node is greater. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Recently, there are a lot of centrality methods [20] have been applied to identify the influential nodes in complex networks, including degree, eigenvector centrality [21], closeness centrality [22] and k-shell decomposition [23]. The degree centrality is based on the number of neighbors connected with a node. Chen et al. [24] defined a local centrality based on the degree information of nearest neighbors and the second nearest neighbors. Poulin et al. [25] proposed cumulated nomination centrality based on the iterative method for solving the problem of the feature vector mapping. Zhang et al. [26] proposed a multiscale measurement by considering the interactions from all the paths a node is involved. Kitsak et al. [23] found that the most influential nodes are those who located within the core of the network by decomposing a network with the k-shell decomposition method. By taking into account the neighbors' k-core values, Lin et al. [27] proposed an improved neighbors' k-core (INK) method to identify the influential nodes with the largest k-core values. Ren et al. [28] proposed an iterative method to identify the influential nodes by taking into account the neighbors' influence. And by considering the k-shell decomposition method and resource iteration, Ma et al. [29] proposed an improved method to identify the influential nodes. Then Zhong et al. [30] proposed an improved iterative method by considering the influence of rate.

The traditional structural centralities mainly consider the local or global structural properties of the node in the network, but the differences and similarities of structure between the two random nodes can also reveal the central nodes [31,32]. Inspired by this idea, we proposed a new eigen-centrality from the differences and similarities of structure (ECDS) to identify the influential nodes in a given network. Then, comparing with the Susceptible Infected Recovered (SIR) spreading process [33,34] for four real networks, the results show that the ranking list generated by the ECDS centrality could identify influential nodes more accurately than the k-shell, closeness, and degree centralities.

2. Method

Normally, an undirected network $G = (N, E)$ with N nodes and E edges could be described by an adjacent matrix $\Omega = \{\delta_{ij}\} \in R^{n,n}$ where $\delta_{ij} = 1$ if node i is connected by node j, and $\delta_{ij} = 0$ otherwise. To illustrate the essence of ECDS centrality, consider the network of Fig. 1. For the blue node, the information, relevance or centrality that brings a particular black node is poor, because the blue node can access to the rest of the black nodes directly, without any intermediary (this is because they have almost the same neighbors). However, the red node is essential for the blue node, because without it, the blue node could not access to the green nodes. One way to quantify this is through a structural similarity measure, e.g., Jaccard similarity [35], given by

$$S_{ij} = \frac{|V^+(i) \bigcap V^+(j)|}{|V^+(i) \bigcup V^+(j)|}, \tag{1}$$

and the structural difference measure can be expressed as follows:

$$D_{ij} = 1 - S_{ij}, \tag{2}$$

that allows us to measure the differences and similarities between the neighborhoods of two random nodes i, j given. And we use a tunable parameter to adjust the relationship between the differences and similarities, namely $\alpha \in [0, 1]$. Then we can weigh the contribution made by each node j in the neighbors of a node i by

$$W_{ij} = \alpha D_{ij} + (1 - \alpha) S_{ij}. \tag{3}$$

This allows us to propose the following heuristic: "the centrality of a node is proportional to the sum of the centralities of the nodes in its neighborhood, weighed by their contributions". Mathematically, that is:

$$c_i = \frac{1}{\lambda} \sum_{j=1}^{n} W_{ij} c_j, i = 1, 2, \ldots, n \tag{4}$$