Chaos, Solitons and Fractals 104 (2017) 257–267

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

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

A new method to identify influential nodes based on relative entropy

Liguo Fei [a], Yong Deng [a,b,c,*]

[a] School of Computer and Information Science, Southwest University, Chongqing, 400715, China
[b] Institute of Integrated Automation, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xian, Shaanxi, 710049, China
[c] School of Engineering, Vanderbilt University, Nashville, TN, 37235, USA

ARTICLE INFO

Article history:
Received 30 June 2017
Revised 12 August 2017
Accepted 12 August 2017

Keywords:
Complex networks
Influential nodes
Centrality measure
Relative entropy
TOPSIS

ABSTRACT

How to identify influential nodes is still an open and vital issue in complex networks. To address this problem, a lot of centrality measures have been developed, however, only single measure is focused on by the existing studies, which has its own shortcomings. In this paper, a novel method is proposed to identify influential nodes using relative entropy and TOPSIS method, which combines the advantages of existing centrality measures. Because information flow spreads in different ways in different networks. In the specific network, the appropriate centrality measures should be considered to sort the nodes. In addition, the remoteness between the alternative and the positive/negetive ideal solution is redefined based on relative entropy, which is proven to be more effective in TOPSIS method. To demonstrate the effectiveness of the proposed method, four real networks are selected to conduct several experiments for identifying influential nodes, and the advantages of the method can be illustrated based on the experimental results.

1. Introduction

In recent years, the applications of complex networks have become more and more extensive in lots of fields [1], including systems science [2–4], management science [5,6], social science [7–9] and computer science [10,11]. In addition, the community structure [12–14] and robustness [15,16] of complex networks are researched extensively to analyze their structures and organization. Simultaneously, a number of mechanisms such as self-similarity [17,18], spreading [19] and uncertainty can be controlled by a few vital nodes in the network. The significance to identify influential nodes in complex networks is enormous, both theoretical and practical, such as in the prevention and control of infectious diseases [20,21] and the optimization of traffic system.

The influential nodes are critical to the structure and function of complex networks [22–24]. Although the number of influential nodes will not be a lot, but their impact can quickly spread throughout the network [25–27]. Therefore, a lot of methods have been presented to evaluate the vital nodes in complex networks, such as Degree Centrality (DC), Betweenness Centrality (BC), Closeness Centrality (CC) [28], Eigenvector Centrality (EC) [29], PageRank(PR) [30], LeaderRank(LR) [31] and many other approaches [32–35].

These centrality measures are applied to many practical work and developed by researchers. A bio-inspired centrality measure model is presented to identify the vital nodes in [36]. To deal with the condition of asymmetric network, Bonacich and Lloyd [29] proposed the eigenvector centrality method. The betweenness centrality measure is improved by Gómez et al. [37] in terms of different dimensions. However, there still exist some problems for the single centrality measure to identify influential nodes in complex networks. For data fusion, Dempster-Shafer evidence theory is an efficient method [38,39], which has been applied widely in many fields [40–42] such as failure mode and effects analysis [43,44], decision making [45–47]and uncertain information modelling [48,49]. And an extended data fusion approach called D numbers theory [50–53] has also received extensive attention.

These centrality methods are used widely to identify influential nodes in complex networks. However, each method also has its own limitations and shortcomings. For example, DC has the advantage of low complexity and high practicability, but the shortcoming is that the local information can be focused on but global information will be ignored. CC will fail when handling the networks with disconnected components, because the unreasonable results will be obtained if two nodes have different components. BC compensates for the limitations of CC and DC, but there still exist some problems, for example, if there are a lot of nodes do not belong to the shortest path of other node pairs, thus, the result of BC will be

* Corresponding author at: School of Computer and Information Science, Southwest University, Chongqing, 400715, China
E-mail address: prof.deng@hotmail.com (Y. Deng).

http://dx.doi.org/10.1016/j.chaos.2017.08.010
0960-0779/

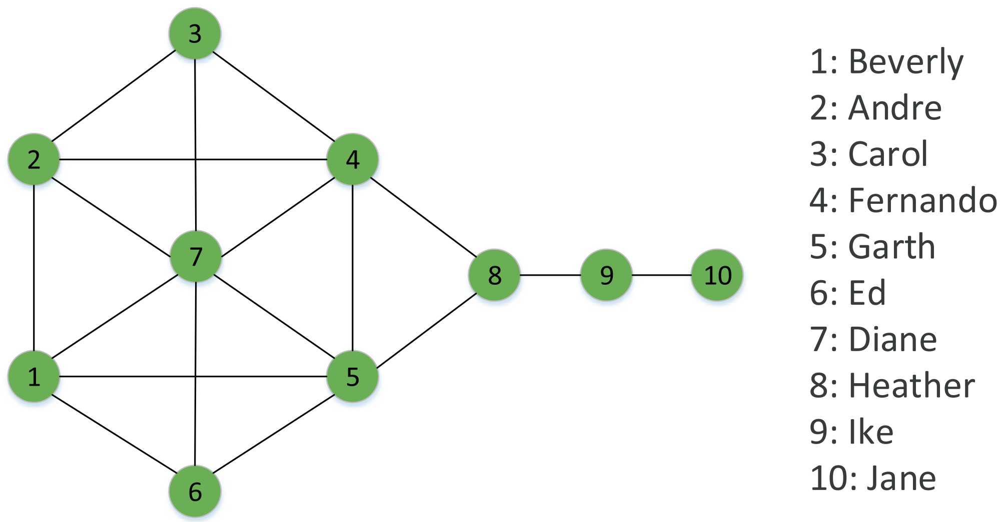

Fig. 1. The classical interpersonal relationship network.

Table 1
The degree, closeness and betweenness values of each node in the classical interpersonal relationship network.

Degree		Closeness		Betweenness	
Node	Value	Node	Value	Node	Value
Diane	6	Fernando	0.0714	Heather	0.5778
Fernando	5	Garth	0.0714	Fernando	0.3556
Garth	5	Diane	0.0667	Ike	0.3333
Andre	4	Heather	0.0667	Garth	0.2889
Beverly	4	Andre	0.0588	Diane	0.0444
Carol	3	Beverly	0.0588	Andre	0.0370
Ed	3	Carol	0.0556	Beverly	0.0370
Heather	3	Ed	0.0556	Jane	0.0000
Ike	2	Ike	0.0486	Ed	0.0000
Jane	1	Jane	0.0345	Carol	0.0000

0 [54]. There is no doubt that this will happen, and inconsistent results can be obtained, if different centrality measures are applied to identify influential nodes in a determined network. To illustrate this problem, an example network of interpersonal relationship is given in Fig. 1 [55].

The values of DC, CC and BC of each node are displayed in Table 1. It can be found that the most influential node using DC is Diane, yet Heather owns the maximal betweenness centrality value, however, its degree centrality value is only 0.333. In addition, the nodes Fernando and Garth are the most important using CC. Thus, ambiguous results can be obtained with different centrality measures in this network.

To solve this issue, in this paper, a novel method is proposed to identify the influential nodes based on relative entropy and TOPSIS method. The centrality values from different measures are considered as multiple attributes to make decision for sorting nodes in complex networks. And the proposed method can obtain the better ranking results, which combines the advantages of existing centrality measures. In addition, the remoteness between alternative and the positive/negetive ideal solution is redefined based on relative entropy, which is proven to be more effective in TOPSIS method. To demonstrate the effectiveness of the proposed method, four experiments are conducted based on four real networks, and the experimental results reveal the superiority of the proposed method.

The organization of the rest of this paper is as follows. Section 2 gives a brief introduction of the definition of graph and some centrality measures. The proposed method to identify vital nodes using relative entropy and TOPSIS method is illustrated in Section 3. Section 4 investigates the effectiveness of the proposed method based on four real networks. Section 5 is the conclusion of this paper.

2. Preliminaries

2.1. The existing centrality measures

A network can be denoted as $G = (V, E)$, where V and E are the set of nodes and edges, respectively. The centrality measures of DC, BC, and CC are defined as follows.

Definition 1. The degree centrality measure of node i, denoted as k_i, is defined as: [56]

$$k_i = \sum_{j}^{N} x_{ij} \tag{1}$$

where i is the central node while j denotes the other nodes which connect to i. And x_{ij} represents the edge between node i and j. The value of x_{ij} is defined as 1 if node i is connected to node j, and 0 otherwise.

Definition 2. The betweenness centrality measure of node i, denoted as b_i, is defined as: [57]

$$b_i = \sum_{j,k \neq i} \frac{g_{jk}(i)}{g_{jk}} \tag{2}$$

where g_{jk} is the number of binary shortest paths between node j and k, and $g_{jk}(i)$ is the number of those paths that go through node i.

Definition 3. The closeness centrality measure of node i, denoted as c_i, is defined as: [57]

$$c_i = \frac{1}{\sum_j^N d_{ij}} \tag{3}$$

where d_{ij} denotes the distance between node i and j.

The above measures have been extended to weighted networks as follows.