Chaos, Solitons and Fractals 103 (2017) 101–110

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 evidential methodology of identifying influential nodes in complex networks

Tian Bian[a], Yong Deng[a,b,*]

[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

ARTICLE INFO

Article history:
Received 1 November 2016
Revised 19 April 2017
Accepted 29 May 2017

Keywords:
Complex networks
Influential nodes
Evidential centrality
Dempster–Shafer evidence theory

ABSTRACT

In the field of complex networks, how to identify influential nodes in complex networks is still an open research topic. In the existing evidential centrality (EVC), the global structure information in complex networks is not taken into consideration. In addition, EVC also has the limitation that only can be applied on weighted networks. In this paper, a New Evidential Centrality (NEC) is proposed by modifying the Basic Probability Assignment (BPA) strength generated by EVC. According to the shortest paths between the nodes in the network rather than just considering local information, some other BPAs are constructed. With a modified combination rule of Dempster–Shafer evidence theory, the new centrality measure is determined. Numerical examples are used to illustrate the efficiency of the proposed method.

1. Introduction

In recent years, the evaluation of node importance in complex networks has attracted much attention because of its great theoretical significance and wide application [1–7], such as in the control of the disease spreading [8–13], creating new marketing tools [14–19] and research on public opinion and rumor dynamics [20–22]. Many centrality measures, the essential tool to identify the centrality of nodes in network analysis, have been used commonly, such as the degree centrality (DC) [23], betweenness centrality(BC) [23–25], closeness centrality (CC) [23,26], eigenvector centrality (EC) [27], PageRank (PR) [28,29], LeaderRank (LR) [30,31] and many other methods [32–39]. The DC method is very simple but of little relevance, since the measure does not take into consideration the global structure of the network. BC and CC are global metrics which can better identify influential nodes, but they are difficult to apply in large-scale networks due to their computational complexity. Another limitation of CC is the lack of applicability to networks with disconnected components: two nodes that belong to different components but do not have a finite distance between them.

It's inevitable to handle the uncertainty in the complex systems [40–42]. As a result, many math tools are developed to address this issue such as fuzzy sets [43,44], evidence theory [45–47] and D numbers [48–50]. The Dempster–Shafer evidence theory (D–S evidence theory), was first proposed by Dempster [51] and then developed by Shafer [52]. This theory is often regarded as an extension of the Bayesian theory, because the Bayesian theory needs stranger conditions. In D–S evidence theory, the probability assigned to each subset is ensured by total belief and the total plausibility for the objects in the subset. That means Bayes method requires the prior information while the D–S evidence theory can deal with the uncertain information under the situation of not knowing the prior probability [53,54]. Due to its ability to deal with the uncertain or imprecise information, the Dempster–Shafer theory has been widely applied in recent years [55,56]. The existing evidential centrality (EVC) [57,58] based on the D–S evidence theory is obtained by the combination of degree and weight strength of each node. However, there is a shortcoming that the EVC centrality measure has ignored the global structure information of the network, which is analogous with the extension of DC - simple but of little relevance. Gao et al. [59] proposed a new evidential semi-local centrality (ESC) by a combination of modified evidential centrality and the extension of semi-local centrality, but it only can be applied in the weighted networks. In this paper, a new evidential centrality measure is proposed by a combination of node's degree and the global structure of network measured by the shortest path. To evaluate the performance of the proposed method, we adopt the Susceptible-Infected (SI) model [60] to examine the spreading influence of the nodes ranked by different centrality measures. The simulations on several real networks are used to show the efficiency of the proposed method.

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

http://dx.doi.org/10.1016/j.chaos.2017.05.040
0960-0779/

The rest of this paper is organized as follows. In Section 2, the existing centrality measures and the evidence theory is introduced. In Section 3, the proposed method for identifying the influential nodes is depicted by an example network. Then, the SI model is adopted to evaluate the performance of proposed method in two example networks and several real complex networks in Section 4. Finally, some conclusions are presented in Section 5.

2. Preliminaries

2.1. Centrality measures

Considering a graph G = (V, E) with n = |V| nodes and m = |E| links. And the node centrality measurement of DC, CC and BC are well defined as follows.

Definition 1. (DC) [61]. The DC of node i, denoted as $C_D(i)$, is defined as

$$C_D(i) = \sum_{j}^{N} x_{ij} \tag{1}$$

where i is the focal node, j represents all other nodes, N is the total number of nodes, and x_{ij} represents the connection between node i and node j. The value of x_{ij} is defined as 1 if node i is connected to node j, and 0 otherwise.

Definition 2. (BC) [61]. The BC of node i, denoted as $C_B(i)$, is defined as

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

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

Definition 3. (CC) [61]. The CC of node i, denoted as $C_C(i)$, is defined as

$$C_C(i) = \left[\sum_{j}^{N} d_{ij}\right]^{-1} \tag{3}$$

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

2.2. Dempster–Shafer evidence theory

In Dempster–Shafer theory, a problem domain is denoted by a finite nonempty set Ω of mutually exclusive and exhaustive hypotheses, called the frame of discernment. Let 2^{Ω} denote the power set of Ω. For completeness of the explanation, a few basic concepts are introduced below [51,52].

Definition 4. (Frame of Discernment). Let $\Omega = H_1, H_2, \ldots, H_N$ be a finite set of N elements, and denote P(Ω) $=\phi, H_1, \ldots, H_N, H_1 \cup H_2, H_1 \cup H_3, \ldots, \Omega$, the power set is called the frame of discernment.

Definition 5. (Basic Probability Assignment (BPA)). For a frame of discernment Ω, a basic probability assignment function is a mapping m: $2^{\Omega} \rightarrow$[0, 1], which is also called the Mass Function, satisfying

$$m(\phi) = 0 \quad and \sum_{A\in 2^{\Omega}} m(A) = 1 \tag{4}$$

where ϕ is an empty set and A is any element of 2^{Ω} and the mass m(A) represents how strongly the evidence supports A. Given two BPAs m_1 and m_2, the Dempster rule can be used to combine them.

Definition 6. (Dempster's Rule of Combination). Dempster's rule of combination, also called orthogonal sum, denoted by m = $m_1 \bigoplus m_2$, is defined as follows

$$m(A) = \frac{1}{1-K} \sum_{B\cap C=A} m_1(B)m_2(C) \tag{5}$$

with

$$K = \sum_{B\cap C=\phi} m_1(B)m_2(C) \tag{6}$$

where A, B, and C are elements of 2^{Ω}, and K is a normalization constant, called the conflict coefficient of two BPAs.

It should be noted that how to determine the BPA is an open issue and the generation of BPA is dependent on the real application [62–64].

2.3. SI model

In this paper the Susceptible-Infected (SI) model [60] is used to demonstrate the efficiency of the proposed method. In SI model, $S(t)$ and $I(t)$ are respectively the numbers of susceptible and infected individuals at time t. In each step, only the individuals which have been infected are able to spread the disease to susceptible individuals, and every susceptible individual gets infected with probability λ from the infected neighbor. $\lambda = (\frac{w_{ij}}{w_M+1})^{\alpha}, \alpha > 0$ [65], where w_{ij} is the weight of edge E_{ij} and α is a positive constant. The total number of infected nodes at time t can be considered as an indicator to evaluate the influence of the initially infected node, namely $F(t)$ [66]. Clearly, the number of cumulative infected nodes increases with time t and eventually reaches a steady value when there is no reachable susceptible node to be infected. For different nodes initially infected, higher $F(t)$ represents a larger influence.

3. New evidential centrality

An evaluation method of importance of the node is established by D–S evidence theory proposed by Wei et al. [57]. The idea of evidential centrality is similar to multi-attribute decision making (MADM). That is, different factors will be combined to obtain final ranking order of each node in the complex networks. It is obtained by the combination of degree and weight strength of each node. However, the evidential centrality is of little relevance since it only considers the degree and weight strength of a node, ignoring the global structure information. In addition, the method has a limitation that can only be applied to the weighted network. In this paper, a modified evidential centrality measure, called the new evidential centrality (NEC) is proposed to identify the influence of the node. Several BPAs of a node are obtained based on the node's degree and the shortest path between the node and other nodes in the network, respectively. For instance, a node with the maximum value of the degree is most important when only the degree of the node is considered. And the importance of another node is represented by the difference of degree centrality value between two nodes. The modified method to get the BPAs of evaluation index is developed below.

3.1. Proposed method

Step 1: Ascertain a frame of discernment θ. Let *high* and *low* be evaluation indices for the influence of each node's attributes. Thus, a frame of discernment θ is given as

θ=(high, low).

Step 2: Create BPAs for each node. $m_{ki}(h)$ and $m_{ki}(l) (i = 1, 2, \ldots, N)$ denote the probabilities of "high" and "low" influence for the degree of node i, respectively. $m_{di}^{j}(h)$ and $m_{di}^{j}(l) (i =$