Information Processing and Management 000 (2016) 1–12

Contents lists available at ScienceDirect

Information Processing and Management

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

Efficient identification of node importance in social networks

Yunyun Yang, Gang Xie*

College of Information Engineering, Taiyuan University of Technology, Taiyuan, Shanxi 030024, People's Republic of China

ARTICLE INFO

Article history:
Received 6 December 2015
Revised 22 March 2016
Accepted 3 April 2016
Available online xxx

Keywords:
Social networks
Node importance
Comprehensive decision analysis method

ABSTRACT

In social networks, identifying influential nodes is essential to control the social networks. Identifying influential nodes has been among one of the most intensively studies of analyzing the structure of networks. There are a multitude of evaluation indicators of node importance in social networks, such as degree, betweenness and cumulative nomination and so on. But most of the indicators only reveal one characteristic of the node. In fact, in social networks, node importance is not affected by a single factor, but is affected by a number of factors. Therefore, the paper puts forward a relatively comprehensive and effective method of evaluation node importance in social networks by using the multi-objective decision method. Firstly, we select several different representative indicators given a certain weight. We regard each node as a solution and different indicators of each node as the solution properties. Then through calculating the closeness degree of each node to the ideal solution, we obtain evaluation indicator of node importance in social networks. Finally, we verify the effectiveness of the proposed method experimentally on a few actual social networks.

1. Introduction

In the era of big data, social networks can be found anywhere. Social networks can be seen as a set of the nodes joined in pairs by edges. In recent years, the analytical approach of the social networks has attracted considerable curiosity from the social and the scientific community. This curiosity can be attributed to a fascinating focus, namely the relationship between social entities, as well as the pattern and meaning of the relationship (Chen & Wang, 2012; Kossinets & Watts, 2006; Van Noorden, 2014; Wasserman, 1994; Watts & Strogatz, 1998). In particular, the degree of importance of each node in social networks is not the same, that is, different nodes have different weights. Therefore, in the actual network, it is valuable to find paramount nodes for controlling the entire network. Consequently, we need to capture and reveal the relationship between the feature of the nodes and network topology so that we understand the impact of a node on system functions, such as reliability, robustness and controllability (Ahlswede, Cai, & Li, 2000; de Camargo, Neto, & Pires, 2012; Hawick & James, 2007).

In social networks, there are a variety of measurement indicators for identifying the most crucial nodes (roles). In general, the measurement indicators (Arulselvan, Commander, & Elefteriadou, 2009; Bhola, Grover, & Sinha, 2010; Borgatti, 2006; Hewett, 2011; Hu, Wang, & Lee, 2010; Xie, Li, & Zhang, 2015) are divided into two categories: from the perspective of local attributes and the perspective of global attributes. For example, degree centrality (Bonacich, 1972) represents that the more the number of neighboring nodes are, the more important the nodes are, which is the simplest index of centrality. Information index (Bonacich, 1987) depends on the amount of information through its propagation path. Besides,

* Corresponding author. Tel.: +86 03516018860, fax: +8618734176603.
E-mail address: xiegang189@163.com (G. Xie).

http://dx.doi.org/10.1016/j.ipm.2016.04.001
0306-4573/© 2016 Published by Elsevier Ltd.

Kitsak, Gallos, and Havlin (2010) proposed k-shell decomposition. Closeness centrality (Freeman, 1979) discussed by Rothenberg et al. is a global measure of centrality based on the distance between nodes in a connected network. Closeness centrality can be seen as a measure of how long it will take information to spread from a given node to others in the network. In addition, there are a host of other concepts such as subgraph centrality (Estrada & Rodriguez-Velazquez, 2005), eigenvector centrality Stephenson and Zelen (1989) and cumulative nomination (Poulin, Boily, & Masse, 2000). In those regards, Ren and Lv (2014), Wang and Zhang (2006), and He, Li, Gan, and Zhu (2008) have done excellent summaries. Further, Sun and Luo (2012), Liu, Ren, Guo, and Wang (2013) also summarized the history of the methods of mining important nodes in complex networks and summarized the research results. They are trying to describe and measure the characteristics of a node from a different point of view, and adapt for different networks (Liu, Tang, & Zhou, 2015; London, Németh, & Pluhár, 2015; Ren, Liu, & Shao, 2013; Zhang, Xu, & Li, 2011). That is, they only reflect one perspective of node importance. But in modern society networks, node importance is not only determined by one attribute, but is also determined by a number of attributes. As an example, individuals' I.Q is determined by education, age, experience, etc. In addition, the proportion of each attribute is not equal. We should give different weights for each attribute. The literature (Yu, Liu, & Li, 2013) combined some indicators to discuss node importance and mainly considered structural characteristics of the network topology. In social networks, node importance should combine with some information concerning the node itself in the actual networks.

Based on this, in this paper from the network topology and the center of one node in social networks, we consider five indicators, including prestige, the amount of information dissemination, the transmitting path of the information, and hence propose a comprehensive evaluation of node importance for social networks.

The rest of the paper is built up as follows. Section 2 introduces the notations and notions related to our problem. Section 3 presents a scheme for efficient identification of node importance in social networks. In Section 4, this developed scheme is applied to identify the most important node(s) in some networks. In Section 5 the concluding remarks are given.

2. The definition of node importance indicators

According to the emphasises of different indicators, from more than one perspective (including the connection status between the nodes, the path of the target node to other nodes, and the perspective of simulation of the flow analysis), we choose five indicators: structural holes, flow betweenness centrality, cumulative nomination, information indicator, and subgraph centrality, to evaluate node importance comprehensively. The paper essentially studies a multi-attribute ranking problem. The definitions of these indicators are described in this section.

First of all, a short review of the required basic graph terminology is presented. For the sake of simplicity, we shall limit ourselves to undirected and unweighted networks. In this situation, if a node i is linked to another node j, then node j is necessarily linked to node i. Besides, all the links have the same weight.

Let G be a graph. A graph G consists of a non-empty finite set $V(G)$ of elements called nodes, and a finite set $E(G)$ of ordered pairs of different nodes called edges. We write $G = (V, E)$ where V and E are the node set and edge set of G, respectively. The size of G is the number of nodes in G, and is denoted by $|G|$. $V = \{v_1, v_2, \cdots, v_N\}$, $|V| = N$; $E = \{e_1, e_2, \cdots, e_M\} \subseteq V \times V$, $|E| = M$.

Next, present some measures typically used to identify node importance of social networks.

Definition 1 (Structural holes (Burt, 2009)). In networks, when there is no direct connection between two nodes, in the same vein, and there is no redundancy relationship between them, the obstacle between them is a structural hole. Burt (2009) gave the expression of structural holes with network constraints coefficient.

$$CH(i) = \sum_j (P_{ij} + \sum_{q \neq i \neq j} P_{iq}P_{qj})^2. \tag{1}$$

Where q is indirect node connecting node i and node j; P_{ij} means the proportion of node i to spend time (energy) on the node j and its total time (energy). The lower $CH(i)$ is, the greater the degree of structural holes is. That is to say the location of the node i is more important.

Definition 2 (Flow betweenness centrality (Freeman, Borgatti, & White, 1991)). In most networks, however, information (or anything else) does not flow only along geodesic paths. News or a rumor or a message does not know the ideal route to take from one place to another more randomly. To work around these issues, in 1991 Freeman et al. raised flow betweenness. Flow betweenness is based on the idea of maximum flow. Imagine each edge in a network as a pipe that can carry a unit flow of some fluid. We can ask what the maximum possible flow then is between a given source node s and target node t through these pipes. In general, the answer is that more than a single unit of flow can be carried between the source and target by making simultaneous use of several different paths through the network. The flow betweenness of a node i is defined as the amount of flow through node i when the maximum flow is transmitted from node s to node t. It is a measure of the influence a node has over the spread of information through the network.

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

where $g_{jk}(i)$ is the number of geodesic paths from node j to node k that pass through node i, and g_{jk} is the total number of geodesic paths from node j to node k.

Please cite this article as: Y. Yang, G. Xie, Efficient identification of node importance in social networks, Information Processing and Management (2016), http://dx.doi.org/10.1016/j.ipm.2016.04.001