Computers in Human Behavior 64 (2016) 621–640

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Full length article

Stochastic graph as a model for social networks

Alireza Rezvanian*, Mohammad Reza Meybodi

Soft Computing Laboratory, Computer Engineering and Information Technology Department, Amirkabir University of Technology (Tehran Polytechnic), Hafez Ave., 424, Tehran, Iran

ARTICLE INFO

Article history:
Received 5 April 2016
Received in revised form 11 July 2016
Accepted 22 July 2016

Keywords:
Complex social networks
Social network analysis
User behavior
Stochastic graphs
Network measures

ABSTRACT

Social networks are usually modeled and represented as deterministic graphs with a set of nodes as users and edges as connection between users of networks. Due to the uncertain and dynamic nature of user behavior and human activities in social networks, their structural and behavioral parameters are time varying parameters and for this reason using deterministic graphs for modeling and analysis of behavior of users may not be appropriate. In this paper, we propose that stochastic graphs, in which weights associated with edges are random variables, may be a better candidate as a graph model for social network analysis. Thus, we first propose generalization of some network measures for stochastic graphs and then propose six learning automata based algorithms for calculating these measures under the situation that the probability distribution functions of the edge weights of the graph are unknown. Simulations on different synthetic stochastic graphs for calculating the network measures using the proposed algorithms show that in order to obtain good estimates for the network measures, the required number of samples taken from edges of the graph is significantly lower than that of standard sampling method aims to analysis of human behavior in online social networks.

1. Introduction

In recent years, online social networks such as *Facebook* and *Twitter* have provided simple facilities for online users to generate and share a variety of information about users' daily life, activities, events, news and more information about their real worlds which results in the online users have become the main features of online social networks and studying how users behave and interact with their friends in online social networks play a significant role for analysis of online social networks. Online social networks similar to many real world networks are usually modeled and represented as deterministic graphs with a set of nodes as users and edges as connection between users of networks. In social network analysis, when networks are modeled as weighted or unweight deterministic graphs, one can study, characterize and analyze the statistical characteristics and the dynamical behaviors of the network and its user behaviors by some network measures (Costa, Rodrigues, Travieso, & Boas, 2007). Popular network measures such as degree, betweenness, closeness and clustering coefficient are originally defined for deterministic graphs. These network measures not only used in characterization of networks but also used as a part of some algorithms such as Girvan-Newman community detection algorithm using high betweenness edges (Girvan & Newman, 2002), overlapping community detection using nodes' closeness (Badie, Aleahmad, Asadpour, & Rahgozar, 2013) and finding the outstanding nodes as a set of seed nodes for maximization of the spread of influence in social network aims to introduce and promote a new products, services, innovations or technologies by ranking important humans (Li, Wu, Wang, & Luo, 2014).

Online social networks are intrinsically non-deterministic and their structures or behaviors such as online activities of users have unpredictable, uncertain and time varying nature and for this reason modeling those using deterministic graph models with fixed weights are too restrictive to solve most of the real network problems. For example in online social networks the online activities of users such as friendship behaviors, liking a comment on a given post in *Facebook* and frequency of taking a comment on a wall post vary over time with unknown probabilities (Jin, Chen, Wang, Hui, & Vasilakos, 2013). Analyzing online social networks with deterministic graph models cannot take into consideration the continuum of activities of the users occurring over time. Even modeling social networks with weighted graphs in which the edge weights are assumed to be fixed weights considers only a snapshot of a real-world network and it is not valid when the activities and

* Corresponding author.
E-mail address: a.rezvanian@aut.ac.ir (A. Rezvanian).

http://dx.doi.org/10.1016/j.chb.2016.07.032
0747-5632/

behavior of users on networks vary with time. Moreover, in analyzing online social networks not only understanding the structure and topology of the network is important but the degree of association among the users in network is also important for analysis of user behaviors in online social networks.

According to the aforementioned points, it seems that stochastic graphs in which weights associated with the edges are random variables is a better candidate as a graph model for real-world network applications with time varying nature. By choosing a stochastic graph as a graph model, every feature, measures and concept of the graph such as path (Beigy & Meybodi, 2006), cover (Rezvanian & Meybodi, 2015b), clique (Rezvanian & Meybodi, 2015a), and spanning tree (Akbari Torkestani & Meybodi, 2012) should be treated as stochastic features. For example, choosing stochastic graph as the graph model of an online social network and defining community structure in terms of clique, and the associations among the humans within the community as random variables, the concept of stochastic clique may be used to study community structure properties.

In this paper, after a brief overview of recent studies for user behavior and human activities in online social network, we first redefine some network measures for stochastic graphs and then design six algorithms for calculating them under the situation that the probability distribution functions of the weights associated with the edges are unknown. The proposed algorithms for calculating network measures by taking samples from the edges of the stochastic graph try to estimate the distribution of the network measures. The process of sampling from the edges of the graph is guided by the aid of learning automata in such a way that the number of samples needed to be taken from the edges of the stochastic graph for estimating the network measures to be reduced as much as possible. In the proposed algorithms, the guided sampling process implemented by learning automata aims to take more samples from the promising region of the graph, the regions that reflects higher rate of changes (e.g., higher rate of user activities), instead of walking around and taking unnecessary samples from non-promising region of the graph.

In order to study the performance of the proposed algorithms for calculating network measures in stochastic graphs, several experimental studies on different synthetic stochastic graphs are conducted. Experimental results show that in order to obtain good estimates for the network measures the number of samples needed to be taken from the edges of the graph by the proposed guided sampling algorithms is significantly lower than the required number of samples needed to be taken from the edges when standard sampling method is used aims to analysis of human activities and user behaviors in online social networks.

The rest of this paper is organized as follows. Section 2 dedicated to material and methods including brief introduction to some of existing network measures for deterministic networks, an overview of recent works about distribution of user behaviors in online social network studies and a brief description about learning automata theory. In section 3, the proposed network measures for stochastic graphs are described. The proposed algorithms for calculating the proposed network measures in stochastic graphs are described in section 4 and section 5 presents the simulation results. In section 6, discusses about this study and results. Finally, section 6 concludes the paper.

2. Material and methods

In this section, to provide the necessary background, we present a brief description of some network measures for deterministic unweighted and weighted networks. We also briefly review the studies performed by the researchers about the distributions of user behaviors and activities of users in online social networks. At the end of this section, learning automata and variable action set learning automata are introduced.

2.1. Network measures for deterministic networks

Network measures and calculating them play a significant role in social network analysis (Borgatti, 2005). Popular network measures such as degree, betweenness, closeness and clustering coefficient not only used for evaluating the node importance in actual complex network studies but also used as a part of some algorithms such as Girvan-Newman community detection algorithm using betweenness (Girvan & Newman, 2002), overlapping community detection using node closeness (Badie et al., 2013). In this section some of well-known network measures for deterministic networks are introduced.

2.1.1. Degree

Degree as a basic network measure has been widely used in many studies and degree of node v_i defined in binary network (also called unweighted network) as follows

$$k_i = \sum_{j \neq i} a_{ij} \tag{1}$$

where j is the index of all other nodes of graph and a_{ij} is 1 if node v_i is adjacent to node v_j, and 0 otherwise. In other words, degree of node v_i is the number of nodes that directly connected to node v_i. Degree centrality is useful in the context of finding the single human which gets affected by the diffusion of any information in the network. It follows from the fact that the human with high degree centrality has the chance of getting affected from many numbers of sources (Freeman, 1979).

2.1.2. Strength

Node strength of node v_i is defined as the sum of adjacent edge weights for weighted network as follows

$$s_i = \sum_{j \neq i} w_{ij} \tag{2}$$

where w_{ij} is greater than 0 if node v_i is adjacent to node v_j and its value indicates the weight of edge between node v_i and node v_j. Similar to degree, a human with high strength centrality is known as popular human with high strength links to other humans; however a human with high strength may not consist of necessarily the maximum number of friends. Strength centrality is useful in the context of finding a human which gets affected by the amount of spreading of any information in the network (Freeman, 1979).

2.1.3. Closeness

Closeness of a node is the inverse sum of shortest paths to all other nodes from that node and defined for binary network with n nodes as follows

$$c_i = \frac{1}{\sum_{j \neq j} d_{ij}} \tag{3}$$

where d_{ij} is the length of shortest path between node v_i and node v_j. Closeness can be regarded as a measure of how long it will take to spread information from that node to all other nodes sequentially. Since, the spread of information can be modeled by the use of shortest paths, in applications such as spread of information, a human with high closeness centrality can be considered as the central point because that human can spread the information faster