Computers in Human Behavior 73 (2017) 64–79

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Presenting novel application-based centrality measures for finding important users based on their activities and social behavior

Ehsan Khadangi, Alireza Bagheri*

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

ARTICLE INFO

Article history:
Received 8 September 2016
Received in revised form
28 November 2016
Accepted 4 March 2017

Keywords:
Social network analysis
Social behavior
Centrality measure
Social media marketing
Activity network

ABSTRACT

There are more important relationships based on users' behavior and the done activities than those of friendship in online social networks. Study of social behavior of users in these networks has many applications. Analyzing online social networks' activity graphs, as a better representation of users' social behavior, may open new perspectives for real applications such as finding important users. Although detecting these influential nodes based on their friendship relationships is studied a lot, finding important nodes using users' behavior and activates has not attracted much attention. In this work, we study users' importance in various Facebook activity networks including like, comment, post, share, and mixed, then compare gained rankings with those of the friendship network and conclude that users influence analysis in activity networks represents very different results. Afterwards, we propose new centrality measures that can present different rankings suitable for different applications, further to have the potential for simultaneous consideration of various activities in a multilayer network. Experimental results highlights the benefits of using the presented methods. To the best of our knowledge, our methods are the first and only proposed centrality measures that can present different rankings for various applications based on users' social behavior.

1. Introduction

Online social networks develop different mechanisms for users' activity and interaction. Facebook as the most famous and popular social networking website is not an exception to this fact. Facebook activities could be classified into the two categories latent and active. Among Facebook active activities the users' behavior on which is visible to other users are posting new contents, taking comment or like on a post or comment, sharing content, tagging photos, joining groups, and using Facebook applications. In addition, Facebook supports latent activities such as chatting, sending instant messages, browsing profiles, and online games. Since different users use some Facebook activities more according to their personality, psychological, and sociological properties, their behavior pattern is not the same as each other. Therefore, it is necessary to present new network analysis methods so as to consider several activities simultaneously besides the friendship network in order to analyze the users' behavior.

Although many works have been presented on user ranking in social networks, ranking Facebook users based on their different activities and interactions requires more examination. Particularly, the vacancy of a method which can measure the influence of nodes for different applications further to the simultaneous consideration of users' activities and interactions is perceived. The main goal of this study is to fill this vacancy. For this purpose, we initially support this idea that different people's importance is different from different viewpoints. Then based on the simultaneous analysis of Facebook users' different behaviors and activities in a multilayer model, an application-based method has been presented which measures users' importance based on the considered application. A literature review of detecting influential nodes and activity network analysis is presented as follows.

1.1. Empirical research

Among the society's people, there are always some who have great power in affecting and guiding different people's thoughts, interests, and beliefs due to personal, scientific, and psychological properties and their social position. Based on the 80/20 rule, also most opinions in sociology, these people are few (Cha, Haddadi,

* Corresponding author.
E-mail addresses: khadangi@gmail.com (E. Khadangi), ar_bagheri@aut.ac.ir (A. Bagheri).

http://dx.doi.org/10.1016/j.chb.2017.03.014
0747-5632/

Benevenuto, & Gummadi, 2010) and are named innovators (Rogers, 2010), salesmen, connectors, hubs, mavens (Point, 2002), or etc. Finding these people is used for influencing other people (Xu, Li, & Song, 2012), marketing and advertising (Li, Lee, & Lien, 2012; Xu et al., 2012), and etc. The theory of "diffusion of innovation" which was presented by Rogers states that only 2.5 percent of the society's people are much brave, risky, and interested in trying new products and thoughts. These people whom Rogers called "innovators" are potential initial adapters for the social network marketing process (Rogers, 2010). By convincing such people, a product, idea, or thought could be spread in social networks extraordinarily.

One of the most important methods of measuring impact in social networks is using centrality measures. Although these measures are applicable in different fields such as gaining the most important roads in a road network (Tsiotas & Polyzos, 2013), finding the important papers in a citation network (Cheang, Chu, Li, & Lim, 2014), and discovering author impact in coauthorship networks (Yan & Ding, 2011; Ding, Yan, Frazho, & Caverlee, 2009; Ding, 2011), one of their most important applications is in the measuring the impact and importance of social networks' users for different applications (Kang, Molinaro, Kraus, Shavitt, & Subrahmanian, 2012; Mochalova & Nanopoulos, 2013; Kiss & Bichler, 2008; Kermani, Badiee, Aliahmadi, Ghazanfari, & Kalantari, 2016).

Node centrality measures calculate the importance of users based on their position in the network. Some works suggest new node centrality measures which have better results (Takes & Kosters, 2011; Chen, Lü, Shang, Zhang, & Zhou, 2012; Alahakoon, Tripathi, Kourtellis, Simha, & Iamnitchi, 2011; Campiteli, Holanda, Soares, Soles, & Kinouchi, 2013); some have focused on presentation of algorithms for increasing the speed of calculating centrality measures especially betweenness and PageRank (Bader, Kintali, Madduri, & Mihail, 2007; McSherry, 2005); some others dealt with finding important nodes for different applications such as targeted social media marketing and advertising (Li et al., 2012; Xu et al., 2012); and some others are involved with the empirical study of the presented measures in different networks particularly online social networks (Cheang et al., 2014; Heidemann, Klier, & Probst, 2010; Hu, Wang, & Lee, 2010; Valente, Coronges, Lakon, & Costenbader, 2008). Furthermore, different generalizations of classic centrality measures have also been presented for weighted networks, including different versions of degree, closeness, betweenness, eigenvector, and PageRank (Opsahl, Agneessens, & Skvoretz, 2010; Yan, Zhai, & Fan, 2013; Xing & Ghorbani, 2004).

The concept of activity network as the network which models the actual interactions between users was first proposed by Chun et al. (Chun et al., 2008). After that, few efforts were made for identifying important users by applying centrality measures to the activity network. For example, Heidemann et al. identified important Facebook users by applying weighted PageRank to Facebook wall-post activity network (Heidemann et al., 2010). Corbellini et al. used centrality measures for suggesting the software engineering group leader. They initially presented an application named Paynal for the software developers' coordination. The software used social network analysis methods in order to analyze interactions among users and achieve high-level knowledge on the development team's members. For example, they used centrality measures in order to reach a node which can manage the team (Corbellini, Schiaffino, & Godoy, 2012).

Although a few works have been done in the field of measuring influence based on the users' activities, to the best of our knowledge, most of them have dealt with only one activity and measure the influence based on that particular activity. In addition, almost in all the works in the field of analyzing users' influence, influence is defined as a non-flexible and rigid idea whereas a person may be important by one application, and unimportant by another one. In this paper, a method has been presented which makes it possible to measure application-based importance of Facebook users by analyzing their activities and behavior simultaneously.

The rest of the paper is organized as follows. In Section 2, we present a literature survey of previous works. Section 3 describes material and methods. In Section 4, we study centrality measures in friendship and various activity networks. Then, we propose an application-based multilayer PageRank in Section 5. Section 6 presents the results of our experiments. Finally, we conclude the paper in Section 7.

2. Material and methods

The methodology of our research consists of five parts. First data collection using BFS network sampling technique, second study of important nodes in different activity networks, third calculating the importance of different activities, forth presenting application-based centrality measures, and ultimately Comparing presented methods with the most popular centrality measures.

Study of important nodes comprises of two parts: first calculating the importance of various activities, then comparing influential users in different activity networks. We also used three measures to compare presented methods with other centrality measures: capability to nodes' differentiation, standard deviation & dynamic range, and robustness over time. These parts are presented in Fig. 1.

In this section, we first present a high level analysis of the used dataset. Then the pre-processing of the collected data is also presented.

2.1. Description of the dataset

We used the collected dataset of Facebook users' activity and friendship networks in our experiments. The dataset include the friendship network of 36204 Facebook users. For friend users, the information about the number of activities including like, comment, post, and share as well as the number of exchanged words in their comments are collected every one month over a period of 3 years from 1 January 2011 to 1 January 2014 (Khadangi, Bagheri, & Shahmohammadi, 2016; Shahmohammadi, Khadangi, & Bagheri, 2016; Khadangi, Bagheri & Zarean, 2017). In our experiments, we usually used a smaller dataset which is a subset of the original one including the information of 8079 Facebook users.

The average degree of the friendship graph is 138. In addition, friendship network shows assortative mixing by degree. The average path length and 90-percentile effective diameter of friendship network are also 4.23 and 5 respectively. In addition, the degree distribution of the friendship follows power-law distribution.

The high level characteristics of different activity networks have also been presented in Table 1. It should be noted that mixed network is a mixture of like, comment, post, and share networks.

High clustering coefficient and low average path length confirm the small-world nature of various Facebook activity networks. In addition, the degree distribution of different activity networks follow power-law or semi-power-law.

2.2. Calculating the importance of different activities

Different activities have different importance for various applications. Accordingly, the presented methods use the importance of different activities for calculating nodes' centrality for different applications.

To calculate the importance of activities, we initially