Information Sciences 293 (2015) 383–389

Contents lists available at ScienceDirect

Information Sciences

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

Revealing the efficiency of information diffusion in online social networks of microblog

Yong Li [a,*], Mengjiong Qian [a], Depeng Jin [a], Pan Hui [b], Athanasios V. Vasilakos [c]

[a] Department of Electronic Engineering, Tsinghua University, China
[b] Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong
[c] Department of Computer and Telecommunications Engineering, University of Western Macedonia, Greece

ARTICLE INFO

Article history:
Received 15 February 2014
Received in revised form 5 April 2014
Accepted 11 September 2014
Available online 20 September 2014

Keywords:
Information diffusion
Power law
Node degree
Online social networks

ABSTRACT

Online Social Networks (OSNs) such as microblogs enable people to simultaneously share information with a large number of peers, and revealing the properties of the formed network structure and the corresponding impacts on information diffusion are important. In this article, we empirically study 10 millions user profiles from the largest Chinese mircoblog, Sina Weibo, and 41.7 millions profiles from Twitter. By studying followers' topology, we find an invariant characteristic that the followers count of users obeys a power law distribution with exponent near 2. We then investigate the reason from the theoretical view and find this network structure achieves the optimal information diffusion in terms of inducing a minimal propagation times. Our study thus reveals the efficiency of information diffusion in OSNs of microblogs, which is validated by experimental examination, theoretical analysis and numerical simulation.

1. Introduction

Online Social Networks (OSNs) have incrementally become as the most popular applications that enable people to share information simultaneously with large number of peers [1]. One of their distinguishing factors and applications is the ability to deliver the content and information to potential interested users [2]. Modelling and understanding their networking structures [3,4], dynamic evolution [5,6], and the formed human relationships [7,8] are hot topics in the area of OSNs. More recently, OSNs of microblogs, such as *Twitter* and *Sina Weibo*, have become as a popular platform for discovering real-time information on the web, such as news stories, current events, and people's opinions. Millions of individual users are sharing the information, which are discovered over microblogs, or to be an important source of spot news, such as revolutions and disasters [9,10]. In such networks, characterizing the information diffusion that depends on the network structures and the scale of nodes is one of the important problems [1,11].

Quantifying causal effects on the information dissemination in-depth requests not only the identification of who influences whom, but also whether individuals would still disseminate information in the absence of social signals about that information [1,12]. Therefore, behaviors of the information diffusion cannot be fully understood without considering the social network structures, over which information is transmitted. In fact, social behaviors shape while are shaped by the networks. Studying large-scale OSNs can be useful to investigate similarities and differences with real-life social networks.

* Corresponding author. Tel.: +86 10 62772387; fax: +86 10 62781038.
E-mail address: liyong07@tsinghua.edu.cn (Y. Li).

http://dx.doi.org/10.1016/j.ins.2014.09.019
0020-0255/

Moreover, this could help to empirically confirm sociological theories such as "Small World" [13], which has been reported in a wide range of situations, including allometric scaling in animals [14], the Gutenberg–Richter law in seismology [15], a purported unified theory of urban structure and living [16], patterns of terrorist and insurgent activities [17], and even the paper publication rates of scientists [18]. In this work, we will validate and investigate the network structures of two of the largest microblog, which brings a great influence on people's life and at the same time it may exhibit some special properties due to the influence of different cultures and activities.

Recently, there exists some related works addressing the network structure of social networks[21,10]. Back to 1999, Barabasi et al. [22] studied the scaling law of some large networks and showed the vertex connectivity follows a scale-free power-law distribution. Information diffusion and influence have been modeled in blogs [23], emails [24], and sites such as Twitter and Digg [25]. Recently, Bakshy et al. [1] examined the role of social networks in information diffusion by randomizes exposuring to announce about friends' information sharing among 253 million subjects. On one hand, some existing work reported that optimal designing for network mechanisms may generate power-law distribution and scale-free structures [19,20]. On the other hand, researchers tried to understand how the scale-free structures optimized the network performance [29,26,28,27]. For example, Jeong et al. [28] concluded that scale-free networks with a conserved network diameter appeared to provide an optimal structural organization, while Viswanathan et al. [27] reported the scale-free mobility behaviors would optimize the success of random searches in animals' forage. However, how does network structure of microblogs influence the information diffusion is not revealed, and it is still an open problem that whether the network structure existed benefits the information diffusion in microblogs.

This work aims to reveal the efficiency of information diffusion in the microblogs in depth. We study large scales of data from the largest Chinese mircoblog, Sina Weibo, and Twitter to measure the formed network structure and the behaviors of information diffusion. In characterizing the existing followers topologies, we treat users in the network as vertices in a directed graph, and if one user is the follower of another user, we connect an directed edge between these two vertices to create a directed graph. Based on the obtained graph, we examine the role of structures on social networks and the efficiency of information diffusion. Firstly, we analyze the distribution of followers count of users, and find it exhibits a power law with exponent near 2. Then, in order to reveal the in-depth reasons, we analyze this phenomenon from theoretical approach to find that the exponent 2 optimizes the network performance in terms of inducing a minimal information diffusion times. Thus, our study reveals the efficiency of information diffusion in OSNs of microblogs, which is validated by experimental examination, theoretical analysis and numerical simulation.

2. Data collection and experimental results

In this section, we introduce our experimental approach for data collection and analyze the network structure formed by the relationships among users and followers.

2.1. Data collection

We focus on two large scale social network sites of microblog. One is the largest microblog, Twitter, and the other is the largest Chinese mircoblog, Sina Weibo. Twitter and Sina Weibo rank 8th and 17th respectively in all the sites on the web, which ensures the significance and importance of our study. Sina Weibo and Twitter both offer Application Programming Interfaces (API) enabling data collection easily, so that we can collect detailed information from users and a list of followers. We crawled and collected over 10 millions users' profiles on Sina Weibo. Each Sina Weibo user keeps a brief profile about himself. The public profile includes the full name, gender, the number of followers, followings and bifollowers, verified type indicating whether the user is a VIP or not, and the number of microblogs of the users. Similar information are also collected from Twitter for 41.7 millions of users. Note that in the collection of these data, instead of using the biased sampling method like breadth first search, we utilize random sampling the profiles from the sites in a random way, which ensures the credibility of our investigation on their network structures.

2.2. Experimental results

We construct directed graphs based on the following relationship and analyze its basic characteristics of node degree. Over 10 and 41.7 million user profiles in the dataset of Sina Weibo and Twitter guarantee we can extract broad attributes of the users. Fig. 1 displays the distribution of the number of followers as the solid line. The y-axis represents Complementary Cumulative Distribution Function (CCDF). From this figure, we can observe that the node degree (the number of followers of users) exhibits a power-law distribution with exponent of 1.94, which is shown by the red dashed line. This result conforms the phenomenon that some real networks including social networks have a power-law exponent between 1 and 3 [9] also exists in microblog networks.

By analyzing the collected data from Sina Weibo, we find its distribution of followers count follows power law distribution with exponent near 2. To validate the generality of this obtained result we also plot the distribution of the number of followers in the Twitter data set as the solid line in Fig. 2. From the results, we can observe the number of followers of users exhibits a power-law distribution with exponent of 2.08 shown by the red dashed line. This result conforms the generality of