Physica A xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Physica A

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

A novel information cascade model in online social networks

Q1 Chao Tong [a,b], Wenbo He [b], Jianwei Niu [a,*], Zhongyu Xie [a]

[a] School of Computer Science and Engineering, Beihang University, Beijing, China
[b] School of Computer Science, McGill University, Montreal, Canada

HIGHLIGHTS

- Discuss the spread of two kinds of users' decisions for city-wide activities.
- Study the scale, scope of cascade subgraph and topological attribute of spread tree.
- Analyze the structure characteristic and occurrence frequency of cascade subgraph.
- Information spread model using equal probability, similarity and popularity of nodes.

ARTICLE INFO

Article history:
Received 22 April 2015
Received in revised form 19 June 2015
Available online xxxx

Keywords:
Information spread model
Cascade subgraph
Node similarity
Node popularity
Social networks

ABSTRACT

The spread and diffusion of information has become one of the hot issues in today's social network analysis. To analyze the spread of online social network information and the attribute of cascade, in this paper, we discuss the spread of two kinds of users' decisions for city-wide activities, namely the "want to take part in the activity" and "be interested in the activity", based on the users' attention in "DouBan" and the data of the city-wide activities. We analyze the characteristics of the activity-decision's spread in these aspects: the scale and scope of the cascade subgraph, the structure characteristic of the cascade subgraph, the topological attribute of spread tree, and the occurrence frequency of cascade subgraph. On this basis, we propose a new information spread model. Based on the classical independent diffusion model, we introduce three mechanisms, equal probability, similarity of nodes, and popularity of nodes, which can generate and affect the spread of information. Besides, by conducting the experiments in six different kinds of network data set, we compare the effects of three mechanisms above mentioned, totally six specific factors, on the spread of information, and put forward that the node's popularity plays an important role in the information spread.

1. Introduction

The spread and diffusion of information has become one of the hot issues in today's social network analysis. In social Q2 networks, people are connected by network links, thus it becomes possible for them to influence each other's decisions and behavior. People are influenced by others: in the opinions they hold, the political positions they support, the products they buy, the technologies they use, the activities they pursue, and many other things. Actually, many phenomena in the life, including the new religion or political movement, the adoption of new science technique, the innovative idea, the success of an invention, grave news and so on, all contain some common characteristics. These phenomena usually begins in a small

* Corresponding author.
E-mail addresses: tongchao@buaa.edu.cn (C. Tong), wenbohe@cs.mcgill.ca (W. He), niujianwei@buaa.edu.cn (J. Niu), dakenz1990@163.com (Z. Xie).

http://dx.doi.org/10.1016/j.physa.2015.10.026
0378-4371/

part of people, early adopters; with the help of the social relationship and organization structure behind people, more and more people become the adopters. As for a certain person, his or her final decision usually depends on the neighbors, friends, colleagues and so on. In a general sense, the information includes the specific information or news, and these actions and idea mentioned above. In the networks, the information spreads according to some certain structures continuously, and form the cascade phenomenon. Namely, information spreads from a small part of network nodes, and is eventually accepted by a wide range of nodes. This terminology, cascade, comes from the work of Banerjee [1]. The concept was also developed around the same time in other works [2,3].

It is natural to observe and consider some of the reasons why cascade occurs. In this paper, the problems that we should pay attention to are: how the information cascades occur in online social network structures? What are the information's characteristics during its spread? How the single node affects each other based on the bottom structure? Can we design a model to describe these phenomena?

In the information and social networks, especially the online social networks, the information spread corresponds to the user's action and the various specific contents released on the web. The action that a Weibo user forwards a micro blog serves as a good example. Researching the information spread in the online social network can not only view the dynamic characteristics of the online social network from a single angle, but also help the web and the application systems to provide better services. The analysis and model of the information spread in the network can be directly applied to advertising, recommend friends to the users, and predict the network's characteristics. The information spread in the networks (or the information diffusion) is an important component of the network's dynamic characteristic. Here the information can be regarded as not only the news or data in traditional meaning, but also the action or decision. The cascade phenomenon has become a widely research topic in sociology, especially the sociologist Everett Rogers proposed the "Diffusion of Innovations" Theory [4]. In recent years, the research efforts about the information diffusion mainly include finding the initial target in the viral marketing [5], searching the vaccination objects in the process of viral infection [6], the topic's trend in the Blogsphere [7] and so on. Most of these works have combined with the information diffusion and the epidemiology, determining the node's spread behavior according to its status in the spread process. In the angle of researching method, research efforts about the information spread can be divided into two categories. One kind of research efforts views the issues in the coarse-grained angle, they do not emphasize the structure of bottom network, but abstract the information spread as the process of interacting each other's decision among the nodes. Then, a series of herding experiments [8–10] and relevant linear cascade models [1,3] are derived from this idea. Corresponding to that, the other kind of research efforts views the issues in the fine-grained angle; they pay attention to the network's structure and the interaction of the information spread, and consider how the nodes' local structure leads to a large scope information diffusion. The network structure's influence on the information diffusion enlarges the network's function, not only the function of linking, but it also serves as the social media. There are some researchers who have analyzed the media attribute of the social network with the work in the information spread's topological structure [11]. The research efforts in the fine-grained angle usually obey two kinds of model, "Models of Collective Behavior" and "Models of Independent Diffusion" [12]. With the help of the model of the infection, some researchers expanded the models of collective behavior and the independent diffusion model, and proposed the sequential influence models [13], linear influence models [14] and other models for estimating the influence. Kwak et al. [11] analyzed the tree structure form by tweeting in the twitter, and found that if a tweet in the twitter is forwarded, the average people covered by it is 1000, and this datum has nothing to do with the person's popularity in the twitter. Götz et al. [15] took the release time of the articles in the blog network into account, and proposed the Zero-crossing model to reflect the time characteristic during spread process. Besides, there are some other cascade models [16–22].

However, these research efforts above mentioned are based on some specific networks, like the blog network. They are not enough for Chinese large online social network. Therefore, in this paper, we focus on the model of information spread and diffusion in the network on the basis of the network structure. Aiming at the biggest Chinese social hobby network and user-generated content network, DouBan [23], we carry research efforts on the process of city-wide activities spread, the characteristics and models of the decision for users to take part in the activities. For instance, we discuss the related attributes of the information spread in the DouBan's user relationship network and the city-wide activity decision data and analyze the spread process of two decisions, "take part in" and "be interested in". We focus on the structure characteristics of cascade subgraph formed during the decision spread process, including the cascade's scale and scope, the cascade subgraph's diameter, density and other structure attributes, the structure of the spread tree, the cascade subgraph structure. We also propose a corresponding model to simulate the phenomenon in the real life. In the model, we introduce 3 mechanisms which can lead to the information diffusion, equal probability, node similarity, and node popularity; and 6 specific mechanisms, including PageRank, Jaccard coefficient between adjacent nodes, the fans popularity and so on. Besides, based on the 6 groups of data set, such as online social network, citation network, Email communication network and so on, we conduct experiments to compare the information diffusion's impact in these 6 mechanisms.

The research efforts about the attributes of the online social network information spread and the information diffusion model introduced in this paper have significant meanings in the applications. According to the structure characteristics of the cascade subgraphs (which are subgraphs of spread tree and will be covered later in this paper) formed in the spread of information, action and any other entities in the network, we can analyze and predict the specific characteristics of the spread's scale, depth, and coverage size. Through cascades formed by different kinds of information spread in the network, we can predict the spread effect of a new kind of information with the help of historical data, and choose the initial node to maximize the information's influence. For instance, in some application circumstances, we can set the key node for the

Please cite this article in press as: C. Tong, et al., A novel information cascade model in online social networks, Physica A (2015), http://dx.doi.org/10.1016/j.physa.2015.10.026