Journal of Network and Computer Applications 86 (2017) 92–102

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Analyzing and modeling dynamics of information diffusion in microblogging social network

Yadong Zhou[a,b], Beibei Zhang[c], Xiaoxiao Sun[a], Qinghua Zheng[a], Ting Liu[a,*]

[a] Ministry of Education Key Lab for Intelligent Networks and Network Security, Xi'an Jiaotong University, PR China
[b] Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, PR China
[c] Xi'an University of Technology, PR China

ARTICLE INFO

Keywords:
Pervasive social networking
Information diffusion
Diffusion model
Retweeting network

ABSTRACT

Among different types of the pervasive social networks, microblogging social network recently provides most efficient services for diffusing information of news, ideas and innovations. The features and models of information diffusion in microblogging social network have attracted many researchers. Compared to other works, our study provides a new perspective of analysis for information diffusion. A multi-level structure is defined to analyze the diffusion process of hot topics. The diffusion of a particular topic is represented as the evolution of the related retweeting network, where retweeting groups and information cascades are growing and interacting. Based on this multi-level structure, interesting features of the *merging effect* between two retweeting groups, the existence of *super group* and the *centralized topology* of information cascades are discovered and analyzed. Furthermore, we find that trend of diffusion in the future is influenced by diffusion in the past, and the main factors of dynamics of retweeting network are also analyzed. From the above analysis, a diffusion model based on cascade model framework is proposed to generate the retweeting network. Based on the real data, the experimental results show that our model could reproduce the diffusion features of the retweeting network effectively and outperforms the most widely used independent cascade model.

1. Introduction

With the rapid growth of pervasive social networking (PSN), various types of PSN platforms play an increasingly important role in sharing and diffusing interesting information in Internet. Among these PSN services, microblogging social network (e.g. Twitter, 2015; SinaWeibo, 2015) provides efficient sharing and communicating services for billions of users over the whole world (Kwak et al., 2010; Hughes, 2014; Haralabopoulos and Anagnostopoulos, 2014) in recent years. With large amount of users and the innovative ways for user interactions (following, retweeting), microblogging has unique patterns of information diffusion dynamics on hot topics (political events, natural disasters, breaking news etc.), and significant impact on social life (Myers and Leskovec, 2014; Feng et al., 2015; Li and Ng, 2013). Its importance also leads to growing research interests in information diffusion over microblogging sites in recent years. Many challenging issues are encountered, specifically:

- How does the topology of diffusion network evolve dynamically in the information diffusion?
- Is the trend of the diffusion in the future influenced by the past diffusion?
- What are the main factors of the diffusion process?
- How to model the information diffusion process to reproduce its unique patterns?

To meet above challenges, a multi-aspect and in-depth analysis for the information diffusion is needed. Since the retweeting mechanism is the main reason of the rapid and wide diffusion of topics over microblogging networks (Pervin et al., 2014), it was the most common starting point for research. However, previous studies usually give general analysis for the diffusion of topics (Dickens et al., 2012; Wei et al., 2012; Lu et al., 2014), which are insufficient to demonstrate the interaction and dynamic evolution among diffusion cascades of tweets related to a same topic. Being able to observe the interactions and evolutions, which can reflect the dynamics of retweeting mechanism, requires a deep understanding of both macro- and micro-level structures involved. Thus, a multi-level structure, including retweeting network, retweeting group and information cascade, is defined in this paper. Based on this, a new perspective of analysis for the dynamics of

* Corresponding author.
E-mail addresses: ydzhou@xjtu.edu.cn (Y. Zhou), bbzhang@sei.xjtu.edu.cn (B. Zhang), xiaoxiaosun@xjtu.edu.cn (X. Sun), qhzheng@xjtu.edu.cn (Q. Zheng), tingliu@xjtu.edu.cn (T. Liu).

http://dx.doi.org/10.1016/j.jnca.2016.09.011
Received 15 April 2016; Received in revised form 30 July 2016; Accepted 26 September 2016
Available online 28 September 2016
1084-8045/

retweeting mechanism based on the real and huge data from Sina microblogging site is studied and presented.

In this paper, the information diffusion of a popular topic is treated as the evolution of a retweeting network, containing users tweeting about the topic and retweeting links between users. Retweeting network can be divided into several weakly connected components, and the set of users in each component is defined as a retweeting group. Tweets are diffused as information cascades in each retweeting group and the retweeting network evolves dynamically. Furthermore, three important features in the evolution of retweeting network are found: (1) *Merging effect*: During the diffusion process, retweeting groups would merge with other groups by new retweeting links. The two groups merged together are defined as a merging pair. Merging effect shows the dynamics of diffusion path. (2) *Super group* phenomenon: The retweeting network is gradually evolving into a highly centralized network with a diffusion center defined as a super group, which is the largest retweeting group with over 30% users in the retweeting network, and attracts newcomers and merges other groups by the influential users. (3) *Centralized topology*: The types of topologies of information cascades consist of star and multi-center star topology, but rarely long chain topology in the retweeting network. It indicates that there could be bursting diffusions around influential users, but rarely long-distance diffusions through ordinary users.

Obviously, the users interested in a topic may tweet it more than once, leading to a repeated diffusion through same users. It is interesting to study whether the trend of the diffusion in the future would be influenced by the past diffusion. We consider this issue as a comparison between two situations for a given merging pair: diffusion when groups are merging and diffusion when groups have merged. The results show that the proportion of users attracted by the diffusion is larger in the second situation than in the first situation, indicating that the diffusion in the future is influenced by the past (Dickens et al., 2012). Namely, the merging effect can promote the information diffusion.

We also noticed the existence of influential users in the diffusion process. To find out the most influential users, we rank users in descending order by three easily obtained features: in-degree (number of followers), out-degree (number of followees) and number of tweets respectively, and add a contrast using the number of retweets proposed by Kwak et al. (2010) and Cha et al. (2010). Then we delete user vertices from the retweeting network to see the importance of each user to the retweeting network. The result shows that the in-degree rank works equally well as the number of retweets rank, but requires no complex calculation of the data. Thus, our study shows that the in-degree, as a common feature of most influential users, can be an indicator of user influence.

Furthermore, a diffusion model with dynamic parameters is proposed to generate the dynamics of retweeting network and reproduce the diffusion features mentioned above. The initial parameters of the model are topic-independent and are estimated continually with the real data. There is a good agreement between the diffusion features in the real microblogging social network and the results of our diffusion model.

The contributions of this paper are summarized as follows:

- Instead of a general analysis for information diffusion, we focus on the interaction and dynamic evolution among diffusion of tweets related to a same topic. Some interesting but not intuitive diffusion features, including merging effect, super group and centralized topology, are found and analyzed by defining a multi-level structure of information diffusion. The findings are valuable to understand the information diffusion over microblogging networks deeply.
- We demonstrate that the trend of the diffusion in the future is influenced by the past diffusion, which is against the implicit assumptions in some traditional diffusion models such as the independent cascade model and the linear threshold model (Kempe et al., 2003).
- We analyze main influence factors of dynamics of retweeting network, which are useful to estimate the tweet probability and retweet probability in the diffusion model. In the analysis, we find that in-degree can be a simple but effective indicator of user influence in the diffusion, which is underestimated in some other works (Cha et al., 2010; Romero et al., 2011).
- We propose a diffusion model based on cascade model framework to generate the retweeting network and reproduce the diffusion features found in this paper. The parameters in the model are estimated dynamically based on Bayesian network over time. The experimental results show that our model could match the important diffusion features of the retweeting network effectively and outperforms the most widely used independent cascade model.

The rest of this paper is organized as follows. Section 2 covers related works and puts our work in perspective. Section 3 presents a brief introduction of Sina microblogging site and the data sets used in this paper. In Section 4, we analyze the diffusion features and main influence factors of dynamics of retweeting network, and study the relevance between the future diffusion trend and the previous diffusion. Section 5 provides a diffusion model with dynamic estimated parameters to generate the retweeting network. In Section 6 we discuss and conclude.

2. Related works

2.1. Microblogging network

Twitter, as the world's most widely used microblogging site, has attracted many studies on its topology features (Kwak et al., 2010; Myers et al., 2014), the retweeting behavior (Pervin et al., 2014), the follow prediction (Hutto et al., 2013), and the interaction between the network topology and the users (Antoniades and Dovrolis, 2014). Furthermore, research on potential applications, such as scientific communication (Letierce et al., 2010) and disaster response (Li et al., 2015), is also popular. In the last two years, Ahmed et al. (2015) explored the differences between tweets mentioned real-world events and tweets not mentioned real-world events. Zaman et al. (2014) analyzed the retweet path and developed a probabilistic model using a Bayesian approach to predict the popularity of tweets. Besides, research articles on Sina microblogging site gradually increased in recent years. Han et al. (2016) analyzed the structural properties of Sina microblogging network and made a comparison between Sina microblog and Twitter. Wang et al. (2012) analyzed the users activities and found possible ways to improve users quality of experience. Bao et al. (2012) provided a method to predict the popularity of posts on sina microblog by the structural diversity of the early adopters. Guan et al. (2014) analyzed the posting and reposting behavior of the users in 21 hot events in Sina microblogging network and found that males were more active.

2.2. Information diffusion over microblogging network

Recently, studies on information diffusion over microblogging networks are increasingly popular. Fan et al. (2011) analyzed the topology and information diffusion process on Sina Microblog. Weng et al. (2013) examined how information diffusion caused the creation of new social links and reshaped the structure of microblogging system. Dickens et al. (2012) presented models of information flow based on the Independent Cascade Model to predict the probabilities of invisible flow on Twitter. Wei et al. (2012) built three diffusion models based on Logistic function for crisis information in micro blog. Bao et al. (2013) analyzed the cumulative effect in information diffusion in Sina microblog. Cheng et al. (2014) analyzed the features that can be used to predict the size and the eventual shape of the cascades. Myers et al.

(2012) presented a model which allowed users to receive information from both network links and external sources, and found that 29 percent of the information volume is from outside network. Feng et al. (2015) pointed out that the diffusion process of popular messages on Sina microblog differed from the spread of disease. Yang et al. (2012) modeled the Twitter network by retweet relations and proposed a variant of the HITS algorithm to find the interesting posts. Li et al. (2016) proved that the information diffusion influenced the link creation on Sina microblog. Kupavskii et al. (2012) trained an algorithm to predict the number of retweets of a given tweet during a fixed time period. Myers and Leskovec (2014) found that the dynamics of network structure can be interrupted by sudden bursts created by the information diffusion in the retweet cascades.

2.3. Indicating influential users

The information diffusion processes in microblogging network are influenced by users. There have been multiple ways to measure user influence. Kwak et al. (2010) presented a comparison of three measures of influence: in-degree, PageRank in the following/follower network and the number of retweets on Twitter. Cha et al. (2010) compared three measures of influence: in-degree, retweets and mentions on Twitter, and indicated that users with high in-degree were not necessarily influential. Julia Heidemann and Klier (2010) provided an algorithm combining PageRank and centrality to evaluate user influence. Romero et al. (2011) presented a novel influence measure involving the passivity of the audience and indicated that the number of followers is a poor measure of influence. Brown and Feng (2011) presented a method based on K-Shell decomposition to identify users with huge influence. Zhang et al. (2011) proposed an influence evaluation model based on user behaviors like retweeting and commenting. Bakshy et al. (2011) analyzed the attributes of Twitter users and quantified the relative influence the users. They found that individuals who have average or less than average influence are the most cost-effective under many plausible assumptions. Shuai et al. (2012) analyzed the indirect influence between two users who are not directly connected and the features that affects the indirect influence.

2.4. Cascade model

Information diffuses as cascades through communication channels. Leskovec et al. (2006) studied the patterns of cascading recommendations in large social networks. Gruhl et al. (2004) analyzed the information diffusion in blog-space and proposed an Independent Cascade model of individual diffusion. Leskovec et al. (2007) analyzed the cascading behaviour over blog-space and provided a simple generative model similar to the susceptible-infected-susceptible (SIS) model (Hethcote, 2000). Xu and Chen (2015) provided a rumor source detection algorithm on Twitter based on the Independent Cascade model.

3. Data collection

Sina Microblog, one of the most popular large-scale online social networks in China, commands more than 212 million active users as of August 2015 and about 100 million messages posted each day (SinaWeibo, 2015). Sina Microblog offers an Application Programming Interface (API) for the data crawling. The data sets used for this work is crawled as a weakly connected sub-graph of the entire Sina Microblog network. It includes 3.13 million user profiles, 127 million social relations and 69.4 million tweets, gathered over a six-month period from the beginning of November 2013 to end of April 2014. We obtain the top 200 words with highest frequency in the tweets, and filter the nonsensical words manually. Then, 38 hot topics are identified based on the remaining words for analysis.

4. Analysis for retweeting network in information diffusion

4.1. Definitions

The microblogging network is considered as a directed graph $G = (V, E)$, where each vertex in V represents a user and each directed edge in $E \subset V \times V$ represents a following relationship. Users can be mentioned when their followees tweet and can decide whether to retweet or not. If vertex u retweets vertex v, the directed edge (u,v) is called a retweeting link. Thus, information diffuses along the directed edges in the microblogging social networks.

In this paper, we focus on the diffusion of a hot topic T_i, defined as a series of tweets related to a hot social event. If u retweets from v about topic T_i, the retweeting link (u,v) is called a retweeting unit of T_i. The retweeting network consists of all retweeting units of T_i. Study on the diffusion process of T_i can be transformed as the study on the evolution of retweeting network for analyzing the dynamics of retweeting mechanism. Here we define three terms as follows.

Definition 1. *Retweeting network* $R_{T_i} = (V_{T_i}, E_{T_i})$ is a sub-graph of G, where each vertex in $V_{T_i} \subset V$ represents a user tweeting about T_i and each edge in $E_{T_i} \subset E$ represents a retweeting link pointing from the follower to the followee tweeting about T_i. R_{T_i} can be divided into several weakly connected components.

Definition 2. *Retweeting group* $g_{T_i,j}$, $j = 0, 1, \ldots, N_{T_i}$ is the set of whole vertices V_{T_i} in one weakly connected component R_{T_i}, $j = 0, 1, \ldots, N_{T_i}$ [21] of the retweeting network R_{T_i}.

Definition 3. *Information cascade* $c_{T_i,j,k} = (V_{T_i,j,k}, E_{T_i,j,k})$, $k = 0, 1, \ldots, M_{T_i}$, $j = 0, 1, \ldots, N_{T_i}$ is a retweeting tree of a particular tweet related to T_i in each retweeting group $g_{T_i,j}$.

An example of the retweeting network is shown in Fig. 1. Retweeting groups consisting of vertices having retweeting relationships. Cascades illustrate the information flow in the retweeting network. The majority of cascades are with three topologies: star, chain and multi-center star (shown in Fig. 2).

From t_0 to t_e, new vertices and retweeting links continuously appear and change the structure of R_{T_i}. Therefore, the definitions above can be formulated as follows:

$$R_{T_i}(t) = (V_{T_i}(t), E_{T_i}(t)), \quad t \in [t_0, t_e] \tag{1}$$

$$g_{T_i,j}(t) = V_{T_i,j}(t), \quad j = 0, 1, \ldots, N_{T_i}, t \in [t_0, t_e] \tag{2}$$

$$c_{T_i,j,k}(t) = (V_{T_i,j,k}(t), E_{T_i,j,k}(t)), \quad k = 0, 1, \ldots, M_{T_i}, j = 0, 1, \ldots, N_{T_i}, t \in [t_0, t_e] \tag{3}$$

The following analysis is based on cases of 38 hot topics. One of the topics is chosen as the example to show the common results. In circumstances where there is no ambiguity, the subscript T_i is omitted.

4.2. Diffusion features of retweeting network

Fig. 3 shows the evolution of retweeting network based on the change of retweeting groups and the number of users. The y-axis in the

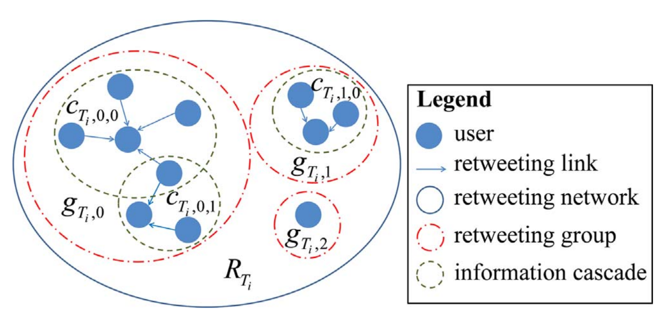

Fig. 1. An example of the retweeting network of topic T_i.

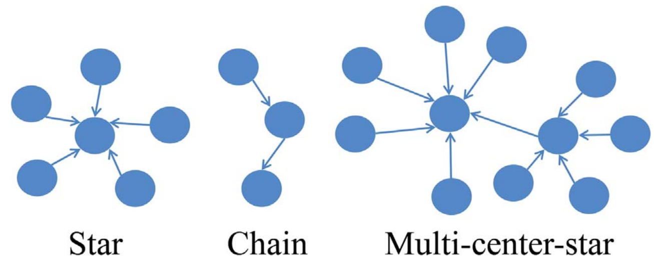

Fig. 2. Major cascade topologies in the retweeting network.

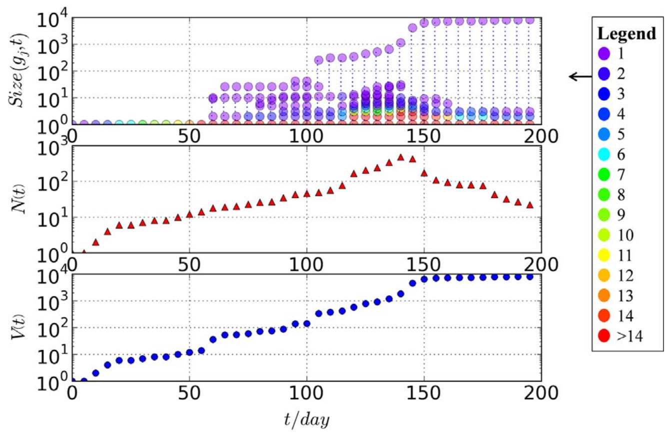

Fig. 3. Evolution of retweeting network based on the change of retweeting groups and the number of users.

top figure represents the distribution of the number of users in each retweeting group $Size(g_j, t)$, and the colors represent the number of groups where $Size(g_j, t) = y$. The y-axis in the middle figure represents the number of all groups $N(t)$ in the retweeting network. The y-axis in the bottom figure represents the number of users $V(t)$ in the retweeting network. The x-axis represents the time in days.

During the evolution of each hot topic, it is easily to distinguish a special retweeting group due to its huge size. A giant group, called the super group g_{max}, is formed during the diffusion process. With the growth of g_{max}, $N(t)$ increases and then decreases. The evolution process can be divided into three periods according to the inflection points during the dynamic growth of the size of g_{max}: appearance period, stabilization period and stabilization period. In different periods, the size of g_{max} would increase at different accelerations. The divisions of the three periods in the example shown in Fig. 3 are the appearance period ($60 \le t < 100$), the scale-up period ($100 \le t < 150$) and the stabilization period ($150 \le t < 195$). Figs. 4–6 show the typical centralized topology and number of cascades during each evolution period. The detailed analysis is as follow.

- *Appearance period*: During this period, the number of users $V(t)$ grows slowly. It is the time when an influential user participates in the topic and establishes the embryonic form of g_{max}. Notice that g_{max} is a combination of newcomers and old groups. Since g_{max} contains large number of users, including several influential users, newcomers are more likely to join g_{max} than other small groups in the following time steps. Fig. 4 shows the typical information cascades during the appearance period. The cascades are only with star topology.
- *Scale-up period*: During this period, the number of users $V(t)$ begins to grow faster. The super group g_{max} attracts newcomers and merges other groups by retweeting links due to the virtue of the influential users in it. As the merging actions become more and more violent, $N(t)$ begins to decrease. Notice that g_{max} becomes the diffusion center and is growing rapidly, showing a huge difference in size with the other groups. Fig. 5 shows the typical information

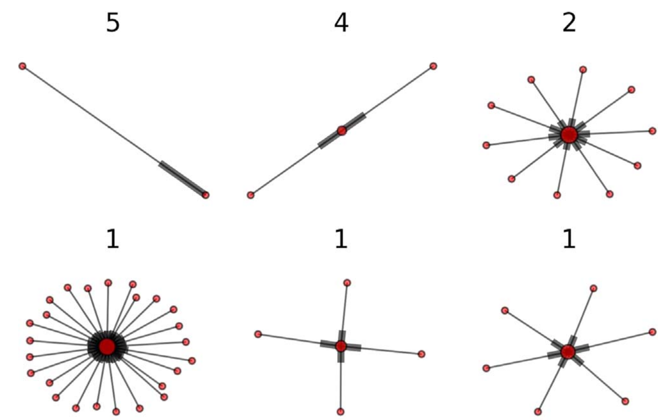

Fig. 4. Typical information cascades in each evolution period: appearance period (the cascades are only with star topology).

cascades during the scale-up period. Cascades with chain topology appear in the network.
- *Stabilization period*: This period is the end of the topic's life cycle. The growth of $V(t)$ gradually stabilizes. A few new users participates in the topic, while the merging actions among retweeting groups are still in progress, leading to a slow reduction of $N(t)$. There are over 600 information cascades in the network and the top 49 are shown in Fig. 6. Notice that cascades with the three major topologies co-exist in the retweeting network.

In addition, the distribution of users is shown in Fig. 7. Notice that g_{max} contains over 30% users in the retweeting network, showing a huge difference with the other groups. The distribution of cascade topology is shown in Fig. 8. Most cascades are with star or multi-center star topology. Cascades with long chain topology are relatively rare. It indicates that there are mainly bursting diffusions around influential users, but rarely long-distance diffusions through ordinary users.

The above analysis shows some interesting phenomena in the evolution of retweeting network. Retweeting groups merge with each other by retweeting links. Retweeting network evolves following the merging effect among groups. The end of the evolution shows a clear polarization: one super group and many small groups. The super group contains about one-third of the users and becomes the diffusion center of the retweeting network. These phenomena are discovered in all cases of hot topics, which are undetected before to our best of knowledge. We give the definitions of the merging effect and the super group as follow.

Definition 4. *Merging effect* is a phenomenon that two retweeting groups are combined by new retweeting links and become one larger group. The two original groups are called a merging pair.

Definition 5. *Super group* g_{max} is a retweeting group with the largest number of users among all the groups. It is the diffusion center of the retweeting network.

Fig. 9 shows the evolution process of the super group g_{max}, where different colors and styles of the edges represent different information cascades.

Fig. 9 indicates that the information cascades are not growing independently in the retweeting network. Some active users with close retweeting relationship participate in more than one cascades, associating the cascades with each other. The phenomenon among each cascade is the *merging effect*. Compressing the merging effect on the time dimension results to the super group g_{max}. From the above analysis, we can identify three kind of users in the retweeting network as follow.

- Popular users who can bring a great amount of retweeting links. These users bring numbers of users for the retweeting network R_{T_i} and become the main part of g_{max}.
- Active users who connect different cascades. These users form the

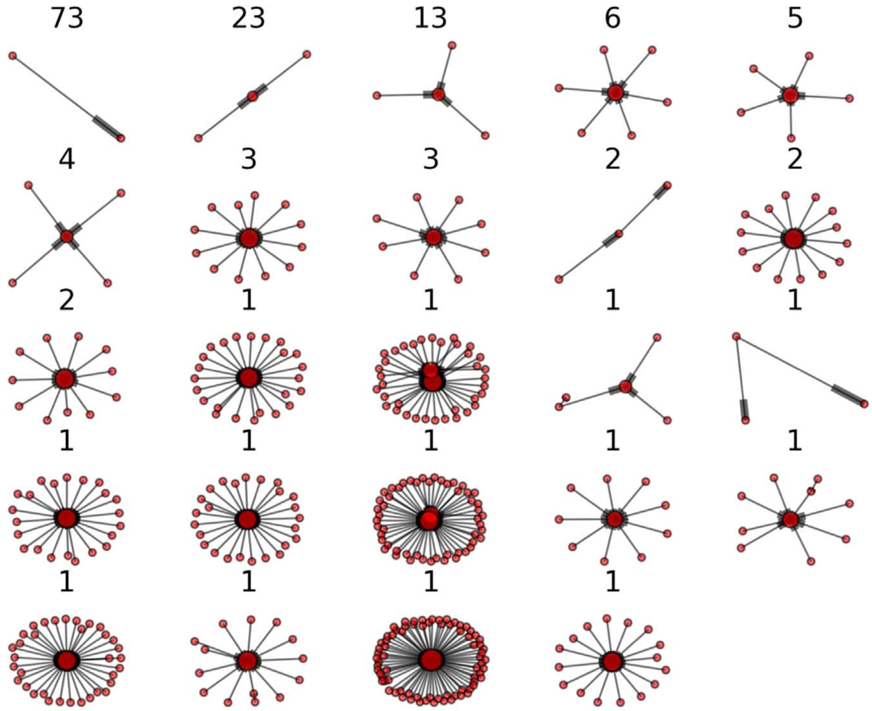

Fig. 5. Typical information cascades in each evolution period: scale-up period (the cascades with chain topology appear in the network).

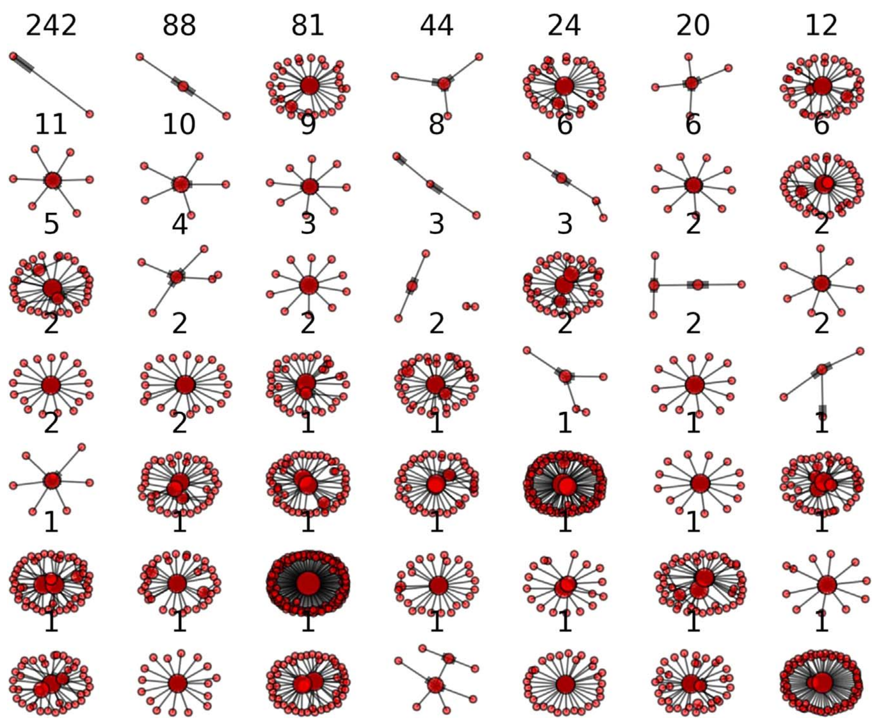

Fig. 6. Typical information cascades in each evolution period: stabilization period (the cascades with three major topologies including star topology, chain topology and multi-center star topology co-exist in the retweeting network).

information channels among the cascades in the retweeting network, leading to the interactive evolution process.

- Inactive users. These users take the most part of the retweeting network, but have little influence on the information diffusion.

Users in many traditional online social networks, such as blog network, need to write long content for most of the posts and rarely post repeatedly on the same topic. However, users in microblogging network often tweet more than once on the same topic. Fig. 10 shows the distribution of the number of tweets that a user participates in a retweeting network. Over 20% of users participate in at least 2 tweets and 10% of users participate in at least 3 tweets. Notice that there are a few users who tweet/retweet about the same topic more than 100 times.

4.3. Does the past diffusion influence the trend of the diffusion in the future?

From the above analysis we can notice that when a hot topic T_i

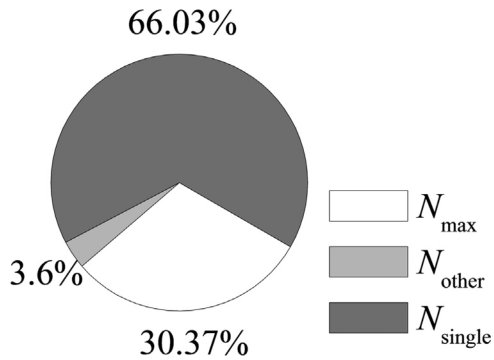

Fig. 7. Distribution of users based on the retweeting groups (N_{max} represents the number of users in the super group g_{max}. N_{single} represents the number of users with no retweeting links. N_{other} represents the number of users in other groups).

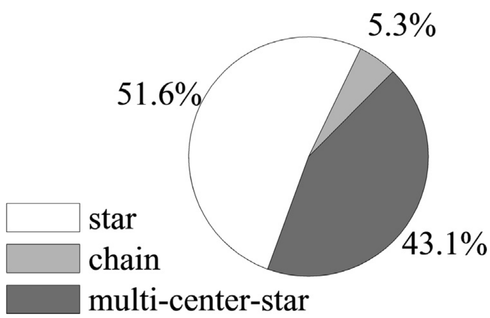

Fig. 8. Distribution of the cascade topologies.

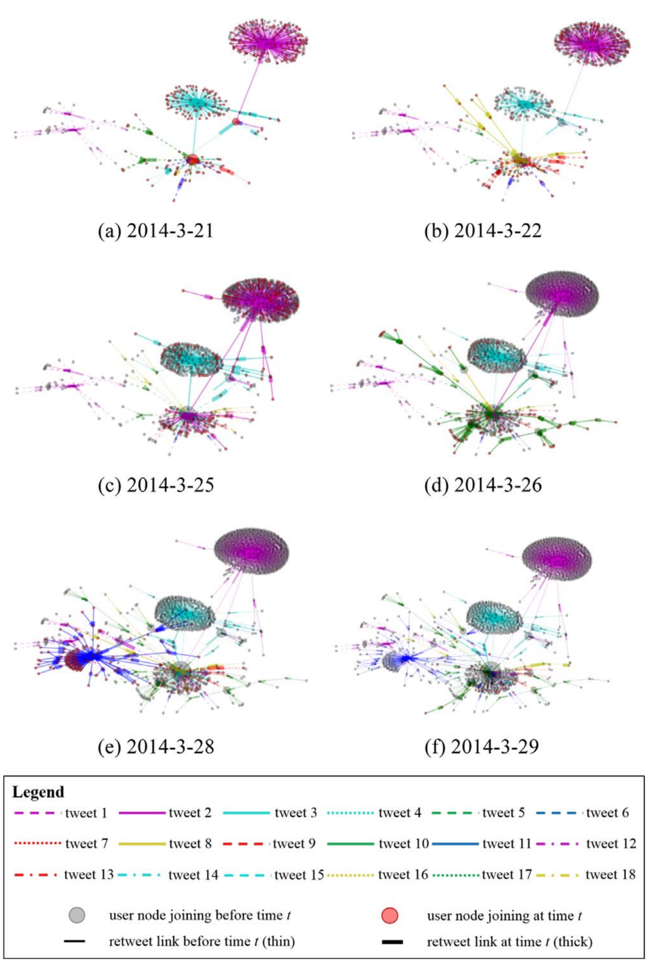

Fig. 9. The evolution process of the super group g_{max} (the size of each node is proportional to the number of tweets of him/her).

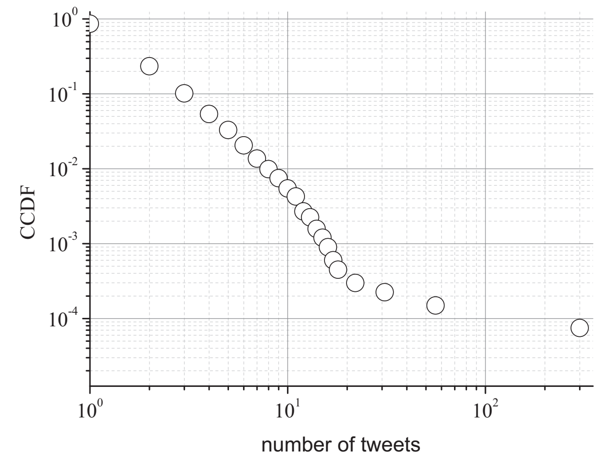

Fig. 10. The number of tweets that the user participate in R_{T_i} (CCDF means complementary cumulative distribution function).

diffusing on the microblogging network G, users who are interest in T_i often tweet/retweet about the topic more than once. The repeated tweeting behavior may cause second retweeting behavior from his/her followers, which let the topic diffuses repeatedly on the same edge of the retweeting network. Fig. 11 shows the repeated diffusion in the retweeting network, where x axis represents the number of occurrences of a retweeting link (u,v) in the same retweeting network and y axis represents the complementary cumulative distribution function. 7% of retweeting links occurs at least twice, and 1.6% of retweeting links occurs at least three times. The distribution of the number of occurrences fits to a power-law distribution with the exponent of 2.12.

Most of the retweeting links occur only once in the same retweeting network as shown in Fig. 11. But considering the large number of inactive users and the small number of active users in the information diffusion process, the repeated diffusion behavior can be considered as a common phenomenon.

The repeated diffusion between two users are often considered to be independent in many typical information diffusion models (such as independent cascade model and linear threshold model), which means that the future diffusion trend will not be influenced by the past diffusion process. Unlike other social networks, the directed connections, the brief information expression and the efficient retweeting mechanism create a different information diffusion mode on the microblogging networks. The above analysis shows that there are some users participating in multiple information cascades, and the different information cascades may share retweeting links. The information

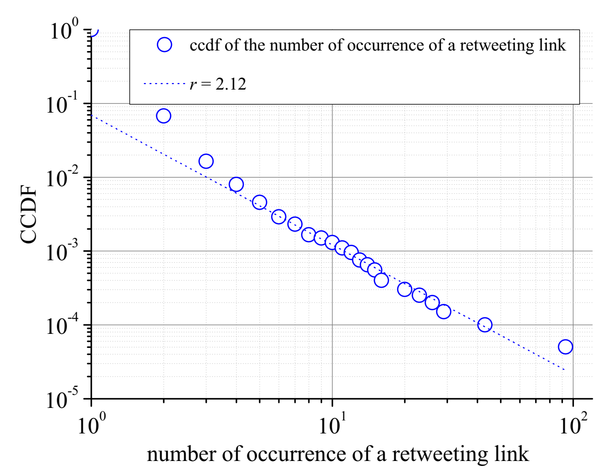

Fig. 11. The number of occurrences of a retweeting link (u,v) in (u,v) in R_{T_i} (CCDF means complementary cumulative distribution function).

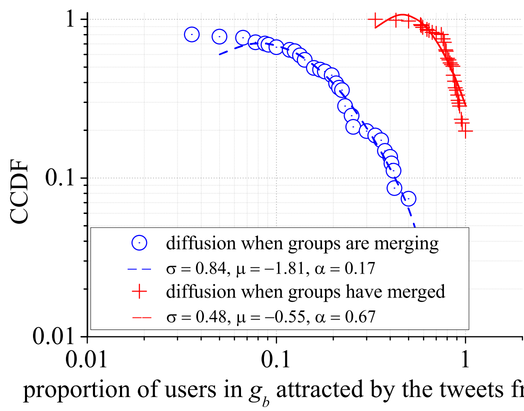

Fig. 12. Proportion of attracted users in two situations (CCDF means complementary cumulative distribution function).

diffusion process may become dependent due to the repeated diffusion phenomenon. We are curious about whether the trend of the diffusion in the future will be influenced by the past diffusion. This question can be transformed to analysing the influence of merging effect.

Given a merging pair g_a and g_b, the merging action happens when a user in $g_a(t_1)$ tweets w_1 and users in $g_b(t_1)$ retweet w_1 at time t_1. It is the first time when information is diffused from g_a to g_b. If a user in $g_a(t_1)$ tweets w_2 later at time $t_2 > t_1$ and some users in $g_b(t_1)$ retweet w_2, the information is diffused again from g_a to g_b. Both diffusions from g_a attract a certain number of users in g_b. The two situations here is described as the diffusion when groups are merging and the diffusion when groups have merged.

We randomly select 100 merging pairs and compare the proportion of users in $g_b(t_1)$ attracted by the tweets from $g_a(t_1)$ in the two situations. The comparison is limited in a small time window to avoid the interference from the changing popularity of the topic. Here, the small time window is chosen as 24 h, and that means we only analyze the behaviors that the users in $g_b(t_1)$ attracted by the tweets from $g_a(t_1)$ during the period from when the groups are merging to 24 h later. Fig. 12 shows the distribution of the proportion of attracted users in each situation. The y-axis represents the complementary cumulative distribution function (CCDF). To our statistics, the proportion is larger in the second situation than in the first situation. Both distributions fit to a log-normal distribution.

The possible reason is as follow. The first information diffusion from $g_a(t_1)$ to $g_b(t_1)$, establishing a diffusion path between the two groups. Users in $g_b(t_1)$ firstly read the tweet from users in $g_a(t_1)$, but may not retweet it. When comes a second tweet from $g_a(t_1)$, users in $g_b(t_1)$ have been familiar with the tweet from users in $g_a(t_1)$, leading to more retweeting.

In summary, our study proves that the trend of the diffusion in the future is influenced by the past. In addition, the result illustrates that the merging effect promotes the information diffusion effectively.

4.4. The main influence factors of dynamics of retweeting network

A large number of users participate in the information diffusion in the microblogging networks, but the influence of each user is different. Only a few users in the retweeting network have huge influence. According to the analysis of Fig. 9, there are two kinds of users who can influence the evolution of the retweeting network R_{T_i} and the formation of the super group g_{max} significantly. The first kind of users create popular tweets. The second kind of users retweet actively and bring more retweeting. The analysis of the influential users takes two important characteristics of influence: the number of tweets that a user participates in R_{T_i} and the number of retweets caused by the user. We calculate the correlation between the two characteristics of influence and the correlation with the user features. The results are shown in Table 1. The number of followers, the number of followees, the number of friends, the number of tweets, the number of original tweets, the number of retweets, and the average daily number of retweets caused by the user are considered as the common user features.

The results shown in Table 1 reveal that the number of tweets that the user participate in R_{T_i} has strong correlation with the number of followees, the number of tweets and the number of retweets respectively. It indicates that the users participating in many information cascades are usually active users in the microblogging networks. They follow many other users, tweet and retweet frequently. The number of tweets that the user participate in R_{T_i} also has correlation with the number of friends, which can represent the size of his/her friend zone in the microblogging network. Table 1 also shows that the number of retweets caused by the user has strong correlation with the number of followers and the average daily number of retweets caused by the user. It indicates that popular users often bring a large number of retweets in R_{T_i}. It is interesting to notice that the two characteristics of influence have little correlation between each other.

Here we give a deeper analysis for the popular users who can bring a large number of retweets in R_{T_i}. We rank users in descending order by three easily obtained features: in-degree (number of followers), out-degree (number of followees) and number of tweets respectively, and add a contrast using the number of retweets proposed by Kwak et al. (2010) and Cha et al. (2010). Then we delete user vertices off the retweeting network to see the importance of each user of the retweeting network. Notice that all users here are active members of hot topics.

Fig. 13 shows the changes of the number of users in the super group $Size(g_{max})$ and the number of all the groups N. The more changing of the $Size(g_{max})$ and N along with the number of nodes deleted, means the more significant influence that the metric has. Among all the metrics, the in-degree rank works the second best. When comparing the in-degree rank to the number of retweets rank that work best, we find a big hop in each result. Both hops are caused by the same user. When this user is ignored, the results of in-degree rank and the number of retweets rank are similar. However, it is noticed that the number of retweets requires complex calculation, which is unnecessary for the in-degree. Therefore, in-degree can be a simple but effective measure of influence among active users. In real cases, users with large in-degree

Table 1
Correlation between user features.

User features	Number of tweets posted or retweeted by the user in R_{T_i}	Number of retweets originated from the user in R_{T_i}
Number of followers of the user	0.09	0.34
Number of followees of the user	0.49	0.08
Number of friends of the user	0.33	0.01
Number of tweets of the user	0.57	0.05
Number of original tweets of the user	0.28	0.08
Number of retweets of the user	0.61	0.02
Number of retweets originated from the user	0.03	0.39
Number of tweets posted or retweeted by the user in R_{T_i}	–	0.002
Number of retweets originated from the user in R_{T_i}	0.002	–

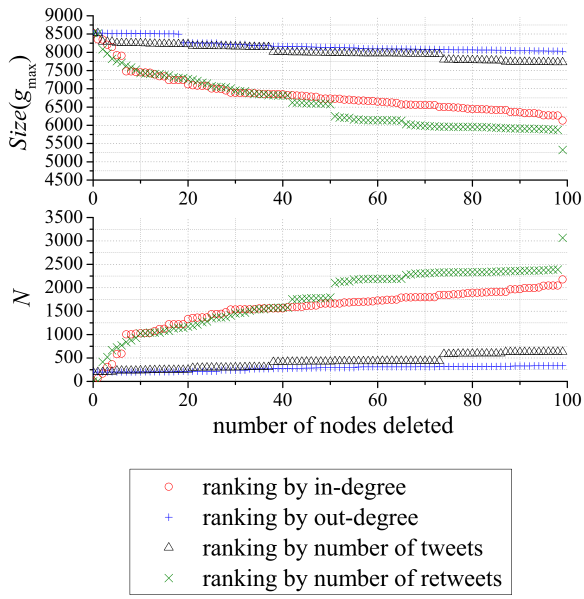

Fig. 13. Delete nodes according to four metrics.

usually are celebrities or news media. Their tweets are popular and can be diffused widely. So, they are effective in diffusing information.

5. Diffusion model

The above analysis indicate that the traditional diffusion models need some changes before applying on the microblogging network. In this section, we build a diffusion model based on a cascade model framework to generate the retweeting network of the information diffusion with properties observed in Section 4.2. In the model, the main problem is how to estimate the tweet probability and retweet probability. According to the analysis in Section 4.3, the past diffusion could influence the trend of the diffusion in the future. Thus, we form the estimation problem of tweet probability and retweet probability as a posterior probability estimation problem, and apply Bayesian network to solve it. Besides, the features of the Bayesian network are selected based on the analysis for main influence factors of dynamics of retweeting network in Section 4.4.

5.1. Model

The microblogging network is described in Fig. 1 as a directed graph $G = (V, E)$. Here we label each edge (u,v) with a retweet probability $P_{u,v}$. Information is only allowed to diffuse along the edges. Each user node u has two states: active (ever tweets or retweets) or inactive (never tweets and retweets). Each edge (u,v) also has two states: retweeting (u ever retweets from v) or no-retweeting (u never retweets from v). The retweeting network $R_{T_i} = (V_{T_i}, E_{T_i})$ of topic T_i is formed by all active nodes and retweeting edges. A real world node is added to G in order to represent the effect of the various media outside the microblogging world (Gruhl et al., 2004). All user nodes follow the real world node with a directed edge labelled with a tweet probability P_u. The real world node and the edges linked to it does not belong to the retweeting network R_{T_i}.

The process of the cascade model is described as follows. Starting with an initial set of user nodes V_0, the process unfolds in discrete steps according to the following rules. At each time step t, real world node tweets with a probability according to the hot of the topic $Hot_{T_i}(t)$. User u can retweet the real world node with a tweet probability P_u, representing the real tweeting behaviour. u can also retweet u's followees with a retweet probability $P_{u,v}$, representing the real retweeting behaviour. u can retweet only once at one time step t. When u retweets the real world node or u's followee v at t, u becomes active and the action is recorded in u's history tweet list. The edge (u,v) turns into the state *retweeting*.

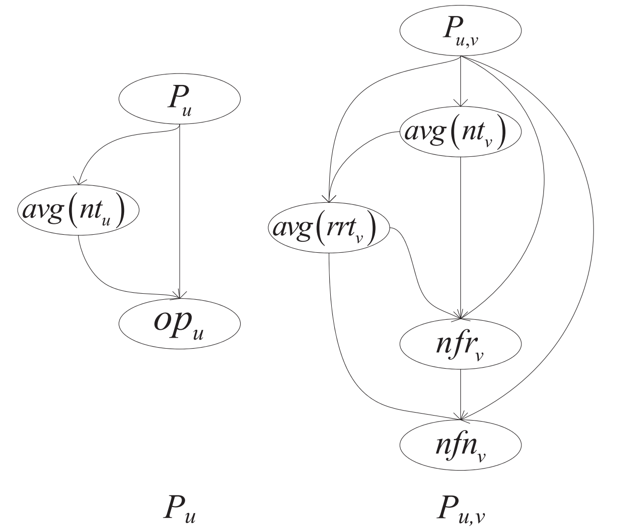

Fig. 14. Bayesian network of the tweet probability P_u and the retweet probability $P_{u,v}$.

5.2. Parameter estimation

The tweet probability P_u and the retweet probability $P_{u,v}$ are defined as follows

$$P_u = P(avg(nt_u), op_u) \tag{4}$$

$$P_{u,v} = P(nfr_v, nfn_v, avg(nt_v), avg(rrt_v)) \tag{5}$$

with $avg(\cdot)$: average number of $\cdot$; $nt_{(\cdot)}$: number of tweet of user $\cdot$; $op_{(\cdot)}$: percent of original tweets of user $\cdot$; $nfr_{(\cdot)}$: number of followers of user $\cdot$; $nfn_{(\cdot)}$: number of friends of user $\cdot$; $rrt_{(\cdot)}$: number of received retweets of user $\cdot$.

Both P_u and $P_{u,v}$ are calculated by Bayesian network. The structures of the Bayesian network are shown in Fig. 14, and the parameterized calculation formulas of P_u and $P_{u,v}$ are shown in Eqs. (6) and (7) respectively. The probabilities P_u and $P_{u,v}$ only depend on the features of the user and are topic-independent:

$$P_u = P(P_u|avg(nt_u), op_u) = P(avg(nt_u)|P_u)P(op_u|P_u) \tag{6}$$

$$\begin{aligned} P_{u,v} &= P(P_{u,v}|avg(nt_v), avg(rrt_v), nfr_v, \\ nfn_v) &= P(avg(nt_v)|P_{u,v})P(avg(rrt_v)|P_{u,v}, avg(nt_v)) \quad *P(nfr_v|P_{u,v}, avg(nt_v), \\ &avg(rrt_v))P(nfn_v|P_{u,v}, avg(rrt_v), nfr_v) \end{aligned} \tag{7}$$

The retweet probability $P_{u,v}$ in (5) is a fixed value, but the real retweet probability will change over time. The probability for u to retweet the message w from v is affected by the time when v tweets w. A tweet appearing at time t will be retweeted at time $t + \Delta t$ with a probability

$$P(t + \Delta t) \propto \Delta t^{-2} \tag{8}$$

It attenuates faster in microblogging network than in blog network (Leskovec et al., 2007). Fig. 15 shows an example of how the retweet probability attenuates over time.

The final retweet probability is as below

$$P_{u,v}(\Delta t) \propto P_{u,v} \times \Delta t^{-2} \tag{9}$$

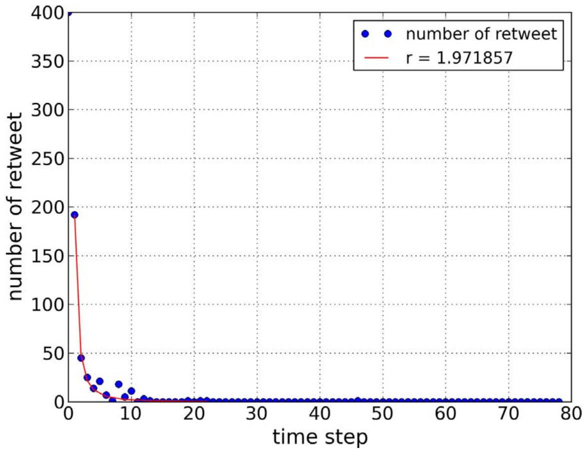

Fig. 15. Retweet probability of a particular tweet changing over time

Table 2
Comparison between the results from real data and from the models.

Methods		Topologies of retweeting-groups		
		N_{max} (%)	N_{single} (%)	N_{other} (%)
Real data		30.37	66.03	3.60
Our model		21.22	76.58	2.20
Independent	β_1	18.70	80.10	1.20
Cascade	β_2	24.70	73.50	1.80
Model	β_3	28.80	66.10	5.10

Table 3
Comparison between the results from real data and from the models.

Methods		Topologies of cascades		
		Star (%)	Chain (%)	Multi-star (%)
Real data		51.60	5.30	43.10
Our model		54.30	5.00	40.70
Independent	β_1	44.40	42.10	13.50
cascade	β_2	66.70	30.80	2.50
model	β_3	45.80	45.80	8.40

Table 4
Comparison between the results from real data and from the models.

Methods		Proportion of attracted users in two situations					
		Groups are merging			Groups have merged		
		σ	μ	a	σ	μ	a
Real data		0.83	−2.12	0.18	0.43	−0.55	0.64
Our model		0.92	−1.89	0.19	0.36	−0.52	0.53
Independent	β_1	\	\	\	0.02	−0.02	0.05
cascade	β_2	1.09	−4.74	0.01	0.03	−0.03	0.06
model	β_3	0.61	−1.40	0.16	0.74	−0.48	0.91

5.3. Validation of the model

In order to prove the rationality and applicability of our model, we compare it to the independent cascade model with a single parameter β (Leskovec et al., 2007). β is calculated by three important user properties respectively. Both our model and the independent cascade model are validated on real data. The results are shown in Tables 2–4.

- *The distribution of users based on the retweeting groups*: Compared to the distribution of real data, the percentage of N_{max} in both our model and the independent cascade model are all smaller. The main reason is that the network rebuilt from the crawled data is a sub-graph of the real network. Because of the large scale of the real network, it is technically impossible for us to crawl all the relationships. Therefore, the g_{max} generated from models has smaller scale than the real g_{max}.
- *The distribution of the cascade topologies*: Notice a good agreement between the distribution of real data and of our cascade model. The independent cascade model has generated too many chains and few multi-center-stars in all three cases of parameter β, which shows a huge difference against the result of real data.
- *Proportion of attracted users in two situations*: The coefficients of the distribution in our model are highly similar to the real data. However, the distributions generated by the independent cascade model are different to the real data and in some cases, the distribution does not fit to a log-normal distribution.

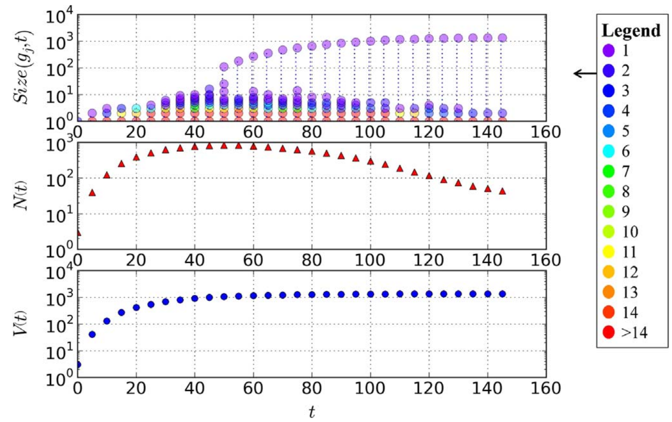

Fig. 16. Evolution of retweeting-groups (model result).

Fig. 16 shows the evolution of the retweeting groups of our model. It is clear to distinguish the three growing periods of g_{max}: the appearance period, the scale-up period and the stabilization period, which have been observed in Section 4.

In summary, Our model matches the important features of the retweeting-groups in real data and outperforms the independent cascade model with a single parameter β.

6. Conclusion

In this paper, we provide a multi-aspect analysis for the information diffusion over microblogging networks. Unlike most previous studies which give general analysis for the diffusion of topics, our study focuses on all diffusion cascades related to a same hot topic, and analyzes the retweeting dynamics in both macro- and micro-level. The multi-level structure (including retweeting network, retweeting group and information cascade) defined in this paper gives the chance to analyze the interaction and dynamic evolution among cascades. Some interesting diffusion features are discovered (i.e. merging effect, super group and centralized topology), which are undetected before to the best of our knowledge.

In previous studies of diffusion models, some common assumptions are given to simplify the model. Our study proves that some of them are not suitable for the information diffusion over microblogging networks. Some traditional diffusion models, such as the independent cascade model and the linear threshold model, assume that each time step in the diffusion process is independent. However, our study in Section 4.3 shows that the trend of the diffusion in the future is apparently influenced by the past diffusion. Meanwhile, the above study also illustrates the inadequacy of using fixed parameters in the diffusion process. Diffusion model with time-varying parameters should be considered in the future related research.

Another interesting result found in our study is that in-degree, as a common feature of most influential users, can indicate user influence simply and effectively. It is inappropriate to underestimate the importance of in-degree.

A diffusion model based on cascade model framework is provided in this paper to generate the retweeting network. The dynamic parameters reflect the dynamics of the evolution process of the retweeting network. The experimental results match the important diffusion features of the retweeting network found in this paper.

For future work, more analysis of the retweeting network are needed. Future research will take more factors in to consideration to improve our diffusion model and provide more accurate prediction. The future work will give us a better understanding of the information diffusion over microblogging social networks.

Acknowledgment

The research presented in this paper is supported in part by the National Natural Science Foundation (61572397, 61502383, 61375040, 61571360), Fundamental Research Project of Natural Science in Shaanxi Province (2015JM6298, 2015JM6299), Fund of Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory (ITD-U15004/KX152600013), China Postdoctoral Science Foundation (2014M562416), The Ministry of Education Innovation Research Team (IRT13035) and Specialized Research Plan Funded Project of Shaanxi Province Department of Education (15JK1505).

References

Ahmed, M., Spagna, S., Huici, F., Niccolini, S., 2013. A peek into the future: Predicting the evolution of popularity in user generated content. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM '13, ACM, New York, NY, USA, pp. 607–616. http://dx.doi.org/10.1145/2433396.2433473

Antoniades, D., Dovrolis, C., 2014. Co-evolutionary dynamics in social networks: A case study of twitter. In: Proceedings of the 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems, SITIS '14, IEEE Computer Society, Washington, DC, USA, pp. 361–368. http://dx.doi.org/10.1109/SITIS.2014.68

Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J., 2011. Everyone's an influencer: Quantifying influence on twitter. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, ACM, New York, NY, USA, pp. 65–74. http://dx.doi.org/10.1145/1935826.1935845

Bao, P., Shen, H.-W., Huang, J., Cheng, X.-Q., 2013. Popularity prediction in microblogging network: A case study on sina weibo. In: Proceedings of the 22nd International Conference on World Wide Web, WWW '13 Companion, ACM, New York, NY, USA, pp. 177–178. http://dx.doi.org/10.1145/2487788.2487877

Bao, P., Shen, H.-W., Chen, W., Cheng, X.-Q., 2013. Cumulative effect in information diffusion: empirical study on a microblogging network. PLoS One 8 (10), 1–7. http://dx.doi.org/10.1371/journal.pone.0076027.

Brown, P.E., Feng J., 2011. Measuring user influence on twitter using modified k-shell decomposition. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.

Cha, M., Haddadi, H., Benevenuto, F., Gummadi, K.P., 2010. Measuring user influence in twitter: The million follower fallacy. In: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, pp. 10–17.

Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J., 2014. Can cascades be predicted?. In: Proceedings of the 23rd International Conference on World Wide Web, WWW '14, ACM, New York, NY, USA, 2014, pp. 925–936. http://dx.doi.org/10.1145/2566486.2567997

Dickens, L., Molloy, I., Lobo, J., Cheng, P.C., Russo, A., 2012. Learning stochastic models of information flow. In: 2012 IEEE 28th International Conference on Data Engineering, 2012, pp. 570–581.http://dx.doi.org/10.1109/ICDE.2012.103

Fan, P., Li, P., Jiang, Z., Li, W., Wang, H., 2011. Measurement and analysis of topology and information propagation on sina-microblog. In: Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on, pp. 396–401. http://dx.doi.org/10.1109/ISI.2011.5984121

Feng, L., Hu, Y., Li, B., Stanley, H.E., Havlin, S., Braunstein, L.A., 2015. Competing for attention in social media under information overload conditions. PLoS One 10 (7), 1–13. http://dx.doi.org/10.1371/journal.pone.0126090.

Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A., 2004. Information diffusion through blogspace. In: Proceedings of the 13th International Conference on World Wide Web, WWW '04, ACM, New York, NY, USA, pp. 491–501. http://dx.doi.org/10.1145/988672.988739

Guan, W., Gao, H., Yang, M., Li, Y., Ma, H., Qian, W., Cao, Z., Yang, X., 2014. Analyzing user behavior of the micro-blogging website sina weibo during hot social events. Physica A: Stat. Mech. Appl. 395, 340–351. http://dx.doi.org/10.1016/j.physa.2013.09.059, URL ⟨http://www.sciencedirect.com/science/article/pii/S0378437113009369⟩.

Han, W., Zhu, X., Zhu, Z., Chen, W., Zheng, W., Lu, J., 2016. A comparative analysis on weibo and twitter. Tsinghua Sci. Technol. 21 (1), 1–16. http://dx.doi.org/10.1109/TST.2016.7399279.

Haralabopoulos, G., Anagnostopoulos, I., 2014. Real time enhanced random sampling of online social networks. J. Netw. Comput. Appl. 41, 126–134. http://dx.doi.org/10.1016/j.jnca.2013.10.016, URL ⟨http://www.sciencedirect.com/science/article/pii/S1084804513002592⟩.

Hethcote, H.W., 2000. The mathematics of infectious diseases. SIAM Review 42 (4), 599–653. http://dx.doi.org/10.1137/S0036144500371907.

Hughes S., 2014. 15 twitter facts and figures for 2014 you need to know, URL ⟨http://www.jeffbullas.com/⟩.

Hutto, C., Yardi, S., Gilbert, E., 2013. A longitudinal study of follow predictors on twitter. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, ACM, New York, NY, USA, pp. 821–830. http://dx.doi.org/10.1145/2470654.2470771

Julia Heidemann, F.P., Klier, Mathias, 2010. Identifying key users in online social networks: A pagerank based approach. In: ICIS 2010 Proceedings, pp. 1–22.

Kempe, D., Kleinberg, J., Tardos, E., 2003. Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03, ACM, New York, NY, USA, pp. 137–146. http://dx.doi.org/10.1145/956750.956769

Kupavskii, A., Ostroumova, L., Umnov, A., Usachev, S., Serdyukov, P., Gusev, G., Kustarev, A., 2012. Prediction of retweet cascade size over time. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, CIKM '12, ACM, New York, NY, USA, pp. 2335–2338.

Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is twitter, a social network or a news media?. In: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, pp. 591–600. http://dx.doi.org/10.1145/1772690.1772751

Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., Hurst, M., Patterns of Cascading Behavior in Large Blog Graphs, pp. 551–556 (Ch. 60). http://dx.doi.org/10.1137/1.9781611972771.60

Leskovec, J., Singh, A., Kleinberg, J., 2006. Patterns of Influence in a Recommendation Network. Springer, Berlin Heidelberg, 380–389. http://dx.doi.org/10.1007/11731139_44.

Letierce, J., Passant, A., Decker, S., Breslin, J.G., 2010. Understanding how twitter is used to spread scientific messages. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line.

Li, H.L., Ng, V.T., 2013. Discovering associations between news and contents in social network sites with the d-miner service framework. J. Netw. Comput. Appl. 36 (6), 1651–1659. http://dx.doi.org/10.1016/j.jnca.2013.04.013, URL ⟨http://www.sciencedirect.com/science/article/pii/S1084804513001240⟩.

Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., Squicciarini, A., Tapia, A.H., 2015. Twitter mining for disaster response: A domain adaptation approach. In: Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management, 2015.

Li, D., Zhang, Y., Xu, Z., Chu, D., Li, S., 2016. Exploiting information diffusion feature for link prediction in sina weibo. Scient. Rep. 6, 20058. http://dx.doi.org/10.1038/srep20058.

Lu, X., Yu, Z., Guo, B., Zhou, X., 2014. Predicting the content dissemination trends by repost behavior modeling in mobile social networks. J. Netw. Comput. Appl. 42, 197–207. http://dx.doi.org/10.1016/j.jnca.2014.01.015, URL ⟨http://www.sciencedirect.com/science/article/pii/S1084804514000599⟩.

Myers, S.A., Leskovec J., 2014. The bursty dynamics of the twitter information network. In: Proceedings of the 23rd International Conference on World Wide Web, WWW '14, ACM, New York, NY, USA, pp. 913–924.

Myers, S.A., Zhu, C., Leskovec, J., 2012. Information diffusion and external influence in networks. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, ACM, New York, NY, USA, pp. 33–41. http://dx.doi.org/10.1145/2339530.2339540

Myers, S.A., Sharma, A., Gupta, P., Lin, J., 2014. Information network or social network?: the structure of the twitter follow graph. In: Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion, ACM, New York, NY, USA, pp. 493–498. http://dx.doi.org/10.1145/2567948.2576939

Pervin, N., Takeda, H., Toriumi, F., 2014. Factors affecting retweetability: an event-centric analysis on twitter. In: ICIS 2014 Proceedings.

Romero, D.M., Galuba, W., Asur, S., Huberman, B.A., 2011. Influence and Passivity in Social Media. Springer, Berlin Heidelberg, 18–33. http://dx.doi.org/10.1007/978-3-642-23808-6_2.

Shuai, X., Ding, Y., Busemeyer, J., Chen, S., Sun, Y., Tang, J., 2012. Modeling indirect influence on twitter. Int. J. Semantic Web Inf. Syst. 8 (4), 20–36.

SinaWeibo, 2015. Sina weibo homepage. URL ⟨http://www.weibo.com⟩.

Twitter, 2015. Twitter homepage. URL ⟨http://www.twitter.com⟩.

Wang, C., Guan, X., Qin, T., Li, W., 2012. Who are active? An in-depth measurement on user activity characteristics in sina microblogging. In: Global Communications Conference (GLOBECOM), 2012 IEEE, Anaheim, CA, USA, pp. 2083–2088. http://dx.doi.org/10.1109/GLOCOM.2012.6503423

Wei, J., Bu, B., Liang, L., 2012. Estimating the diffusion models of crisis information in micro blog. J. Inf. 6 (4), 600–610. http://dx.doi.org/10.1016/j.joi.2012.06.005, URL ⟨http://www.sciencedirect.com/science/article/pii/S175115771200048X⟩.

Weng L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., Schifanella, R., Menczer, F., Flammini, A., 2013. The role of information diffusion in the evolution of social networks. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13, ACM, New York, NY, USA, pp. 356–364. http://dx.doi.org/10.1145/2487575.2487607

Xu, W., Chen, H., 2015. Scalable rumor source detection under independent cascade model in online social networks. In: 2015 11th International Conference on Mobile Ad-hoc and Sensor Networks (MSN), pp. 236–242. http://dx.doi.org/10.1109/MSN.2015.36

Yang, M.-C., Lee, J.-T., Lee, S.-W., Rim, H.-C., 2012. Finding interesting posts in twitter based on retweet graph analysis. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12, ACM, New York, NY, USA, pp. 1073–1074.

Zaman, T., Fox, E.B., Bradlow, E.T., 2014. A bayesian approach for predicting the popularity of tweets. Ann. Appl. Stat. 8 (3), 1583–1611. http://dx.doi.org/10.1214/14-AOAS741.

Zhang, M., Sun, C., Liu, W., 2011. Identifying influential users of micro-blogging services: a dynamic action-based network approach. In: Pacific Asia Conference on Information Systems (PACIS) 2011 Proceedings.