Journal of Universal Computer Science, vol. 22, no. 3 (2016), 360-374
submitted: 15/10/15, accepted: 23/1/16, appeared: 1/3/16

A Dynamic Model of Reposting Information Propagation Based on Empirical Analysis and Markov Process

Gui-Xun Luo
(School of Communication and Information Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
12111007@bjtu.edu.cn)

Yun Liu
(School of Communication and Information Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
liuyun@bjtu.edu.cn)

Zhi-Yuan Zhang
(School of Communication and Information Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
13111005@bjtu.edu.cn)

Abstract: In this paper, based on abundant data from Sina Weibo, we perform a comprehensive and in-depth empirical analysis of repostings and draw some conclusions. First, in regards to quantity, reposting takes up a large proportion of daily microblog activity. Second, the depth of repostings follows an exponential distribution and the first three orders of repostings hold 99 percent of the total amount of reposting, which provides an important foundation for solving the question of Influence Maximization. Third, the time interval for repostings also obeys exponential distribution. Therefore, we have built a dynamic information propagation model in terms of conclusions drawn from Weibo data and the Continuous-Time Markov Process. Due to the basis of the temporal network, our proposed model can change with the time and structure of a network, thus giving it good adaptability and predictability as compared to the traditional information diffusion model. From the final simulation results, our proposed model achieves a good predictive effect.

Keywords: Continuous-Time Markov Process, Reposting, Information Propagation Model
Categories: H.1.0, L.1.0

1 Introduction

In recent years, social network analysis has gained extensive attention from many aspects, including economics, biology, social psychology, physics, etc. One of the most attractive issues of social network analysis is information diffusion, which is a process through which a new idea extensively spreads through communication channels (Rogers & Rogers, 1983). This field has been broadly studied from social psychology, economics, and epidemiology perspectives (Strang & Soule, 1998; Pastor-Satorras & Vespignani, 2002; Kempe; Kleinberg et al., 2003; Huberman;

Leskovec et al., 2005). Over the past decade, the explosion of online social networks, a new field of large-scale information diffusion, was established. In online social networks, information diffusion occurs when a user reads and shares a post of another user, then the shared post is read and shared by other users, and so on (Rodriguez, Leskovec et al., 2012). Compared to traditional social networks, online social networks can provide mass data that includes user relationships, network structure, detail of topics, degree of users, etc. Based on these explicit relationships, the liner threshold model, the independent cascade model (Kempe, Kleinberg et al., 2003), and the general cascade model have been proposed for modelling the information diffusion. The main idea is that nodes may influence one another under certain conditions. On the other hand, Jiwoon et al. studied information dissemination through non-explicit relationships. They found that BlogCast is one of the main causes of information diffusion between bloggers who have no explicit relationships (Ha, Kim et al., 2015).

Some scholars researched information diffusion from the aspect of social network structures in recent years. Bakshy et al. examined the role of social networks online in information diffusion with a large-scale field experiment. They suggested that weak ties may play a more important role in the dissemination of information online than the stronger ties (Bakshy, Rosenn et al., 2012). Li et al. revealed the relationship between the efficiency of information diffusion and the network structures in microblogs in depth. They found that the followers obey a power law distribution with an exponent of near 2 and demonstrated that it can optimize network performance (Li, Qian et al., 2015). Donetti and Hurtado suggested that optimal designs for network mechanisms may generate power-law distribution and scale-free structures (Donetti, Hurtado et al., 2008). Researchers (Viswanathan, Buldyrev et al., 1999; Albert, 2000; Barthélemy, 2003; Lin & Li, 2010) also tried to understand how scale-free structures optimized network performance.

As a representative example of an online social network, the microblog becomes a popular communication tool and plays an important role in information diffusion. In microblogs, short messages of a maximum of 140 characters are posted by users. Then, other users read the message and repost it when the users find it interesting. Apparently, reposting can be considered an efficient way of information propagation when an original post is propagated to other users. Through analysing the Weibo list, we found the reposting takes up a great proportion of overall messages because users are more likely to learn a particular topic by reposting (Yu, Asur et al., 2012).

Research on Weibo reposting has strong practical significance (Pham & Klamma, 2011). Among the Weibo list, the messages with larger amounts of reposts can be called hot topics in microblogging. The hot topics can attract a lot of attention, such as unexpected political events, natural disasters, celebrity scandals, and so on. Hot topics can be seen as a magnifying glass to social problems or as a barometer of social climate. Therefore, deep research on reposting can provide more knowledge about social and public sentiments, and can prevent the public from impact and injury due to sudden and sensitive affairs.

As reposting has become a main information diffusion method, it is significant to explore reposting mechanisms and to know its characteristics. Recently, a lot of work has been done to study reposting behavior (Luo, Liu et al., 2014; Luu & Thomas, 2015). Boyd et al. explored various motivations for reposting on Twitter (Boyd, Golder et al., 2010). Kwak and Park tracked statistics about retweets, and revealed that any retweeted tweet is able to reach a certain quantity no matter how many followers the original tweet had (Lee, Kwak et al., 2010). Suh et al. investigated the relationships between the tweet features and their retweetability (Suh, Hong et al., 2010). Among the content features, they found that URLs and hashtags have strong relationships with retweetability. Naveed et al. (2011) found that general, public topics are more likely to be retweeted than the narrow, personal topics. Further, tweets about bad news seem to spread fast in Twitter.

Based on the function and features of retweeting, some researchers started to build models to predict the likelihood of a tweet being retweeted. Naveed et al. (2011) developed a prediction model based on the influential content features, while Suh (2010) built a predictive retweet model using the Generalized Linear Model. Adopting two data-mining approaches, the Logistic Regression and the Conditional Random Fields, Juarez and Lucas (2012) estimated the reposting of messages. Peng and Zhu (2011) proposed using conditional random fields (CRFs) to predict the retweet patterns. They also investigated partitioning the social graphs and constructed appropriate network relations for improving prediction effectiveness and efficiency. However, most of them depend on a static network structure instead of dynamic information diffusion process over the network, and are thus essentially descriptive models instead of predictive models.

In this paper, we will continue to study information diffusion in online social networks from the respect of reposting. Sina Weibo, as the first and most famous microblog in China, has more than 500 million registered users， which provides plenty of useful data for analyzing information propagation. As the main approach for information diffusion, we will focus our research on reposting from the respect of speed, response time, pathway of reposting, etc. Based on our findings, we propose a dynamic reposting information diffusion model. Additionally, the amount of reposting is the most important form to demonstrate a user's influence. Therefore, our proposed model can effectively solve the open question of influence maximization. In the real world, the process of information is a continuous-time model. However, most models above are discrete-time models, and therefore it is difficult to define an appropriate time step t and predict the information diffusion or user's influence accurately. Due to its dynamic and continuous-time model, our model has a huge advantage in predicting reposts and a user's influence.

The remainder of the paper is structured as follows. In the next section, we carry out an in-depth empirical analysis of reposting on Sina Weibo. In the third section, based on our findings and the Continuous-Time Markov Process, we propose a dynamic reposting information diffusion model. Section four evaluates the prediction power of our proposed dynamic information diffusion model. Finally, we draw various conclusions and note the directions for further studies.

2 Empirical analysis of reposting on online social media

Microblogging services, some of the most popular social media sites on the Internet, have a high influence on the real world. Examples of influential posts include Twitter messages and the U.S. presidential election results (Shamma, Kennedy et al., 2009), and Facebook and the Arab spring (Wolfsfeld, Segev et al., 2013). With the number of users on each site, they can diffuse information very effectively. In this paper, we analyze data from Sina Weibo, which is one of the most powerful microblogging services in China, and focused on the analysis of reposting.

Reposting plays a crucial role in information diffusion. Only followers can read posts written by the people they follow, according to Sina Weibo's privacy rules. If the follower finds the post interesting and reposts, then the follower's follower can read the post. Therefore, the reposted post can be read by even more users in Sino Weibo. The reposting process spreads the information throughout the Weibo user base. Until now, many scholars have done research on reposting and have emphasized its importance in information diffusion (Guo, Chen et al., 2015, Wang, Song et al., 2015). However, most of these studies lack in-depth analysis, especially from the aspect of quantity.

Sina Weibo provides APIs for web developers to acquire basic user and post information. However, the official APIs restricts the amount of user information available and the speed at which the data can be obtained. Therefore, we adopted a method that combines APIs with web crawlers programmed in Python. Finally, we collected data from 2,567,899 users and approximately 9 million posts, from users registered in June 2014. In this paper, firstly we analysed the proportion of reposting in users' post lists. In the Sina Weibo list, the posts roughly can be divided into two categories: what the users themselves posted and what they reposted from others' posts. We calculated the proportion of posts reposted in the user's posts list and drew a statistical result as shown in [Fig. 1] below.

With a reposting proportion of 0.1 to 0.9, the number of users gradually increased. In particular, as the ratio is more than 90%, the user quantity takes on an explosive growth. From this result, we can obviously find that the ratio of reposting is rather high, which illustrates that the reposting activity is very popular and is of great significance to disseminating information. However, in the interval of 0-0.1, there is an uncommon phenomenon. There are two possible explanations: 1) there is so much media on Sina Weibo, users usually directly publish information instead of reposting others' news; and 2) a certain number of zombie fans exist in Sina Weibo, and they have only posted a few times. Another noticeable feature of [Fig. 1] is that the percentage of reposting proportion at about 0.5 has an even growth with the increase of microblogging activity. That illustrates that people who frequently use microblogs have more balanced performance.

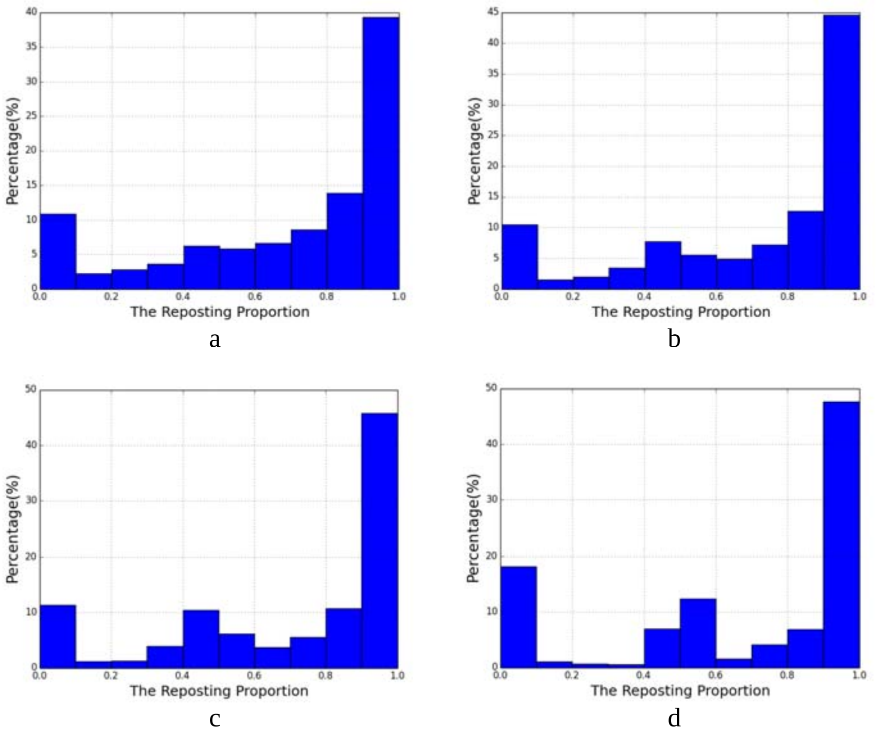

Figure 1: The histogram distribution of reposting proportion. For the four diagrams, the difference is the scope of the statistics. In [*Fig. 1*] *(a), all users' microblog list contains more than 50 items of reposting. By that analogy, the number of microblog list is more than 100，150，300 respectively in (b), (c), and (d).*

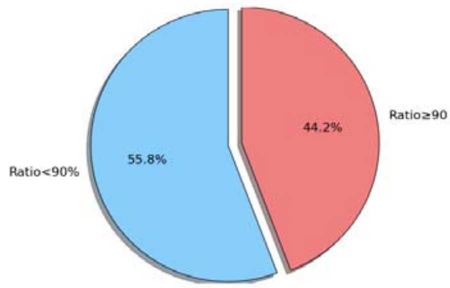

Figure 2: The pie distribution of reposting ratio. The difference of pies is cut-off point: (a) is 50%, (b) is 90%.

More intuitively, we take the ratio of 50% and 90% as a cut-off point for reposting statistics. The number of users whose reposting proportion is more than 50% takes up to 74.1 percent of total Sina Weibo users, while the ratio of users whose activity mostly includes reposting is up to 30.3. The results state that reposting accounts for a large proportion and plays a crucial role in the spread of information. This also provides an effective approach to evaluate a user's influence. When we estimate a user influence, the efficiency of reposting has a more important quantitative index in addition to the amount of his/her followers and friends. The previous research has a lack of evidence to prove the importance of reposting from the aspect of quantity, and our research makes up for it.

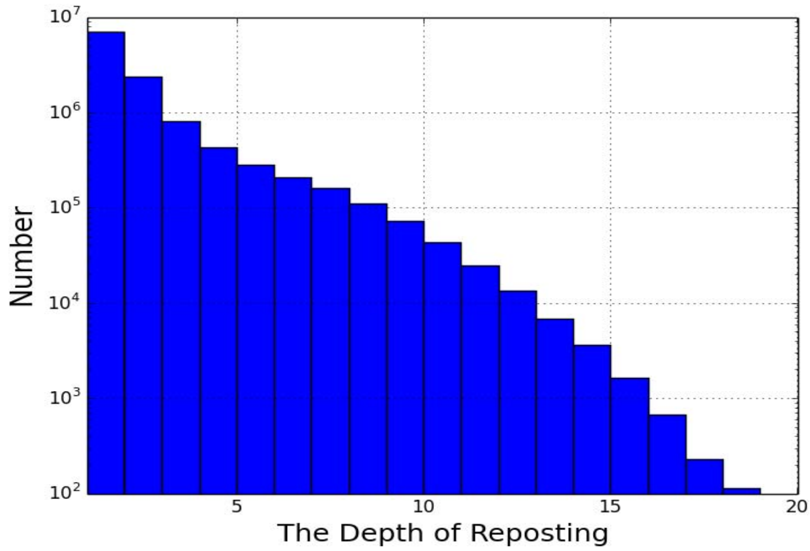

Figure 3: The histogram distribution of the depth of reposting.

When a piece of information is generated on the microblog, it will spread in a complex way. The reposting actually provides the dissemination channels to deliver information to far more than a source's immediate followers. Therefore, knowing the reposting pattern is very important for studies on information diffusion. Many researchers apply graph theory knowledge and describe the reposting pattern as a diffusion tree (Thij, Ouboter et al., 2015). However, studies on each specific Weibo post seem meaningless. We instead focus on an extensive statistical analysis of the reposting pathway. In other words, when a user sees a post, the post goes through how many users reposted.

虽然不知道自己何时会看，但还是 mark 了。。。//@Hikari_老伙计们掩护我好吗：深蹲局最爱了 //@关爱智障儿童成长:这几天少发微博也是正在怒补中@乳不贫何以平天下

Figure 4: An example of microblog post in Sina Weibo and translated as English below: although don't know when will see, still mark it…//@Hikari_老伙计们掩护我好吗：love the SHIELD //@关爱智障儿童成长:post message less to catch up it those days@乳不贫何以平天下

The example post in [Fig. 4] is reposted two times before being seen by the current user. These two times are referred to as the depth of reposting. We have a lot of statistics to address this issue and the result as is shown in [Fig. 3]. From the figure, we find the depth of reposting obeys exponential distribution. It illustrates that for the original post, most of the behavior of reposting takes place by his/her direct followers, which we call first-order reposting. For most of the posts, the number of first, second, and third order reposting takes 99 percent of the total amount of reposts. In light of this finding, when we predict or estimate the quantity of reposting for a specific post, we need only to calculate the number of first, second, and third order reposts.

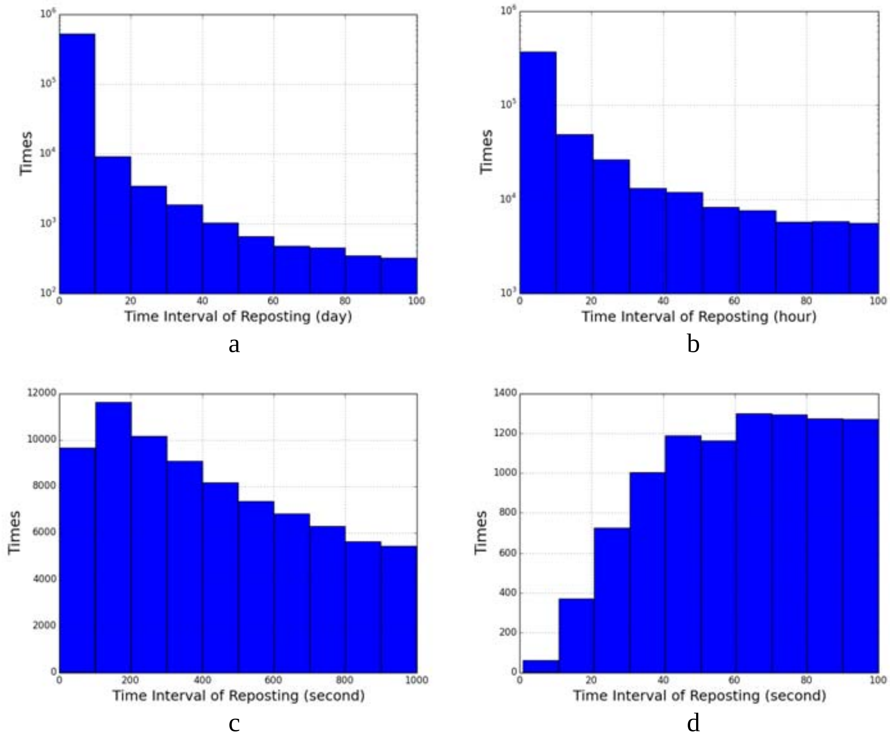

Figure 5: The histogram distribution of time interval of reposting. The statistical time interval for the four diagrams is 100days, 100hours, 1000seconds and 100 seconds.

Next, we focus on the question of time interval of reposting. In general cases, how much time do users spend finding and reposting Weibo posts from when the original post was published? What distribution will the time interval of reposting take on? We calculate the time interval of 20 million posts, and [Fig. 5] shows the statistical results. From [Fig. 5] (a)(b), we know that the time interval of reposting also obeys exponential distribution. [Fig. 5] (a) tells us that most of the reposting takes place within ten days. When we adjust the statistical interval to one hundred hours, the time of reposting still exponentially decreases. However, as the statistical interval is under one thousand seconds, we find that the interval of 100s to 200s has the most number of reposts, which is different from the previous statistic. The combination of [Fig. 5(c)] and [Fig. 5(d)] indicate that people need some time to know the original post, and they repost it when they find it interesting. On the whole, the time interval of reposts in Sina Weibo is subjected to exponential distribution, which can provide a foundation with which to model information diffusion for the next study. In the Markov Process, the residence time before transiting to the next state has a memoryless property and follows exponential distribution.

In this section, we carry on a large-scale empirical analysis of reposting in Sina Weibo. First of all, we reaffirm the importance of reposting as the main approach of information diffusion. It is practically significant to have a good understanding of reposting. Second, we perform quantitative research on the pattern of reposting, which provides a basis for estimation or prediction of user influence. Finally, we take a statistical analysis of the time interval of reposting that approximately obeys exponential distribution.

3 Model

After analyzing the characteristics of reposting, we will build a dynamic reposting information diffusion model. In order to describe reposting behavior more precisely, we establish a temporal reposting network by taking into account continuous time. For an original post posted by a user i, if his/her follower j finds it interesting and reposts it, there will be a link from user i to j, like the red line shown in [Fig. 6]. Using that analogy, if the follower of j reposts the post, a link from j to his/her follower will be generated. A link represents one repost between two users in [Fig. 6]. In [Fig. 6(a)] there is one time of reposting among all the users, where there exists many reposts between users in [Fig. 6(b)]. Therefore, the temporal network can be denoted as G (V, E, T(E)), where V $=\{V_0,V_1,\ldots\ldots V_n$, represents users who posted or reposted a post. $E=\{V_j \rightarrow V_i \mid V_j$ reposted the post of $V_i\}$. So the information flow is along the paths form the followees to followers over time. The function $T(V_j \rightarrow V_i) = \{t^0_{ij}, t^1_{ij}, \ldots\ldots t^n_{ij}\}$, where t^n_{ij} denotes the time-interval of reposting that user j reposted user i's post.

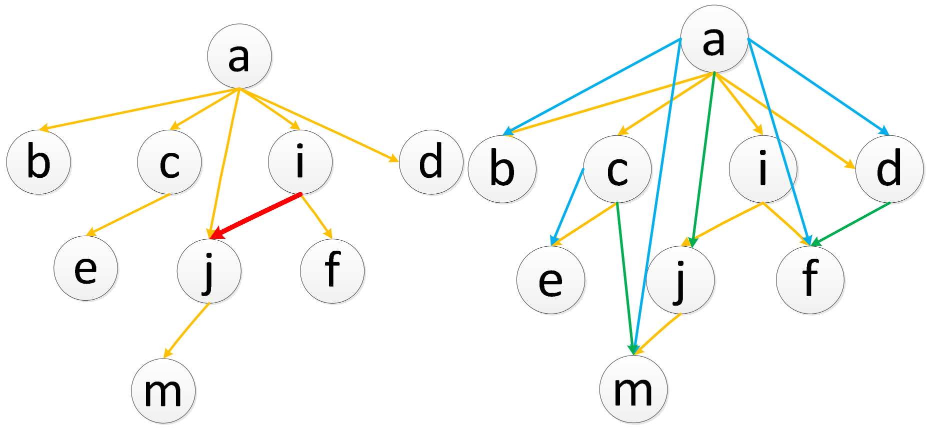

Figure 6: Temporal reposing network. The small circles with a,b,c… represent users in the process of reposting. The links represent reposting relation between users. Left figure denotes one time of reposting, while the right one is state diagram through many times of reposting.

Let {X(t), t>=0} denoting the users who reposting the post at time point x. Because of the exponential distribution of the reposting time interval, X(t) can form a Continuous-Time Markov Process. Based on the Markov property, the user who will repost the post depends only on the present user and is independent of the whole history of the reposting process. That can be expressed as:

$$\begin{aligned} P_{ij}(\mathrm{t}) &= \mathrm{P}\{\mathrm{X}(\mathrm{t}+\gamma) = \mathrm{j} \mid \mathrm{X}(\gamma) = \mathrm{i}, \mathrm{X}(\mathrm{u}) = \mathrm{x}(\mathrm{u}), 0 \le u < \gamma\} \\ &= P\{\mathrm{X}(\mathrm{t}+\gamma) = \mathrm{j} \mid \mathrm{X}(\gamma) = \mathrm{i}\} \end{aligned} \tag{1}$$

where P(t) is the transition probability from i to j within time *t*, i is the current user who posed or reposted a post, and j is the user who will repost the specific post. X(u) is the history of the reposting process before time point γ. We suppose this Markov process of reposting is time-homogeneous, that is, the transition probability P(t) is independent of the starting time of reposting:

$$\begin{aligned} P_{ij}(\mathrm{t}) &= P\{\mathrm{X}(\mathrm{t}+\gamma) = \mathrm{j} \mid \mathrm{X}(\gamma) = \mathrm{i}\} \\ &= P\{\mathrm{X}(\mathrm{t}) = \mathrm{j} \mid \mathrm{X}(0) = \mathrm{i}\} \end{aligned} \tag{2}$$

The theory of continuous-time Markov chains is much more intricate than the theory of discrete-time Markov chains. It is very difficult to calculate directly the transition probability matrix P(t). Therefore, we introduce the transition rate matrix Q, then the P(t) can be estimated form Q.

In probability theory, the transition rate matrix Q, also known as infinitesimal generator matrix, is an array of numbers describing the rate at which a continuous time Markov chain moves between states. We define the transition rate matrix as:

$$Q = \begin{pmatrix} q_{0,0} & q_{0,1} & \cdots \\ q_{1,0} & q_{1,1} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{3}$$

$$q_{ij=} \lim_{\Delta t \to 0} \frac{P\{X_{t+\Delta t} = j \mid X_t = i\}}{\Delta t} = \lim_{\Delta t \to 0} \frac{P_{ij}(\Delta t)}{\Delta t} \quad (i \neq j) \tag{4}$$

as the probability per time unit that the CTMC makes a transition from state *i* to state *j*. In our model, it denotes the transition rate to repost the post of user i by user j.

Define the out-state rate as: $q_i = \sum_{j \neq i} q_{i,j} = -q_{i,i}$ which means when the chain leaves states *i* with rate q_i, it must enter some other state. This model denotes the rate of user *i* in propagating the post to other users.

From [Fig. 5], we know the time interval of reposting is subject to exponential distribution. Thus, we assume the time of user i to propagate a post to the other users follows an exponential distribution without the state rate q_i. According to the property of exponential distribution, the expected value of an exponentially distributed random variable T_i with rate q_i is given by (Feller 1971). $E[T_i] = \frac{1}{q_i}$

According to the actual data of the reposting time interval, the mean of T_i can be calculated.

Based on the construction of the Continuous-Time Markov chain, a user posts a message, then the other user reposts the message in a time interval that follows an exponential distribution. The process that leaves state *i* after an exponentially distributed time with mean $1/\ q_i$ and then jumps to another state *j* can form an embedded Markov chain with a one-step transition probability m_{ij}, and satisfies $\begin{cases} \sum_{j \neq i} m_{i,j} = 1 \\ m_{i,i} = 0 \end{cases}$. Also, we have one important function: $q_{ij=} q_i m_{i,j}$ (Tijms 2003). Given the out-state rate, we can estimate the one-step transition probability from user *i* to user *j* as:

$$m_{i,j} = \sum_c q_i \exp(-q_i t_{ij(c)}) \tag{5}$$

where c is the total number of user *j* reposts posted by user *i*, $t_{ij(c)}$ denotes the time interval from user *i* to user *j* on the *c*th message.

Now, we can calculate the transition rate matrix Q. According to the Kolmogoroff's forward differential equations (Gardiner 1985), when the state space is finite, we can estimate the transition probability matrix P(t) by solving:

$$P'_{i,j}(t) = q_i \times \sum_{i \neq k} P_{ik}(t) \times P_{kj}(t) - q_i \times P_{ij}(t) \tag{6}$$

The above equation can be reformed as the matrix form: $P^{'}(\mathrm{t}) = \mathrm{QP(t)}$. The solution of this system for differential equations is given by:

$$P(\mathrm{t}) = \mathrm{e}^{tQ} = \sum_{n=0}^{\infty} \frac{t^n}{n!} Q^n, \quad t \geq 0 \tag{7}$$

For the large enough n we can use the Taylor expansion, so the P(t) is estimated by:

$$P(\mathrm{t}) = \lim_{n \to \infty} (\mathrm{I} + \mathrm{Qt}/\, \mathrm{n})^n \tag{8}$$

Based on the matrix P(t), we can establish a dynamic information flow network where we find the source, process, and destination of information flow. In contrast to a traditional static network, we can obtain various temporary networks by adjusting the time interval in order to make the network more effective. In this way, we can predict the size of reposting for networks and the precise number of reposts for a user. $p_{\mathrm{ij}}(t)$ describes the relationship of user i and user j. For estimating the influence of user i, the index of $p_{\mathrm{ij}}(t)$ is more important, because reposting means information diffusion. Besides, we can acquire different strength relationships between two users according to different time intervals.

The contribution and innovation of this paper is mainly in three aspects: Firstly, the paper takes a methodical and in-depth research empirical analysis on the behavior of reposting on Sina Weibo. Secondly, our model is proposed based on dynamic networks instead of static networks. Third, the model adopts a Continuous-Time Markov Process to describe the relationships between users.

4 Simulation

The main feature of our proposed model is predictive; therefore, we will test its performance compared with the Autoregressive Integrated Moving Average Model (ARIMA), which is a forecasting technique that projects the future values of a series based entirely on its own inertia. An ARIMA model can be viewed as a "filter" that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts. In our simulation, we use the Weibo data from the last month to train our model and the ARIMA model, and then predict the spreading coverage in the next month. Note that the optimal ARIMA is always selected based on Akaike information criterion (AIC) and Bayesian information criterion (BIC) for comparison. For our proposed model, we can obtain the reposting relationship p_{ij}(t). According to the previous research, we know that most reposting takes place within ten days. Therefore, we only predict the number of reposts for the next 10 days after the original post is posted.

In order to evaluate the prediction performance, we need to calculate one actual value and one predicted value. For the actual value, suppose a user u posts a message m_1 at time t_1, and n users repost the message at time t_2, then we call the actual value of spreading coverage of u from t_1 to t_2 about the message m_1 is n. Through the data statistics, we can acquire it in day 1, day 2 through day 10 by:

$$\textit{The actual value of } u = \sum_{m=1}^{m=M} n_m \tag{9}$$

where M is total number of messages posted by user u. For the predicted value of user u, the calculation is:

$$\text{The predicted value of } u = \sum_{M}\sum_{j=0}^{N} p_{i,\mathrm{j}}(\tau) \tag{10}$$

where N is total number of users who reposted the message, τ is a different reposting interval. For example, user i posted ten messages, based on his/her history of microblogging, utilizing our model to calculate relationship $p_{i,\mathrm{j}}$ between he/she and his/her follower j in order. Then, according to formula 10, we figure the predicted value.

In simulation, we use the same dataset as in Section 2 from Sina Weibo to compare the performance of our proposed model with the ARIMA model. We chose 500 users randomly with two months, and utilized three famous metrics for examining prediction accuracy: MAE (Mean Absolute Error), RMSE (Root Mean Square Error), and MASE (Mean Absolute Scaled Error).

Methods	**MAE**	**RMSE**	**MASE**
Our model	3.95	4.33	1.17
ARIMA	4.83	5.67	1.52

Table 1: The average values of three metrics

As shown as in [Tab. 1], the performance of our proposed model is better than ARIMA, because the ARIMA model fits the overall trend of the time series data and does not consider the underlying cascading network causing the change of the spreading coverage. The simulation results prove that the advantage of our proposed model is obvious, because it is a continuous-time dynamic model based on the temporal reposting network. It comes closer the reality of a reposting network in a specific period of time.

5 Conclusions

In this paper, we take a systematical and in-depth research empirical analysis on the behavior of reposting in Sina Weibo. First, we found that the behavior of reposting takes up a large proportion in daily microblog activity and prove the significance of research on reposting from terms of quantity. Second, we discovered the depth of reposting behavior’s adherence to exponential distribution. Additionally, the first three orders of reposting hold 99 percent of the total amount of reposting. It is important for the question of Influence Maximization. As a matter of quantitative indicators of user influence, we only need to calculate the number of the first three orders of reposting to measure user influence. Thirdly, we know the time interval of reposting also follows exponential distribution. In addition, most of the reposting takes place in a very short time. This provides the foundation for modeling reposting propagation and prediction. In the Continuous-Time Markov Process, the residence time on the state before transiting to the next state has a memoryless property and thus follows exponential distribution.

Based on the empirical analysis and the property of the Markov Process, we propose a dynamic reposting information propagation model. In previous research, most models depended on a static network structure instead of a dynamic information diffusion process over a network, and are thus essentially descriptive models instead of predictive models. In terms of the temporal network, our proposed model can change over time. Therefore, it has good adaptability and predictability. Finally, from the simulation results, our proposed model has good performance compared to the ARIMA model. The next step is to consider more factors of reposting to improve our proposed model and its prediction accuracy. In addition, our proposed model can apply to other microblogging sites, such as Twitter, Facebook, Instagram, and LinkedIn. This paper and our model focus on the aspect of reposting. In other scenes, our model can analyze relationships of friends, comments, and so on. What's more, we hope that this paper will be helpful in the further study of reposting information propagation.

Acknowledgements

This work has been supported by the National Natural Science Foundation of China under Grant 61172072, 61401015, 61271308.

References

[Albert, (2000)] Albert, R., The large-scale organization of metabolic networks. Nature, 2000. 407(6804): p. 651-654 (5 October 2000) | doi :10.1038/35036627.

[Bakshy, E., et al, 2012] Bakshy, E., I. Rosenn, C. Marlow and L. Adamic (2012). The Role of Social Networks in Information Diffusion. Proceedings of the 21st international conference on World Wide Web.

[Barthélemy, 2003] Barthélemy, M., Crossover from scale-free to spatial networks. Epl, 2003. 63(6): p. 915-921.

[Boyd and Lotan, 2010] Boyd, D., S. Golder, and G. Lotan. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. in System Sciences (HICSS), 2010 43rd Hawaii International Conference on. 2010.

[Donetti and Muñoz, 2008] Donetti, L., P.I. Hurtado, and M.A. Muñoz, Network synchronization: optimal and pessimal scale-free topologies. Journal of Physics A Mathematical & Theoretical, 2008. 41(22): p. 4539-4539.

[Feller, 1971] Feller, W., An introduction to probability theory and its applications. Vol. II. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[Gardiner, 1985] Gardiner, C.W., Handbook of Stochastic Methods. Stochastic Processes in Physics & Chemistry Elsevier, 1985.

[Guo et al., 2015] Guo, B., H. Chen, Z. Yu, X. Xie, S. Huangfu and D. Zhang FlierMeet: A Mobile Crowdsensing System for Cross-Space Public Information Reposting, Tagging, and Sharing. IEEE Transactions on Mobile Computing, 2015: p. 1-1.

[Ha et al.,2015] Ha, J., S. W. Kim, S. W. Kim, C. Faloutsos and S. Park An analysis on information diffusion through BlogCast in a blogosphere. Information Sciences, 2015. 290: p. 45–62.

[Huan-Kai and Dongzhen, 2011] Huan-Kai Peng, J.Z., Dongzhen Piao, Rong Yan and Joy Ying Zhang, Retweet Modeling Using Conditional Random Fields. Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference, 2011.

[Huberman, 2005] Huberman, B.A., J. Leskovec, and L.A. Adamic, The Dynamics of Viral Marketing. Acm Transactions on the Web, 2005. 1(1): p. 5.

[Júnior and Almeida, 2012] Júnior, J.P.S., et al. An investigation on repost activity prediction for social media events. in Proceedings of the 13th international conference on Web Information Systems Engineering. 2012.

[Kempe and Kleinberg 2003] Kempe, D., J. Kleinberg, and E. Tardos, Maximizing the Spread of Influence Through a Social Network. Kdd Proceedings of the Ninth Acm Sigkdd International Conference on Knowledge Discovery & Data, 2003: p. 137--146.

[Lee et al.,2010] Lee, C., H. Kwak, H. Park and S. Moon What is Twitter, a social network or a news media? Www Proceedings of International Conference on World Wide Web, 2010: p. 591-600.

[Li et al.,2015] Li, Y., M. Qian, D. Jin, P. Hui and A. V. Vasilakos Revealing the efficiency of information diffusion in online social networks of microblog. Information Sciences, 2015. 293: p. 383-389.

[Lin, 2010] Lin, M. and N. Li, Scale-free network provides an optimal pattern for knowledge transfer. Physica A Statistical Mechanics & Its Applications, 2010. 389(3): p. 473–480.

[Luo et al.,2014] Luo, G. X., Y. Liu, Q. A. Zeng, S. M. Diao and F. Xiong A dynamic evolution model of human opinion as affected by advertising. Physica A Statistical Mechanics & Its Applications, 2014. 414(10): p. 254-262

[Luu and Thomas, 2015] Luu, M.D. and A.C. Thomas, Beyond Mere Following: Mention Network, a Better Alternative for Researching User Interaction and Behavior. 2015: Springer International Publishing. 362-368.

[Naveed et al.,2011] Naveed, N., T. Gottron, J. Kunegis and A. C. Alhadi Bad News Travel Fast: A Content-based Analysis of Interestingness on Twitter. uni, 2011.

[Pastor-Satorras and Vespignani, 2002] Pastor-Satorras, R. and A. Vespignani, Epidemics and immunization in scale-free networks. Bornholdt S & Schuster H G Handbook of Graph & Networks, 2002: p. 111-130.

[Pham and Klamma, 2011]Pham M, Cao Y, Klamma R, et al. A Clustering Approach for Collaborative Filtering Recommendation Using Social Network Analysis[J]. Journal of Universal Computerence, 2011, 17(4):583-604.

[Rodriguez and Leskovec,2012] Rodriguez, M.G., J. Leskovec, and B. Schölkopf. Structure and Dynamics of Information Pathways in Online Media. in Proceedings of the sixth ACM international conference on Web search and data mining. 2012.

[Rogers and Rogers, 1983] Rogers, E.M. and E.M. Rogers, Diffusion of Innovations (3rd ed.). Communication & Development Critical Perspectives, 1983: p. 29–38.

[Shamma and Churchill, 2009] Shamma, D.A., L. Kennedy, and E.F. Churchill. Tweet the debates: Understanding community annotation of uncollected sources. in In WSM ’09: Proceedings of the international workshop on Workshop on Social. 2009.

[Strang and Soule, 1998] Strang, D. and S.A. Soule, Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills. Annual Review of Sociology, 1998. 24(1): p. 265-290.

[Suh and Pirolli, 2010] Suh, B., et al., Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. Proceeding SOCIALCOM '10 Proceedings of the 2010 IEEE Second International Conference on Social Computing, 2010: p. 177-184.

[Thij et al., 2015] Thij, M. T., T. Ouboter, D. Worm, N. Litvak, H. V. D. Berg and S. Bhulai Modelling of trends in Twitter using retweet graph dynamics. Lecture Notes in Computer Science, 2015.

[Tijms, 2003] Tijms, H.C., A First Course in Stochastic Models. John Wiley & Sons Ltd Chichester, 2003: p. x.

[Viswanathan et al., 1999] Viswanathan, G. M., S. V. Buldyrev, S. Havlin, L. M. Da and E. P. Raposo Optimizing the success of random searches. Nature 401: 911–914. Nature, 1999. 401(6756): p. 911-914 (28 October 1999) | doi :10.1038/44831.

[Wang et al., 2015] Wang, X. J., M. Song, S. Z. Guo and Z. L. Yang, Information spreading in correlated microblog reposting network based on directed percolation theory. Acta Physica Sinica, 2015. 64(4): p. 44502-044502.

[Wolfsfeld, 2013) Wolfsfeld, G., E. Segev, and T. Sheafer, Social Media and the Arab Spring. International Journal of Press/politics, 2013. 18: p. 115-137.

[Yu and Huberman, 2012] Yu, L., S. Asur, and B.A. Huberman, Artificial Inflation: The True Story of Trends in Sina Weibo. Computer Science - Computers and Society, 2012.