A New Information Diffusion Model in Weibo: UA-SCIR

Hai-long ZHU$^{1,2,3,*}$ and Jun YANG$^4$

$^1$Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093
$^2$School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049
$^3$National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029
$^4$Department of Automation, Tsinghua University, Beijing 100084

*Corresponding author

Keywords: Information diffusion model, User attribute, Weibo.

Abstract. To illustrate how information spreads in Weibo, this paper proposed a user attributes based diffusion model (UA-SCIR) which defined four node states: susceptible, contacted, infected and refractory. The contacted state was first introduced to describe the nodes who had read the message but had not decided whether reposting or ignoring. Moreover, we recalculated the transfer probability between various states based on user attributes. Experiments on real-world dataset show that our UA-SCIR model can demonstrate the information diffusion progress more accurately than the existing models.

Introduction

In recent years, with the development of mobile internet, many social network services (SNS) had rapidly integrated into and changed people's lifestyles. Sina Weibo was the most important microblog services in China. According to the “2017 Weibo User Development Report” [1], the monthly active users totaled 376 million by September 2017. In addition, “The 40th China Statistical Report on Internet Development” [2] showed that the utilization ratio of Weibo reached 38%, which was the third most important social network service in China. The popularity of Weibo had profoundly changed the way of information diffusion among people. In traditional media, information dissemination was unidirectional transmitted from a small number of central nodes, such as radio and television stations, to millions of users. However, in social networks, each user can publish information, and can also forward what he received. The dissemination was carried out in the form of “nuclear fission”. Recently, many important news had been firstly published and spread in Weibo. Because of the speed and range of information propagation in Weibo, research on the information diffusion model in Weibo played an important role in management and utilization of SNS.

Lots of work had been devoted to information diffusion modeling, of which the epidemic models was representative method. In the epidemic models, the nodes were considered in different states, such as susceptible, infected and refractory. The susceptible-infected (SI) model had two node states, in which the susceptible nodes can be infected by their neighbors with an infection rate, but the infected nodes cannot be recovered. So all nodes in the network would become infected in the end. Unlike the SI model, the susceptible-infected-susceptible (SIS) model [3] allowed the infected nodes being restored as a susceptible node again. Another commonly used was the susceptible-infected-refractory (SIR) model [4], which introduced the refractory status. Infected nodes can be recovered to refractory state with a refractory rate and immune to the disease. All nodes in the network would eventually become refractory. The SIR model was most wildly used in information diffusion modeling.

Many researchers had improved the SIR model, some of which optimized the calculation of the infection rate[5,6,7,8]. Such as LIU et al. [9] redefined the infection rate in the SIR model based on the weak relationship and influence of nodes. ZHANG[10] suggested that the nodes’ degree can be
applied to evaluate the infection rate. Similarly, WANG et al. [11] proposed the RWSIR model, in which the mutual impact function between users (infection rate) was defined by users’ degree. Alternatively, some other researches defined new node states in the model [12,13,14,15]. For example, ZHAO et al. proposed SIHR model[16] by adding a new node status: Hibernated(H). GU[17] and WANG[18] presented two different SEIR models with introduction the latent(E) node from different ways. ZAN et al. disclosed SICR[19] model with Counterattack(C) nodes. XIONG did similar work[20] with the introduction of contact(C) nodes and demonstrated the SCIR model.

Though studies on diffusion models for social network had proliferated in recent years, effort for analysing message propagation in Weibo was lacking. Moreover, little research had been conducted in studying user attributes. This paper presented an improved information diffusion model to illustrate the process of information dissemination in Weibo.

The rest of this paper was organized as follows: Section 2 presented a new information diffusion model: UA-SCIR. In Section 3, the effectiveness of the model was verified through experiments with real-world dataset data. Section 4 summarized the work of this paper and introduced the future direction of research.

The Model

When a user $u_i$ posts a microblog $w$, the fans of $u_i$ could see $w$ with a certain probability after logged in, and the fans can repost or ignore $w$. If a node had ignored the message, he may choose to repost the message again because many of his neighbors had forwarded $w$. Moreover, Weibo was a time-sensitive media. As time goes by, users will lose interest in messages whose spreading would gradually decline. According to the propagation mechanism of Weibo, this paper proposed a new information diffusion model which defined four node states: susceptible, contacted, infected and refractory. Susceptible nodes(S) are the users who had not seen the message, and contacted nodes(C) had seen the message but had not decided whether repost or ignore. The nodes reposted message are defined as infected (I), while the nodes who ignored are called refractory(R). The contacted is temporary state and the other status are final states. The diffusion rules in UA-SCIR as follows:

1. All nodes are initially in S state, When a node posts a message, it turns into I state;
2. S node connected with I could see the message with probability $\alpha$ and transfer to C state;
3. C node repost the message with probability $\beta$ and become I state, while turn into R state with probability $(1-\beta)$;
4. R node connected with I may repost the message and change into I state with probability $\gamma$.

The four node states are illustrated in Figure 1.

Figure 1. The UA-SCIR model.

We defined $S(t), C(t), I(t), R(t)$ as the density of susceptible, contacted, infected and refractory nodes at time $t$, obviously $S(t)+C(t)+I(t)+R(t)=1$. In the initial condition, $S(0)=1$, $C(0)=0$, $R(0)=0$, $I(0)=0$. According to dissemination rules in UA-SCIR, the $S(t)$, $C(t)$, $I(t)$, $R(t)$ changing rate as follows:

$$\begin{align}
\frac{dS(t)}{dt} & = -\alpha S(t) I(t) \\
\frac{dC(t)}{dt} & = \alpha S(t) I(t) - \beta C(t) - (1-\beta) C(t) \\
\frac{dI(t)}{dt} & = \beta C(t) + \gamma R(t) I(t) \\
\frac{dR(t)}{dt} & = (1-\beta) C(t) - \gamma R(t) I(t)
\end{align}$$

(1)
Contacting probability ($\alpha$): We firstly calculated the average contacting probability $p$, which reflects the user’s activity in Weibo. And the contacting probability ($\alpha$) for a node $u$ was related to the user’s activity degree ($\text{Deg}(u)$) and number of followings ($\text{Fol}(u)$). $\text{Deg}_{\text{Avg}}$ and $\text{Fol}_{\text{Avg}}$ were the average values of user activity and followings, respectively.

$$a = p \times \frac{\log(\text{Deg}(u))}{\log(\text{Deg}_{\text{Avg}})} \times \frac{\log(\text{Fol}_{\text{Avg}})}{\log(\text{Fol}(u))} \tag{2}$$

Reposting probability ($\beta$): For the weibo $\langle u_i, u_j, w \rangle$, we found that the probability($\beta$) of $u_j$ forwarding $w$ depended on the hotness ($h$) of $w$, the activity degree of $u_j$, the fans amount($\text{Fen}(u_i)$) of $u_i$ and the average reposting probability $q$.

$$\beta = q \times h \times \frac{\log(\text{Fen}(u_i))}{\log(\text{Fen}_{\text{Avg}})} \times \frac{\log(\text{Deg}(u_j))}{\log(\text{Deg}_{\text{Avg}})} \tag{3}$$

Reposting again probability ($\gamma$): Similarly, the reposting again probability ($\gamma$) was determined by the number of neighbor I nodes ($\text{CountI}(u_i)$) and mean reposting again probability($r$).

$$\gamma = \log(\text{CountI}(u_i)) \times r \tag{4}$$

Aging factor ($\theta$): The aging factor ($\theta$) was computed by hotness ($h$) of message and basic aging factor $s$. The three probabilities will be multiplied by the aging factor in each iteration of the model.

$$\theta = \frac{s}{h} \tag{5}$$

The $p$, $q$, $r$, $s$ were parameters of UA-SCIR model, and $h$ was a parameter of Weibo messages.

Experiments Results

Through several experiments, we analyzed the evolution of various node densities in the UA-SCIR. As showed in Figure2, the susceptible nodes showed a monotonically decreasing trend because they would become contacted nodes with probability $\alpha$, and then transferred into infected nodes or refractory nodes. Infected nodes wouldn’t become refractory so the density of infected nodes was unidirectional rising. As there was a mechanism for refractory nodes to repost again in the model, it can be seen that refractory nodes density slightly decreased in the final stage.

![Figure 2. Nodes Density in UA-SCIR.](image)

To gain more insight, we verified the effectiveness of the UA-SCIR through diffusion process of real-world Weibo messages, and compared the results with SIR and SCIR. We selected two representative messages $w_1$ and $w_2$, both of which were about hot topics. The fans amount of $w_1$ source nodes was small, and which of $w_2$ was big. The Figure 3(a) showed the actual propagation process of $w_1$ and simulations of three models. We can found that the simulation result of the UA-SCIR were most similar to the actual progress. Furthermore, we contrasted the errors between infective nodes predicted in the three models and the actual nodes, and the RMSE as the evaluation index. Compared to the other two models, the error of the UA-SCIR model was greatly reduced. In addition, the error gradually decreased as time goes by.
We conducted the same experiments on Weibo $w_2$ and the results were showed in Figure 4. Compared with $w_1$, we observed that the simulation for $w_2$ in UA-SCIR was more accurate. The reason may be the original node of $w_2$ had more fans and the propagation process was determined by group behavior, so the macro analysis can reduce the influence of individual random behavior.

Conclusion and Future Work

This paper analyzed the process and characteristics of information dissemination in Weibo, and proposed an information diffusion model (UA-SCIR) based on the epidemic models. This model described the fact that information overloading lead many users fail to see all messages pushed to them, introduced the contacted state and optimized the transfer probabilities between different states. In particular, it first demonstrated the mechanism that user changed from refractory to infected state. Moreover, the new model calculated the contacting and reposting probability based on different user attributes, abandoning the unified probability in the existing models. Experiments on real-world dataset showed that the model can describe the process of information diffusion more accurately.

The results of this paper provided some interesting insight for the information spreading in social network. However, in view of the complexity of messages dissemination in Weibo, the research in this paper needs further improvements: First, the activity of Weibo users was periodically changed at different time, the contacting and reposting probability were affected by the wall clock and the UA-SCIR model ignored this factor. Second, it was also necessary to further study the influence of user interest, interaction frequency between users on the reposting probability. Third, some users would repost the message twice or more, the current model has not yet adapted to this situation, and the underlying mechanisms still need further research.

Acknowledgement

This research was financially supported by the National Nature Science Foundation of China (61572496).
References