

A Information Propagation Model Based on Various Emotions and Heterogeneous Mean Field in Social Networks

Tingting Wang¹, Min Hu², Lan Kou³
{ s170101198@stu.cqupt.edu.cn¹, humin@cqupt.edu.cn², koulau@cqupt.edu.cn }

School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications^{1,2,3}

Abstract. Information propagation in social networks can be effected by various factors, such as connected relationships, interactions between users, and so on. Previous studies mainly focused on analyzing the impact of physical connections on information propagation, and rarely studied the effects of different emotions on information propagation. Aiming to solve this problem, this paper proposes an emotion-based susceptible-infected-recovered information propagation model(E-SIR). The model primarily researches the impacts of different emotions and diverse connections on the information propagation process. We introduced the emotional transmissibility and the information transmissibility to describe the infection abilities of different emotions and the possibility of information transmission between different users, respectively. In addition, the dynamic equations are established based on the heterogeneous mean field. The burst threshold and spreading scale are theoretically analyzed and verified by experiments. In general, this paper focuses on the impacts of various emotions and different connections on information propagation.

Keywords: Information propagation, emotions, connected relationships, social networks

1 Introduction

With the rapid development of Internet technologies, online social medias, such as Twitter, Facebook and Sina Weibo (like Twitter), have emerged and quickly integrated into people's daily life. It has profoundly changed the ways of people communication and of information transmitted[1]. Nowadays, people customarily share the information with friends, colleagues, or people owning common interests in online social networks. The contents of the information sharing involve all aspects of life, politics, and emotions and so on. As a results, a large amount of data are available for collection and analysis, which brings new opportunities for the study of information propagation in complex networks[2]. Meanwhile, the researches on information dissemination in social networks has great significance for public opinion control, viral marketing, and finding influence nodes[3]. Therefore, it have been an important topic for scholars to study in recent years.

Due to the similarities between information diffusion and epidemic propagation, scholars introduce epidemic propagation model into the researches of information diffusion and establish

the information diffusion models according to epidemic disease propagation models[4,5]. And the differences between information propagation and epidemiology have been mainly considered, such as network structures [6], the influences between individual nodes [7], etc., to model and analyze the information dissemination process.

However, these are mainly considered the topology of the networks, that is, the physical connections, ignoring the intrinsic properties of users in the social networks. Previous studies have shown that emotions are key factors affecting human behaviors and play an essential role in people's daily decision-making[8]. Most studies on the dissemination of emotional information merely considered the transmission of positive and negative information[9] but rarely considered the fine-classification of emotions. Because the classification of emotions is too simple, previous studies cannot correlate and compare different emotional information. Until recent years, researchers began to finely classify emotions and study the effects of different emotions on information dissemination [10,11]. But those ignored the connected relationships of the nodes and the influence of homogeneity [12] on the propagation of information diffusion. Aiming to fit this gap, we consider the impacts of homogeneity between users. Homogeneity is not only reflected in the characteristics of demographics, such as age, race, hometown, common friends and hobbies, but also in the states of mind, such as happiness [13]. Meanwhile, the heterogeneity of the networks also affects the dissemination of information. The users for various relationships establish different connections on social networks, thus forming various communities. Previous studies have shown that nodes in the same community spread information more frequently, and in different communities spread information rarely, or never[6]. Therefore, disseminating information in the same community is easier than in different communities. Hence, this paper combines the connections and the emotional characteristics of users to model and analyze the information propagation process, and studies the impacts of different emotions and various connected relations on the information propagation process.

The main contributions of this article are summarized as follows.

1. An information propagation model is proposed, which is based on various emotions and a heterogeneous mean field. This model considers different emotional features of users and the heterogeneity of real social networks. It can meticulously depict the dynamic process of information dissemination.
2. Homogeneity is considered that the more similar the users are, the greater the probability of contacts between users. Based on this, this paper proposes emotional transmissibility, and judges the similarity of user's historical emotions. The higher the historical sentiment similarity, the greater the probability of information dissemination will be.
3. The information transmissibility is used to calculate the possibility of information to be transmitted between two nodes. The heterogeneity of the network structures makes the information transmissibility in the same community and between communities different. The heterogeneity of the network structures leads to the two different information transmissibilities in the same community and between communities. It can be seen that the network structures are also key factors affecting information dissemination.

The rest of this paper is organized as follows. The second section introduces the problem statement and related definitions. The information propagation model proposed in this paper is introduced in detail in the third section. In the fourth section, the results of the paper are analyzed and discussed. The fifth section summarizes the main work and research results of this paper.

2 Preliminaries and problem statement

In this section, some definitions and problem statement will be introduced. We use a graph $G=(V,E)$ to represent an online social networks, where $V=\{V_1,V_2,\dots,V_n\}$ is the user set, E represent the user's connection relationships.

Definition 1: Emotion represents the subjective attitude of the user. This article classfyemotion into five categories, and $x=\{0=\text{no emotion},1=\text{anger},2=\text{joy},3=\text{disgust},4=\text{sadness}\}$ represents various emotions.

Definition 2: Emotional transmissibility $S_{u,v}^x$ represents the ability of the emotion x to spread between the user u and the user v . Psychology shows that the spread of different emotional information has different propagation features. J. Berger's research shows that emotions with high arousal (such as anxiety and anger) can promote information dissemination more than emotions with low arousal (such as sadness) [14]. Therefore, the abilities to spread various emotions may be different. According to emotional consistency and homogeneity, the user's emotions are consistent to the same event before and after, and the more similar the user's historical emotions, the more likely they are to generate emotional infections. Based on this, this paper uses the correlation of historical emotions between two users to measure the possibility of an emotion spreading between two individuals. Its defined as follows:

$$S_{u,v}^x = sim(x_u, x_v) = \frac{2 \log[\min_{i \in x_u \cap x_v} f(i)]}{\log[\min_{i \in x_u} f(i)] + \log[\min_{i \in x_v} f(i)]} \quad (1)$$

where x_u and x_v represents the emotions of the user u and v respectively. i is a certain kind of emotions, and $f(i)$ is the proportion of a certain kind of emotion i .

Definition 3: Information transmissibility. ε denotes the information transmissibility. There are various connections between users in real life, which leads to the formation of communities in social networks. If users belong to different communities or same communities, the information transmissibilities would be vary. Users in the same community are more likely to diffuse than those users who are in different communities. Information transmissibility in same community are expressed in symbol of ε_{in} , and information transmissibility between different communities are expressed in symbol of ε_{out} .

Problem Statement As shown in Figure 1, there are two infected nodes (information spreader), one recovered node (It no more propagate information), and the rest of nodes are susceptible nodes(Those nodes turn into infected nodes with a possibility). After a period of information propagation, the susceptible nodes may become a infected nodes, and the infected nodes may become an recovered nodes. The main purpose of this paper is to construct an information diffusion model based on the emotional transmissibility and information transmissibility between users, and predict the variation trend of the three types of nodes in the process. At the same time, we will analyze the impacts of two factors on the information epidemic threshold and the outbreak scale.

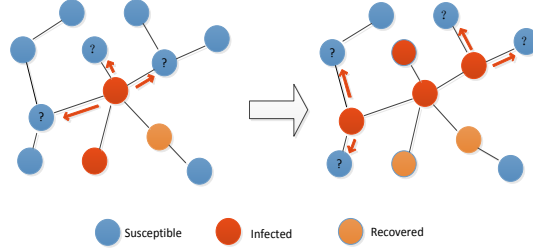


Fig. 1. Propagate process

3 Emotion-based suspicious-infected-recover model

This paper proposes an emotion-based susceptible-infected-recovered information propagation (E-SIR) model to simulate the process of information dissemination. We consider emotional transmissibility and information transmissibility into the model to predict the propagation process. The SIR (susceptible-infected-recovered) epidemic propagation model divides nodes into three categories, susceptible nodes, infected nodes, and recovered nodes. The susceptible node indicates a node that has not received the information but may become the information spreader after receiving the information, and are represented by the symbol S , $S \subseteq V$. Infected nodes represent those nodes that receive information and propagate. They are denoted by the symbol I , $I \subseteq V$. Recovered nodes are those that receive information but are not willing to propagate, and are represented by the symbol R , $R \subseteq V$, $|S| + |I| + |R| = N$.

Traditional epidemiological models assume that all types of nodes are evenly mixed, but this is not inconsistent with the real social networks. In reality, the complex social networks are generally heterogeneous, that is, various types of nodes are not evenly distributed. In order to make our model more realistic, this paper will use the heterogeneous mean field theory to construct the equation of transformation between different types nodes. According to the degree of different nodes, the mean field theory expresses the density of various nodes as $S_k(t)$, $I_k(t)$, and $R_k(t)$. $S_k(t)$, $I_k(t)$, and $R_k(t)$ represents the proportion of susceptible, infected, and recovered nodes, $S_k(t) + I_k(t) + R_k(t) = 1$. According to [15], the heterogeneous mean field equation can be written as:

$$\begin{cases} \frac{dS_k(t)}{dt} = -k\beta S_k(t)\theta_k(t) \\ \frac{dI_k(t)}{dt} = k\beta S_k(t)\theta_k(t) - \gamma I_k(t) \\ \frac{dR_k(t)}{dt} = \gamma I_k(t) \end{cases} \quad (2)$$

Where, β is the probability that the susceptible nodes be infected after encountering the infected nodes, and β_c is the epidemic threshold. When $\beta \geq \beta_c$, the infected nodes in the networks will grow exponentially, which means that information will spread widely in the

networks. Conversely, when $\beta \leq \beta_c$, the information will not propagate through the networks. γ is the probability that an infected nodes will become a recovered nodes after a period of time, and $\theta_k(t)$ is the probability that a susceptible nodes is connected to an infected nodes.

In this paper, the emotional transmissibility between users is introduced into the basic model as the weight that impacts information dissemination. As shown in Figure 2[a], after encountering an infected node, the susceptible node will become an infected node after a period of time with the probability β . However, when we consider the influence of user's emotion on information dissemination as weight, the probability of propagation should be expressed as $\beta S_{u,v}^x$, which represents the probability of propagation of x -type emotional information between nodes V_u and V_v . In addition, the information propagation probability is impacted by the user's connected relations in the networks. Therefore, according to the connected relations, the probability of information propagation can be expressed as $\varepsilon\beta$, where the value of ε is determined according to the community in which the nodes are located. They are in the same community, then $\varepsilon = \varepsilon_{in}$, otherwise $\varepsilon = \varepsilon_{out}$. The process is shown in Figure 2[b]. As shown in Figure 2[c], infected nodes become recovered nodes with probability γ after a period of time.

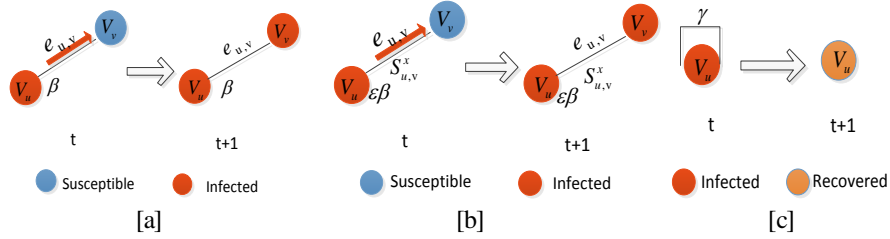


Fig. 2. Nodes change process([a]Regardless of other factors [b] Considering emotions and connections [c]change process of infected nodes)

After considering the above factors comprehensively, the equations can be written as:

$$\frac{dS_k(t)}{dt} = -kS_{u,v}^x \varepsilon \beta S_k(t) \Theta(t) \quad (3)$$

$$\frac{dI_k(t)}{dt} = k\beta S_{u,v}^x \varepsilon S_k(t) \Theta(t) - k\gamma I_k(t) \quad (4)$$

$$\frac{dR_k(t)}{dt} = k\gamma I_k(t) \quad (5)$$

Where, $\Theta(t) = \sum_n P(n|k) I_n(t)$, $P(n|k) = \frac{nP(n)}{\langle k \rangle}$, $P(n)$ is the degree distribution, $\langle k \rangle$ is

the average degree.

The meaning of equation (3) is that the S state nodes be infected by connected I state nodes with the probability of $S_{u,v}^x \varepsilon \beta$ at time t . That is, the density of susceptible nodes decreases at

a rate proportional to the probability of $S_{u,v}^x \varepsilon \beta$. Equation (4) represents the rate of change of the I state nodes at time t , which includes both nodes added from the S state and subtract the nodes that turns into R state with the probability γ . Similarly, equation (5) represents the rate of change of the R state at time t , and the R state is only transformed from the I state, therefore, it is only one term.

In the initial time, there are no nodes of the S state and the R state in the networks, so the initial density of the I state is 1. We calculated from (3):

$$S_k(t) = e^{-\int_0^t k S_{u,v}^x \varepsilon \beta \Theta(t') dt'} = e^{-k S_{u,v}^x \varepsilon \beta \Phi(t)} \quad (6)$$

Where

$$\Phi(t) = \int_0^t \Theta(t') dt' = \frac{\sum_n n P(n) \int_0^t I_n(t') dt'}{\langle k \rangle} \quad (7)$$

It can be calculated by equation (5), $R(t) = \gamma \int_0^t I(t') dt'$. Therefore, the expression of $\Phi(t)$ can be rewritten as:

$$\Phi(t) = \int_0^t \Theta(t') dt' = \frac{\sum_n n P(n) \frac{R_n(t)}{\gamma}}{\langle k \rangle} \quad (8)$$

The physical meaning of $\Phi(t)$ is the probability that the other end of any sides of the networks is the R state nodes at time t . When the dynamics in the networks reaches steady state, whether there is global outbreak of information can be clearly analyzed by $\Phi(t)$.

In order to facilitate the theoretical calculation, we assume that the time is continuous. According to equation (7), we can get an expression of the density of infected nodes. It can be expressed as follows:

$$\begin{aligned} \frac{d\Phi(t)}{dt} &= \frac{\sum_n n P(n) I_n(t)}{\langle k \rangle} = \frac{\sum_n n P(n) (1 - R_n(t) - S_n(t))}{\langle k \rangle} \\ &= 1 - \gamma \Phi(t) - \frac{\sum_n n P(n) e^{-k S_{u,v}^x \varepsilon \beta \Phi(t)}}{\langle k \rangle} \end{aligned} \quad (9)$$

When the network dynamics reaches steady state, $t \rightarrow \infty$, there are no infected nodes in the networks, $I_\infty \rightarrow 0$, $\frac{d\Phi(t)}{dt} = 0$. According to formula (9), it can be calculated:

$$0 = 1 - \gamma \Phi_\infty - \frac{\sum_n n P(n) e^{-k S_{u,v}^x \varepsilon \beta \Phi_\infty}}{\langle k \rangle} \quad (10)$$

$\Phi_\infty = 0$ is a constant solution to this equation. Construct a function $f(\Phi_\infty) = 1 - \gamma \Phi_\infty - \frac{\sum_n n P(n) e^{-k S_{u,v}^x \varepsilon \beta \Phi_\infty}}{\langle k \rangle}$. To have a non-zero solution, due to $f(\frac{1}{\gamma}) < 0$, $0 < \gamma < 1$, must have $\frac{df(\Phi_\infty)}{d\Phi_\infty} |_{\Phi_\infty=0} > 0$, which is $\frac{\sum_n n P(n) \langle k \rangle S_{u,v}^x \varepsilon \beta}{\langle k \rangle} - \gamma > 0 \Leftrightarrow \frac{\langle k^2 \rangle S_{u,v}^x \varepsilon \beta}{\langle k \rangle} > \gamma$. It can be obtained that when $\beta > \frac{\gamma \langle k \rangle}{S_{u,v}^x \varepsilon \langle k^2 \rangle}$, Φ_∞ has a non-zero solution, so the burst threshold equation can be expressed as:

$$\beta_c = \frac{\gamma \langle k \rangle}{S_{u,v}^x \varepsilon \langle k^2 \rangle} \quad (11)$$

Since the propagation reaches steady state, there will only be two states of S and R , that is, when $t \rightarrow \infty$, $I_\infty \rightarrow 0$; R_f is the final infection density, which is the final scale of information diffusion.

$$R_f = \sum_k P(k) R_k(\infty) = \sum_k P(k) (1 - S_k(\infty)) = \sum_k P(k) (1 - e^{-k S_{u,v}^x \varepsilon \beta \Phi_\infty}) \quad (12)$$

The equation (12) carry out Taylor expansion to Φ_∞ :

$$R_f \approx S_{u,v}^x \varepsilon \beta \Phi_\infty \sum_k P(k) k \quad (13)$$

A second-order Taylor expansion for equation (10) can be written as:

$$\gamma \Phi_\infty = 1 - \frac{\sum_n n P(n) \left(\frac{(-k S_{u,v}^x \varepsilon \beta \Phi_\infty)^0}{0!} + \frac{(-k S_{u,v}^x \varepsilon \beta \Phi_\infty)^1}{1!} + \frac{(-k S_{u,v}^x \varepsilon \beta \Phi_\infty)^2}{2!} + o(\Phi_\infty) \right)}{\langle k \rangle}$$

Therefore,

$$\Phi_\infty = \frac{2(\langle k^2 \rangle S_{u,v}^x \varepsilon \beta - \gamma \langle k \rangle)}{\langle k^3 \rangle (S_{u,v}^x \varepsilon \beta)^2} \quad (14)$$

According to the equation(13) and equation(14),it can be calculated:

$$R_f \approx \frac{2(\langle k^2 \rangle S_{u,v}^x \varepsilon \beta - \gamma \langle k \rangle)}{\langle k^3 \rangle S_{u,v}^x \varepsilon \beta} \langle k \rangle \quad (15)$$

From equation(11) and equation(13) we can conclude that both emotions and network structures have the impacts on the burst threshold and the scale of the information spreading.

The formula (11) shows that the stronger the emotional transmissibility $S_{u,v}^x \uparrow$, the smaller the threshold $\beta_c \downarrow$, and the more likely the information erupt. The greater the information transmissibility $\varepsilon \uparrow$, the smaller the threshold $\beta_c \downarrow$ and the more likely the information will break out. This shows that if the influence of emotions is high, it will spread more easily, and the more closely related information will spread more easily. Equation (15) shows that the stronger the emotional transmissibility $S_{u,v}^x \uparrow$, the greater the scale of information propagation $R_f \uparrow$. In the same way, the greater the information transmissibility $\varepsilon \uparrow$, the greater the scope of communication will be $R_f \uparrow$. From this, we can analyze the influence of emotional factors and network structures on the scale of information dissemination.

Table 1. Symbolic representation

Symbol	Interpretation
β	Probability of a susceptible nodes turning into an infected nodes
$S_{u,v}^x$	Emotional transmissibility emotion x between user u and user v
ε	Information transmissibility
γ	The probability of the infected nodes turning into the recovered nodes
R_f	The scale of the final spreading
β_c	Propagation threshold

4 Result analysis

4.1 Experimental data

This paper verifies the model of this paper in artificial network and real network respectively. The artificial networks are a small world networks and a Barabási–Albert(BA) scale-free networks generated by python's networkx toolkit, and both networks are 5,000 nodes. The small world networks has 10,000 edges, and the average degree is 4. The BA networks have 10,000 edges, and the average degree is 5.996. About real networks, this article uses the data set taken from Sina Weibo (like twitter), and the data can obtain from <https://doi.org/10.6084/m9.figshare.4311920.v2>. The time is range from September 14 to March 2015. The number of nodes are 88532, and the edges are 7490687. The average degree of the social networks is 169.23. Those degree distribution are shown in Figure 3, the horizontal axis represents the degree of the nodes, that is, the number of nodes connected to other nodes. The greater the degree, the greater the influence of the nodes. The vertical axis represents the occurrence probability of nodes with different degrees, and the greater the probabilities, the more the number of the nodes with different degrees. From Figure 3[a], we can see that the degree distribution of the small world networks obeys the normal distribution, which is obviously not consistent with the reality. The BA networks obeys the power law distribution, although this is not very realistic, but it is more realistic than the small world network. Therefore, the verification discussion in this paper is mainly carried out on the BA network. It can be seen from Figure 3[b] that although the degree distribution of Sina Weibo is not strictly consistent

with the power-law distribution, the overall trend is similar and has a significant long tail effect. In other words, the nodes with great influences are generally a few, and the influences of most nodes are not large..

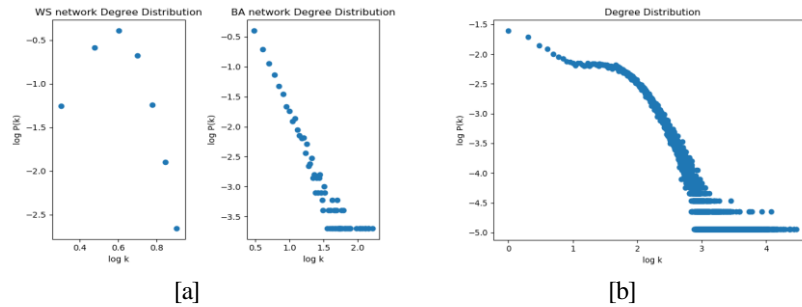


Fig. 3. Degree distribution [a] is the artificial network degree distribution, the left side is the small world network degree distribution, the right is the BA network degree distribution, and the figure [b] is the Sina microblog degree distribution. (In the figure, the horizontal axis is the degree of each nodes, and the vertical axis represents the occurrence probability of different degree, and those curves represent the distribution of different degrees. By this graph, the occurrence probability of different influence nodes can be visually reflected.)

4.2 Analysis of information dissemination process

The initial state has only one infected node in the networks, and the rest are susceptible nodes. At this time, the influence of emotions and connected relations on information dissemination is not considered. We set the initial infection probability to 0.2 and the recovery probability to 0.01. The analysis results of the trend of the three types of nodes, such as susceptible nodes, infected nodes and recovered nodes, in the social networks are shown in Figure 4. The horizontal axis represents time step and the vertical axis represents the proportion of the three types types of nodes. From the figure we can see that the number of initial recovered nodes is zero. As the infected node starts to spread the message, the susceptible nodes begin to decrease, and more nodes become infected nodes to start spreading messages. As the number of nodes propagating messages increasing, recovered nodes also begin to increase. When the infected nodes reaches the peak, the susceptible nodes begins to approach zero, and the infected nodes begins to decrease, meanwhile, more recovered nodes appear. when the infected nodes approaches zero, the recovered nodes tends to be stable, and the information propagation process ends.

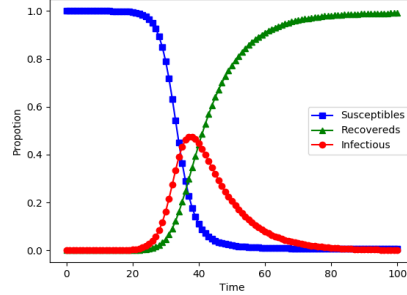


Fig. 4. Information dissemination process (The horizontal axis is the time step, the vertical axis is the proportion of the three types of nodes, and the three curves represent the changes of the three types of nodes in the propagation process.)

4.3 Relationships between emotions and information dissemination

Psychology believes that different emotions have different influences on information diffusion. In article [16], it is confirmed that the correlation of different emotions is indeed different, among which the correlation of anger is the highest and the correlation of sadness is the lowest. In this paper, the Bayesian classifier is used to classify the content of Twitter, which is divided into four categories: anger, joy, disgust and sadness. The correlation of emotions between users is calculated by using the Pearson coefficient. As described in the second section, the user's emotions are represented by x . The proportion of each type of emotion in these four types of emotions are represented by f_x . For one of the emotions x ($x=0,1,2,3,4$) and a pair of users (u, v) with information communication relationships, we put the information spreader u 's f_x^u into the sequence O_x . Meanwhile, the recipient v 's f_x^v is put into the sequence T_x . At this time, the Pearson correlation coefficient of a certain kind of emotion between users can be expressed as $C_p^x = \frac{1}{l-1} \sum_{i=1}^l \left(\frac{O_i - \langle O_x \rangle}{\sigma_{O_x}} \right) \left(\frac{T_i - \langle T_m \rangle}{\sigma_{T_m}} \right)$, where l represents the sequence length, $\langle O_x \rangle = \frac{1}{l} \sum_{i=1}^l O_i$ represents the mean, and $\sigma_{O_x} = \sqrt{\frac{1}{l-1} \sum_{i=1}^l (O_i - \langle O_x \rangle)^2}$ is the standard deviation. The correlations calculated as anger, happiness, disgust, and sadness were 0.41, 0.35, 0.04, and 0.03, respectively.

This will also be introduced into the information dissemination model of this paper as key parameters. Firstly, we consider only the impacts of emotional factors on information dissemination. As shown in Figure 5[a], the horizontal axis represents the time step and the vertical axis of the graph represents the proportion of the infected nodes, which are the information spreaders. Meanwhile, the different curves represent the changes of the information spreaders with different emotions. In Figure 5[b], the horizontal axis represents the time step and the vertical axis of the graph represents the proportion of the recovered nodes, which do not transmit information again. The curves in the graph represent the changes of recovered nodes with different emotions. From the Figure 5 we can see that except for the speed and the spreading scale of sadness are both similar to the situation without considering emotions, and the rest of

the emotions will promote the spread of information. It can be seen that emotions do play an important role in the process of information dissemination. From Figure 5[a] we can see that happiness spreads fastest and can affected more nodes in the shortest time than other emotions, and the second is anger. This also verifies that the high-valence emotions in psychology are more easily transmitted than low-valence. The disgust spreads faster than sadness, and the speed of sadness spreads and the number of people who communicate is similar to the situation without emotions. Thus it can be seen that in social networks, sadness has little effect on information dissemination. It can be seen from Figure 5[b] that the spread of happiness and anger is very close, and happiness spreads a little wider than anger. The spreading scale of disgust is broader than sadness or no emotions, and the spreading scale of sadness is similar to no emotions. As shown in Figure5[b], anger and happiness have the greatest impact on the burst threshold. Disgust and sadness have little effect on the burst threshold of the propagation, but the extent of disgust to the final spread is still wider than sadness.

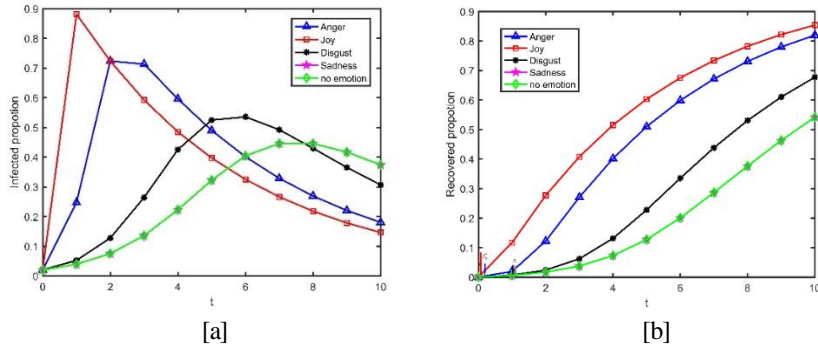


Fig. 5. The spread of different emotions(The horizontal axis of both graphs represents the time step. The vertical axis of the graph [a] represents the proportion of the infected nodes, which are the proportion of information spreaders, and the different curves represent the changes of the information spreaders with different emotions. The vertical axis of the graph [b] represents the proportion of the recovered nodes, which do not to transmit information again. The curves in the graph represent the changes of recovered nodes with different emotions. The curves can be used to analyze the range and the burst threshold of information dissemination.)

4.4 Impacts of connected relationships on information dissemination

This paper considers the impacts of network heterogeneity on information dissemination from the perspective of community structure. The GN algorithm is used to classify the community by the classical network structures division method, and statistical analysis is carried out information diffusion probability on the internal and external of the community. This paper considers the impact of retweet on information dissemination. By linear regression, the probability of occurrence of retweet behavior in the same community is 0.35, and the probability of occurrence between different communities is 0.13[17]. They are also key parameters affecting the propagation model proposed in this paper. The information dissemination in same community or not is analyzed according to the probability that the retweet behavior occurs in same community or not. As shown in Figure 6[a], the horizontal axis represents time step and the vertical axis represents the proportion of infected nodes in the social networks, and the information dissemination in the same community is faster and spreads wider than between

different community. However, the time of disappearing is also faster, which is in line with the psychological characteristics of people more interested in new things. Users in the same community are connected by certain common characteristics, such as hobbies and friendships. As a result, there may be a lot of interest in a certain type of information, which leads to the phenomenon that information can be quickly transmitted in the community. However, since people are more interested in new things, when they see that everyone is spreading a certain message, the enthusiasm for communication will be diminished, which is the case that the number of infected people has dropped rapidly. For users in different communities, because everyone's contact is not so close, the number of people who spread information will not suddenly increase, but also decrease slowly. Figure 6[b] can illustrate the phenomena mentioned above. In addition, we can see that the coverage of the same community information will be wider than that of different communities. From this, we can see the impacts of the connected relationships on information dissemination. The connection is closer, the speed of information dissemination is faster and the scale is wider, but the faster it disappears. The connection is relatively not intimate, and the speed of information spreading and disappearing is slower than previous situation, and the range of transmission is relatively limited.

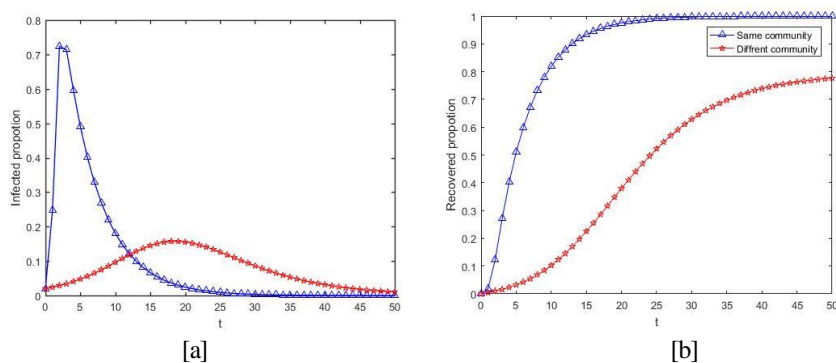


Fig.6. Communication of different community information(The horizontal axis of the two graphs both are time steps, and the vertical axis of graph [a] represents the proportion of infected nodes, namely, the proportion of information spreaders. In this graph, the two curves represent the changes of spreaders in same community and between different communities, respectively. In Figure [b], The vertical axis represents the change of the recovered nodes, namely, the change of those nodes which do not transmit information again. The two curves represent the scope of information dissemination in same community and between different communities, respectively.)

4.5 Model comparison

We select the event "315 party" to verify the propagation process. In order to verify the objective performance of the proposed model, the SIR(susceptible-infected-recovered) model and the SIS(susceptible-infected-susceptible) model were selected for verification. The fitting experimental results of the proposed model, of SIR(susceptible-infected-recovered), of SIS(susceptible-infected-susceptible) model and of actual information dissemination are shown in Figure 7. In this figure, the horizontal axis represents time step and the vertical axis represents the proportion of information spreader in social networks. It can be seen from Figure 7 that the

proposed model is more fitted to the actual data than the SIR(susceptible-infected-recovered) and SIS(susceptible-infected-susceptible) model. The SIS(susceptible-infected-susceptible) model divides the nodes into two categories, the susceptible nodes and the infected nodes. From Figure 7, we can see that the SIS(susceptible-infected-susceptible) model is still highly fitted to the real data at the initial time. Because two types of nodes are only considered to transform each other in the SIS(susceptible-infected-susceptible) model, it is not considered that the user may not want to propagate or turn into other nodes after contacted. SIR(susceptible-infected-recovered) considers an ideal situation. The user is transformed from a susceptible node to a infected node, and the probability of the infected node becoming an recovered node is certain. The mutual influence between users is not considered, so the degree of fitting to real data is not very high. In this paper, the influences of user's emotions and connections on information dissemination is considered, so that the propagation of real data can be better fitted. However, we can see from the figure that there are many changes in the real propagation situation, but the overall trend can be predicted by the propagation model. From Figure 7, we can see that although the SIR(susceptible-infected-recovered) model can predict the trend of information dissemination, it cannot accurately predict the burst time of information and the maximum proportion of the infected population. Although the model proposed in this paper does not accurately predict the burst time of information, it can predict the maximum proportion of infected people, thus predict the maximum scale of information dissemination.

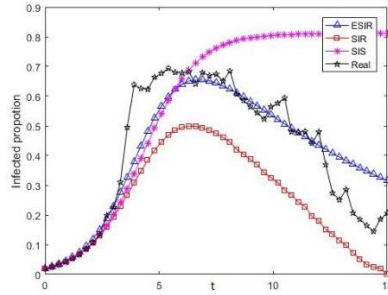


Fig. 7. Comparison of three type models and real data fitting(The horizontal axis is the time vertical axis, and the vertical axis represents the proportion of infected nodes(information disseminators). The curves represent the changes of information disseminators in different models and real data.)

5 Conclusions

This paper proposed an emotion-based susceptible-infected-recovered information propagation model(E-SIR) model to analyze the impacts of different emotions and relations between users on the information burst threshold and propagation range. Firstly, we considered the correlation between user's emotions to measure the degree of emotional connection between different users, in order to determine the emotional transmissibility and its impact on the process of information dissemination. Secondly, this paper considered the influence of the user's own connection relationships on information dissemination, and introduced the information transmissibility to analyze the influences of the connected relationships on information dissemination. The influences of the above two factors on information dissemination are theoretically analyzed. Finally, the theoretical derivation is proved by experiments. At the same

time, it is found that happiness and anger can promote the spread of information and have a greater impact on the threshold of information explosion. Compared with the situation without emotion, joy and anger will advance the time information explosion, and the scale of influence will be wider. Disgust make the information spread more widely than sadness, but it not has a same impact on information dissemination as joy and anger. Sadness has little effect on the process of information dissemination. In addition, we also found that the connection is closer, the speed of information dissemination is faster and the scale is wider, but the faster it disappears. Conversely, the contact is not closer, the speed of information dissemination is slower, but the duration of the spread is longer and the disappearance is slower. This paper can help us to understand the law of information dissemination in social networks and the impacts on emotional factors and connections between users on information dissemination. It not only provides reliable evidence for public opinion monitors, but also provides basis for taking relevant measures.

Funding. This work was supported in part by the National Natural Science Foundation of China under Grant No. 61871062, and in part by the Scientific Research Foundation of CQUPT under Grant No. A2018-07.

References

- [1] Persia, F. D'Auria, D.: A Survey of Online Social Networks: Challenges and Opportunities, 2017 IEEE International Conference on Information Reuse and Integration (IRI). San Diego. CA. pp. 614-620 (2017)
- [2] Zhang, ZK. Liu, C. Zhan, XX. et al.: Dynamics of information diffusion and its applications on complex networks. *Physics Reports* (2016)
- [3] Huang, HC. Lai, LC. Hu, M. et al. Information Propagation Control Method in Social Networks Based on Exact Controllability Theory. *Journal of electronics & information technology*. pp. 1707-1714 (2018)
- [4] Liu, Y. Wang, B. Wu, B.: Characterizing super-spreading in microblog: An epidemic-based information propagation model. *Physica A-statistical mechanics and its applications*. pp. 463. 202-218 (2016)
- [5] Wang, XY. Zhao, TF.: Model for multi-messages spreading over complex networks considering the relationship between messages. *Communications in nonlinear science and numerical simulation*. pp. 43. 63-69 (2017)
- [6] Garg, Rajiv. Smith, Michael D. Telang, Rahul.: Measuring Information Diffusion in an Online Community. *Journal of management information systems*. pp. 11-37 (2011)
- [7] Huang, HC. Sun, XR. Hu, M.: An Information Diffusion Model of Social Network Based on Node Attitude. *Advanced Engineering Sciences*. pp. 113-119 (2018)
- [8] Stieglitz, Stefan. Dang-Xuan, Linh.: Emotions and Information Diffusion in Social Media-Sentiment of Microblogs and Sharing Behavior. *Journal of management information systems*. pp. 294 (2013)
- [9] Brady, William J. Wills, JA. Jost, JT. et al.: Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the national academy of sciences of the united states of America*. pp. 114-28 (2017)
- [10] Wang, QY. Lin, Z. Jin, YH. et al.: ESIS: Emotion-based spreader-ignorant-stifler model for information diffusion, *Knowledge-based systems*. pp. 46-55 (2015)
- [11] Wang, ZP. Zhao, JC. Xu, K.: Emotion-based Independent Cascade model for information propagation in online social media, 2016 13th International Conference on Service Systems and Service Management (ICSSSM). Kunming. pp. 1-6 (2016)
- [12] Miller, M. Lynn, S.-L. James, MC.: Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27. pp. 415-444 (2001)
- [13] J. Berger.: Arousal increases social transmission of information. *Psychological Science*. pp. 891-893 (2011)
- [14] Bollen, J. Goncalves, B. Ruan, GC. Mao, HN.: Happiness is Assortative in Online Social Networks. *Artificial life*. pp. 237-251 (2011)
- [15] Wang, W. Tang, M. Stanley Eugene H. Braunstein A Lidia.: Unification of theoretical approaches for epidemic spreading on complex networks. *Reports on progress in physics*. pp. 80-1 (2017)
- [16] Fan, R. Xu, K. Zhao, JC.: An agent-based model for emotion contagion and competition in online social media. *Physica a-statistical mechanics and its applications*. pp. 245-259 (2018)
- [17] Xiong, X. Li, YY. Qiao, SJ.: An emotional contagion model for heterogeneous social media with multiple behaviors. *Physica A-statistical mechanics and its applications*. pp. 185-202 (2018)