Research on Rumor Propagation Simulation Based on Behavior-Attribute

Hailiang Chen
college of systems engineering, national university of defense technology
Changsha, China
chenhailiang_15@163.com

Bin Chen
college of systems engineering, national university of defense technology
Changsha, China
chenbin06@nudt.edu.cn

Chuan Ai
college of systems engineering, national university of defense technology
Changsha, China
aichuan@nudt.edu.cn

Mengna Zhu
college of systems engineering, national university of defense technology
Changsha, China
Monazhu1998@163.com

Kaisheng Lai
School of Journalism and Communication, Jinan University
Guangzhou, China
kaishenglai@126.com

Lingnan He
The School of Communication and Design, Sun yat-sen University
Guangzhou, Chin
lingnan.he@qq.com

***Abstract*—The rapid development of information technology has provided a hotbed for rumors, and the study of the characteristics of rumors propagation is essential for taking intervention measures. This paper proposes the NF-S (LIR) model which considers users' behavior and attributes separately at the individual level. The data are collected in the form of a questionnaire, and two sets of experiments are conducted using simulation methods to verify the rationality of the model and predict the effects of different interventions in different scenarios. The mechanism of rumor spreading is studied in our work, and the effects of government interventions are testified in the experiments.**

Keywords—Rumor propagation, NF-S(LIR), simulation

I. INTRODUCTION

The rapid development of the Internet, smartphones and information technology has given online social networks a booming opportunity. Through online social networks, information can spread quickly. The propagation of Internet rumors accounted for the major propagation of information, which has emerged with the development of the information age. The rumors were produced and propagated with the help of network tools and platforms, and their influence was amplified. The research on the spreading mode of rumors helps the government to make rumors-breaking strategies in time, to reduce the harm of rumors to society.

A large number of scholars have borrowed infectious disease models to study information propagation. Classical evolution models mainly include SIRS, SEIR, SEIRS, etc. The SIRS model mainly considers that the immunized person will have a certain chance to return to the susceptible state[1] [2]. Anderson and May believe that there is a certain latent state to be transformed in the propagation process, so they proposed the SEIR model which adds the latent identity E[3] [4] based on the SIR model. With the SIRS model and the SEIR model combined, a new information propagation model, SEIRS, is produced[5]. Since then, many scholars have optimized and improved the model based on their research problems and considerations. Leskovec et al. built a topic propagation model in the blog network based on the SIS model[6]. Jiuchang et al. proposed three models of information dissemination in the study of crisis information dissemination[7]. Borrowing the SI model[8], Zhao et al. introduced individual fitness parameters and proposed a propagation model for sudden topics in the network. Gruhl et al. used SIRS to establish a network topic propagation model[9]. One of the key areas of research in information dissemination is the study of the evolution of Internet public opinion in public opinion dissemination. Chen Fuji et al. proposed the SEIRS network public opinion propagation evolution model which was joined the topic derivation and they used specific cases for data simulation[10]. Wang Chao proposed an information dissemination model in social networks based on the SEIR model[11]. Lin Qin and Guo Dongqiang took into account the user's psychological characteristics and established a social network public opinion propagation model based on the SIS model. They proposed prevention, control methods and they conducted simulation verification[12]. G. Chen et al. set the ignorant to continuously flood into the network and convert them into other state nodes according to different probabilities. They establish the ILSAR model by introducing the latent person L and the attenuator A to describe the process of spreading panic emotions under sudden social events[13]. Lin Xiaojing etc. constructed a network public opinion propagation and diffusion model based on the SEIR model which introduced the participation of new netizens and saturated contact rates[14]. Based on the SIR model, Xiang Zhuoyuan et al. added bystanders, skeptics and opponents to form a total of six role states, and constructed a SIR-CO microblog rumor propagation model for analyzing the influence of rumors and anti-rumor laws[15]. Scholars such as Xiao Renbin established a game evolution model of the government and netizens in the process of information dissemination and proposed government countermeasures according to the different scales of network mass events[16]. Some scholars use the infectious disease model to simulate the process of public negative attitude transmission[17]. Li Kejia et al. introduced thermal propagation nodes in social networks to improve the SIR model, demonstrating that network pushers have powerful information dissemination capabilities[18]. Wang Xiwei added the comparison of new media and traditional media in the research of information dissemination and proposed the MSIR model based on SIR[19]. Jia F introduced environmental noise and proposed a random rumor propagation model[20]. In the classic SIR model, Ren Ning et al. added the role of the opponent and proposed the SICR model[21].

The previous research only considered the user's information attribute state transition, which attached their behavior to the information attribute and did not consider the user's behavior and attribute separately. However, in the process of rumors propagation, the user's information attributes will change to the same state more than once, but the user's behavior will not be repeated. For example, a person

978-1-7281-9558-2/20/$31.00 ©2020 IEEE

DOI 10.1109/DSC50466.2020.00041

can't forward the same rumor multiple times. Therefore, there are differences between information attributes and behavior, which should be modeled separately. Therefore, this paper proposes a rumor propagation model based on user behavior-attribute. In the model, the behavior considers Non-forwarding and Forwarding, and the attribute considers its information attribute Sealed state, Latent state, Infection state and Removal state (NF-S (LIR)). The age is also considered in our work. First, the data are collected through empirical investigation, and then the model parameters are built. Then the Pregel-based architecture for simulation experiments. The proposed model is testified.

The rest of the article is organized as follows: Section 2 proposes the NF-S (LIR) model in detail. Section 3 describes the experimental data of the empirical investigation and constructs parameters of the model. Section 4 uses simulation experiments to test the model and predict the effects of different intervention efforts in different scenarios. The paper is concluded in section5.

II. Model Introduction

We propose the NF-S (LIR) model to describe the spread of rumors in social networks at the individual level. This model is based on the classic SEIR model, but this model considers the user's attitude to rumors and forwarding behavior separately and incorporates the influence of forwarding messages by neighbor nodes. So, it has practical significance. And the conversion mode between different attributes is also different from the classic model. In the NF-S (LIR) model, each user is treated as a node in the network, and the relationship between the users is represented by connected edges. The user has a behavior state and an information attribute state.

A. Behavior state

There are two user behavior states, forwarding and non-forwarding shown in Fig 1:

Fig. 1. Behavior state Transition

If a user does not receive a message, he/she is in the N state. When receiving a message from a neighbor, users have a certain probability of becoming an F state. This probability is affected by its attributes. After the user sends the message, the behavior state immediately changes to the N state.

When the user is in the F state, the neighbors will be affected, resulting in changes in both behavior and attributes. When the user is in the N state, he/she will not affect the neighbors.

The user's behavior state is expressed as $S^b \in \{N, F\}$, and the transition probability from N to F is shown in formula (1).

$$p_{N \to F} = p(S_i^b = \mathrm{F} | S_{i-1}^{attr} \in \{L, I, R\}, S_{i-1}^b = \mathrm{N}) \tag{1}$$

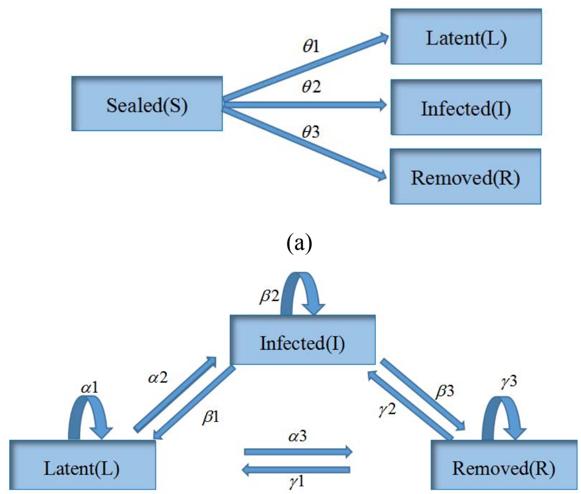
Among them, S_{i-1}^{attr} represents the attribute state of the user at time $i-1$.

B. Attribute State

In the process of rumor propagation, each user has one of the following states at any time:

- S (Sealed): The user has not received the message, and its information attribute is sealed.
- L (Latent): The user receives the message, but he/she is uncertain about the credibility of the message. His/her information attribute is latent.
- I (Infected): The user receives the message and he/she tends to agree with the rumor. His/her information attribute is infected.
- R (Removed): The user receives the message but he/she considers that the message is not credible, which is likely to be a rumor. His/her information attribute is removed.

The state transition diagram of the S(LIR) model is as Fig 2.



(a)

(b)

Fig. 2. Attribute State Transition

The initial information attribute of each node is S. As the difference of individual attributes, people's cognitive state of a message is different. When users receive a message for the first time, their information attributes change into different states (L, I, R) according to different probabilities $\theta 1, \theta 2, \theta 3$ respectively as showing in Fig 2(a).

When a message is forwarded multiple times by a person's neighbors, it will cause this person's information state attribute to transfer among L, I, R. The transfer method and probability are showing in Fig 2(b). The arrow represents the direction of transition, and the variable next to the arrow represents the probability of transition between states. For example, $\alpha 1$ represents the probability that the user keeps the latent state unchanged. $\alpha 2$ indicates the probability of the user transfer from the latent state to the infected state. Others are similar. Thus, the state transition matrix can be obtained, as in Equation 2.

$$P_{people} = \begin{bmatrix} \alpha 1 & \alpha 2 & \alpha 3 \\ \beta 1 & \beta 2 & \beta 3 \\ \gamma 1 & \gamma 2 & \gamma 3 \end{bmatrix} \tag{2}$$

For each user, after receiving the message for the first time, at time i, his/her information state is $S_i^{attr} = [x1 \quad x2 \quad x3]$, where $x1 \quad x2 \quad x3$ indicates the probability that the user is in L, I, R respectively. For example, if the user is in state L, the value of $x1$ is 1, and the rest is 0. The state at the next moment is obtained by the equation $S_{i+1}^{attr} = S_i^{attr} * P_{people}$.

III. DATA DESCRIPTION

A. Questionnaire Data

Through the online questionnaire survey platform, push the questionnaire to netizens to obtain research data. The rumors set in the questionnaire are untrue information texts that are closely related to people 's interests and high ambiguity. As for the rumors, there is no more other information.

According to the rumor definition, it is a kind of uncertain information that has utility and ambiguity for people. Therefore, according to a question in the questionnaire "Do you think this message is credible?", The answer is 5 points, one for very untrustworthy, two for a little untrustworthy, three for a little uncertain, four for a little trustworthy, five for absolutely trustworthy. And then based on the answer to the question "Whether to forward", we filtered out "very definitely not" "unsure" and "very definitely" samples. The combination of "whether credible" and "whether to forward" can filter out real samples, that public response an uncertain message. Therefore, this study divided the users into three types, latent persons (L), infected persons (I), and removers (R) according to the NF-S (LIR) model.

In this experiment, two survey data were obtained, and a total of 15625 valid questionnaires were obtained. The first survey is taken at the beginning of the rumor propagation for the state of information attributes that users are in, while the second survey is taken at the end of the spread of rumors. The result is shown in TABLE I.

TABLE I INFORMATION ATTRIBUTE PROPORTION

Information attribute	L	I	R
Initial proportion	19%	4%	77%
Final proportion	28%	18%	54%

Combining the two survey data, the user's forwarding probability under different information attributes can be obtained, as shown in TABLE II. The population proportion is shown in TABLE III.

TABLE II RATE OF USER FORWARDING

Information attribute	L	I	R
Forward probability	50%	80%	10%

TABLE III PROPORTION OF POPULATION

age	young	adult	old
proportion	61%	31%	8%

B. Data Analysis

Treat all users as one system. It is assumed that when the user's information attribute changes in (L, I, R) three states, the next information attribute only related to the current information attribute, not the historical information attribute,. Therefore, the stochastic process of transformation of information attributes in the system can be regarded as a Markov chain. At each step of the Markov chain, the system can change from one state to another according to the probability distribution, or it can maintain the current state. The change of state is called transition, and the associated probability is called transition probability. So the system transfer matrix is P_{whole} . Since the state transition matrix P_{people} of each person in the system is consistent, the system transition matrix is equal to the individual transition matrix that is $P_{whole} = P_{people}$.

Assuming that $X_i = [x_1 \quad x_2 \quad x_3]$ is the proportion of users in the population at the time i at {L, I, R}, the proportion of users at the time $i+1$ can be expressed as X_{i+1}, and there is a relationship: $X_{i+1} = X_i * P_{whole}$. When the information attribute tends to be stable among the crowd, there exists $\lim_{i\to\infty}(X_{i+1} = X_i)$, namely:

$$X_\infty = X_\infty * P_{whole} \tag{3}$$

At the same time, the sum of the probabilities is 1, as shown in Equation 4.

$$\begin{aligned} \alpha1 + \alpha2 + \alpha3 = 1 \\ \beta1 + \beta2 + \beta3 = 1 \\ \gamma1 + \gamma2 + \gamma3 = 1 \end{aligned} \tag{4}$$

Considering the differences in the processing of rumor information at different ages, when calculating the transfer matrix, users are divided into three ages, young (10-39 years old), adult people (40-49 years old)), and the old (over 50 years old). The classification of age is not unique and it is just an example. Details are shown in TABLE IV

Combining with Equation (3) (4) and selecting the parameters within a reasonable range, the transition matrix(5)(6)(7) can be obtained.

$$P_{whole}^{young} = \begin{bmatrix} 0.3 & 0.1 & 0.6 \\ 0.3 & 0.512 & 0.188 \\ 0.3 & 0.1 & 0.6 \end{bmatrix} \tag{5}$$

$$P_{whole}^{adult} = \begin{bmatrix} 0.1 & 0.1 & 0.8 \\ 0.3 & 0.6 & 0.1 \\ 0.3 & 0.1 & 0.6 \end{bmatrix} \tag{6}$$

$$P_{whole}^{old} = \begin{bmatrix} 0.5 & 0.36 & 0.14 \\ 0.3 & 0.47 & 0.23 \\ 0.1 & 0.01 & 0.89 \end{bmatrix} \tag{7}$$

The transfer matrix above is for young, adult and the old people respectively.

TABLE IV PROPORTION OF L I R POPULATION IN DIFFERENT AGE GROUPS

age	Young			Adult			Old		
Information attribute	L	I	R	L	I	R	L	I	R
Initial proportion	16%	3%	81%	24%	3%	73%	26%	10%	64%
Final proportion	30%	17%	53%	25%	20%	55%	22%	16%	62%

IV. EXPERIMENT

This experiment is based on Spark computing engine and Pregel distributed graph computing framework and is programmed in scala language. Two sets of experiments are designed in this section. Experiment 1 uses the survey results data to reproduce the actual survey scene. Experiment 2 predicted the spread of rumors spread by different groups of people and compared the results of the effects of different interventions. Social networks meet power-law characteristics, which play an important role in the spread of online rumors. So in the experiment, BA scale-free network was adopted, with 10,000 nodes and 130,753 edges.

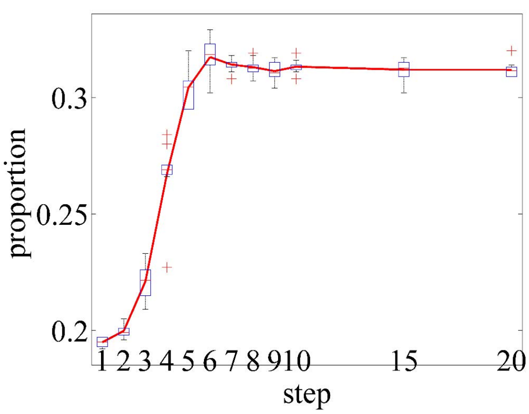
A. Experiment 1 Reproduction of the survey results

The experiment is carried out at the individual level, and the information attributes of user nodes are updated by sending rumors, merging rumors, and receiving rumors. Under the conditions of the third part of the data, which are TABLE I to TABLE IV and P_{whole}^{young} , P_{whole}^{adult}, P_{whole}^{old}, we take 10 times for simulation and each simulation is iterated for 20 steps. the simulation results can be obtained as TABLE V.

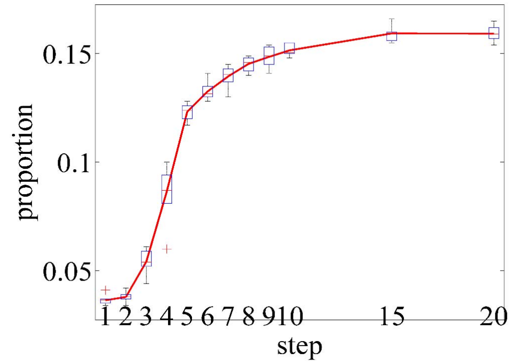
TABLE V THE RESULT OF EXPERIMENT 1

Information attribute	L	I	R
Survey proportion	28%	18%	54%
Simulation proportion	31.2%	15.9%	52.9%

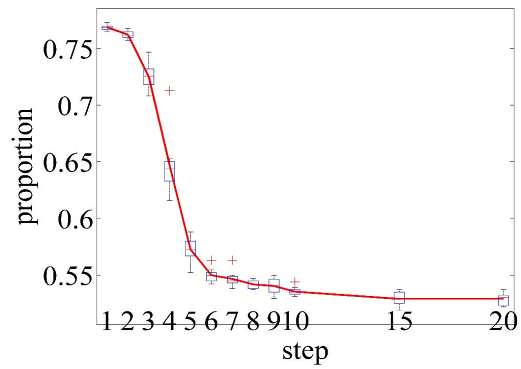
Figure 3 combines the box chart and the line chart, reflecting the average results of 10 experiments, which shows the change in the proportion of people with different information attributes, (a) for the latent person, (b) for the infected person, and (c) for the removed person. When the simulation is iterated to the 20th step, the information attribute proportion of the population in the system has stabilized. From the results, we can find that the simulation results are very close to the survey proportion, which shows that our simulation model is practical and can reproduce the real scene to a certain extent.



(a)Proportion change of L



(b)Proportion change of I



(c)Proportion change of R

Fig. 3. Population proportion change chart in experiment 1

B. Experiment 2: Rumor prediction and intervention effect

According to different rumors, three scenarios were set up to study the spread of rumors in different groups. Also, intervention strategies were set up. Intervention is designed from the following three aspects:

- Probability of Role Conversion (PRC): According to different strategies, the probability of the latent person (L) and the immune person (R) transforming into an infected person(I) and the probability of the infected person (I) staying the same is reduced accordingly.
- Probability of N to F(PNF): According to different strategies, the probability of conversion from N to F is reduced accordingly.
- The propagation Probability of people in different information attributes(PP): According to different rumors, the forwarding probability of the latent, infected, and immunized people is reduced accordingly.

The strength of the three different intervention strategies is shown in TABLE VI.

TABLE VI Strategic Strength

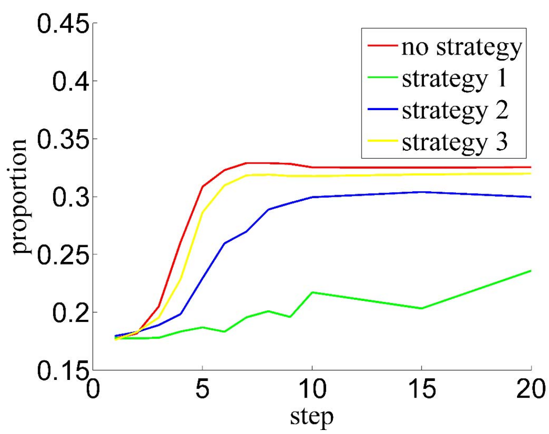
	PRC	PNF	PP
Strategy 1	50%	50%	50%
Strategy 2	30%	30%	30%
Strategy 3	10%	10%	10%
No strategy	0%	0%	0%

The numbers in the table represent the proportion of the corresponding probability drop.

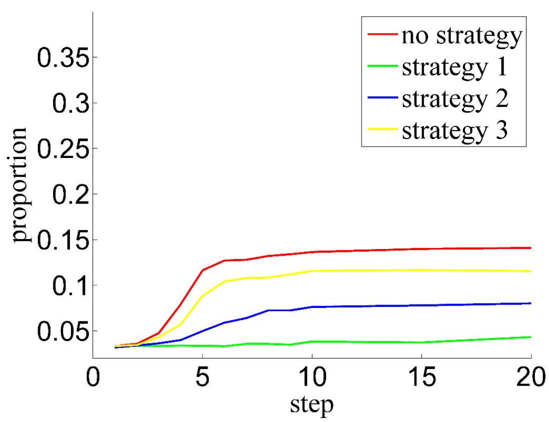
Scenario 1: The propagation of rumors that young people are concerned about.

Distribution of age attributes in experiment 1 is set as young people are 61%, adult people are 31% and the old people are 8%.In this scenario simulation, it is adjusted to that young people account for 80% of the total. The rest of the people are adults and old people but the proportion remains the same between them. That is, 80% of young people, 15.9% of adult people, and 4.1% of the old. The remaining parameters are the same as experiment 1.

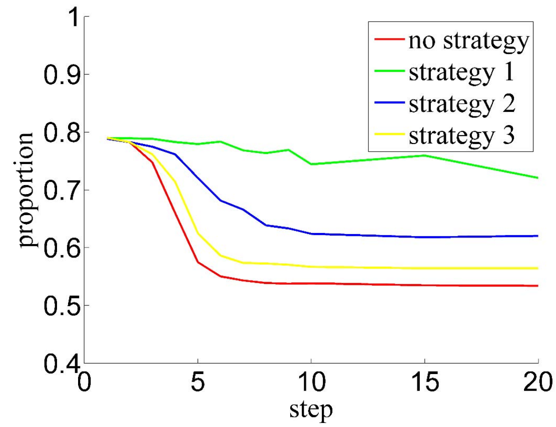
Fig. 4 shows the proportion of users with different information attributes in scenario 1, where (a) is for the latent, (b) is for the infected, and (c) is for the removed.



(a) Changes in the proportion of L



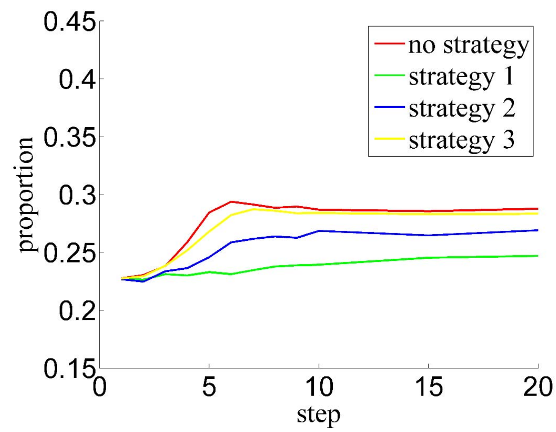
(b) Changes in the proportion of I



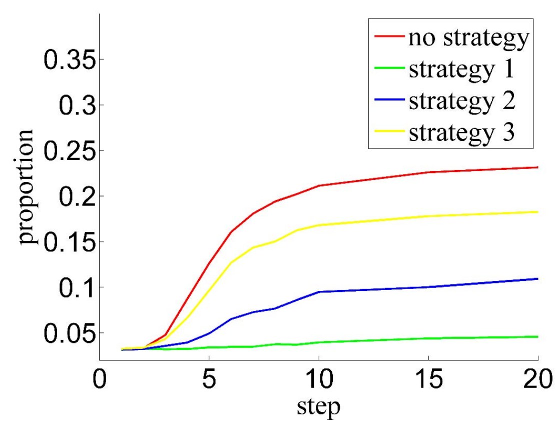
(c) Changes in the proportion of R

Figure 4 Comparison of Different Intervention Strategies in Scenario 1

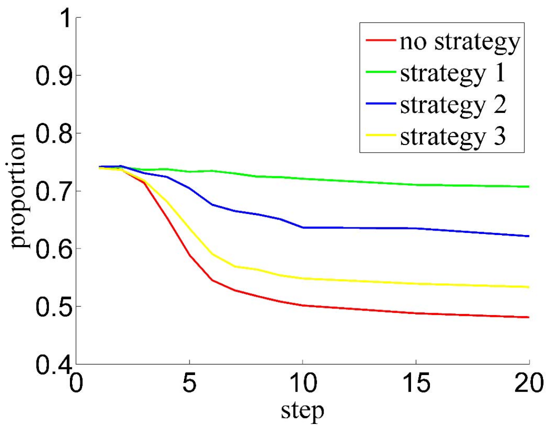
Scenario 2: The propagation of rumors that adult people are concerned about.



(a) Changes in the proportion of L



(b) Changes in the proportion of I



(c) Changes in the proportion of R

Figure 5 Comparison of Different Intervention Strategies in Scenario 2

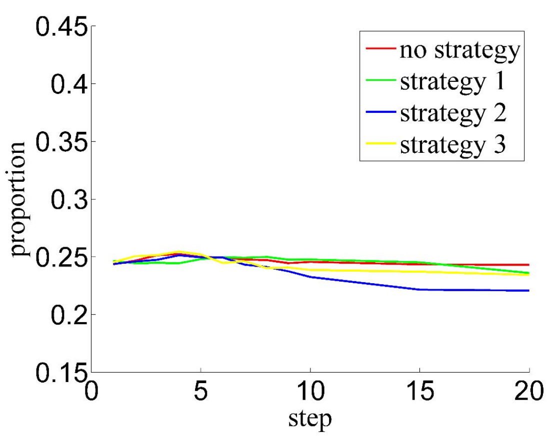
In scenario 2, the age attribute is adjusted: adult people account for 80% of the total. The rest are young and the old but the proportion remains the same between them. That is, 17.7% of young people, 80% of adult people, and 2.3% of old people. The remaining parameters are the same as experiment 1.

Fig. 5 shows the proportion of users with different information attributes in scenario 1, where (a) is for the latent, (b) is for the infected, and (c) is for the removed.

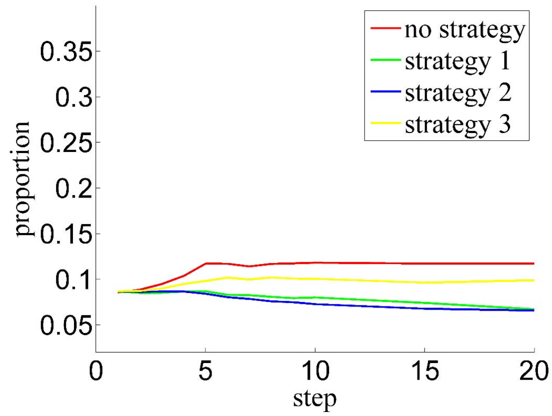
Scenario 3: The propagation of rumors that the old people concerned about.

In this scenario simulation, the age attribute is adjusted: the old account for 80% of the total. The rest are young and adult but the proportion remains the same between them. That is 13.3% of young people, 6.7% of adult people, 80% of old people. The remaining parameters are the same as experiment 1.

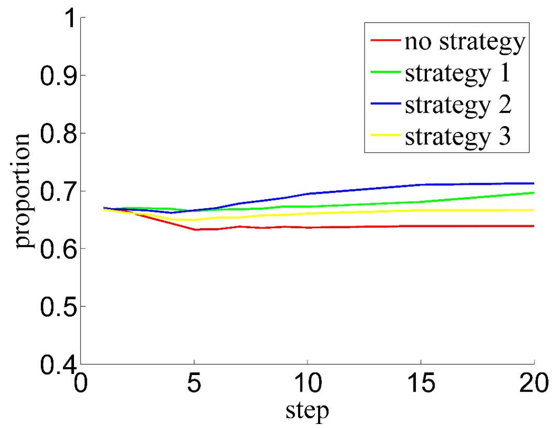
Fig. 6 shows the proportion of users with different information attributes in scenario 1, where (a) is for the latent, (b) is for the infected, and (c) is for the removed.



(a) Changes in the proportion of L



(b) Changes in the proportion of I



(c) Changes in the proportion of R

Figure 6 Comparison of different intervention strategies in scenario 3

TABLE VII CONVERGENT STATE COMPARISON OF L I R WITHOUT INTERVENTION

	L	I	R
Scenario 1	32.5%	14.1%	53.4%
Scenario 2	28.8%	23.1%	48.1%
Scenario 3	24.3%	11.7%	64.0%

From the comparison of Figures 5, 6, and 7, it can be found that with the increase of the intervention strength, the effect of preventing the spread of rumors is more and more obvious. Under the intervention strategy 1, the proportion of the population's LIR remains almost unchanged. In the process of spreading rumors of concern to the old, the proportion of latent people did not change much under the condition of no intervention, so the effects were not obvious under different intervention strategies.

V. CONCLUSION

Rumors are extremely harmful to the public. In order to clarify the characteristics of rumors transmission, the rumor propagation is modeled from the individual user level, which considers the user's behavior of forwarding rumors, and the attributes of attitudes to rumors separately. There are two types of user behaviors in the model, Forwarding and Non-forwarding, and four types of information attributes, sealed, latent, infected and removed. The user's behavior will affect the attribute state of its neighbor nodes. At the same time, its behavior state is affected by its attribute state.

First of all, we collected the age distribution proportion and cognitive attitudes of the target population in the form of questionnaires, which is whether the message is credible and whether he/she would forward. Through that, we obtained the parameters needed in the simulation experiment indirectly. Then, based on the spark which is a computing engine and Pregel which is a distributed graph computing framework, simulation experiments were conducted on the BA scale-free network. The experimental results and the survey results are consistent within the tolerable range. This proves the rationality of the NF-S (LIR) model. Finally, by setting up simulation experiments with different intervention intensity in different scenarios, the intervention effects of different intervention strengths are predicted, which are used to support the government to choose the right interventions.

ACKNOWLEDGMENT

This study is supported by National Key Research & Development (R&D) Plan under Grant No. 2018YFC0806900 and the National Natural Science Foundation of China under Grant Nos.71673292, 21808181, 61673388, 71673294 and National Social Science Foundation of China under Grant No.17CGL047 and Guangdong Key Laboratory for Big Data Analysis and Simulation of Public Opinion.

REFERENCES

[1] Enatsu Y, Nakata Y, Muroya Y. Lyapunov functional techniques for the global stability analysis of a de‑ layed SIRS epidemic model[J]. Nonlinear Analysis: Real World Applications, 2012,13(5):2120-2133.

[2] CH LI, CC TSAI, SY YANG. Analysis of epidemic spreading of an SIRS model in complex heteroge ‑ neous networks[J]. Communications in Nonlinear Science and Numerical Simulation, 2014,19(4):1042-1054

[3] Anderson R, May R. Infectious diseases of humans: dynamics and control [M]. Oxford: Oxford University Press, 1991: 99.

[4] Witbooi P J. Stability of an SEIR epidemic model with independent stochastic perturbations[J]. Physica A: Statistical Mechanics and its Applications, 2013,392(20):4928-4936

[5] Denphedtnong A, Chinviriyasit S, Chinviriyasit W. On the dynamics of SEIRS epidemic model with transport-related infection[J]. Mathematical Biosci‑ ences, 2013,245(2):188-205.

[6] Leskovec J, McGlohon M, Faloutsos C, et al.Patterns of cascading behavior in large blog graphs[C]//Proceedings of the SIAM International Conference on Data Mining. New York:ACM Press, 2007, 7:551-556.

[7] Wei J, Bu B, Liang L.Estimating the Diffusion Models of Crisis Information in Micro Blog[J]. Journal of Informetrics, 2012, 6 (4) : 600-610.

[8] Zhao L, Yuan R, Guan X, et al.Propagation modeling and analysis of incidental topics in blogosphere[M]//Online communities and social computing.Berlin Heidelberg: Springer, 2009:401-410.

[9] Gruhl D, Guha R, Liben-Nowell D, et al.Information dif- fusion through blogspace[C]//Proceedings of the 13th Inter- national Conference on World Wide Web, 2004:491-501.

[10] Chen Fuji, Chen Ting, et al. Research on the behavior of network public opinion propagation on a new S E I R S model [J]. Intelligence and Information Work, 2 0 1 4, (4): 6 2-6 7 .

[11] Wang Chao, Yang Xuying, Xu Ke, et al. SEIR-based social network information dissemination model [J]. Journal of Electronics, 2014, 42 (11): 2325-2330.

[12] Research on SIS model-optimized social network public opinion communication ——Based on user psychological characteristics.

[13] GH CHEN,HZ SHEN,GM CHEN, et al. A New Kinetic Model to Discuss the Control of Panic Spreading in Emer‑ gency[J]. Physica A: Statistical Mechanics & Its Applica‑ tions, 2014,(C):345-357.

[14] Lin Xiaojing, Zhuang Yaming, Sun Liling. Research on SEIR Internet Public Opinion Propagation Model with Saturated Contact Rate [J]. Information Journal, 2015, 34 (3): 150-155.

[15] Xiang Zhuoyuan, Chen Yuling. Research on the spread model and influence evaluation of rumors on Weibo [J]. Scientific Research Management, 2016, (1): 39-47.

[16] Xiao Renbin, Zhang Yaofeng. Evolutionary Game Analysis of Information Dissemination of Internet Group Events [J]. Complex Systems and Complexity Science, 2012, 9 (1): 1-7 .

[17] Zhong Huiling, Li Wei, Zhang Guanxiang, Research on the Evolution of Internet Public Opinion on the "Neighborhood" Conflict. Journal of Information,2016,35(3):111-117.

[18] Li Kejia, Wang Yikang, LIKejia, et al. Application of improved SIR model in social network information dissemination [J]. Electronic Technology, 2013, 26 (8): 168.

[19] Wang Xiwei, Zhao Dan, Li Jiaxing, Yang Mengqing. Research on the Evolution Model and Simulation of Internet Public Opinion in the New Media Environment_Based on the Information Ecology Perspective [j]. Journal of Information, 2016,35 (10): 1011-1021.

[20] Jia F, Lv G. Dynamic analysis of a stochastic rumor propagation model[J]. Physica A-statistical Mechanics and Its Applications, 2018,490: 613-623.

[21] Ren Ning, Li Jinxian. Rumor spreading model with opposition mechanism [J]. Journal of Yunnan University for Nationalities (Natural Science Edition), 2019, 28(01): 67-71