Research on Information Communication Model of Internet Public Opinion in Risk Events ——A sample of the Microblog event “Genetic Editing Baby” in November 2018

Yuqiao Yan, Ziqi Liu, Henan Diao

Beijing Normal University, China

Keywords: risk perception, information communication model, risk communication, gene editing; public opinion analysis

Abstract: the frequent occurrence of the current controversial events in Chinese society has aroused heated discussions and debates in various circles of society. In order to grasp the trend of public opinion in a timely manner and try to avoid the emergence of mass incidents, it is necessary to predict the heat and direction of public opinion. This study is the psychology-based risk perception measurement school theory, combined with the communication science of the big data analysis, and uses the calculation method of information science information communication model to establish an information communication model for controversial scientific and technological events, and the "world's first genetic editing baby born in China" event in November 2018 for the actual control sample was used as the model reference. It aims to predict the communication heat of this type of event before the outbreak of public opinion to make a preliminary prediction, so as to buy time for relevant parties to conduct timely public opinion intervention.

1. Research background

In the Chinese society that has been in the period of social transformation since the new century, controversial events concerning science have emerged one after another. From nuclear power plants, PX projects to smog governance issues, and genetically modified food safety issues, there has been an endless stream of debates in various sectors of society. Some mishandling incidents have even led to fragmentation and hostility between certain groups. As a researcher and analyst of public opinion, how to predict, monitor and analyze of public sentiment in controversial events, and use emergency plans to intervene and control in time, so as to establish a public opinion environment for benign communication and avoid the purpose of vicious incidents is the focus of research on such events. At the level of public opinion management, what is more needed is a method that can predict the trend of public opinion of controversial events, so as to intervene before the outbreak of public opinion events to avoid greater social adverse consequences caused by this scientific and technological problem. Risk Communication refers to a process that provides the public with information to reduce anxiety and panic and recommendations to help them cope with the crisis. Effective risk communication can quickly improve people's awareness and skills to cope with and prevent the harm of public emergencies, eliminate unnecessary panic and maintain social stability. The paradigm for the adoption of risk communication is based on empirical data from psychological experiments and quantitative analysis of social surveys. It is also based on this attribute of the field, and some event has occurred for measurement and analysis results can be as an important experience of public opinion to predict reference, to accurately predict model is set up the idea became possible.

At present, public opinion research is no longer limited to the monitoring and post-mortem analysis of public opinion, but more focused on the prediction research of public opinion. The public opinion prediction in social media is mainly based on social network analysis through mathematical modeling. Researchers in the field of information science, by referring to the mathematical study of infectious disease communication law in biology and the innovation diffusion principle in sociology, put forward the information communication model for public

opinion that combines process model and influence model. With the continuous development of the network era, the carrier of information communication model research is gradually transferred to the network, and the proposal of many communication models can almost be attributed to the problem of influence. As a psychologist, Paul Slovich's series of studies visualize the abstract problem of risk perception and lead it from macro to micro. The core concept of Slovic's research is the personal “risk perception”, the effect of different dimensions on the individual's perception of risk events, a series of empirical measurements are used to validate and improve risk-aware factor analysis and measurement scenarios. Slovich proposed different dimensions of risk perception measurement and designed measurement methods and scales based on these dimensions, which are concentrated in the book "Perception of Risk". Slovic's greatest contribution lies in the establishment of the risk perception measurement school, which quantifies the risk perception of individuals that are difficult to measure in the previous humanities and social sciences, allowing researchers to visually observe the level of risk perception.

For the simulation of public opinion on Micro-blog, the current research results are mainly divided into two categories: first, the influence of public risk perception on public opinion is simulated by means of system dynamics model; secondly, infectious disease model is used to simulate the communication process of risk perception and its impact on social crisis. In addition, some scholars explore the evolution, influencing factors and effects of risk perception. Wang lian and jia jianmin analyzed the public risk after the earthquake from two dimensions of time and space. The evolution law of perception is studied. Li Feng used the Logit model to study the influencing factors of public risk perception evolution in the 5.12 Wenchuan earthquake. Cheng Junhui and Zhao Jinlou studied the impact of risk perception on the steady state of public opinion communication. However, in Zhao Wenbing's "Simulation Research on Social Network Information Communication in Web2.0 Environment", it is pointed out that the infectious disease model is not suitable for describing information communication under social networks, and thus constructs an information acceptance-browsing-sharing model. For the overall situation of public opinion communication of Micro-blog, He Xing combines and improves the two graph models, namely the IC model and the LT model, in the "Improvement and Verification of Micro-blog Influence Communication Model", but the specific measurement value of Micro-blog influence is based on 1-2 sample direct assignment calculations and the source parameters do not give an empirical basis, the validity of the model needs to be improved.

From the current research, although the achievements in the field of communication cannot be quantitatively calculated, it provides a theoretical basis for this research; the modeling method of information science provides a research paradigm, but there is a problem of insufficient validity due to the lack of empirical measurement of parameter assignment. However, although there are few macroscopic studies on the promotion of psychological measurement scheme to the whole field of public opinion, the current results provide the calculation method of individual value of communication node for this study. Therefore, based on the existing models of communication theories, psychological experiments and information science, this research will conduct information mining of public opinion Microblog based on big data, improve and propose a new information communication model and verify it.

2. Research objectives

The objective of this study is to obtain the communication choice (communication or non-communication and its threshold, the amount of communication, the probability of communication, etc.) of an individual user in social media with different degrees of risk perception through quantitative psychological measurement of risk perception, and take the behavior result as parameter value for modeling. This model is used to describe the number of people who participated in the communication process for a certain period of time in a social media communication of a controversial scientific event, in an ideal state (when public opinions are not controlled).

Therefore, ideally, when you know a risk science and technology event information, initial communication situation and audience risk perception, you can help the relevant by observing the level of the risk perception value of the audience to see how many people become the communicators, so as to help relevant organizations to formulate the public opinion control scheme. After the model is determined, the model is substituted into relevant controversial scientific and technological events, and the actual data and model prediction data are used for testing.

3. Psychological measurement: the relationship between event risk perception and user communication behavior

3.1 Pretest experiment: event risk value

Before making formal measurement values, according to the different dimensions of the audience's perception of risk, the news events with different risk values for the audience at different risk levels are designed, and the risk values of the events are measured as one of the reference variables to measure the selection of the audience's communication behavior. In this model, the risk perception of events is divided into “fear risk dimension” and “unknown risk dimension”, besides this model, time dimension and emotion dimension are also involved. Considering that this experiment belongs to the pretest experiment used to screen events and independent variables, it is necessary to control the demographic variables and subject background of the subjects. Therefore, 15 subjects were selected in this experiment, aged 20-24 years old, all of whom were senior undergraduate students or first-year graduate students of Beijing normal university, and had not participated in journalism/psychology/information science related courses.

After the 15 subjects arrived, they entered the laboratory in turn, and 64 events numbered 1-64 appeared one by one in a disordered order. The reading time of each event was 15 seconds. After reading an event, the subjects immediately filled in the scale questions. The scale for each event is filled in for a limited time of 30 seconds. Finally, the “directness of effect” is taken as the reference quantity, the knowledge score of the risk is a, the severity score of the consequence is b, the time-dependent score of risk is c, and the controllable score of risk is d. Then the calculation method of R1 (risk value of event) is: R1 (mean) = (reversely scored) a + b + (reversely scored) c + d.

3.2 Experiment on the relationship between risk perception status of events and user behavior communication

3.2.1 Experimental procedure

This experiment adopts the single-factor within-subject experimental design. The independent variable is the risk perception of the subjects in the face of different events. The dependent variable is the willingness of the subjects to transfer the risk events. Among them, in the total experimental data, we control the interference of event communication such as Event Forwarding and number of compliments, so as to measure the willingness of individuals to spread the risk events under different risk perceptions. In the process of studying risk events, Paul Slovich used the method of measuring the individual's risk perception by using the scale method. In 1998, Xie Xiaofei developed the “Research questionnaire on risk perception in general social situations” based on the actual situation of China based on the measurement results, this study selected the five dimensions of predecessors' risk perception scale measurement, and asked the subjects to score the risk perception of different events through the Richter 7-level scale, so as to obtain the risk perception of different subjects under different events, and the communication willingness and the number of stimulations as reference quantities.

3.2.2 Data Results

In the experimental results, the following variables (data) are output:

Table 1. Symbols corresponding to the contents of variables

variable	symbol
Event number	i
Event risk score	H
Risk perception status of participants in this event (the first four rows in the scale are averaged and added)	F
Number of times of being transmitted that are willing to forward	t
Willingness to spread of the event subjects (average score in line 5 of the scale)	P

(1) Event risk score and subject's communication willingness

Firstly, the correlation between H and P is directly tested, and the significance is calculated as follows:

Table 2. Variable correlation analysis results

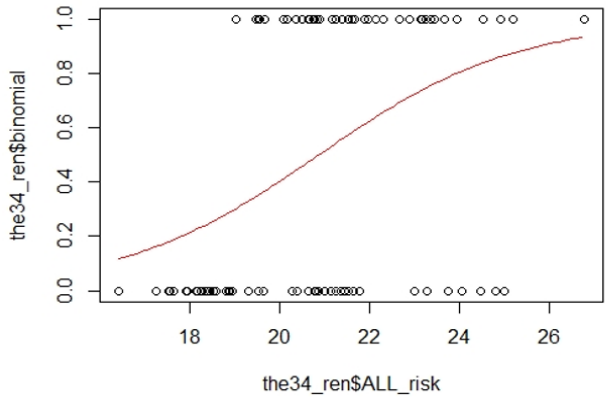
Correlation		Willingness to spread P
Pearson Correlation	Event risk score H	.367*
Sig. (2-tailed)		.023
N		38

The Pearson correlation coefficient was 0.367 and the double-tailed significant p value was 0.023. From the results, it can be seen that event risk score H is not significantly correlated with the subject's communication willingness P. The risk perception status F of the subjects was introduced, and the regression was further calculated and analyzed.

(2) Exploring the relationship between risk perception and communication willingness

Since the willingness to spread directly takes the threshold value of communication or non-communication in the model, directly using the scores of the seven-level scales 1-7 in the questionnaire not only poorly fitted, but also interferes with the activation threshold value to some extent. Therefore, the two-level scale is adopted in the scheme. The grading scale, which is divided into grading limits by 3.5, is divided into two levels (communication /non- communication).

Taking the subjects as observation points, the F" and P" test correlations are obtained and fitted (that is, the data of the number of rows of the subject), and the data were dualized (>3.5 means "willing to forward", < 3.5 means "unwilling to forward").




```
Call:
glm(formula = the34_ren$binomial ~ the34_ren$ALL_risk, family = "binomial")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.0075  -0.9059  -0.0681   1.0533   1.5458

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -9.4598     2.6349  -3.590 0.000330 ***
the34_ren$ALL_risk  0.4533     0.1260   3.596 0.000323 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 127.54  on 91  degrees of freedom
Residual deviance: 111.15  on 90  degrees of freedom
AIC: 115.15

Number of Fisher Scoring iterations: 4
```

Figure 1. Logistics curve fitting results of scheme three

As a result, it can be seen that with the subject as the observation point, F" is related to P" in the 99.9% confidence interval and the correlation is strong. Specifically, the two are related to the logistic curve, and the fitted curve is:

$$y = \frac{1}{1+\exp(-0.45x+9.46)}$$

This relationship is very important for the construction of the next part of the communication

model.

3.3 Experimental conclusions

It can be seen from the above three schemes that the establishment of a relevant communication model requires the subject to be fitted as an observation point, that is, the individual's willingness to forward depends not only on the risk value of the event itself, but more importantly, the risk perception of the individual to the event. Based on the experimental results, a model is built for calculation.

4. Design and calculation of communication model

4.1 Assumptions

In order to simplify the model and make the model clearer and achievable, the following assumptions are made:

(1) The user relationship, the total number of users, and the user characteristics are unchanged during the communication process, which is a static network.

(2) The influencing factors are independent of each other

4.2 Construction of communication network

(1) Constructed an undirected network

A BA scale-free network is constructed with each user as a node and the number of fans per user as the degree of the node.

(2) Initial activation node

By assigning the attribute s (t) to the node, it indicates whether the node knows to forward the event after knowing, that is, whether the node is activated. The value of s (t) is as follows:

$$s(t) = \begin{cases} 1 & \text{Activated} \\ 0 & \text{Not activated} \end{cases}$$

4.3 User's communication willingness

From the conclusion of 4.3, we use the probability model to express the user's communication willingness:

$$\begin{cases} P(t) = \dfrac{1}{1+e^{-0.45\theta+9.46}} \\ \theta = H\sum_{i=1}^{n(t)} w_i \end{cases}$$

Where H is the risk value of the event; n (t) represents the number of neighbors n (t) in the active state at time t, and wi represents the influence of the neighbor i in the active state. Because the influence of different levels of users in the actual situation is different, the influence of the fans perception of the same user is also different. Therefore, the influence of the nodes in the model is represented by a random quantity fluctuating near the constant. The constant is determined according to the willingness. The relationship between the number of times of communication t and the event risk value H and the communication willingness is determined ((H-communication willingness threshold)/t).

That is,

$$w_i = \alpha \pm \zeta, \qquad \zeta \ll \alpha$$

4.4 Communication process

The communication process is based on the constructed probability model and uses the

Independent Cascade Mode (IC model) as the idea. At each moment in the communication process, each activated node attempts to activate its neighbor nodes, and each inactive node calculates the probability of being activated according to the probability model, when P (i, t) > 0.5 is satisfied, that is activated. That is, the threshold that is activated is P=0.5. At time t+1, the above process is repeated. After the communication simulation is completed, the proportion of activated nodes in the total nodes at each moment is calculated.

5. Data acquisition and fitting

5.1 Data Collection and Analysis: public opinion status of the first genetic editing baby event

The first genetic editing baby event is the actual data sample for model validation in this study, so it is first necessary to have a grasp of the overall public opinion status of the event.

The main content of the event is:

On November 26, He Jiankui, a scientist from Shenzhen, China, announced the day before the 2nd International Human Genome Editors Summit that a pair of genetic editing babies named Lulu and Nana were born in China in November. The genes of the twins have been modified to make them naturally resistant to AIDS after birth. This is the world's first genetic editing baby with immune AIDS.

——People’s Daily online reported on November 26, 2018

The incident involved a large-scale discussion and controversy in a short period of time due to medical risks, genetic engineering, potential threats to the human gene pool, and huge ethical issues, and the overall outbreak of fermentation within 36 hours after the outbreak state of the incident. There is no obvious tendency to reverse. This study selects Sina's official big data public opinion platform for event data, and collects and processes the information of the event. The number of Microblog information changes as shown in Figure 2:

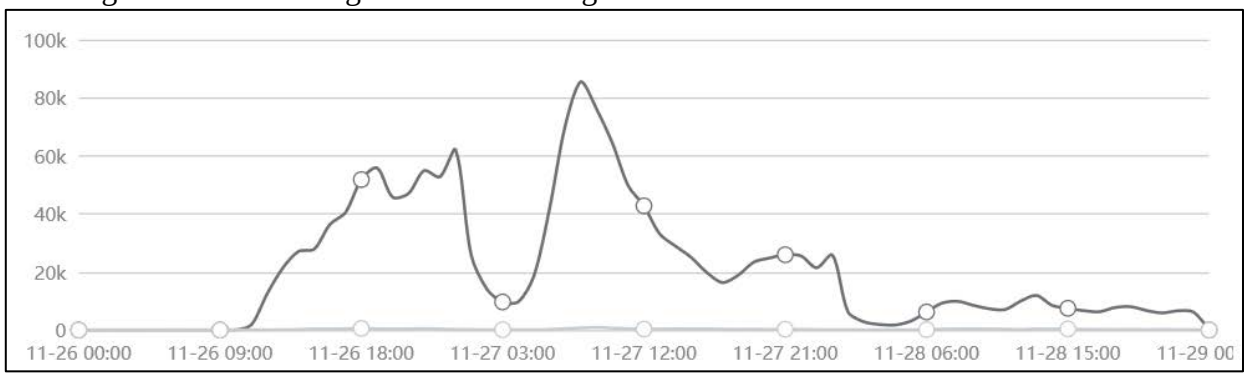


Figure 2. Changes in the number of Microblog information from November 26th to November 29th

It can be seen from the figure that during this period of time, the change in the amount of information in the Microblog as a whole is basically consistent with the changes in the news media reports. Within 72 hours, the total amount of information was 1579,096.

The event started around 10:00 am on November 26, and the amount of information disseminated increased rapidly thereafter. On November 27th, the information valley area between 0:00 and November 27th and 7:00 is the night time period, and after 7:00, it starts to increase according to the previous rising value. The peak area is from 8:00 to 10:00 in the morning of November 27, and the number of Microblog information reaches the highest peak point during the hour from 8:30-9:30, a total of 85,694. Since the model of this study predicts the rapid communication of information until the communication of the peak point activates the communication condition, the sample time period needs to select the time segment from the origin to the information fall. Based on the overall statistical results of the total time of the event, the time point of observation event from the source to the rapid communication period needs to be selected around 10 o 'clock on November 26, so the interception time is 09:55 to 09:55 on November 27 as the sample time period, which is the communication heat growth period.

Source Microblog is: "The Associated Press: He Jiankui, a researcher at Southern University of Science and Technology, announced the completion of the world's first genetic editing baby." (November 26, 2018, 10:52:02)

In this incident, the total number of Microblog is 884,359. At the highest peak point, the number of Microblog is 85,694. During this period, the changes of newly added information on Microblog are shown in figure 3:

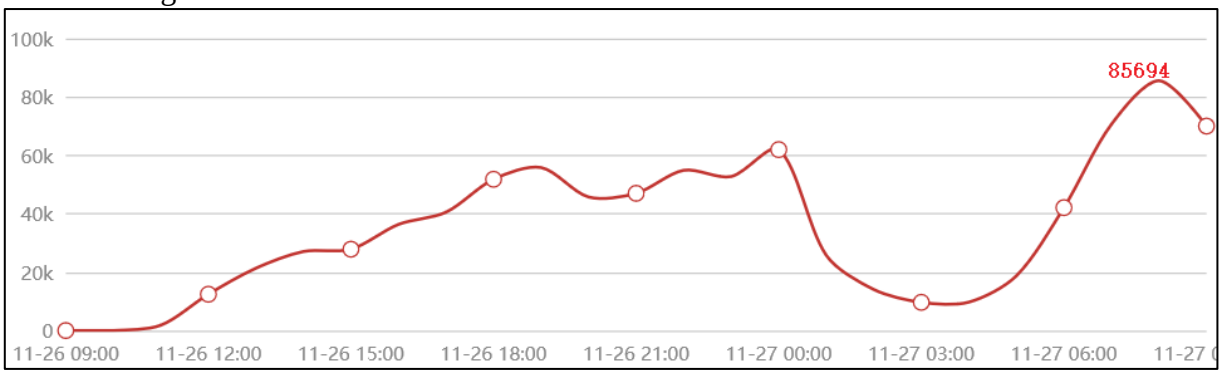


Figure 3. Microblog information increments during the sample interception time period

During the sample interception period, the number of sensitive attribute Microblogs was 386,241, accounting for 43.67% of the Microblog information. The quantity change is shown in Figure 4.

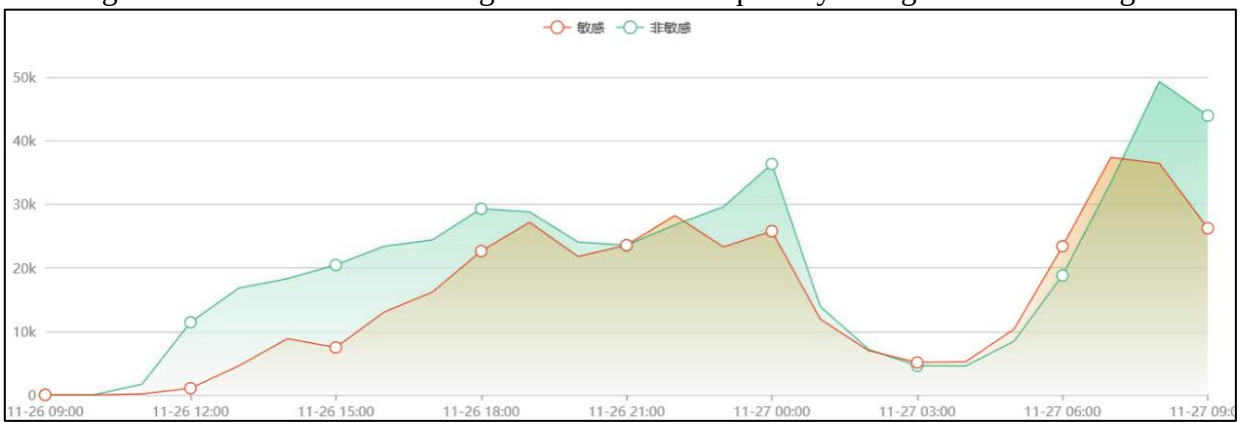


Figure 4. Changes in the number of sensitive/non-sensitive information during the sample time period

5.2 Simulation experiment

According to the Microblog data of the whole network, its network structure conforms to BA scale-free network characteristic. The fan number distribution (that is, the node degree distribution) is fitted, and the distribution is found to be consistent with the Pareto distribution, which is consistent with the characteristics of the BA scale-free network. Furthermore, we also fit the fan distribution to the users involved in this event, which is also consistent with the characteristics of the BA scale-free network. The result of the fit is:

$$y=1503491*x-2.34$$

Where, x is the number of fans and y is the corresponding number of fans. Goodness of fit R2 = 92.4%. In this incident, 1775307 people were involved in the communication. Considering factors such as computer performance and computational security, a BA scale-free network is constructed with a node number of 10000 and a power exponent of -2.44. After constructing the network, set the node attribute, that is, the node influence. Set to a random amount that fluctuates around the constant with a constant of 0.25.

The event broke out between 11:00 on November 26 and 23:59 on November 27, 2018. Select the original user who is not forwarded from 11:00-11:59 on the 26th, a total of 12 digits, mapped to

the constructed network. The mapping method was to map the ranking of the number of fans of these original users in this communication network to the corresponding ranking nodes in the node degree of the constructed network.

5.3 Simulation Results

As shown in Figure 5, the activation of some nodes (degrees > 10), and the active node is red. The network communication simulation of the first 4 degrees > 10 nodes is as follows.

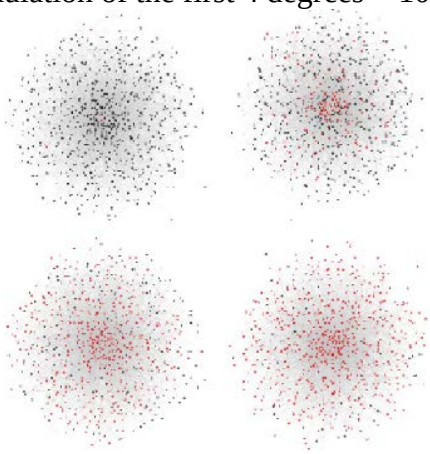
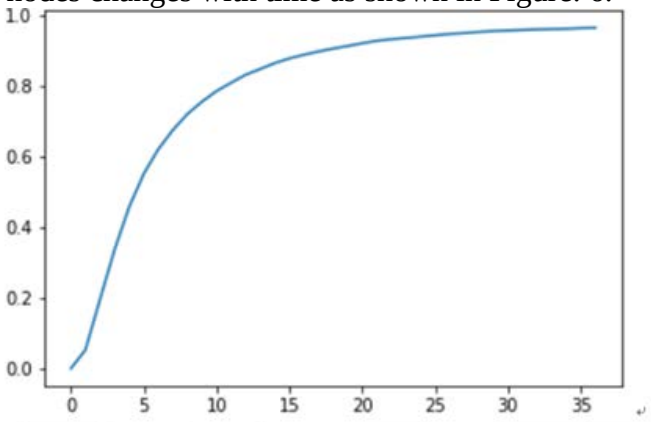


Figure 5. Partial node network communication simulation in the first four times

After 4 communications, it spreads to a larger range. At the same time, the proportion of activated nodes to total nodes changes with time as shown in Figure. 6.



0	1	2	3	4	5	6	7	8	9
0.10%	5.35%	19.64%	33.71%	45.76%	55.14%	62.11%	67.55%	72.09%	75.58%
10	11	12	13	14	15	16	17	18	19
78.57%	80.92%	83.19%	84.82%	86.48%	87.80%	88.85%	89.76%	90.58%	91.32%
20	21	22	23	24	25	26	27	28	29
92.08%	92.81%	93.21%	93.59%	93.96%	94.35%	94.73%	94.96%	95.30%	95.55%
30	31	32	33	34	35	36			
95.70%	95.87%	96.01%	96.10%	96.19%	96.37%	96.45%			

Figure 6. Changes of the relationship between the proportion of activated nodes in total nodes and the communication times

From the simulation results, it is seen that the number of communications has gradually increased exponentially in the early period, and the speed of communication has gradually slowed down. After 36 times of communication, the network coverage reached 96.45%. The growth trend of the simulation curve is in good agreement with the overall trend of the actual situation described in FIG. 4 (excluding the period of low activity at night). The fluctuation in the actual situation can be regarded as the disturbance of other factors.

5.4 Data fitting summary

Overall, the simulation results are consistent with the actual situation. But in terms of effect, there is still an inevitable deviation between the simulation results and the actual situation. The main reasons for these deviations value are: First, when the initial communication source is set in the experiment and mapped to the simulation network model according to the number of its fans, it should be considered from the whole network, but since the data of this scale cannot be known, A rough estimation method is therefore used, which has an impact on the communication rate (Especially the rate at which the communication begins). Second, because the model is mainly based on the risk value of the event and the user's risk perception, due to the limitation of space, other influencing factors such as the user's communication habits and preferences, and the characteristics of the subject and object of communication are not added. This is where the model itself needs improvement. In the psychological experiment, the subject observed the event, Microblog belongs to the ideal state, which controls the degree of communication involvement of the subject, and likes other variables such as the number of forwarding, but the variables in the actual Microblog communication are more complex than the simulated state, so some deviations are caused. Third, in actual communication, communication delay and time measurement are also very important factors, in this experiment, communication evolves indiscriminately according to the number of times, and thus there will be some visual errors in the result curve.

6. Research conclusions

To observe whether the risk value of an event has a very important influence on the communication willingness of users, the user's perception of the event risk value needs to be used as an intermediate variable, and the two are connected in series, and There is a logistic curve relationship between the backward and the communication willingness, and the forward and the risk value of an event have a linear correlation after taking the logarithm. In addition, the communication situation is also related to network structure, influence of communicators contacted, and the source of the communication, etc. For the communication phenomenon of this kind of event, this paper cuts into the model from the perspective of risk perception, and the increase improves the risk perception dimension of the future model.

In the aspect of risk perception measurement, the relationship between risk perception status and communication willingness is obtained. In terms of model, the information communication model of controversial scientific and technological events on Microblog is established on the basis of risk perception and under the circumstance of random disturbance of other communication influencing factors, and the effect is better. Under the condition that the risk value of an event and the communication environment are known, the model can predict the event communication better. The results of the simulation experiment are basically consistent with the actual communication situation, and the model establishment is basically established. When some major events break out, the relevant departments can timely use the improved model for event communication prediction, and take corresponding measures according to the communication willingness of different users.

At the same time, it is also necessary to see that the model is mainly based on the risk value of the event and the user's risk perception, owing to space limitations, other influencing factors such as the user's communication habits and preferences, and the characteristics of the communication between the subject and the object are not added. Therefore, the model established in this study needs to be improved in terms of prediction accuracy. In this study, the improved model scheme is a preliminary prototype, which can greatly improve the practical application effect of the model after improving the measurement scheme and simulation means.

References

[1] Meng Bo, Liu Mao, Li Dongxue. Quantitative calculation method and application of public risk perception [J]. Journal of Safety and Environment, 2011, 11 (06): 219 - 222.

[2] Miao Weishan, Jia Hepeng. Scientific Communication: Dissolving Reality Disputes and Its Research Prospects —— Discussing the Hot Spots of Science Communication [J]. Journalist, 2016 (12): 44 - 51.

[3] Lin Qin, Guo Dongqiang. Research on Social Network Public Opinion Communication Based on Optimized SIS Model——Based on User Psychological Characteristics [J]. Information Science, 2017, 35 (03): 53 - 56+75.

[4] Cai Shuqin, Yuan Qian, Zhou Peng. Study on the negative threshold linear propagation model under enterprise response [J]. Journal of Systems Engineering, 2017, 32 (02): 145 - 155.

[5] Cao Ning. Research on network public opinion government intervention based on stratified infectious disease model [D]. Guangdong University of Technology, 2018.

[6] Wang Jiajie. Research on risk communication of environmental group events under the perspective of social media [D]. East China Normal University, 2018.

[7] Li Liang. Research on topic discovery and heat prediction of microblogging esoteric events [D]. Xi'an University of Technology, 2018.

[8] He Xing. Improvement and verification of microblog influence propagation model [D]. University of Science and Technology of China, 2014.

[9] Zhao Dan. Research on microblogging public opinion communication in mobile environment based on information ecological theory [D]. Jilin University, 2017.

[10] Emily Ying Yang Chan, Holly Ching Yu Lam, Phoebe Pui Wun Chung, Zhe Huang, Tony Ka Chun Yung, Kelvin Wai Kit Ling, Gloria Kwong Wai Chan, Cheuk Pong Chiu. Risk Perception and Knowledge in Fire Risk Reduction in a Dong Minority Rural Village in China: A Health-EDRM Education Intervention Study [J]. International Journal of Disaster Risk Science, 2018, 9 (03): 306 - 318.

[11] Wang Zhiying, Liang Jing, Liu Xiaodi. A review of public risk perception research in emergencies [J]. Journal of Information, 2018, 37 (10): 161 - 166.

[12] Wang Lingning. An Empirical Study on the Influence of Media Science Communication on College Students' Scientific Literacy——Taking Transgenic Communication as an Example [J]. Journal of Information University, 2018 (05): 77 - 83+149.

[13] Cheng Junhui, Zhao Jinlou. Research on Social Network Public Opinion Communication Model Based on Information Risk Perception [J]. Journal of Information, 2015, 34 (01): 134 - 138.

[14] Wang. Effects of Brand Awareness and Social Norms on User-Perceived Cyber Privacy Risk [J]. International Journal of Electronic Commerce, 2019, 23 (2).

[15] Estevan, García-Massó, Molina García, Barnett. Identifying profiles of children at risk of being less physically active: an exploratory study using a self-organized map approach for motor competence [J]. Journal of Sports Sciences, 2019, 37 (12).