

# STUDY OF THE EARLY WARNING MODEL FOR NETWORK PUBLIC OPINION RISKS BASED ON THE INTEREST PREFERENCES OF INTERNET USERS

Xiao Li\*, Sarah Bauch

*School of Information Management, Nanjing University, Nanjing 210023, China.*

**Abstract:** In the context of big data, from the perspective of network ecology, this article divides Weibo public opinion subjects into four groups based on their interest motivations: image managers, platform operators, public opinion consumers and network promoters. According to the four groups Extract Weibo public opinion risk factors based on the interest preferences of each community to construct a Weibo public opinion risk model. The model was verified using Weibo real-time data as a sample, and through logistic regression analysis, it was concluded that the prior cognition, information preference, behavioral preference, information person preference and other interest preferences of the Weibo public opinion subject community have an impact on the development trend of Weibo public opinion. The degree of influence coefficient is used to calculate the early warning level, thereby obtaining a Weibo public opinion early warning model that can be dynamically adjusted according to case updates. The research has important theoretical significance for deepening and enriching the research on online public opinion, and at the same time provides new methodological support for realizing public opinion risk early warning.

**Keywords:** Internet public opinion; Logistic analysis; HHM risk framework; Risk modeling

## 1 INTRODUCTION

Currently, the Internet has entered the era of big data, and Weibo has become a mainstream application of communication media. Different from traditional media, Weibo is based on nodes and uses social networks to spread information. Each node has the ability to comment, forward, etc. on information. In this mode, the development of public opinion on Weibo

Development is largely affected by the personal interests, preferences and behavioral habits of Weibo users. Therefore, this article intends to start from the interests and preferences of netizens in the use of Weibo, establish an HHM Weibo public opinion risk early warning system, and establish a Weibo public opinion risk early warning model.

This article disassembles Weibo public opinion into four Weibo public opinion subjects, and establishes a Weibo public opinion risk early warning model based on their preference characteristics. It is elaborated from the four aspects of prior cognition, information preference, behavioral preference and information person preference of the four main bodies: image management main body, platform operation main body, public opinion consumption main body and network promotion main body.

### 1.1 Public Opinion Subjects

This article divides Weibo public opinion subjects into four categories based on their different interest demands. Among them, the main body of public opinion consumers is a group whose purpose is entertainment and Internet surfing. Image management subjects are groups whose purpose is to safeguard their own interests and create a positive image. The platform operating entity is a community that provides messaging services to other communities. The main body of online promotion is a community that actively promotes and creates public opinion for the purpose of profit [1-2].

### 1.2 Prior Knowledge

Prior knowledge refers to Weibo users' basic interest in a certain type of public opinion, that is, users' natural points of interest without the influence of public opinion. Internet hot spots are the mapping of hot social events on the Internet. Netizens with similar interests tend to pay more attention to the same type of events. Therefore, the keywords in Weibo information are a manifestation of netizens' prior cognition.

### 1.3 Information Preferences

Information preference refers to the degree of preference of Weibo users for Weibo information topics and forms. Due to different identity backgrounds and knowledge composition, they have different acceptance levels of different information formats and information topics. Therefore, the topic classification and information format of Weibo information are a manifestation of information preferences.

## 1.4 Behavioral Preferences

Behavioral preferences refer to inertial behaviors that express attitudes, emotions, cognitions, and opinions. Driven by different psychological backgrounds and interests, Weibo users express their emotions by forwarding, commenting, and liking Weibo information. Therefore, the number of Weibo forwardings, comments, and likes reflects the behavior of netizens to a large extent. preferences.

## 1.5 Informant Preferences

Informant preference refers to Weibo users' preference for public opinion groups and the degree of influence by ethnic groups. Since Weibo has two functions: "fans" and "follows", different "ethnic groups" are formed because of their different positions and points of interest, and interactive discussions are conducted based on ethnic groups. Ethnic groups will also influence them through this mutual fan relationship. Therefore, the number of fans and followers of Weibo users is a reflection of the information person's preference.

## 2 HHM METHOD RISK FACTOR IDENTIFICATION

### 2.1 HHM Method

The HHM method specifically refers to decomposing a complex system or process into multiple levels or levels. Each level or level is a complete system or process from a specific perspective, so as to understand the different levels of the system or process. question. This method is very effective for modeling large-scale, complex structural systems. Weibo public opinion covers a wide range of areas, has a lot of content, and has complex influencing factors. Therefore, this article will identify the risk factors of the Weibo subject community based on HHM.

### 2.2 HHM Framework for Weibo Public Opinion Risk Identification

This article focuses on the risk identification of Weibo subject communities, analyzes the risk factors of Weibo public opinion from three aspects: subject community, preference and behavior, and constructs the following HHM sub-framework for risk identification as a direct framework model for risk identification .

## 3 DETERMINATION OF RISK FACTORS OF WEIBO PUBLIC OPINION

Starting from the four perspectives of the subject layer, this article analyzes the behavioral preferences of different subject groups, thereby analyzing their possible behavioral activities and extracting the risk factors of the Weibo public opinion early warning model, and finally conducts modeling analysis through their behavioral data.

### 3.1 Weibo Public Opinion Risk Indicator System

According to HHM's analysis of Weibo public opinion from four perspectives, the Weibo public opinion risk indicator system has two layers, the preference layer and the data monitoring layer. The preference layer includes four preferences: prior cognition, information preference, behavioral preference, and information person preference. The data monitoring layer includes 8 basic indicators such as F1 topic category, F2 fan effect, F3 headline account, F4 information format, F5 forwarding volume, F6 like volume, F7 comment volume, and F8 topic post count.

### 3.2 The Main Perspective of Image Management

The image manager's prior cognition and information preference are reflected in the attention to the topic, of which the information format can be ignored. The image manager's external publicity department often focuses on its own related topic categories and keyword searches. The search keywords of each image manager are different, so the risk factor F1 topic category is extracted. The number of image managers accounts for a small number of the total Weibo users, so their behavioral preferences have little impact on the overall Weibo. However, due to fan benefits, image managers are often industry leaders in a certain industry and have certain authority, so their influence is mainly reflected in the number of image managers. on its fan appeal. So the risk factor F2 fan effect was extracted [3-5].

### 3.3 Perspective of Platform Operating Entity

The number of platform operators accounts for a small number of the total Weibo users, so their behavioral preferences have little impact on the overall Weibo. However, due to the resonance among platform operators, a Weibo public opinion spreads from a single media platform to a multimedia platform, making Weibo Public opinion amplifies like ripples. So the risk factor F2 fan effect was extracted.

### 3.4 The Perspective of Public Opinion Consumers

The large number of public opinion consumers is the basis for the development of Weibo public opinion. The factors that affect the interests of public opinion consumers are the factors that affect Weibo public opinion. First, Weibo users' browsing of a certain topic can reflect the user's prior cognition and information preferences. The topic category of public opinion, sensitive keywords, being influenced by surrounding users or attracted by media headlines, and the information format of the public opinion all have a great impact on Weibo users' browsing choices. Therefore, the influencing factors F1 topic category, F3 headline number, and F4 information format are extracted.

Public opinion consumers have fewer fans, so their posting behavior can only affect their own small circle and does not need to be considered. And their behavior of forwarding, liking, and commenting on other people's Weibo can promote the development of that Weibo. Therefore, the influencing factors F5 retweets, F6 likes, and F7 comments are extracted. The fans of public opinion consumers are often other users in the real social network circle. The number is small and will not be considered for now.

### 3.5 The Main Perspective of Online Promotion

Internet promoters use Weibo public opinion to achieve their own goals, so there is no prior cognition, information preference and information person preference. As for their behavioral preferences, they often put hot topic names in the title of their posts to increase their own views, or to increase the number of retweets and likes in order to promote the development of a certain public opinion. Since the number of retweets and likes is difficult to distinguish between online promoters and public opinion consumers, this article only considers the behavioral preference of posting hot topics. This behavior can lead to an increase in the number of posts on a certain topic, but a smaller number of truly relevant posts. So the number of influencing factors F8 topic posts was extracted [6-8].

## 4 CONSTRUCTION OF RISK EARLY WARNING MODEL BASED ON LOGISTIC

The risk early warning system consists of three parts, namely the case library indicator monitoring subsystem, the risk identification subsystem and the early warning and forecasting subsystem. First, build a case library and build an indicator monitoring subsystem. For different types of cases, there are natural differences in the interests and preferences of Weibo users. In order to narrow this difference and improve the early warning level of the early warning system for such cases, it is necessary to replace and supplement the database cases based on research, so that the case database cases remain highly homogeneous. The indicator monitoring subsystem is divided into monitoring layer indicators and description layer indicators. The monitoring layer tracks and records 8 basic Weibo public opinion data, and calculates 4 characteristic layer indicators through content mining, structure mining and behavior mining. Characteristic layer indicators can add mining data and determine the weight of each indicator based on the importance of the four communities in such cases, thereby determining the monitoring layer indicators that need to be mined [9-11]. Secondly, determine the risk identification threshold and build a risk identification subsystem. The risk identification subsystem is divided into six sub-modules: early warning setting, indicator setting, weight interval setting, calculation, comparison and discrimination, and alarm. Among them, the early warning setting module selects the monitoring description layer indicator object according to the case type. The indicator setting module edits the description layer indicator algorithm for the early warning setting module. The weight interval module determines the indicator selection range of the indicator setting module. The calculation module combines the indicators. Perform calculations to obtain the final early warning monitoring indicators, determine the early warning level through the early warning setting module, and finally input the calculation results into the early warning and forecasting subsystem [12-15].

Finally, the early warning and forecasting subsystem takes the highest number of views of the case library case as a reference and is set to 1. Convert the pageviews of other cases to the highest pageviews into a percentage, which is the popularity ratio of the case. Use the popularity ratio as the exposure rate to perform logistic regression analysis on the monitoring layer indicators, and calculate the OR value of the indicator in the process of forming public opinion on Weibo, thereby determining the warning level warning value of each early warning indicator, based on the early warning preset by the risk identification module. Forecasts are issued based on level.

### 4.1 LOGISTIC Risk Early Warning Model

The logistic model is suitable for situations where the explained variable is a dummy variable (that is, the value is 1 or 0). In a statistical sense, this model generally measures the probability when an event occurs (the value is 1) through the logarithmic distribution of the function. The model expression is shown in Equation 1:

$$\ln\left(\frac{P}{1-P}\right) = g(x) \quad (1)$$

Among them  $g(x) = w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n$ , P is the probability that the dummy variable takes 1, and 1-P is the probability that the dummy variable takes 0. X1-Xn are factors affecting dummy variables. W1-W nThe coefficient of

each influencing factor. This article mainly analyzes the OR value of this model to determine the degree of influence of each factor on the dummy variables.

#### 4.2 Indicator Selection

In the research, the reliability of the model depends to a large extent on the quality of the early warning analysis indicators. Considering that the indicators collected directly using Weibo alone are too large, which may lead to research distortion, this article combines the number of likes, the number of retweets, 8 bottom-level indicators such as page views are calculated, and secondary indicators are obtained for calculation. Specifically, the Weibo public opinion indicators selected in the study are shown in Table 1.

**Table 1** Variable assignment method table

Variable Code	Variable Name	Assignment Method
X1	fan effect	Weibo public opinion internet celebrity index
X2	Toutiao account	Weibo public opinion headline index
X3	message format	Image-based (photos, short videos) is 1; (comics, radio) is 0 text-based
X4	Forward volume	The total number of retweets of all posts under this topic
X5	Number of likes	The total number of likes for all posts in this topic
X6	Number of comments	The total number of comments on all posts under this topic
X7	Number of topic posts	Original posts posted by participating users on this topic
y	Public opinion popularity value	Take the day's public opinion high page views as a reference value

#### 4.3 Input Layer Settings

This article uses multinomial logistic regression in SPSS 18.0 to analyze the impact of various factors on public opinion. The input layer settings are shown in Table 2.

**Table 2** SPSS 18.0 input layer settings

Options	Input Value	Options	Input Value
dependent variable	Hot public opinion Weibo ranking	Classification	Classification code: 1 Critical value: 0.5
covariate	X1-X6	save	Predicted value: probability Residuals: Standardized contains the covariance matrix

The analysis path is as follows:

(1) First, calculate the quartile level division of Y. SPSS 18.0 gives the degree of quartile analysis, namely Analysis → Descriptive Statistics → Frequency. After the ratio is obtained, Y is expressed as a categorical value, that is, 1 represents high, 2 represents relatively high, 3 represents medium, and 4 represents low. This article mainly studies the impact of risk factors on public opinion popularity, so this article uses 1, the highest value of public opinion popularity, as the reference variable.

(2) To perform multinomial logistic regression analysis, in SPSS 18.0, enter Analysis → Regression → Multinomial logistic regression.

#### 4.4 Interpretation of Results

This article explains the output values of each factor. The list of output values that need to be explained is shown in Table 3.

**Table 3** SPSS 18.0 output layer explanation table

Output Value	Explain	Output Value	Explain
	OR value: describes the independent variable versus the dependent variable		
Exp(B)	Quantity of influence. The OR value is greater than 1, which is a promoting variable. If the OR value is less than 1, it is a protected variable.	Sig	P value: P<0.05 is significant; P>0.05 is not significant

## 5 EMPIRICAL ANALYSIS

### 5.1 Data Acquisition and Processing

This article selects the data object as social topics on Sina Weibo, and the data acquisition method is JAVA programming technology. Starting from the original Weibo, it traverses to obtain the following field data: user ID, user name, information content, user attributes, number of forwarded comments and likes, status Update time, etc., and establish a database of social public opinion cases.

As of 21:00 on March 27, 2017, starting from 1218 original Weibo posts on 47 public opinion topics, a total of topic categories, hosting media IDs, number of posts, number of fans of hosting media, information content, and number of likes were obtained, page views and other 8 types of data. Weibo Public Opinion is a public opinion monitoring platform. Users can customize all-media public opinion monitoring information and reports including terminals by combining keyword schemes. This article uses the Weibo public opinion platform to obtain the public opinion index. The acquisition source type is set to Weibo, and the time span is from 0:00 to 24:00 on March 27, 2017.

Data sorting is carried out based on the obtained basic data, mainly including: removing duplicate values and missing values; based on the number of content forms such as pictures, videos, voices, texts, expressions, etc. in the information content field, the corresponding information format preference fields are assigned values of 0 and 1; The cleaned data was numerically calculated according to the research variable formula; the public opinion popularity ratio was calculated; and finally a Weibo public opinion early warning case library matrix with 47 rows and 8 columns was constructed.

### 5.2 Case Selection

This article selected 47 social Weibo public opinion cases that occurred on March 27, 2017 on Sina Weibo for risk early warning analysis to test the interests of image maintainers, platform operators, public opinion consumers and online promoters in the early warning analysis. The accuracy of the impact of preferences on Weibo public opinion.

### 5.3 Analysis Results

This article uses the method of logistic regression analysis to conduct empirical analysis using 47 social case modeling of Sina public opinion, sorting out the interest preferences of image maintainers, platform operators, public opinion consumers and online promoters in social Weibo public opinion. The influence of public opinion on Weibo. Based on the model calculation results (such as Table 5), the corresponding conclusions are drawn:

(1) Based on the role of image maintainers, the following conclusions are drawn: The degree of radicalization of internal netizens changes significantly higher than the degree of personal profit-involvement of netizens. Due to the "group instinct", individual netizens will subconsciously coordinate with each other to achieve consistency of will and behavior. In addition, the current Internet provides virtual identities for netizens, making netizens express their opinions more freely and the "group effect" is more significant. .

(2) Based on the effect of platform operators, it is concluded that the influence of the media is higher than the persuasive effect of the media, and the persuasive effect of the media is significantly higher than the influence of the emergency itself. The main way for netizens to obtain information is through media reports. The media has a direct impact on netizens' emotions by processing and rendering event information. Netizens tend to pay attention to the large number of media reports and forget the reality and ignore the information source itself.

**Table 4** Weibo public opinion data statistics table

Serial Number	Browse	Participate	Forward	Comment	Like	Posting Media	Focus On	Fan
1	8156000	4	138	1555	587	Gao Ji Chinese website	418	3710
2	63148000	10	2943	21798	12475	Reporter Zhang Zhizhong	1827	3346
3	1101000	3	370	95	1404	Gansu website registration	208	242
4	897000	3	624	4073	2641	Capital Experts Weibo244 Group		80
5	1984000	2	37	129	88	Chongqing Morning News	1577	840000
6	4438000	1	163	481	315	China News Weekly	2339	37670000
7	21776000	8	3899	8594	11457	Jiangsu News	852	2780000
8	310000	3	46	76	20	Nanjing zero distance	1075	3460000
9	634500	1	9	18	7	Suzhou micro life	1570	810000
10	157000	2	13	23	12	Legal Network	753	470000

11	240000	1	92	137	73	modern express	1049	1900000
12	291000	1	0	1	2	Hefei Lifestyle	150	180000
13	250000	1	32	51	16	Sina Jiangsu	3834	820000
14	107000	1	32	82	298	personal	2340	7040000
15	2881000	2	3	55	148	Zhejiang Evening News	1254	100000
16	386000	1	22	169	201	West China Metropolis2004 Daily		4020000
17	578000	4	110	594	130	Sina Jiangxi	2974	52000
18	43680	1	5	2	17	personal	120	128
19	527000	1	175	140	557	personal	14	10
20	39751	1	0	15	55	Impression Yantai	495	110000
21	5739000	2	1	7	11	Today's headlines	1101	45
22	57420	1	1	0	9	Suzhou Fire Protection	195	15
23	684000	1	15	15	42	international online	926	4140000
24	6827000	6	10	69	15	chinaxiaokang.com	1254	100000
25	438000	6	15	29	66	personal	104	350000
26	450000	2	88	95	188	Queen of Stocks	2995	27879
27	578000	1	53	26	102	Jinding Strategy Rubik's92 Cube		20527
28	158900	1	1	5	10	Ancient sages discuss the165 present		427
29	264000	3	67	22	297	Shuishui Liu Xuan	301	5521
30	49000000	12	9026	15405	2872	Sina Sichuan	3287	1480000
31	5720000	11	4623	67014	66234	Sichuan Public Security396 Bureau		290000
32	839000	2	82	798	138	Released in Chengdu	860	6320000
33	26400	3	1618	8917	4042	CCTV News	1975	48900000
34	313000	3	6324	23473	12852	personal	2606	530000
35	93800	1	68	154	81	Refuting rumors and the1628 truth		350000
36	1573000	2	38	260	132	personal	2064	8479
37	86000	2	123	124	945	personal	535	5350
38	300000	2	3048	16826	9465	Sina Education	1910	4500000
39	1751000	7	1509	5910000	50354	beijing morning news	1831	3238
40	69000	1	13	252	83	Sina Justice	3702	34493
41	9401000	8	1062	3757	1513	China News Network	647	3143
42	415000	1	53	97	129	Sina stocks	1621	2640000
43	879000	1	2	5	4	Xiamen Evening News	344	2270000
44	775000	2	22	43	23	modern express	1052	904
45	1291000	3	30	120	33	Yangzi Evening News	1568	1241
46	587000	1	14	44	12	Sina Xiamen	4001	127
47	377000	2	111	199	162	China News Weekly	2421	37650000

(3) Regarding the role of government work, the conclusion is drawn: The effect of the government in handling crises is significantly higher than the role of the government in guiding emotions. When the government is dealing with a crisis, it can speed up its response, improve its work transparency, and enhance its credibility, which can reduce the intensity of the spread of netizens' emotions and weaken the crisis more than just focusing on online reports and making false reports.

(4) Based on the effect of online promoters, it is concluded that online promoters have a greater influence. It is appropriate for online promoters to "fill", "hype" a single topic, or "spoil" a hot topic by placing commercial advertisements. The fever of online public opinion has a great impact. To "reduce" the fever of online public opinion, it is necessary to further regulate the behavior of online promoters.

## 6 CONCLUSION

The rules summarized in this article are applicable to social Weibo public opinion events, but the model does not extract all risk factors in Weibo public opinion. There are still many areas that need improvement, and further work needs to be carried out.

- (1) This article selected 47 cases for verification using the logistic model, but the number of cases selected was small, and the convincingness of the proof needs to be strengthened.
- (2) The risk warning model of this study is based on the behavioral interests and preferences of netizens. The extracted indicators need to be refined based on psychology, behavior, etc. The determination of variables and their weights also needs to be based on cases. Adjustment, and the disadvantage of this method is that the update speed of the case library restricts the prediction accuracy, making the prediction results inaccurate.

**Table 5** Parameters in the equation of model calculation results

	B	SE	WALD	DF	SIGNIFICANCE	EXP(B)
Step 1 User identity						
message format	0.283	.000	.000	1	.987	1.920
Number of fans	0.178	.000	.536	1	.464	1.510
Toutiao account	0.082	.000	.495	1	.482	1.208
Forward volume	0.129	.001	.833	1	.362	1.347
Number of comments	0.053	.000	.419	1	.518	1.130
Number of likes	0.076	.000	1.055	1	.304	1.193
Number of topic posts	0.231	.279	3.650	1	.056	1.703
Step 2 Constant	1.288	.790	2.659	1	.103	.276

In short, the prediction research of online public opinion still has a lot of room for development in the future, and the research on the combination of netizens' interest preferences and public opinion needs to be further deepened.

## COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

## REFERENCES

- [1] Yuan Ye, Lan Yuexin, Xia Yixue. Research on public security Weibo influence enhancement strategies for online public opinion. *China Public Security (Academic Edition)*, 2018(01):85-91.
- [2] Liu Qi. Research on the role of government Weibo in emergencies. Jiangxi Normal University, 2013.
- [3] Du Yangqin, Huo Youguang, Suo Zhihai. Empirical analysis of the micro-social network structure of government Weibo - based on the perspective of structural hole theory. *Journal of Intelligence*, 2013, 32(05): 25-31.
- [4] Du Yangqin, Huo Youguang, Suo Zhihai. Analysis of the social network structure of Weibo based on complex network modularization - taking the "Shanghai Release" government affairs Weibo as an example. *Library and Information Knowledge*, 2013, (03): 81- 89+121.
- [5] Wang Xiwei, Xing Yunfei, Zhao Dan, Li Jiaying. Research on the dissemination of online public opinion information in the mobile environment based on social network analysis - taking the "smog" topic on Sina Weibo as an example. *Library and Information Work*, 2015, 59(07): 14-22.
- [6] Liu Xiaoyan, Zheng Weixiong. Analysis of the effectiveness of corporate social media marketing communication - taking the Weibo diffusion network as an example. *Journalism and Communication Research*, 2015, 22(02): 89-102+128.
- [7] Liu Lucuan, Sun Kai. The relationship between emotional experience and satisfaction of social media users - taking Weibo as an example. *Journal of Library Science in China*, 2015, 41(01): 76-91.
- [8] Xu Zhiming, Li Dong, Liu Ting, Li Sheng, Wang Gang, Yuan Shulun. Similarity measurement of Weibo users and its application. *Journal of Computer Science*, 2014, 37(01): 207-218.
- [9] Zhang Xue. Research on content marketing of paper media on Weibo platform. Shanghai International Studies University, 2014.
- [10] Wang Yimin, Ding Yi, Zhu Ruixun. Government Weibo: Opening up a new e-government ecosystem. *Journal of Yunnan Administration Institute*, 2013, 15(06): 78-81.
- [11] Zheng Jie. The role of Weibo in promoting the development of traditional media. *China Municipal Newspaper*, 2013, (09): 30-31.
- [12] Chen Jiangliu. Weibo for political inquiry: Analysis of Weibo application based on e-government. *News World*, 2013, (07): 158-160.
- [13] Zhu Chenfei, Huang Shuhua, He Hangsong, Kang Youchun. Internet public opinion crisis early warning based on BP\_Adaboost algorithm. *China Public Security (Academic Edition)*, 2017(04):95-101.
- [14] Tan Yizhen, Liu Mo, Li Yang, Zhu Lin. Research on risk assessment of online public opinion in border areas. *China Public Security (Academic Edition)*, 2017(04):107-113.

- [15] Nominzaya, Lan Yuexin. Empirical analysis and countermeasure research on the government's response to online public opinion in violent terrorist incidents. *China Public Security (Academic Edition)*, 2017(01):105-111.