

Research Article

Novel Network Public Opinion Prediction and Guidance Model Based on “S-Curve”: Taking the Loss of Contact with “Malaysia Airlines”

Xiangdong Liu ¹, Axiao Cao ¹ and Chuyang Li²

¹School of Economics, Jinan University, Guangzhou 510632, China

²College of Computation Science, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China

Correspondence should be addressed to Xiangdong Liu; tliuxd@jnu.edu.cn

Received 28 May 2021; Accepted 13 July 2021; Published 22 July 2021

Academic Editor: Weilin Xiao

Copyright © 2021 Xiangdong Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is of great significance for the government to control the network public opinion in time and maintain social stability to predict the network public opinion in emergency. This paper proposes a novel improvement method to “S-curve” theory in the context of big data and establishes three novel network public opinion prediction models. These models take into account the proliferation trend of initial and follow-up network public opinion over a long period of time when emergencies are formed and the objective environment suddenly changes, based on the information diffusion model conforming to the traditional “S-curve” theory. The novel improvement and establishment allow our model to have more accurate predictions than other scholars’ models that mainly study the first network public opinion in a shorter period of time. And it is more applicable to real social conditions, in line with the public’s cognition of reality, and provides more reference for the government to take preventive and corresponding positive guiding measures in advance. To better establish the model, we obtained the 24-day Weibo data associated with the incident of “Malaysia Airlines” loss of contact from big data for model establishment, public opinion prediction, and comprehensive evaluation. The result innovatively shows that, in addition to the initial public opinion that is worthy of attention, the follow-up public opinion is also noteworthy and proves that our model has more practical value.

1. Introduction

Public opinion is the abbreviation of “the situation of public opinion.” It is the summation of the beliefs, attitudes, opinions, and emotions expressed by many people about various phenomena and problems in the society. “From the perspective of traditional sociological theory, public opinion is a comprehensive reflection of public opinion and belongs to the category of public opinion theory. From the theory of public opinion in modern society, in addition to the simple generalization of the law of public opinion, public opinion is more focused on the description of “public opinion and its role and the rule of governing and political orientation” [1]. Network public opinion is a special form of public opinion; that is, network public opinion is a collection of feelings, attitudes,

opinions, and opinions of the majority of netizens, with the network as the carrier and events as the core. The expression of network public opinion is fast, the information is diverse, and the way is interactive. The openness and virtuality of the network also determine the characteristics of directness, arbitrariness, diversification, abruptness, concealment, and deviation of the network public opinion. Emergency is the inducing factor of network public opinion, as a kind of unconventional event, and it has the characteristics of instantaneous, accidental burst point, instant gathering, harm to the subject and society, and the crisis of the development trend. The sum of related comments, reports, opinions, and sentiments and attitudes formed in cyberspace after an emergency occurs is the network public opinion of the emergency, which is a special form of network public opinion [2].

Weibo, or microblog, is a broadcast-style social network platform that shares brief and real-time information through a following mechanism. In China, it is difficult for any public opinion that may spread across the whole society to circumvent Weibo as a communication carrier. Sina Weibo is currently one of the largest online platforms in China that can satisfy ordinary users to deliver information to the country. According to the annual financial results released by Weibo, as of December 2020, the number of monthly active users reached 521 million, an increase of about 5 million over the same period last year; the average number of daily active users reached 225 million, a net increase of about 3 million users over the same period last year. Sina Weibo has a large number of active users, a rapid growth in monthly active users, and a high degree of openness on the platform. Now, Weibo has become an important platform for ordinary people to express their views on news hotspots and has strong information dissemination and emotional contagion capabilities [3]. Therefore, Sina Weibo plays an important role in public opinion monitoring of emergencies.

Public opinion research is based on the research of information diffusion mode. Foreign scholar Hagerstrand was the first one to study the traditional information diffusion mode of geographical entity space. He pointed out that the accumulation and change of information adoption over time conformed to the S-shaped logistic curve. He also pointed out that the information diffusion of traditional geographic entity space has multiple effects and will be affected by communication and interaction [4]. Rogers elaborated on the “S-curve” theory in his book “Diffusion of Innovation,” combined with this theory, to point out that the innovation and diffusion of new things and new ideas require the use of certain social networks to achieve [5]. Fiona Duggan proposed the “crisis information dissemination mode” that combined the “S-curve” theory [6]. The existing domestic research on the network public opinion model for emergencies in China is mainly concerned with the initial public opinion in a short period of time. There is a lack of models on the emergence of unclear or undisclosed emergencies as the follow-up network public opinion of the event. After an emergency breaks out, the unknown nature of the matter may put people in a state of excessive speculation and psychological panic. The release of updated news may push network public opinion to a climax, and the scale of public opinion on the Internet at this time will even exceed the public opinion when the emergency breaks out. The fierce follow-up network public opinion is actually a reflection of the public’s attitude towards emergencies, which may be accompanied by or lead to extreme behavioral reactions, which seriously affect social stability. Therefore, the government’s supervision of follow-up public opinion is also crucial. According to the established novel model, the government can take corresponding measures to guide people’s thoughts and behaviors to avoid derivative emergencies.

Many studies have shown that the trend of information diffusion caused by the formation of the initial public opinion of an emergency is in line with the “S-curve” theory, but it is rarely discussed that the emergence of subsequent

public opinion leads to the diffusion of information. Therefore, this article mainly discusses the incident of the “Malaysia Airlines” loss of contact based on the corresponding 24-day Weibo data and defines the network public opinion of emergency, including the initial network public opinion and the follow-up network public opinion. Based on the traditional information diffusion model that conforms to the “S-curve” theory, this article further discusses the information diffusion model of the initial and subsequent network public opinion by adding factors that occur when the objective environment changes suddenly in cyberspace and emergency formation. That is, three public opinion prediction models are established based on the traditional curve fitting method to perform data fitting, and the original model is reasonably explained and improved based on the characteristics of the network public opinion and the actual situation during the fitting process. Then, we innovatively put forward the influence of follow-up public opinion on the network public opinion of emergent events. Then, the mean absolute percentage error (MAPE) was used to evaluate and compare three different models simultaneously, and MAPE is the average percentage of the sum of each absolute error (the absolute value of the deviation between the observed value and the true value) and the true value. And finally, it was concluded that, with the subsequent sudden changes in the objective environment, the traditional “S-curve” theory cannot be mechanically applied in the discussion of network public opinion. The network public opinion in the new environment should conform to the public’s perception of reality to analyze and predict objective development. Therefore, when predicting the network public opinion of emergencies, in addition to the initial public opinion, the follow-up public opinion is also worthy of attention. Moreover, by predicting network public opinion, the government can guide positive public opinion forecasting and analysis when emergencies occur to promote harmonious and healthy development and take timely measures against negative public opinion analysis and forecasting to control network public opinion and maintain social stability.

The research content of this paper is mainly divided into three parts. The first part introduces public opinion and Weibo big data platform and then determines the research direction and methods by combining the relevant literature of domestic and foreign scholars on public opinion information. The second part mainly explains the selection and processing of data, and this paper mainly analyzes the information diffusion model of network public opinion through the relevant public opinion data of the loss of contact with “Malaysia Airlines.” The third section is also the most important part, which introduces the theoretical improvement and the establishment of three types of models on the basis of “S-curve” and then makes a comprehensive analysis and evaluation of the three models.

2. Data Selection and Specific Processing

2.1. Selection of Data. The research data in this article comes from the official big data platform of Sina Weibo (see micro index: <https://data.weibo.com/index>). To better combine the

actual situation to discuss the network public opinion. This article selects the platform's data related to the loss of contact with "Malaysia Airlines" from March 8, 2014, to March 31, 2014 (24 days in total), as the data basis for the establishment of the model. At 2:40 am on March 8, 2014, Malaysia Airlines flight number is MH370. The plane that was originally scheduled to fly from Kuala Lumpur to Beijing lost contact. That day, it caused heated discussions with Weibo. At 10 p.m on March 24, 2014, Malaysian Prime Minister Najib announced in Kuala Lumpur that Malaysia Airlines' missing flight MH370 had crashed in the southern Indian Ocean. No one survived on the plane. This caused a Weibo heat on the same day; Figure 1 shows the single-day word frequency curve of the keywords "Malaysia Airlines" and "Malaysia" during the observation period. The single-day word frequency of the keywords "Malaysia Airlines" and "Malaysia" during the observation period increased rapidly on the day of the sudden incident of "Malaysia Airlines" loss of connection (March 8). Taking the emergency of March 8 as the formation event, a large-scale change in the frequency of Japanese words was triggered when the news of the crash was officially confirmed (March 24). Taking the emergency of March 8 as the formation event, the official confirmation of the crash (March 24) triggered a large-scale change in the word frequency on a single day. The date of the obvious change in the data coincides with the date of the key event described in the background introduction of the loss of contact with "Malaysia Airlines."

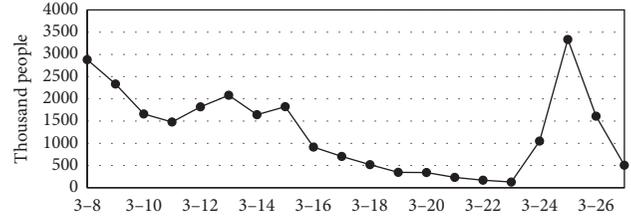


FIGURE 1: The single-day word frequency curve of the keywords "Malaysia Airlines" and "Malaysia" during the observation period.

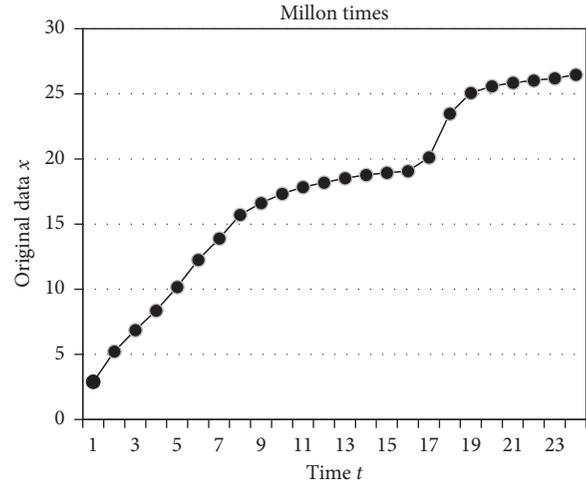


FIGURE 2: x and t data image.

2.2. *Processing of Data.* First, we use numbers 1 to 24 to represent each piece of data between March 8, 2014, and March 31, 2014, expressed in a as the original daily word frequency. To weaken the randomness of the original data and highlight the law of the original data. We accumulate the original data a and store the accumulated data as the new variable x (the relationship between x and t is shown in Figure 2). Logarithmically transform the variable x , and store the data in the variable $\log(x)$ (the relationship between $\log(x)$ and t is shown in Figure 3). The logarithmic transformation is the Box-cox transformation when $\lambda = 0$. There are two reasons for the logarithmic transformation of the original data: ① the scale of variables is compressed, and the absolute value of data is reduced to make it change between relatively small data ranges, which is convenient for calculation; ② the nature and correlation of the data will not be changed after the logarithm transformation.

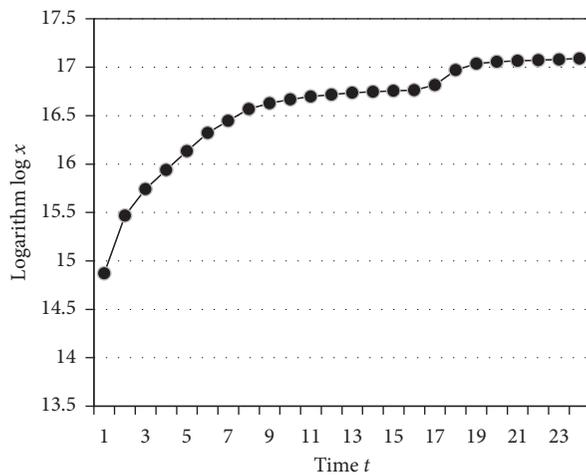


FIGURE 3: $\log(x)$ and t data image.

After the data is preprocessed, it is summarized in Table 1. Table 1 summarizes the variables a , x , and $\log(x)$. Three numbers are selected randomly from the column of $\log(x)$ to form a test set to test the fitting effect of the model. The remaining 21 numbers constitute the training

set to build the model. The three samples randomly selected in this paper are the corresponding data at $t = 10, 22,$ and 12 .

TABLE 1: Data frame.

t	a	x	$\log(x)$
1	2878813	2878813	14.87
2	2334038	5212851	15.47
3	1658542	6871393	15.74
4	1478279	8349672	15.94
5	1818663	10168335	16.13
6	2081481	12249816	16.32
7	1641772	13891588	16.45
8	1821670	15713258	16.57
9	912939	16626197	16.63
10	701404	17327601	16.67
11	518800	17846401	16.70
12	346422	18192823	16.72
13	343224	18536047	16.74
14	232773	18768820	16.75
15	168282	18937102	16.76
16	128418	19065520	16.76
17	1047551	20128509	16.82
18	3336808	23465317	16.97
19	1610509	25075826	17.04
20	505760	25581586	17.06
21	275780	25857366	17.07
22	166097	26023463	17.07
23	162715	26186178	17.08
24	278691	26464869	17.09

3. Model Establishment and Theoretical Improvement Based on “S-Curve”

The research of public opinion is based on the research of information diffusion mode. The change of information diffusion mode in physical space over time conforms to the trend of “S-curve.” Therefore, based on the “S-curve” theory, this paper makes theoretical improvements, establishes three models on the basis of information diffusion, studies the prediction of network public opinion of emergencies, and evaluates and compares the fitting effects of different models on data.

3.1. Introduction to “S-Curve”. The Logistic curve (also known as the “S-curve”) was first proposed by Belgian biomathematician Pierre Francois Veluler when he was studying the subject of population growth. The basic assumption is that there is no difference among individuals in the population, they have the same growth rate, and the relative growth rate is proportional to the remaining space resources; space resources are limited, with a certain saturation value, and the saturation value has remained unchanged. The characteristic of the curve change is as follows: the initial stage is roughly exponential growth; then, as it becomes saturated, the increase slows down; finally, the increase stops when it reaches maturity (see Figure 4).

This law is abstracted as a mathematical expression:

$$x(t) = \frac{k}{1 + ae^{-rt}} \quad (k, a > 0). \quad (1)$$

The population growth rate expression is

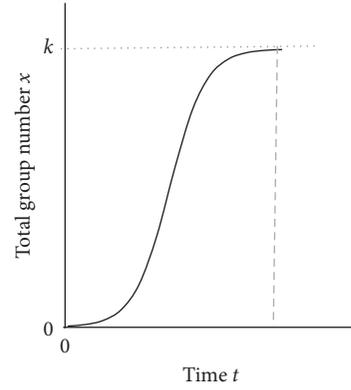


FIGURE 4: Logistic model.

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{k}\right). \quad (2)$$

The relative population growth rate expression is where k is the space saturation capacity, r is the innate rate of increase, and $r > 0$ represents the maximum increase of each individual in the population when it is not suppressed and reflects the inherent characteristics of the species [7]. The growth rate has a quadratic relationship with the population size x (see Figure 5).

$$\frac{dx}{xdt} = r \left(1 - \frac{x}{k}\right), \quad (3)$$

3.2. “S-Curve” and Information Dissemination in Physical Space. In the well-known spatial diffusion “average information domain” model [8], Hagerstrand simulates the probability distribution and empirical study of information diffusion by Monte Carlo method and draws an important conclusion that the accumulation of information users over time accords with the S-shaped logistic curve. The image of information diffusion rate varies with time as a whole, showing a “low-high-low” parabola shape “from low to high and then from high to low” (see Figure 5). Hagerstrand studies the information diffusion mode in the field of physical space; that is, in the physical diffusion space and environment, information diffusion depends on the ancient channels and ways of face-to-face communication. That is to say, the information diffusion of an event is accompanied by the formation of public opinion, and the object described by the information diffusion model in physical space is defined as the captured traditional social public opinion. However, generally speaking, in the physical space, the formation of traditional social public opinion is relatively slow.

3.3. Research on the Applicability of the “S-Curve” Network Public Opinion Prediction Model for Emergencies. In the following content, we apply the “S-curve” theory to the microblog public opinion analysis of the emergent formation event of the loss of contact with “Malaysia Airlines” and establish model 1 (m1) with the following expression:

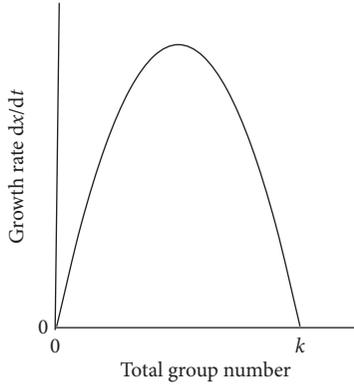


FIGURE 5: Logistic model (dx/dt) and x curve.

$$\log x_i = \frac{k}{1 + ae^{-rt_i}} + \varepsilon_i \quad (k, a > 0, i = 1, 2, \dots, 24), \quad (4)$$

where k , a , and r are unknown parameters and ε_i is a random error variable and satisfies

$$\begin{aligned} E(\varepsilon_i) &= 0, \\ \text{Var}(\varepsilon_i) &= \sigma_j^2, \\ \text{Cov}(\varepsilon_i, \varepsilon_j) &= 0 \quad (i \neq j), \end{aligned} \quad (5)$$

where $E(\varepsilon_i)$ is the expected value of ε_i and $\text{Var}(\varepsilon_i)$ is the variance of ε_i . In general, it is usually assumed that $\sigma_j^2 = \sigma^2$ (constant). $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ means that there is no correlation between observation errors; that is, the variation of errors is random and completely accidental. Under the assumption that $\sigma_j^2 = \sigma^2$ (constant), the ordinary least squares method can be used to estimate the parameters, but the remainder will be ignored when the model is linearized, and the estimation error will be large. Therefore, the nonlinear least squares method [9, 10] (unconstrained optimization) is used to estimate the parameters to make the error smaller; that is, $Q(k, a, r) = \sum_{i=1}^n (\log x_i - (k/(1 + ae^{-rt_i})))^2$ has a minimum value. The corresponding R language function is nls().

From the results of the program, model 1 and its various parameters have passed the significance test. The Residual Standard Error (RSE1) is 0.09769. We save the prediction result of model 1 on the test set in the variable test1. The expression of Model 1 is

$$\log x(t) = \frac{16.991837}{1 + 0.164634e^{-0.223025t}}. \quad (6)$$

A graph of the relationship between the fitted line of model 1 and the original data is shown in Figure 6.

It can be roughly seen from Figure 6 that the fitting effect of model 1 is general, the data points from $t = 12$ to $t = 17$ are not well fitted, and there is a significant deviation from the original data, and after $t = 18$, the fitting of the final saturation value is also not ideal. Therefore, to make the fitting effect more ideal and more in line with the actual situation of the original data, this paper improves the theory of the ‘‘S-curve’’ and establishes a new model to explore the fitting effect of the network public opinion data.

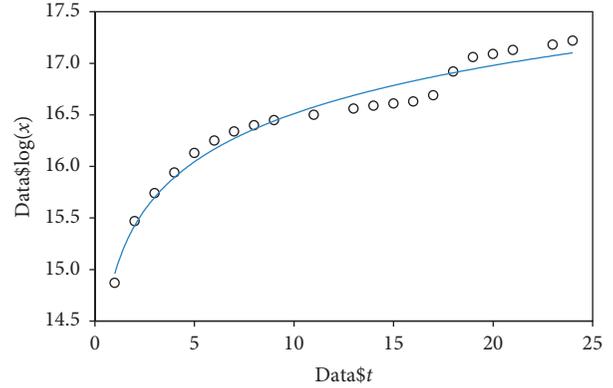


FIGURE 6: Fitted image of model 1 $\log(x)$ and t .

3.4. Theoretical Improvement of ‘‘S-Curve’’. The traditional ‘‘S-curve’’ theory is suitable for the analysis and study of social public opinion of information diffusion in physical space. With the development of the Internet, social democracy is constantly dependent on the Internet. There is the occurrence of emergency, then inducing the network public opinion. At this time, the traditional ‘‘S-curve’’ theory can no longer accurately analyze and predict the network public opinion, so it is necessary to improve and innovate the ‘‘S-curve’’ theory.

3.4.1. Cyberspace Information Diffusion and Network Public Opinion. Traditional social public opinion is fleeting or difficult to capture. Some of them are daily discussions on the streets and alleys, while others only exist in the ideas of people. Therefore, the object described by the information diffusion model of physical space is only the captured traditional social public opinion. With the popularization of the Internet, information diffusion in cyberspace has gradually become an important research object. Information diffusion in cyberspace is a process, in which information continues to expand its scope of adoption and application in cyberspace over time. It is based on proliferating behavioral participants (diffusion nodes), supported by the network environment, and driven by information potential. Conditioned by information circulation and transmission, it is a complex process involving many factors [11].

The spread of network information for a certain event can also promote the formation of social public opinion, and the formation speed is faster than the physical space. The Internet breaks through the space constraints and reduces the time cost of information transmission, thereby reducing intermediate links [12]. Modern news media, enterprises, and the government also make full use of this feature of the Internet to publish as soon as a major event occurs, so that the masses can get news in a short time. The self-media era has also made information dissemination popular and autonomous, and the contents with dissemination values published by personal accounts will also be quickly disseminated. At the same time, as an equal, free, and open information communication platform, the Internet makes the flow and interaction of information possible, and both

sides of information communication can interact with the other party equally. While people can quickly obtain information on the Internet, it also means that people can instantly express or share opinions, express attitudes, and vent their emotions on the Internet, that is, network public opinion; or they transfer the received network information to the physical space in a traditional way to promote the formation of public opinion in the physical space.

The formation of network public opinion and the promotion of public opinion in physical space are realized in the process of network information diffusion. Both are public opinions that can be captured and are part of modern social public opinion (see Figure 7). Therefore, the monitoring of the network public opinion has attracted increasing attention from the government.

3.4.2. “S-Curve” Applied to the Theoretical Improvement of Network Public Opinion in Emergency. The network information diffusion mode still conforms to the S-shaped logistic curve. However, in the “low-high-low” parabolic mode, where the information diffusion rate changes over time from low to high and then from high to low (see Figure 5), the transition of “low-high” from “low to high” is relatively rapid. For emergencies, the rapid transition of the “low-high” process of the network information diffusion model is more obvious. In the public opinion monitoring of the day, the data may only reflect the “high-low” pattern of “high to low.” This is because, in the information dissemination of physical space or the network information dissemination of ordinary events, it may take three to five days or longer to form public opinion, while the network public opinion of emergencies can be formed within two days or even one day [13]. The “high-low” mode is determined by the momentary gathering of people and the destructiveness of behavior. Due to the sudden occurrence of emergencies and a higher degree of damage than conventional events, public opinion on emergencies can be formed extremely quickly with the help of the online platform. However, the dissemination of Internet public opinion in emergencies essentially still undergoes a process of “from low to high and then from high to low” with the time-varying information diffusion rate. The formation and extinction of public opinion still follows the mechanism of the formation and extinction of public opinion in physical space, that is, conforming to the logistic curve.

After the accumulated daily data, the images of the information diffusion mode of the physical space and the information diffusion mode of the emergency event are shown in Figures 8 and 9, respectively, where $k_1 < k_2$, $t_1 > t_2$; this is because, in the information network society, the Internet has a wide range of audiences, and knowledge and information are important sources of power. After emergencies, with the help of the Internet as a medium, the original accumulation of information with the help of the power of the network has reached an unprecedented degree of concentration, and the speed of information diffusion is faster than that of the physical space, so it takes a short time for the total cumulative information to reach the saturation value.

4. Theoretical Improvement and Model Establishment of Double “S-Curve”

4.1. Raising the Question. The “S-curve” is based on the assumption that the space saturation capacity k remains unchanged and does not take into account the sudden changes in the objective environment of public opinion. If the saturation capacity changes, which is reflected in the actual problem as a sudden change in the objective environment, the “S-curve” model cannot accurately describe the actual problem. However, in the real world, the constant change of the space saturation capacity is a frequent phenomenon. For example, technological innovation, productivity improvement, and even the release of updated news will bring about major structural changes [14]. At this time, it is more appropriate to use the double “S-curve” model to describe the reality.

4.2. Double “S-Curve” and Network Public Opinion of General Events. There are two situations in which external influencing factors lead to changes in the space saturation capacity: one is a jump-type mutation, that is, when the original curve is saturated or close to the saturation value due to the sudden addition of new influencing factors such as technological innovation. This causes the curve to reenter a new growth development period starting from the change point in time (see Figure 10). The other is a continuous and gradual change in the saturation capacity of the space; that is, a certain influencing factor has appeared in the early and mid-term of the curve development and continues to affect the development of the curve over time, resulting in the continuous expansion of the saturation capacity of the curve (see Figure 11). The above two situations describe the development law of the jumping double “S-curve” and the gradual double “S-curve,” respectively. For the former, the curves before and after the change of the space saturation capacity are relatively independent and can be studied separately and can be merged into the original curve until the new growth reaches the new space capacity. For the latter, the curve change is an overall process, and changes at various moments are closely related and difficult to separate. Generally speaking, from the image, the two types of double “S-curves” have obvious jumps compared with the “S-curves.”

After certain events in the physical space occur, if the reason for their occurrence is unclear or undisclosed, the original event is likely to exist in the form of an event when the official announcement of the reason or confirmation of the occurrence of the event occurs. And this original event will trigger a new round of public opinion, that is, follow-up public opinion. The formation of follow-up public opinion will change the saturation capacity of the original public opinion. Therefore, public opinion images that generally form events can be described by two types of double “S-curves.”

4.3. Improving the Double “S-Curve” and Network Public Opinion of Emergencies. In cyberspace, when the cause of an event is unclear or undisclosed as a formative event, or the official announcement of the cause of the event or

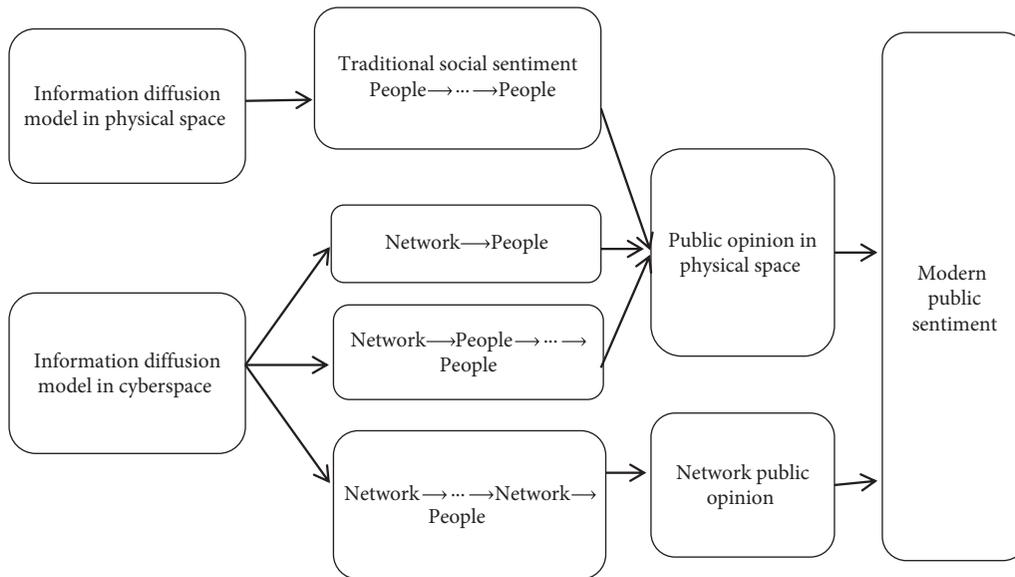


FIGURE 7: The relationship between traditional public opinion and modern public opinion.

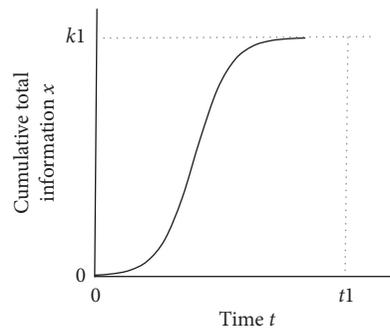


FIGURE 8: Information diffusion mode in physical space.

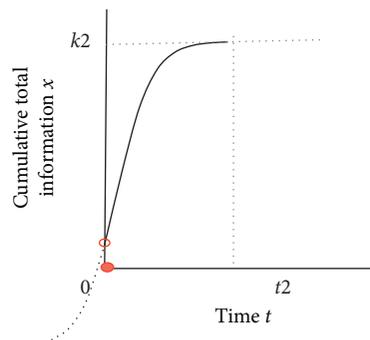


FIGURE 9: Network information diffusion mode of emergencies.

confirmation of the occurrence of the event may again cause major public opinion, that is, the overall network public opinion will undergo major structural changes, the public opinion, when public opinion is raised again, may be affected by an emergency that has occurred, or it may treat new news as a new event, and not be affected by an emergency that has occurred. Therefore, the network public opinion of emergencies can also be described by two types of double “S-curves.”

However, in cyberspace, public opinion is formed faster than in physical space. Network public opinion for emergencies with unclear or undisclosed causes is formed faster than public opinion in physical space (initial public opinion formation), and such emergencies, such as forming events, once again form public opinion in cyberspace (follow-up public opinion formation). Therefore, in general, the formation of public opinion on the Internet for emergent events is faster than that for general events in physical space.

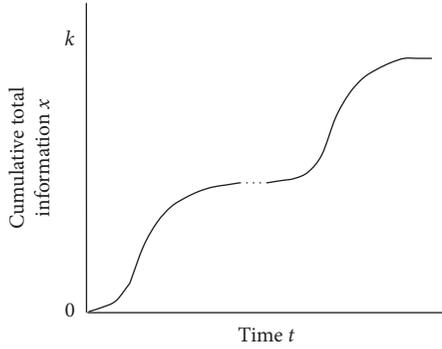


FIGURE 10: Information diffusion mode of general formation events (jump type).

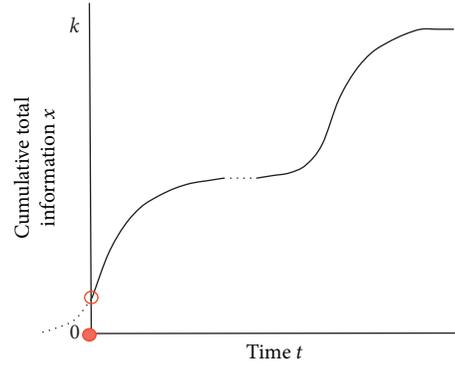


FIGURE 12: Network information diffusion mode of emergencies (jump type).

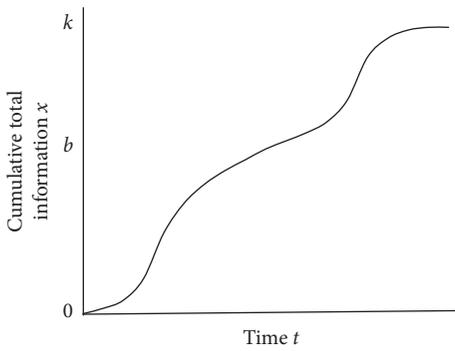


FIGURE 11: Information diffusion mode of general formation events (gradual type).

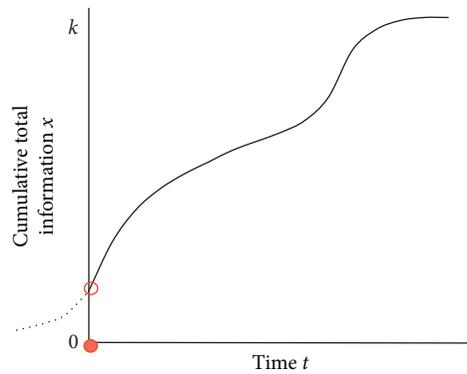


FIGURE 13: Network information diffusion mode of emergencies (gradual type).

The theoretical improvement in applying the double “S-curve” to the network public opinion of emergencies (Figures 10–12, Figures 11–13) is similar to the theoretical improvement in applying the “S-curve” to the network public opinion of emergencies (Figures 8 and 9).

The world is changing rapidly, and when emergencies with unclear or undisclosed causes are used as forming events, there may be many public opinions afterwards, the space capacity of the total amount of information will continue to increase, and the images will show more jumps. The development of network public opinion in emergencies may present a nonlinear development model of multiperiod, multiparameter, and dynamic multi-“S” shape development

mechanisms. This article takes the “Malaysia Airlines” loss of connection in March 2014 as an example and establishes a double “S-curve” emergency event prediction model.

4.4. Establishing a Model Based on the Improved Jump-Type Double “S-Curve” Theory. The network information diffusion mode of emergency with jump change, that is, the improved jump double “S-curve,” can be considered formed by the combination of two independent improved “S-curves” (see Figure 12). By using the improved jump double “S-curve” theory, we can establish model 2 (m2):

$$\log x_i = \sum_{h=1}^2 \left(I(t_i \in T_h) \frac{k_h}{1 + a_h e^{-r_h t_i}} \right) + \varepsilon_i \quad (k_h, a_h > 0, h = 1, 2, i = 1, 2, \dots, 24). \tag{7}$$

$H = 1, 2$ represent the first and second independent “S-curves,” respectively. And use nonlinear least squares to estimate the parameter $k_h, a_h, r_h (h = 1, 2)$ of model 2, so that $Q(k_1, a_1, r_1, k_2, a_2, r_2) = \sum_{i=1}^n (\log x_i - \sum_{h=1}^2 (I(t_i \in T_h) (k_h / (1 + a_h e^{-r_h t_i}))))^2$ has a minimum value.

From the x and t image in Figure 2, and the $\log(x)$ and t image in Figure 3, it can be seen that the “Mahan Airlines” event has a major jump at $t = 17$. Therefore, the split point is

$t = 17$ to divide the data into two parts $t = 1$ to 16 and $t = 17$ to 24 for fitting. The corresponding R language function is `nls()`. From the results of program operation, Model 2 and its various parameters have passed the significance test. The Residual Standard Error (RSE2) is 0.03591. Store the prediction result of model 2 on the test set in the variable `test2`. The expression of Model 2 is shown in the following way:

$$\log x(t) = I(t \leq 16) \frac{16.791925}{1 + 0.168621e^{-0.300460t}} + I(t \geq 17) \frac{17.082978}{1 + 0.036851e^{-0.848520t}} \quad (8)$$

$T \leq 16$ means that, after the data is processed according to the number 1–24, the data of the first 16 days are displayed, and $t \geq 17$ can be expressed in the same way. Make a graph of the relationship between the fitted line of model 2 and the original data (see Figure 14). Figure 14 shows that the fitting effect of model 2 is very good, and the RSE2 is smaller than RSE1. This means that the fitting effect of model 2 is better than that of model 1. It indicates that the traditional information diffusion model of physical space is not so good in describing the network public opinion of the sudden formation of events. It means that when the objective environment of network public opinion in emergency changes suddenly, the traditional “S-curve” theory cannot be mechanically applied. Instead, the analysis and prediction of network public opinion under new circumstances should be more in line with objective development.

4.5. Establishing a Model Based on the Improved Gradual Double “S-Curve” Theory

4.5.1. *Gradual Double “S-Curve” Is Applied to the Public Opinion of General Events.* The model we begin to build in this chapter is based on a differential equation model of our predecessor [6]. This kind of differential equation model can be used to describe the public opinion of general forming events in physical space (see Figure 10). This section will briefly introduce the process of establishing the model.

In practical applications, “S-curve” describes the information diffusion mode of physical space. If the gradual double “S-curve” expansion is carried out in physical space, firstly, according to the differential form of the “S-curve,” it is a parabolic function with an opening downward (see Figure 5), and its relative growth rate image is a descending straight line. From this analogy, the differential curve of the gradual double “S-curve” should be “M-shaped” (see Figure 15); that is, there are two maximum values and one minimum value. It can be seen from the figure that the relationship of (dx/dt) and x is approximately a fourth-degree polynomial, and the differential form of the double “S-curve” satisfies the following equation:

$$\frac{dx}{dt} = rx \left| 1 - \frac{x}{k} \right| \left| a \left(1 - \frac{x}{b} \right)^2 + c \right| \quad (k, b \neq 0), \quad (9)$$

where b is the space saturation capacity when the curve grows from the first stage to the second stage, and a and c are the parameters that control the shape of the curve.

Then, divide each point on the “M-shaped” growth rate curve by the value of x to get the equation that the relative growth rate dx/xdt satisfies:

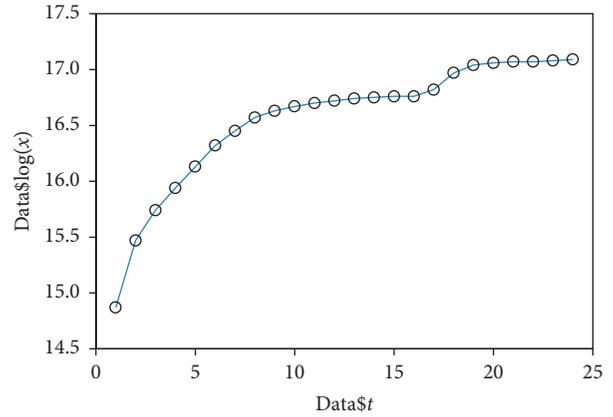


FIGURE 14: Fitting diagram of model 2 $\log(x)$ and t .

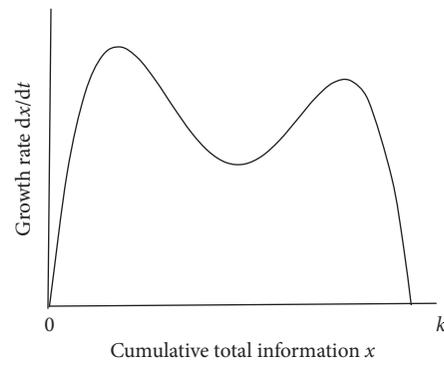


FIGURE 15: The (dx/dt) and x curve of the gradual double “S-curve” model.

$$\frac{dx}{xdt} = r \left| 1 - \frac{x}{k} \right| \left| a \left(1 - \frac{x}{b} \right)^2 + c \right| \quad (k, b \neq 0). \quad (10)$$

We use approximate methods to replace dx/xdt with actual data. Use the values of dx/xdt and x , then use the least square method to estimate the values of the parameters r, k, a, b and c of equation (10), and solve the differential equation (9) to obtain x at each time.

4.5.2. *The Theoretical Improvement of the Application of Gradual Double “S-Curve” in the Network Public Opinion of Emergencies.* The mode of the gradual double “S-curve” to form the network public opinion prediction model of unexpected events should be similar to Figure 13. Therefore, the image of (dx/dt) and x should be as shown in Figures 16 and 17. As for Figure 16, the differential curve of the improved gradual double “S-curve” is still “M-type” (quartic function), and the differential form dx/dt of the double “S-curve” satisfies equation (9). The relative growth rate dx/xdt satisfies equation (10). As for Figure 17, the differential curve of the improved gradual double “S-curve” is more suitable for fitting with cubic functions, so the differential form dx/dt satisfies equation (11); the relative growth rate dx/xdt satisfies equation (12).

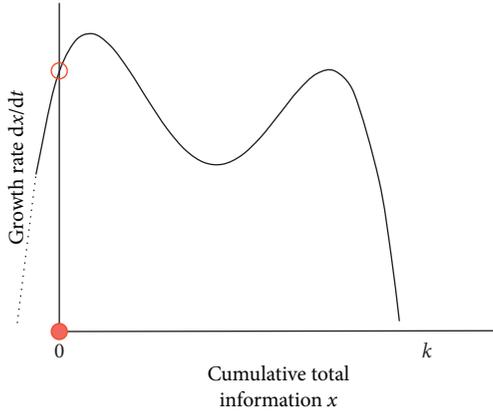


FIGURE 16: The (dx/dt) and x relationship of the improved gradual double “S-curve” model is a fourth-order image.

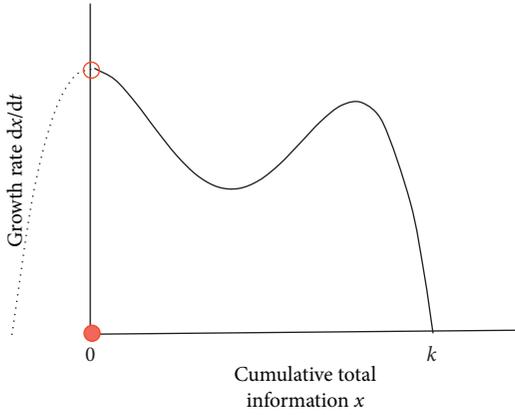


FIGURE 17: The (dx/dt) and x relationship of the improved gradual double “S-curve” model is a cubic image.

$$\frac{dx}{dt} = rx \left| 1 - \frac{x}{k} \right| \left| 1 - \frac{x}{b} \right| \quad (k, b \neq 0), \quad (11)$$

$$\frac{dx}{xdt} = r \left| 1 - \frac{x}{k} \right| \left| 1 - \frac{x}{b} \right| \quad (k, b \neq 0). \quad (12)$$

4.5.3. Improving the Gradual Double “S-Curve” and Network Public Opinion Prediction Model for Emergencies. Use the improved gradual double “S-curve” theory to establish a network public opinion prediction model for emergencies, which is recorded as Model 3. The detailed steps are as follows:

(i) Differential processing

The $\log(x)$ in the original data is differentially processed, and the derivative is approximately replaced by the central difference quotient to obtain the approximate value of the differential dx/dt and the relative growth rate dx/xdt (13), where x_t is the accumulated information data at a time t , and $dx/xdt|_{t=i}$ is the relative growth rate of the total amount of information at time t .

$$\frac{dx}{xdt} \Big|_{t=i} \approx \frac{x_{t=i+1} - x_{t=i-1}}{2x_{t=i}} \quad (i = 2, \dots, 23). \quad (13)$$

Store the approximately calculated dx/dt and dx/xdt together with the data in rows 2–23 of the original data in the new data frame and put all data belonging to the training set ($i \in \text{train}$) in data 2, used for the next parameter estimation.

(ii) Parameter Estimation

For different emergencies, when using curves to fit public opinion, a suitable function model should be selected according to the shape of the differential curve. From the training set $dx/dt|_{t=i}$ and $\log(x)$ (see Figure 18), it can be seen that the differential curve of the “Mahan” event is more suitable to be described by a cubic function. In this paper, the cubic function is used for fitting, and the four-degree fitting can be obtained by analogy.

Equation (12) is a cubic polynomial, converting it to a mathematical formula:

$$\frac{dx}{xdt} = Ax^2 + Bx + C. \quad (14)$$

To facilitate parameter estimation, the parameters A, B, C have a certain quantitative relationship with r, k, b in equation (12) and can be converted. Use the values of dx/xdt and x to establish a transition model (m3):

$$\frac{dx}{xdt} \Big|_{t=i} = Ax_i^2 + Bx_i + C + \varepsilon_i \quad (A, B, C \neq 0, i \in \text{train}). \quad (15)$$

The least square method is used to estimate the values of the parameters A, B , and C of equation (15). From the results of program operation, the transition model (m3) and its various parameters have passed the significance test.

(iii) Data Fitting

Substitute the estimated value of each parameter $A = 0.009175, B = -0.313680, C = 2.682796$ into equation (14), using MATLAB to solve the numerical solution of ordinary differential equations based on the Runge–Kutta method, and to obtain the estimated value of $\log(x)$ at each time ($t = 1$ to 24), namely, $\widehat{\log(x)}$, we store the predicted data belonging to the test set in the first 24 values solved in the variable test 3.

Draw the relationship between the fitted line of model 3 and the original data (see Figure 19). It can be seen initially from the figure that the fitting result of model 3 is better than that of model 1 in the period $t = 12$ to $t = 17$, and after $t = 22$, the final saturation value fits better than model 1. However, from the image point of view, compared with model 2, the fitting effect of model 3 is not ideal. To some extent, this is due to the irrational group behavior of

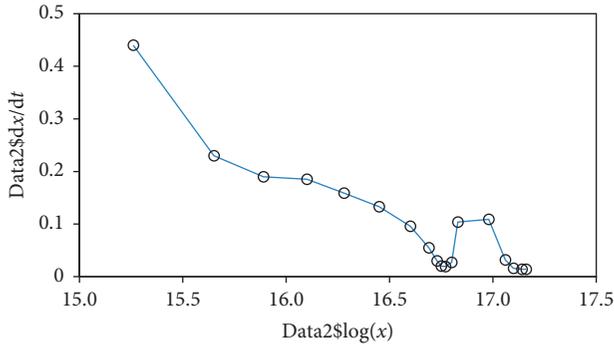


FIGURE 18: dx/dt and $\log(x)$ relationship diagram.

the sudden formation of network public opinion of events driven by the Internet. However, the description of model 3 is more consistent with the public’s cognition of reality, reflecting the internal relations and internal laws of things; that is, the change of public opinion is an integral process, and the changes of each moment are closely related and difficult to be separated.

5. Comprehensive Evaluation of the Models

This article is based on predecessors’ researches on the network public opinion of emergencies and obtains Weibo data related to the loss of contact with the emergent “Malaysia Airlines” from the big data platform. It innovatively puts forward the novel concept of network public opinion in emergency, that is, the sum of network public opinion caused by emergencies whose cause is unclear or undisclosed, and the follow-up network public opinion caused by the official announcement of the cause of the incident or the confirmation of the incident with this type of emergency as the formative event. Three types of prediction models are established for this network public opinion, and a comprehensive evaluation of the models is carried out at the same time. We use a more novel curve fitting method to establish a public opinion forecasting model. In the fitting process, the original model is reasonably explained and novel improved based on the characteristics of network public opinion and the actual situation. The three models fitted in the article are model 1 based on the “S-curve” theory, model 2 of the improved jump double “S-curve” theory, and model 3 of the improved gradual double “S-curve” theory.

Since the transition model (m3) was used when model 3 was established, the transition model (m3) is used to estimate that the intermediate parameters cannot represent model 3. Therefore, the RSE of the transition model (m3) is not comparable to the RSE of model 1 (m1) and model 2 (m2), so RSE cannot be used as a common standard for comprehensive model evaluation. To solve this problem, this paper uses the MAPE to evaluate different models. MAPE is expressed as a percentage and is dimensionless, eliminating the influence of the level of time series data and the unit of measurement. Therefore, it can be used to evaluate the same set of data in different models and measure the deviation between the observed value and the true value.

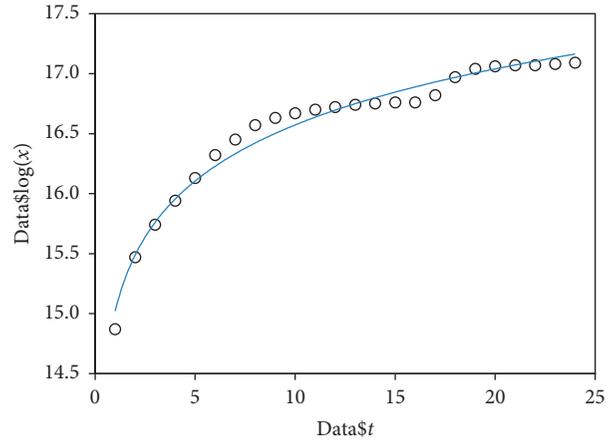


FIGURE 19: Fitted image of model 3 $\log(x)$ and t .

The expression is as follows:

$$MAPE = \frac{\sum_{t=1}^n |(observed_t - predicted_t)/observed_t|}{n} \times 100\% \tag{16}$$

The original $\log(x)$ of the test set and the prediction results of each model on the test set are gathered before the evaluation and comparison of the prediction effects of models 1, 2, and 3. Then, calculate each MAPE.

We obtained $MAPE1 = 0.42703\%$, $MAPE2 = 0.04263\%$, $MAPE3 = 0.15782\%$. From the perspective of MAPE, the prediction situation from good to bad is model 2 > model 3 > model 1. And from the fitting curve of each model (see Figures 6, 14, and 19), the fitting situation from good to bad is model 2 > model 3 > model 1. Finally, we summarize the advantages and disadvantages of the three models, as shown in Table 2.

The results of the comprehensive evaluation and analysis of the three models from the perspective of fitting and prediction are of great significance to reality. First, the fitting and prediction effect of model 1 is inferior to models 2 and 3. Although the “S-curve” theory is suitable for forecasting in any period, its nonlinear fitting ability is weak, and the fitting effect is not ideal; this indicates that the traditional physical space information diffusion model is not suitable for describing the emergencies in the network public opinion. It means that when the objective environment of network public opinion in emergency changes suddenly, the traditional “S-curve” theory cannot be mechanically applied. Instead, the analysis and prediction of network public opinion under new circumstances should be more in line with objective development. Second, the important basic assumption for the establishment of model 2 is that the curve changes before and after the space saturation capacity changes are relatively independent and can be studied separately before the new growth reaches a stable level. The fitting effect and prediction effect of this model are very good, but the segmentation point needs to be established in advance; third, model 3 uses differential processing, which reduces the fitting progress to a certain extent; when mid- and long-term predictions are made, the deviation is large;

TABLE 2: Advantages and disadvantages of the three models.

Model	Principle	Advantage	Disadvantage
Model 1 (m1)	“S-curve”	Suitable for forecasting in any period	The nonlinear fitting ability is weak, and the fitting effect is not ideal
Model 2 (m2)	Improved jump double “S-curve”	Fitting effect and prediction effect are very good	The establishment of the model is based on the independence of local laws, and the segmentation point needs to be established in advance
Model 3 (m3)	Improved gradual double “S-curve”	It is suitable for forecasting in any period and can reflect the inner relationship and inner law of things	The use of differential processing reduces the fitting progress to a certain extent; the deviation is large when mid- and long-term prediction is made; the analytical formula of the differential equation is not easy to obtain, and the numerical solution should be used instead

the analytical formula of the differential equation is not easy to obtain, so it needs to be replaced by a numerical solution. Although the fitting effect of model 3 is not as good as that of model 2, the description of model 3 is more in line with people’s perception of reality, reflecting the internal relations and internal laws of things; that is, the change of public opinion is an overall process, and the changes at each moment are closely related and difficult to separate. However, the fitting and prediction effect of model 2 is better than that of model 3, to a certain extent, indicating that the network public opinion of emergencies is irrational under the drive of the Internet. Under this circumstance, when the official announcement of the cause of the incident or confirmation of the occurrence of the incident is done, the network public opinion caused by the emergent event as a forming event can be treated as a new round of network public opinion that is inevitable, but the time of occurrence is uncertain. After an emergency occurs, the government has to not only deal with the panic caused by the current network public opinion, but also prepare for more intense network public opinion that may occur in the future. Before the release of updated news related to emergencies or the announcement of the truth of the matter, the government should use appropriate means to actively guide people’s emotions in advance to reduce the peak of secondary public opinion on the Internet and to avoid excessive ideological reactions and extreme behavior of the public when the truth is restored, and to avoid emergencies derived from this. Therefore, when predicting the network public opinion of emergencies, in addition to the initial public opinion, the follow-up public opinion is also worthy of attention. This provides more practical value for subsequent supervision and guidance of public opinion.

6. Concluding Remarks

In this paper, we obtained 24-day data related to the loss of contact with “Malaysia Airlines” from the official Weibo platform to study the information diffusion mode under the network public opinion. According to the difference between network public opinion diffusion and physical space public opinion diffusion, as well as the occurrence of emergencies, on the basis of the traditional S-curve model,

three novel public opinion prediction models are innovatively determining the specific calculation procedures to carry out data fitting analysis and prediction: model 1 based on the “S-curve” theory, model 2 of the improved jump double “S-curve” theory, and model 3 of the improved gradual double “S-curve” theory. The main contribution of this paper is that, through the novel explanation and improvement of the model, it provides a better analysis and prediction method for the information diffusion mode of network public opinion. Considering the formation of emergencies, changes in the objective environment, and rapid transmission of information in the context of cyberspace, not only the initial public opinion within a short period of time but also subsequent public opinion within a long period of time is taken into account. It solves the problem of public opinion prediction in the big data network. Then, specific calculation programs are given in these models, and the trend of information diffusion is predicted. We overcome the spread of network public opinion under the cognitive feedback of the public to the reality under the sudden change of objective environment to deal with the problem. Finally, through data fitting, MAPE is used to eliminate the influence of data level and units of measurement in time series, and the comprehensive evaluation shows that the novel improved model is representative of the information diffusion mode under network public opinion, with more effectiveness and higher practicability, in line with the development of modern society. At the same time, the government can use the novel improved “S-curve” model to timely and effectively predict and analyze public opinion. With positive public opinion forecasts, the government can step up publicity and send signals to the public in all aspects to guide them. For negative public opinion forecasts, the government can properly release relevant signals to the public in advance, so that the public can be psychologically prepared to receive them to prevent the sudden outbreak of the public after the incident, in case it is out of control.

Data Availability

The data used to support the findings of this study are available from the micro index (<https://data.weibo.com/index>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (No. 71471075), Fundamental Research Funds for the Central University (No. 19JNLH09), Innovation Team Project in Guangdong Province, P.R. China (No. 2016WCXTD004) and Industry-University-Research Innovation Fund of Science and Technology Development Center of Ministry of Education, P.R. China (No. 2019J01017), and the authors gratefully acknowledge this support.

References

- [1] L.-h. Wang, "A brief discussion on the study of "public opinion"," *Tianjin Social Sciences*, vol. 24, no. 2, pp. 78–81, 2004, in Chinese.
- [2] C.-h. Yi and Z. He, "Study on the regularities of formation and evolution of internet public opinions caused by emergent events," *Journal of Xiangtan University (Philosophy and Social Sciences)*, vol. 38, no. 2, pp. 74–78, 2014, in Chinese.
- [3] Y. He, Y. Zhang, and M. Xiao, "Etc. Infection relationship research on the emotions of different microblogging community: with the topic of "air asia losing contact" as an example," *Statistics & Information Forum*, vol. 32, no. 8, pp. 110–116, 2017, in Chinese.
- [4] T. Hagerstrand, *Innovation Diffusion as a Spatial Process*, The University Chicago Press, Chicago, IL, USA, 1967.
- [5] E. M. Rogers, *Diffusion of Innovations*, Simon and Schuster, New York, NY, USA, 2003.
- [6] W.-z. Zhu, "New characteristic of crisis information communication model and its management strategies in digital age," *Journal of Modern Information*, vol. 29, no. 2, pp. 60–63, 2009, in Chinese.
- [7] J.-q. Liu and W.-y. Yuan, "Updated logistic model: double S-shape curve model," *Mathematics in Practice and Theory*, vol. 38, no. 17, pp. 60–65, 2008, in Chinese.
- [8] Y.-r. Deng, *Research on Information Diffusion Based on Network Dimension-Force*, Harbin Engineering University, Harbin, China, 2008, in Chinese.
- [9] X.-b. Wang, "The method of nonlinear least square fitting logistic curve and its application in diffusion of computer technology," *Journal of Shanghai University (Natural Science Edition)*, vol. 1, no. 4, pp. 369–375, 1995, in Chinese.
- [10] M.-l. Cheng, "Extension and application of logistic curve," *Operations Research and Management*, vol. 12, no. 3, pp. 85–88, 2003, in Chinese.
- [11] Y.-r. Deng and X. f. Xu, "Analysis of information diffusion under the network environment," *Modern Management Science*, vol. 27, no. 2, pp. 59–61, 2008, in Chinese.
- [12] X.-d. Liu, Y.-t. Cao, and L.-m. Li, "Predicting model in network impact: a monitoring and warning system for public opinion in universities under big data framework," *Journal of Shenzhen University (Humanities & Social Sciences)*, vol. 32, no. 4, pp. 156–160, 2015, in Chinese.
- [13] X.-d. Liu and J.-l. Chen, "A differentiated strategy of tourism e-commerce product promotion based on provincial tourism development difference: based on large data and data visualization method," *Statistics and Information Forum*, vol. 32, no. 8, pp. 31–38, 2017, in Chinese.
- [14] X.-d. Liu and W.-j. Li, "Study on jump volatility of financial high-frequency data: based on the method of big-data kernel functions SVM," *Statistics & Information Forum*, vol. 33, no. 9, pp. 23–30, 2018, in Chinese.