Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
COVID-19 information contact and participation analysis and dynamic prediction in the Chinese Sina-microblog

Fulian Yin a , Hongyu Pang a , Xinyu Xia a , Xueying Shao a , Jianhong Wu b , *

a College of Information and Communication Engineering, Beijing, 100024, PR China
b Fields-CQAM Laboratory of Mathematics for Public Health, Laboratory for Industrial and Applied Mathematics, York University, Toronto, M3J1P3, Canada

A R T I C L E   I N F O

Article history:
Received 27 April 2020
Received in revised form 16 October 2020
Available online 29 January 2021

Keywords:
COVID-19
Dynamic model
Reading and forwarding
Sina-microblog

A B S T R A C T

The outbreak of a novel coronavirus (COVID-19) aroused great public opinion in the Chinese Sina-microblog. To help in designing effective communication strategies during a major public health emergency, we analyze the real data of COVID-19 information and propose a comprehensive susceptible–reading–forwarding–immune (SRFI) model to understand the patterns of key information propagation considering both public contact and participation. We develop the SRFI model, based on the public reading quantity and forwarding quantity that denote contact and participation respectively, and take into account the behavior that users may re-enter another related topic during the attention phase or the participation phase freely. Data fitting using the real data of both reading quantity and forwarding quantity obtained from Chinese Sina-microblog can parameterize the model to make an accurate prediction of the COVID-19 public opinion trend until the next major news item occurs, and the sensitivity analysis provides the basic strategies for communication.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Up to 24:00 on April 16, 2020, 2,352,198 cases have been affected by a novel coronavirus pneumonia (COVID-19) worldwide since it has been successively found in Wuhan. Major news items combined have generated quite strong fluctuations in public opinions. For example, on January 28, 2020, Nanshan Zhong, a well-known expert in infectious disease control, emphasized that people should not go out at present [1], this appeal attracted wide attention and warning that people said they could go out only if Nanshan Zhong admitted [2]. Sina-microblog is the most popular social network in China [3] and the outbreak-related topics about COVID-19 grew exponentially on that platform. As reading quantity and forwarding quantity of “Nanshan Zhong” have reached almost 9.40 billion and 2.05 million on Sina-microblog, understanding how these emerging public contact and participation spread in social media to alter the public behaviors is important to help to design effective communication strategies for rapid implementation of public health interventions.

Fig. 1 shows the schematic diagram of COVID-19 information propagation in Sina-microblog, and the nodes represent users in different states. Many original post owners (green nodes) can publish single or multiple epidemic topics subordinate to different news. Take Headline News and Pear Video for example. While Headline News reports on multiple topics, and Pear Video focuses on a particular topic, and both of them can be read or forwarded by users attracted by

* Corresponding author.
E-mail address: wujhhida@hotmail.com (J. Wu).

https://doi.org/10.1016/j.physa.2021.125788
0378-4371/© 2021 Elsevier B.V. All rights reserved.
both topics. Especially, reading users (blue nodes) can choose to be silent then leave and also can forward after reading becoming forwarding users (orange nodes). When users finish one topic, they may join others. And the relevant forwarding users can resume in the topics later and forward multiple messages becoming cross-forwarding users (pink nodes), leading to a multi-level information diffusion process. All users (readers and forwarding users) can choose to read or forward only one or multiple topics and therefore information propagates through one or multiple topics. This promotes the COVID-19 information dissemination rapidly.

To our best knowledge, there is no appropriate model framework that can be used to analyze public contact propagation and public participation propagation together during a major public health emergency. In consideration of the urgent need of developing theoretical knowledge and practical technologies to help effective communications of public health interventions, we propose a comprehensive susceptible–reading–forwarding–immune (SRFI) model based on both reading quantity and forwarding quantity that represent public contact and participation respectively to analyze the public opinion propagation of the COVID-19. In particular, we consider the characteristic user behavior that users may participate repeatedly in the reading or forwarding on different topics of the COVID-19 information dissemination.

Traditionally, considering rumor is similar to epidemiology in several propagation ways, many scholars used susceptible–infected (SI) model [4,5], susceptible–infected–recovered (SIR) model [6,7], susceptible–exposed–infected–recovered (SEIR) model [8,9] and susceptible–infected–susceptible (SIS) [10] model to represent rumor propagation and address relevant issues. Generally, information propagation in social networks was analyzed by stratifying users into three classes: heard rumor (ignorants), actively spreading rumor (spreaders) and no longer spreading rumor (stiffers) [11]. Then, scholars introduced new modules into classical models to understand the process of information dissemination better and then achieve various research purposes. In 2011, Zhao et al. [12] provided a more detailed and realistic description of the rumor spreading process with combination of forgetting mechanism and the SIR model of epidemics on an online social blogging platform called LiveJournal. In 2012, Zhao et al. [13] extended the classical SIR rumor spreading model by adding a direct link from ignorants to stiffers and a new kind of people—Hibernators in order to reduce the maximum rumor influence, which was called susceptible–infected–hibernator–removed (SIHR) model. In 2012, Xiong et al. [14] proposed a susceptible–contacted–infected–refractory (SCIR) diffusion model, which contained four possible states to characterize information propagation on online microblogs. In 2014, Zhao et al. [15] added the refutation mechanism in homogeneous social networks to the basic model using the Runge–Kutta method, which could help authorities reduce the maximum influence of the rumor. Zhang et al. [16] emphasized a special rumor spreading characteristic called “the cumulative effects of memory” and added the memory mechanism, meanwhile simulated the rumor spreading process on Sina-Microblog. Rui et al. [17] proposed a susceptible–potential–infective–removed (SPIR) model introducing a potential spreader set, which made the state-changing mechanism more reasonable and accurate for the diffusion process.

In addition, Zhang et al. [18] used an improved SIR model to posit that a coupled network comprised two categories of nodes, then made use of the data collected from Weibo and WeChat of an actual news event to visualize the information spread process in the cross-network dissemination case of public opinion. Huang et al. [19] established a human dynamics model for deducing retweeting behavior and investigated it by gathering data through Sina API, which revealed that the distribution of the probability of a message to be browsed over time presented power-law characters. In 2018,
Zan [20] studied the double rumors spreading with different launch times and introduced two kinds of models: double-susceptible–infected–recovered (DSIR) model and comprehensive-DSIR (C-DSIR) model, which focused on the interaction from old rumor to new rumor and the propagation of two rumors posted successively. And in 2019, we [21] proposed an epidemic model called susceptible–forwarding–immune (SFI) to capture a single information propagation trend in the Sina-microblog considering the forwarding quantity of users. Trpevski D et al. [22], Qian et al. [23], Wang et al. [24] also proposed many improved models for the spread of information. Especially Tanaka M et al. [25] added a new module to the traditional model using the datasets from the Japanese Mixi and Facebook.

After summarizing a lot of literature, we find that most of the experimental data used for simulation are from well-known social platforms in the world, such as Twitter and Facebook. Furthermore, scholars from different countries may choose a local social platform in order to obtain data more convenient, especially Sina-microblog in China. By analyzing different data sources from other papers, we discover that existing studies have used forwarding quantity of one Weibo or multiple Weibos, the content of rumors and browsing behavior of users. To our best knowledge, they have not introduced the new module by combining with reading quantity and forwarding quantity of topics which are significant indicators to measure public contact and participation.

The paper is organized as follows: in Section 2, we analyze the public opinion data of crucial moment on the COVID-19; in Section 3, we introduce mathematical model definition for information contact and participation, fit the model with the real data of two typical topics, make a staged prediction of the overall public opinion trend, then conduct a parameter sensitivity analysis and give effective intervention strategies about the information contact and participation; and in Section 4, we draw a conclusion.

2. COVID-19 information contact and participation analysis

2.1. Information reading and forwarding topology

The urgency is often accompanied with a number of topics related to the COVID-19 and the behavior of reading and forwarding usually prompts the propagation of the event. Reading is a kind of behavior that reflects users’ contact in information propagation and forwarding is a kind of behavior that reflects users’ participation in information propagation. In order to more clearly show the propagation process of both public contact and participation, the network topology is shown in Fig. 2 which describes the state of each node in the network at a certain moment in the process of information dissemination. Taking the propagation of three original post owners under three topics as an example, the user’s overall state of integrated information propagation is given.

The information posted by the original post owners (red nodes), and can be read separately by single readers with interest (blue nodes) in which readers can then participate in the forwarding about the outbreaks (black nodes) or choose to be silent (yellow nodes). Especially, some co-spreaders between different topics repeatedly read and forward the related information successively then become the cross-reading users (pink nodes) and cross-forwarding users (green nodes) because of the correlation about the epidemic. Of course, there will also be many readers who contact information choose to be silent, such as the un-forwarding users (yellow nodes). In the real-world, the number of information is dynamically changing and cannot be clearly calculated.

The reading quantity and forwarding quantity of the whole COVID-19 are composed of many topics with multiple information. Different from the traditional public hot events, the outbreak is causing great public concern. With the continuous development of the COVID-19, there is a high level of repetition in public reading and forwarding on different topics. As reading is a measure of contact and forwarding is a measure of participation for information dissemination, in this paper, we build the comprehensive susceptible–reading–forwarding–immune (SRFI) dynamics model with considering the repeated behavior including “re-reading” and “re-forwarding” on the impact of public opinion propagation.

2.2. Information contact and participation analysis

Since the outbreak of COVID-19, around 7900 major topics appeared in Sina-microblog. Fig. 3 shows the cumulative reading quantity and forwarding quantity from December 31, 2019 to April 16, 2020, where the ordinate is the logarithm of the cumulative reading and forwarding quantities. It can be roughly seen that during the period from December 31, 2019 to January 16, 2020, there was only a certain amount of hot topics of COVID-19, however, from January 17, 2020, the hot topics about the epidemic gradually increased, especially after Nanshan Zhong confirmed that the COVID-19 could be transmitted from human to human on January 20, 2020, both the two quantities kept increasing dramatically, and then we will analyze the contact and participation data of COVID-19 information from January 17, 2020 to April 16, 2020 and Table 1 gives the specific values of the cumulative reading quantity and forwarding quantity.

Fig. 4 shows the box diagram of reading quantity and forwarding quantity from January 17, 2020 to April 16, 2020, which reflect the user’s contact and participation in information, respectively. Although in terms of the order of magnitude, the reading quantity is larger than the forwarding quantity and its range is relatively wide, in particular, the reading quantity and the forwarding quantity vary within about $3 \times 10^{10}$ and $7 \times 10^{6}$ respectively, except for the outlier forwarding quantity caused by several emergencies, these two propagation trends are consistent, both in line with the
laws of dynamics development. Besides, there are also different features in reading quantity and forwarding quantity such as the median, so it is necessary to build a model combining these two attributes.

In order to analyze the public opinion trend more clearly, Fig. 5 gives the reading quantity, forwarding quantity and the number of topics together with line charts and histograms from January 17, 2020 to April 16, 2020. We can see that since January 17, 2020, the number of reading and forwarding of COVID-19 has gradually increased, and the emergence of some important events or hot topics has a major impact on public opinion and the overall development can be divided into four stages so far based on this. The first stage, from January 17, 2020 to February 19, 2020, which both two quantities
increased gradually, is the outbreak stage. In the beginning, topic #5 rumors of viral pneumonia in Wuhan # attracted public attention, then Wuhan began its closure on January 23, 2020, with the emergence of topics such as # Wuhan Bus Metro Suspension of Operation #, the reading and forwarding quantities of COVID-19 began to increase dramatically and reached the highest value and the number of topics began to gradually increase leading public opinion enter the outbreak stage. Besides, topic #Doctor Wenliang Li passed away # posted on February 7, 2020 and # Easy epidemic prevention station # posted on February 19, 2020 have aroused great attention from people, causing both two quantities to reach the extreme again and public opinion continues to ferment. The second stage is from February 19, 2020 to March 16, 2020, in which both two quantities start to decrease with the stabilization and improvement of the domestic cases, and the overall number of topics increased and stabilized in a certain range in the later period. The third stage is another outbreak stage from March 16, 2020 to April 3, 2020, the overall trend is on the rise as the disease becomes more serious abroad, and the number of topics has risen again. With the release of # support Hubei Medical Team to evacuate # on March 17, 2020 and other topics related to COVID-19 of the global, both two quantities increased dramatically again, and public opinion entered into another outbreak stage. The fourth stage, from April 3, 2020 to April 16, 2020 is the occasional fluctuation stage. April 4, 2020 is Tomb-Sweeping Day of China, some topics for mourning the anti-epidemic heroes began to increase such as # Three minutes of silence in all China # which caused a lot of users to read and forward and led to some occasional fluctuations in the reading quantity and forwarding quantity. The division of the public opinion of COVID-19 helps us understand the trend caused by the entire epidemic hot topics.
The propagation dynamics model based on the reading quantity and forwarding quantity of COVID-19 constructed in this paper is shown in Fig. 6. Here, we only consider the accessible population in the process of information propagation and pay attention to both the information diffusion caused by users’ reading behavior and forwarding behavior. Assuming that the number of users (N) who can contact information in the process of propagation on Sina-microblog remains unchanged, we stratify the population into four states: the susceptible state (S), in which the users unaware of but susceptible to the information of the event; the reading state (R), in which the users have read information and are susceptible to forward it; the forwarding state (F), in which the users have forwarded the information actively to influence other users; and the immune state (I), in which the users have read or forwarded the information, but are no longer read or forward the information even if receive them again.

A susceptible user can read one information with an average exposure rate $\beta$ and a user in reading state will leave and become other states with a deactivation rate $\gamma$. The forwarding users can become immune users who are inactive to the event with an average inactive rate $\alpha$, with $1/\alpha$ being the average duration where an F-user remains active in being contacted.
The core of our model is to study the role of both repeated reading and forwarding because users exposure to different information about COVID-19. Hence, with forwarding probability $p$ from reading users to forwarding users and immunity probability $q$ denotes those who keep silent in the event and go straight to the immune state, we use $(1 - p - q)$ to represent the “re-reading” probability for a reading user returns to the susceptible state. Besides, we use parameter $\theta$ to describe the “re-forwarding” probability for a forwarding user who can return a new round of susceptible state of COVID-19.

In particular, each user may have a unique state, that is, at the same time, each user can be only one of the susceptible, reading, forwarding or immune states. We obtain the following SRFI dynamics model:

\begin{align}
S'(t) &= -\beta S(t)F(t) + (1 - p - q)\gamma R(t) + \theta \alpha F(t), \quad (1) \\
R'(t) &= \beta S(t)F(t) - \gamma R(t), \quad (2) \\
F'(t) &= p\gamma R(t) - \alpha F(t), \quad (3) \\
I'(t) &= q\gamma R(t) + (1 - \theta)\alpha F(t), \quad (4)
\end{align}

where $' = d/dt$ is the derivative with respect to $t$. The behavior transformation and state transition of the masses can also be interpreted as follow:

**Reading:** Since an active forwarding user will contact an average number of $\beta N$ users per unit time, the probability of a normal user is a susceptible user is $S(t)/N$, and there are $F(t)$ active forwarding users in total, then the number of new reading users is $\beta S(t)F(t)$. **Forwarding:** Some reading users will inactive with the deactivation rate $\gamma$ and participate in the forwarding state with the forwarding probability $p$, the number of new forwarding users is $p\gamma R(t)$. **Re-enter:** As events unfold, there are two ways to generate repeated behavior and initiate a new round of reading and forwarding: users who contact one information and yearn for getting more information about COVID-19 will re-enter to the susceptible state from reading state, the average number of re-entered users is $(1 - p - q)\gamma R(t)$ per unit time; users who have forwarded one information and be interested in related information about COVID-19 will re-enter to susceptible from forwarding state, the average number of the re-entered users is $\theta \alpha F(t)$ per unit time. **Immune:** Some reading users will not participate in the forwarding and become the immune users directly because they want to keep silent in the event, and the number of direct immune users is $q\gamma R(t)$. And some forwarding users will go to the immune state out of active time, the number of inactive users is $(1 - \theta)\alpha F(t)$.

The Sina-microblog provides the number of cumulative reading population and forwarding population which is the total times of reading and forwarding within a topic about COVID-19, and we calculate the sum of the whole event, given by

\begin{align}
C'_R(t) &= \beta S(t)F(t), \quad (5) \\
C'_F(t) &= p\gamma R(t). \quad (6)
\end{align}

The corresponding differential equation can be expressed as:

\begin{align}
C_R(t) &= \int_0^t \beta S(t)F(t) \, dt, \quad (7) \\
C_F(t) &= \int_0^t p\gamma R(t) \, dt. \quad (8)
\end{align}

Considering the initial condition: $R_0 = C_{R0} \ll N$, $F_0 = C_{F0} \ll N$, $I_0 = 0$ and $S_0 = N - R_0 - F_0$. The final condition from Eqs. (4)-(6), it follows that $I(t)$, $C_R(t)$ and $C_F(t)$ are all increasing since $I'(t) = q\gamma R(t) + (1 - \theta)\alpha F(t) > 0$, $C'_R(t) = 0$.
\( \beta S(t)F(t) > 0 \) and \( C_F(t) = pyR(t) > 0 \). Therefore, the final states are \( I_{\infty} = \lim_{t \to \infty} I(t) < N \), \( C_{R_{\infty}} = \lim_{t \to \infty} C_R(t) < N \), \( C_{F_{\infty}} = \lim_{t \to \infty} C_F(t) < N \), \( R(t) \) and \( F(t) \) are tending to 0 (\( R_{\infty} = 0, F_{\infty} = 0 \)), and \( S_{\infty} = N - I_{\infty} \). Here \( C_{R_{\infty}} \) and \( C_{F_{\infty}} \) are the final size of the COVID-19 reading and forwarding. In addition, the number of maximal reading users and maximal forwarding users are \( \hat{R}_{\text{max}} = \max(R(t)) \) and \( \hat{F}_{\text{max}} = \max(F(t)) \), respectively.

We define the reproduction ratio \( \Re_0 \) to measure whether the outbreak forwarding quantity was likely to break out. In the initial post of the COVID-19, the forwarding outbreak is given by \( F'(0) = (p\beta S_0 - \alpha)F(0) < 0 \) due to the decreasing of \( S \). Then we deduce \textbf{Public opinion reproduction ratio} \( \Re_0 \):

The reproduction ratio \( \Re_0 \) is defined to measure whether public opinion was likely to break out. We use the calculation method of basic reproduction number developed in [26], and rewrite our model as follows:

\[
\dot{x} = M(x) - V(x),
\]

where \( x = (R(t), F(t))^T \) and

\[
M(x) = \begin{bmatrix}
\beta S(t)F(t) \\
pyR(t)
\end{bmatrix},
\]

\[
V(x) = \begin{bmatrix}
yR(t) \\
\alpha F(t)
\end{bmatrix}.
\]

Calculate the derivatives \( M \) and \( V \) at no information propagation equilibrium \( F_0 = (S_0, 0, 0, 0) \), we can obtain

\[
M = \begin{bmatrix}
0 & \beta S_0 \\
py & 0
\end{bmatrix},
\]

and

\[
V = \begin{bmatrix}
y & 0 \\
0 & \alpha
\end{bmatrix}.
\]

The roots of the characteristic equation can deduce the eigenvalues of the matrix \( MV^{-1} \):

\[
|\lambda E - MV^{-1}| = \begin{vmatrix}
\lambda & -\frac{\beta S_0}{\alpha} \\
-p & \lambda
\end{vmatrix} = 0.
\]

Because \( \Re_0 \) is not negative, we have

\[
\Re_0 = \lambda = \sqrt{\frac{p\beta S_0}{\alpha}}.
\]

Here, based on the extension of \( \Re_0 \), we define the (effective) reproduction number \( \Re_e \) to describe the outbreak of public opinion at each time \( t \), and it has more practical significance in the dynamic development process. Then we have

\[
\Re_e = \sqrt{\frac{p(t)\beta(t)S(t)}{\alpha(t)}}.
\]

The \( \Re_e \) represents the propagation capability of each period and it is time-varying, which is determined by the average exposures rate \( \beta \), the average inactive rate \( \alpha \), the forwarding probability \( p \), and the susceptible users \( S(t) \). When \( \Re_e < 1 \), it means that the comprehensive public opinion will decline which implies the propagation can never break out. The \( \Re_e > 1 \) indicates that the comprehensive public opinion grows exponentially initially.

3.2. Data fitting

\textbf{Parameter estimation method}:

To use our SRFI model to explore some distinctions of qualitative behaviors for prediction, we use the LS method to estimate the model parameters and the initial data of our SRFI model. The parameter vector can be set as \( \Theta = (\beta, p, q, \alpha, \theta, \gamma, S_0) \), and the corresponding numerical calculation based on the parameter vectors for \( C_R(t) \) and \( C_F(t) \) are denoted by \( f_{C_R}(k, \Theta) \) and \( f_{C_F}(k, \Theta) \), respectively. The LS error function

\[
LS = \sum_{k=0}^{T} |f_{C_R}(k, \Theta) - C_{Rk}|^2 + \sum_{k=0}^{T} |f_{C_F}(k, \Theta) - C_{Fk}|^2
\]

is used in our calculation, where \( C_{Rk} \) and \( C_{Fk} \) denote the actual cumulative reading quantity and forwarding quantity, and \( k = 0, 1, 2, \ldots \) is the sampling time. In our paper, we use DEDiscover software to solve this LS problem.

\textbf{Data description}:
In order to analyze the public opinions with different characteristics, the following two typical events are selected from the whole public opinion outbreak duration. Table 2 shows the reading quantity and forwarding quantity of # Refuse to eat wild animals # with a slow outbreak. And Table 3 shows the reading quantity and forwarding quantity of # Real-time broadcast of joint prevention and control of epidemic situation # with a fast outbreak.

**Data fitting results:**

As shown in Fig. 7, we performed data fitting on the real data of # Refuse to eat wild animals# in Table 2, where the red star and blue star denotes the actual cumulative number of reading and forwarding population, respectively. It can be seen that the initial outbreak of the topic is slow and our SRFI model achieves accurate estimation.

Table 4 gives some important values of early period parameter estimation of # Refuse to eat wild animals #. We can see relative to the topic # Real-time broadcast of joint prevention and control of epidemic situation # the initial susceptible users $S_0$ is smaller meaning that there are fewer users are susceptible at the beginning of the information explosion, leading the topic outbreak slowly. Besides, the parameter $\alpha$ is larger, which means the active period of forwarding users is short so that they cannot influence other susceptible people for a long time, leading a smaller final size of cumulative reading and forwarding quantities.

As shown in Fig. 8, we performed data fitting on the real data of # Real-time broadcast of joint prevention and control of epidemic situation # in Table 3, where the red star and blue star denotes the actual cumulative number of reading and forwarding population, respectively, the red line and blue line denotes the estimated cumulative number of reading and forwarding population, respectively. It can be seen that the initial outbreak of topic # Real-time broadcast of joint
Table 5 gives some important values of early period parameter estimation of # Real-time broadcast of joint prevention and control of epidemic situation #. We can see the initial susceptible users $S_0$ is larger meaning that more users become susceptible at the beginning of the information explosion, leading the topic outbreaks rapidly. In addition, the parameter $\alpha$ is smaller, indicating more users will remain active and affect other susceptible users, which leads it quickly increases to a larger final size of cumulative reading and forwarding quantities.

3.3. COVID-19 information contact and participation prediction

We have divided the main development of COVID-19 information dissemination into four stages according to the analysis of information contact and participation from January 17, 2020, to April 16, 2020. Although we cannot control the occurrence of emergency incidents and the dissemination of information, it is very important that, in each stage, we
can predict the trend of public opinion based on the existing data before the emergency comes. Fig. 9 shows the prediction of reading and forwarding quantities of COVID-19 and Table 6 gives the estimated parameters with DEDiscover software.

At the first stage, topic #5 rumors of viral pneumonia in Wuhan# attracted public attention and the cumulative reading and forwarding quantities of COVID-19 began to increase. We predict the public opinion trend of COVID-19 with the data from January 17, 2020 to January 25, 2020 and it achieves good data fitting with the actual data until January 28, 2020, as shown in Fig. 9(a1). Thus, we extend the data to January 31, 2020 and predict again, fortunately, we get a satisfactory result until the end of this stage, as shown in Fig. 9(a2). It can be seen from the forecast curve of predicted reading population and forwarding population that both the public contact and participation are increasing at the beginning of the epidemic.

With the gradual stabilization of COVID-19 in China, the cumulative reading and forwarding quantities began to grow slowly and the public opinion enters the second stage. We use the data between February 19, 2020 and February 25, 2020 to estimate the parameters and predict the trend of public opinion of the next days, as shown in Fig. 9(b1). Besides, we extend the data until February 26, 2020 to predict the rest of the public opinion trend in this phase and we have achieved very good prediction results, as shown in Fig. 9(b2). At this stage, the predicted reading and forwarding population have a downward trend, which means the public contact and participation are close to saturation.

Unexpectedly, the second stage ended since the seriousness of overseas COVID-19 and public opinion enters the third stage. We use the data between March 16, 2020 and March 22, 2020 to estimate the parameters and predict the trend of public opinion, as shown in Fig. 9(c1). In addition, we extend the data until March 23, 2020 to predict the rest of the public opinion trend in this phase, as shown in Fig. 9(c2). At this stage, the predicted reading and forwarding population have a big value at the starting point and then it starts to fall. With the decrease in reading and forwarding, we can infer that the dramatic outbreak of public opinion only appears at the beginning of this stage.

---

**Table 6**

Parameter results.

|    | $\beta$  | $\alpha$ | $\gamma$ | $p$      | $q$  | $\theta$ | $S_0$       |
|----|----------|----------|----------|----------|------|----------|-------------|
| Fig. 6(a1) | 0.0049    | 0.0018   | 0.3553   | $6.0042 \times 10^{-4}$ | 0.0307 | 0.3437   | $6.1797 \times 10^{5}$ |
| Fig. 6(a2) | 0.0088    | 0.0310   | 0.4841   | $2.9000 \times 10^{-4}$ | 0.0392 | 0.5191   | $5.2450 \times 10^{5}$ |
| Fig. 6(b1) | 0.0152    | 0.9306   | 0.0450   | $7.9350 \times 10^{-5}$ | 0.3230 | 0.4370   | $2.7494 \times 10^{7}$ |
| Fig. 6(b2) | 0.0150    | 1.2250   | 0.0381   | $8.6383 \times 10^{-5}$ | 0.2831 | 0.5673   | $3.4533 \times 10^{7}$ |
| Fig. 6(c1) | 0.0360    | 0.7140   | 0.0265   | $1.0531 \times 10^{-4}$ | 0.2330 | 0.1050   | $4.5804 \times 10^{8}$ |
| Fig. 6(c2) | 0.0099    | 0.7056   | 0.0140   | $1.5568 \times 10^{-4}$ | $2.7685 \times 10^{-4}$ | 0.0043 | $1.0001 \times 10^{7}$ |
| Fig. 6(d1) | 0.0280    | 2.0401   | 0.1041   | $6.0000 \times 10^{-5}$ | 0.8191 | 0.3230   | $1.1055 \times 10^{5}$ |
| Fig. 6(d2) | 0.0450    | 0.4304   | 0.0110   | $2.4700 \times 10^{-4}$ | 0.4040 | 0.6360   | $1.9527 \times 10^{5}$ |
Then public opinion entered the fourth stage due to occasional fluctuations in the reading quantity and forwarding quantity, we use the data between April 3, 2020 and April 9, 2020 to realize a good prediction for the next three days, as shown in Fig. 9(d1). Then we extend the data until April 10, 2020 and have a good prediction until the end of this stage, as shown in Fig. 9(d2). At the beginning of this stage, the predicted reading and forwarding population have a maximum value and show a downward trend. With the decrease in reading and forwarding, we can give the conclusion that the dramatic outbreak of public opinion only appears at the beginning of this stage.

Table 7 gives the results of the predicted effective reproduction ratio \( R_e \) at each time using the estimation results of the last time. We can see the reading quantity, forwarding quantity and \( R_e \) per day together from January 17, 2020 to April 16, 2020 more clearly, and the three curves have similar trends at each stage as shown in Fig. 10. Our SRFI model predicts it has the greatest reproduction ratio \( R_e = 6.5656 \) and breaks out quickly in the early stage of COVID-19. With the development of the epidemic, \( R_e \) starts to be less than 1 in the second stage, meaning that public opinion is gradually calming down. Then it started to increase again at the beginning of the third stage, but it is still less than 1, indicating that although reading and forwarding users have increased at first, public opinion will not continue to erupt. With the occurrence of emergency \( R_e \) suddenly rises greater than 1 which indicates that in the future, under the trend of overall stability, the information on COVID-19 will continue to wave burst with the occurrence of emergencies.

### 3.4. COVID-19 information contact and participation sensitivity analysis and intervention strategies

To further analyze the different parameters responsible for the comprehensive SRFI model, we use the partial rank correlation coefficients (PRCCs) [27] based on 1000 samples for various input parameters against the threshold condition to evaluate the sensitivity. According to the histogram and scatter diagram of \( R_{10}^1 \) dependence, when the correlation is positive, it means that with the increase of the value of the parameter, the value of corresponding index will increase. On the contrary, when the correlation is negative, the index will decrease as the parameter decreases.

Fig. 11 shows that the values of the public opinion reproduction ratio \( \mathcal{R}_0 \) is strongly positively affected by parameters \( \beta, p \) and the initial susceptible users \( S_0 \), and negatively affected by parameter \( \alpha \). This result confirms the correctness of our derivation of \( \mathcal{R}_0 \) in Formula (15). Therefore, the average exposure rate \( \beta \), the forwarding probability \( p \), the probability of inactivation after forwarding \( \alpha \), and the initial value \( S_0 \) are the key factors in determining the event outbreak. Thus, if we want to make public opinion explode, such as positive topics of the COVID-19, we can achieve it by increasing the value of \( \beta, p, S_0 \) and decreasing the value of \( \alpha \). Since the parameter \( \beta \) is the average exposure rate for a user to contact the information and the \( S_0 \) is the initial value of the susceptible population, we can increase these two values by persuading some opinion leaders to participate in the information propagation. Since each opinion leader will motivate a large number of new susceptible users, it can effectively increase the value of \( \beta \) and \( S_0 \). Besides, we can make the content richer and more interesting to attract people to participate in the forwarding of topics and keep users active in the forwarding for a longer time to increase the value of parameter \( p \) which is influenced by users’ interest in topics and decrease the value of \( \alpha \). Correspondingly, if we do not want public opinion to erupt, such as rumor topics of the COVID-19, we need to reduce the value of parameter \( \beta \) and \( S_0 \). Thus, we can motivate the platform to delete the relevant topics of the rumors about the COVID-19, which can effectively reduce the new reading and curb the outbreak of public opinion.

In addition, the final size \( C_{REC}, C_{F\infty} \) and the maximum value \( R_{\infty} \) and \( F_{\infty} \) are also our concern. \( C_{REC} \) and \( R_{\infty} \) denote the final size of cumulative reading and the high peak of reading quantity which can reflect the users’ contact. From Fig. 12
Table 7
The results of predicted public opinion reproduction ratio $\Re_e$.

| Data     | 2020.1.17 | 2020.1.18 | 2020.1.19 | 2020.1.20 | 2020.1.21 | 2020.1.22 | 2020.1.23 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 6.5656    | 6.5444    | 6.3993    | 6.1255    | 5.6959    | 5.1075    | 4.4263    |

| Data     | 2020.1.24 | 2020.1.25 | 2020.1.26 | 2020.1.27 | 2020.1.28 | 2020.1.29 | 2020.1.30 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 3.7758    | 3.2554    | 2.8779    | 2.6011    | 2.3918    | 2.2262    | 2.0905    |

| Data     | 2020.1.31 | 2020.1.32 | 2020.1.33 | 2020.1.34 | 2020.1.35 | 2020.1.36 | 2020.1.37 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 1.9776    | 1.8819    | 1.7994    | 1.7273    | 1.6635    | 1.6066    | 1.5557    |

| Data     | 2020.2.7  | 2020.2.8  | 2020.2.9  | 2020.2.10 | 2020.2.11 | 2020.2.12 | 2020.2.13 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 1.5101    | 1.4686    | 1.43      | 1.3956    | 1.36      | 1.3337    | 1.3058    |

| Data     | 2020.2.14 | 2020.2.15 | 2020.2.16 | 2020.2.17 | 2020.2.18 | 2020.2.19 | 2020.2.20 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 1.28      | 1.2563    | 1.2339    | 1.2162    | 1.1926    | 1.0644    | 0.1452    |

| Data     | 2020.2.21 | 2020.2.22 | 2020.2.23 | 2020.2.24 | 2020.2.25 | 2020.2.26 | 2020.2.27 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 0.2336    | 0.3384    | 0.4409    | 0.5284    | 0.5973    | 0.65      | 0.6902    |

| Data     | 2020.2.28 | 2020.2.29 | 2020.3.1  | 2020.3.2  | 2020.3.3  | 2020.3.4  | 2020.3.5  |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 0.7209    | 0.7447    | 0.7633    | 0.7781    | 0.7898    | 0.7993    | 0.8069    |

| Data     | 2020.2.36 | 2020.2.37 | 2020.2.38 | 2020.2.39 | 2020.2.40 | 2020.2.41 | 2020.2.42 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 0.8131    | 0.8181    | 0.8223    | 0.8257    | 0.8285    | 0.8309    | 0.8328    |

| Data     | 2020.3.13 | 2020.3.14 | 2020.3.15 | 2020.3.16 | 2020.3.17 | 2020.3.18 | 2020.3.19 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 0.8344    | 0.8357    | 0.8368    | 0.9704    | 0.9617    | 0.9506    | 0.9361    |

| Data     | 2020.3.20 | 2020.3.21 | 2020.3.22 | 2020.3.23 | 2020.3.24 | 2020.3.25 | 2020.3.26 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 0.9173    | 0.9392    | 0.8620    | 0.8223    | 0.7726    | 0.7114    | 0.6385    |

| Data     | 2020.3.27 | 2020.3.28 | 2020.3.29 | 2020.3.30 | 2020.3.31 | 2020.3.32 | 2020.3.33 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 0.5553    | 0.4656    | 0.3751    | 0.2907    | 0.2178    | 0.1591    | 0.4602    |

| Data     | 2020.4.3  | 2020.4.4  | 2020.4.5  | 2020.4.6  | 2020.4.7  | 2020.4.8  | 2020.4.9  |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 2.3455    | 0.162     | 0.1976    | 0.2397    | 0.2884    | 0.343     | 0.402     |

| Data     | 2020.4.10 | 2020.4.11 | 2020.4.12 | 2020.4.13 | 2020.4.14 | 2020.4.15 | 2020.4.16 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| $\Re_e$  | 0.4628    | 0.5223    | 0.577     | 0.6246    | 0.6636    | 0.6941    | 0.7167    |

we can see that the parameter $\beta$ and the initial susceptible users $S_0$ have a positive impact on both two indexes. Similarly, $C_{F_{\infty}}$ and $F_{\max}$ denote the final size of cumulative forwarding users and the high peak of the forwarding population which can reflect the users' participation. From Fig. 12 we can see that the parameter $\gamma$ and $p$ have a positive impact and $\beta$ has a small negative impact on both two indexes. The increase in parameter $\beta$ leads to an increase in reading quantity, and in the stage of a rapid increase in forwarding quantity, parameter $\gamma$ and $p$ have a relatively large value.

Combined with PRCC results, here we take topic # Real-time broadcast of joint prevention and control of epidemic situation # as an example to analyze the specific effects of parameters in our SRFI model. Fig. 13 depicts the effects of parameter $\beta$ and initial data $S_0$ on $R$-population and $C_R$ respectively. It shows that both the parameter $\beta$ and initial data $S_0$ have a positive effect on the number of reading users ($R$) and cumulative reading population ($C_R$). Comparatively speaking, the larger the parameter $\beta$ is, the earlier the reading peak appears and the shorter the outbreak duration. And the greater the initial data $S_0$ is, the larger the reading peak value will be without changing the outbreak duration is. Furthermore, it
shows that the initial data \( S_0 \) has a much important influence on the outbreaking behaviors of the topic reading related to COVID-19.

Fig. 14 depicts the effects of parameters \( p \) and \( \gamma \) on \( F \)-population and \( C_F \) respectively. It shows that both the parameters \( p \) and \( \gamma \) have a positive effect on the number of forwarding users \( (F) \) and cumulative forwarding population \( (C_F) \). The larger the parameter \( p \) is, the greater the forwarding peak value and the final size of forwarding users will be without changing the outbreak duration. By contrast, the parameter \( \gamma \) has a more subtle impact on forwarding related to COVID-19, the greater the parameter \( \gamma \) is, the earlier the forwarding peak appears and the slower the outbreak velocity and the propagation decline velocity are.

Thus, if we want to increase users’ attention on topics of the COVID-19, we can achieve it by persuading some opinion leaders who have a large number of fans to participate in the information propagation to increase the value of \( \beta \) and \( S_0 \). Correspondingly, if we want to decrease users’ attention on topics of the COVID-19, we can motivate the platform to delete the relevant topics, which can effectively reduce the reading quantity. In addition, if we want to increase users’ participation in topics of the COVID-19, we can achieve it through increasing the value of \( \gamma \) and \( p \) by making content more innovative to attract people to participate in the forwarding of topics. Here, we have similar strategies as the sensitivity analysis of opinion reproduction ratio \( R_0 \).

4. Conclusions

In this paper, we proposed a comprehensive susceptible–reading–forwarding–immune (SRFI) dynamics model based on the reading quantity and forwarding quantity in Chinese Sina-Microblog to understand both the contribution of users’ contact and participation behavior to information propagation about the COVID-19. The particular feature mechanism of social network that users may re-enter to the susceptible state to have more chance to contact information from reading state or from forwarding state subjectively is discussed, which prompt an active understanding of the epidemic.
We analyzed the public opinion data of crucial moment about the COVID-19 from January 17, 2020 to April 16, 2020 on Chinese Sina-microblog and stratified events development into different stages according to the disease outbreak development. We performed the numerical simulation on two typical topics about the COVID-19 based on both cumulative reading and forwarding quantities to verify the effectiveness of our model. Then in each stage, we used a small amount of data for parameter estimation and then used the parameterized model for trend prediction which agreed with both the real data well until the next event occurred. For characteristic parameters, a PRCC sensitivity analysis was completed that provides some perceptions in design of some effective strategies. We hope this paper could provide a tool efficiently for predicting the direction of public opinion and stabilizing public emotions with on-going COVID-19 development.

**CRediT authorship contribution statement**

**Fulian Yin**: Conceptualization, Methodology, Software, Project administration, Funding acquisition. **Hongyu Pang**: Formal analysis, Software, Data curation, Writing - original draft. **Xinyu Xia**: Validation, Visualization, Investigation, Supervision. **Xueying Shao**: Software, Validation. **Jianhong Wu**: Writing - review & editing.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Acknowledgments**

The work was supported by the National Natural Science Foundation of China (Grant numbers: 61801440), the Natural Science and Engineering Research Council of Canada, the Canada Research Chair Program (JWu), the Fundamental Research Funds for the Central Universities, PR China and the High-quality and Cutting-edge Disciplines Construction Project for Universities in Beijing, PR China (Internet Information, Communication University of China), State Key Laboratory of Media Convergence and Communication, Communication University of China.

**References**

[1] https://weibo.com/2803301701/1rBy3Ddr4?type=comment#_rnd1581399415713.
[2] https://weibo.com/6362632222/1r7QJC1mL?type=comment.
[3] Q. Gao, F. Abel, G.J. Houben, et al., A comparative study of users’ microblogging behavior on Sina Weibo and twitter, in: International Conference on User Modeling and Adaptation and and Personalization, Springer, Berlin and Heidelberg, 2012, pp. 88–101.
[4] F. Chen, A susceptible-infected epidemic model with voluntary vaccinations, J. Math. Biol. 53 (2) (2006) 253–273.
[5] Z. Lu, S. Gao, L. Chen, Analysis of an SI epidemic model with nonlinear transmission and stage structure, Acta Math. Sci. 23 (4) (2003) 440–446.
[6] W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics-I. 1927, Bull. Math. Biol. 53 (1–2) (1991).
[7] L. Stone, B. Shulgin, Z. Agur, Theoretical examination of the pulse vaccination policy in the SIR epidemic model, Math. Comput. Modelling 31 (4–5) (2000) 207–215.
[8] D. Xiao, S. Ruan, Global analysis of an epidemic model with nonmonotone incidence rate, Math. Biosci. 208 (2) (2007) 419–429.
[9] M. Li, J.S. Muldowney, Global stability for the SEIR model in epidemiology, Math. Biosci. 125 (2) (1995) 155–164.
[10] C. Xia, S. Sun, F. Rao, SIS model of epidemic spreading on dynamical networks with community, Front Comput. Sci. China 3 (3) (2009) 361–365.
[11] D.J. Daley, D.C. Kendall, Epidemics and rumours, Nature 204 (4963) (1964) 1118.
[12] L. Zhao, Q. Wang, J. Cheng, et al., Rumor spreading model with consideration of forgetting mechanism: A case of online blogging LiveJournal, Physica A 390 (13) (2011) 2619–2625.
[13] L. Zhao, J. Wang, Y. Chen, et al., SIHR rumor spreading model in social networks, Physica A 391 (7) (2012) 2444–2453.
[14] F. Xiong, Y. Liu, Z. Zhang, et al., An information diffusion model based on retweeting mechanism for online social media, Phys. Lett. A 376 (30–31) (2012) 2103–2108.

[15] L. Zhao, X. Wang, J. Wang, et al., Rumor-propagation model with consideration of refutation mechanism in homogeneous social networks, Discrete Dyn. Nat. Soc. 2014 (2014) 1–11.

[16] Y. Zhang, J. Xu, A rumor spreading model considering the cumulative effects of memor, Discrete Dyn. Nat. Soc. 2015 (2015) 1–11.

[17] X. Rui, F. Meng, Z. Wang, et al., SPIR: The potential spreaders involved SIR model for information diffusion in social networks, Physica A 506 (2018) 254–269.

[18] L. Zhao, X. Wang, J. Wang, et al., Rumor-propagation model with consideration of refutation mechanism in homogeneous social networks, Discrete Dyn. Nat. Soc. 2014 (2014) 1–11.

[19] J. Huang, Q. Su, A rumor spreading model based on user browsing behavior analysis in microblog, in: 2013 10th International Conference on Service Systems and Service Management, Vol. 2013, IEEE, pp. 170–173.

[20] Y. Zan, DSIR double-rumors spreading model in complex networks, Chaos Solitons Fractals 110 (2018) 190–203.

[21] F. Yin, X. Shao, J. Wu, Nearcasting forwarding behaviors and information propagation in Chinese Sina-Microblog, Math. Biosci. Eng.: MBE 16 (5) (2019) 5380–5394.

[22] D. Trpevski, W.K.S. Tang, L. Kocarev, Model for rumor spreading over networks, Phys. Rev. E 81 (5) (2010) 056103.

[23] Z. Qian, S. Tang, X. Zhang, The independent spreaders involved SIR rumor model in complex networks, Physica A 429 (2015) 95–103.

[24] J. Wang, L. Zhao, R. Huang, SIRaRu rumor spreading model in complex networks, Physica A 398 (2014) 43–55.

[25] M. Tanaka, Y. Sakamoto, M. Aida, et al., Study on the growth and decline of SNSs by using the infectious recovery SIR model, in: 2015 10th Asia-Pacific Symposium on Information & Telecommunication Technologies, 2015, pp. 1–3.

[26] X. Feng, X. Huo, B. Tang, et al., Modelling and analyzing virus mutation dynamics of chikungunya outbreaks, Sci. Rep. 9 (1) (2019) 1–15.

[27] U.L. Abbas, R.M. Anderson, J.W. Mellors, Potential impact of antiretroviral chemoprophylaxis on HIV-1 transmission in resource-limited settings, PLOS One 3 (2007).