Research article

Nearcasting forwarding behaviors and information propagation in Chinese Sina-Microblog

Fulian Yin[1], **Xueying Shao**[1] **and Jianhong Wu**[2,*]

[1] College of Information and Communication Engineering, Communication University of China, Beijing, 100024, P.R. China

[2] Laboratory for Industrial and Applied Mathematics, York University, Toronto, M3J1P3, Canada

* **Correspondence:** Email: wujh@yorku.ca; Tel: +14167365243; Fax: +14167365698.

Abstract: As the largest social media in China, the Sina-Microblog plays an important role in public opinion dissemination. Despite intensive efforts in understanding the information propagation dynamics, the use of a simple outbreak model to generate summative indices that can be used to characterize the time series of a single Weibo event has not been attempted. This work fills this gap, and illustrates the potential of using a simple outbreak model in conjunction with the historical data about the cumulative forwarding users for nearcasting the propagation trend.

1. Introduction

The Chinese Sina-Microblog is reported to have 446 million active users per month (32% of the Chinese population) at the end of 2018 and is particularly popular among the young generations. Take the event (Weibo) "304 car theft case" as an example, around 185 thousand of forwarding users were generated into 36 hours since it was posted [1]. This important platform presents some unique characteristics different from other microblogs for public opinion diffusion [2, 3] that a systematic approach towards understanding the diffusion patterns and nearcasting the information propagation is called for.

The information propagation of a microblog is created through the action of "follow" connecting as an interaction network, where a following user is called "follower" and a followed user is called "followee" [4, 5]. There are a large amount of literature and there seem to be three typical and important information diffusion models developed for microblogs [6] according to the spreading process, influence ability [7, 8] and forwarding factors [9, 10]. The model we are exploring is based on

the propagation process as we focus on the dynamic changes of the status of users during the spreading process.

Our approach follows the classical rumor model that evolved from the epidemic models, in which the population is stratified into three mutually exclusive and exhaustive classes: heard rumor (ignorants), actively spreading rumor (spreaders) and no longer spreading rumor (stiflers). The propagation of information (rumor) occurs through the transitions of infections from ignorants to spreaders, and removals from spreader to stiflers [11]. This is very much similar to epidemic models for infectious disease spreading in the population. Extensions from this classic rumor model to information spreading in microblogs include propagation transition models [12–16] and population classification models [17–21]. In particular, taking into account of different behaviors of differential users in different platforms, Huang et al. [12] extended the rumor spreading model to characterize browsing behaviors of users and examine differentiating rumor refuting effects. This model involves a parameter similar to the birth rate in a standard epidemic model to describe the entry of new susceptibles. Liu et al. [13] proposed a modified rumor spreading model (SIRe), where two contact parameters are used to describe the stifler's broadcasting effect and social intimacy degree. Su et al. [14] developed a Microblog-Susceptible-Infected-Removed (Mb-SIR) model for information propagation by explicitly considering the incomplete reading behaviors of users using the probability that a newly posted or retweeted message will be read by its followers. These models also utilized real data from the Chinese Sina-Microblog in their numerical simulations.

Our framework follows the work of Borge-Holthoefer et al. [15] that considered the situation where spreaders are not always active and an ignorant is not interested in spreading the rumor. Their numerical simulations were based on data from Twitter. A similar model for an information dissemination network was proposed in [16], where different transition probabilities from the spreader stage to stifler stage and from the ignorant stage to stifler stage were used. This set a theoretical foundation to compare microblog information dissemination and epidemic disease spreading. There have been other efforts to incorporate further states during the propagation process, in addition to the states of ignorant, spreader and stifler. In particular, reflecting some information features and diffusion characteristics of Weibos in the Chinese Sina-Microblog, Li et al. [17] incorporated the number of fans of infectious and validated the modified SIR model using actual event data. Liu et al. [18] considered a dynamic model to characterize the super-spreading phenomenon in tweet information propagation. An ignorant spreaders, super-spreaders, stiflers model (ISJR model) was proposed in [19] that considered the role of super-spreaders to show how super-spreaders can accelerate the information dissemination and amplify information influence in microblogging networks. Other models have also been proposed to incorporate the recovery state of users in different platforms including the Chinese Microblogs, Japanese Mixi, and Facebook [20, 21].

Here we consider the capacity of using a simple compartmental model for information propagation of Microblog for nearcasting the trend of information propagation in the Chinese Microblogs. Nearcasting is an important issue for assessing public opinions, that aims to project the forwarding trend at the earliest possible stage of a Weibo outbreak so interventions for the information propagation and/or rapid response to the public opinions can be designed and implemented effectively. Our ultimate goal is to develop nearcasting technologies for a group of Weibos (similar to "Tweet" in Twitter) in the ecosystem of Chinese Microblogs. To achieve this goal, we need to develop computable summative indices to characterize each Weibo, and see how these indices can be

effectively calculated or at least estimated from the public available information at the early stage of information propagation of the Weibo. Here,we formulate the basic compartmental model (SFI) and introduce the Weibo propagation indices in the Chinese Sina-Microblog (Section 2), and then contrast two methods to use this SFI-model for nearcasting.

2. The SFI model for forwarding behavior

2.1. Model description

We consider a population of Chinese Sina-Microblog users, stratified in terms of three distinct states: the susceptible state (S), in which users are unaware of but susceptible to the Weibo; the forwarding state (F), in which users have been forwarding the Weibo actively to influence other users; and the immuned state (I), in which users have already forwarded the Weibo, but are no longer forwarding the Weibo and even if receiving it again.

Assuming the populations susceptible to the Weibo is sufficiently large and denoting the total numbers of users in the susceptible, forwarding and immune state at time $t \geq 0$ by $S(t)$, $F(t)$ and $I(t)$, respectively, we obtain the following susceptible-forwarding-immune model:

$$S'(t) = -\beta S(t)F(t), \tag{2.1}$$

$$F'(t) = p\beta S(t)F(t) - \alpha F(t), \tag{2.2}$$

$$I'(t) = (1 - p)\beta S(t)F(t) + \alpha F(t), \tag{2.3}$$

where $' = d/dt$ is the derivative with respect to t. The model is illustrated in Figure 1.
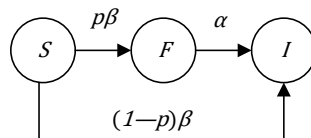

Figure 1. A schematic diagram to illustrate the Weibo spreads in the population with three different states: susceptible (S), forwarding (F), immuned (I).

In the model, β is the average number of exposures to the Weibo of a susceptible user, p is the probability that the exposed user will forward the Weibo and α is the rate at which a user in the F state becomes inactive to forwarding. The mass action term can also be interpreted as follows: an active forwarding user will contact an average number of βN users per unit time, among which $p\beta N$ will choose to forward the Weibo and $(1 - p)\beta N$ will not. Since the probability of a contacted user to be a susceptible user is $S(t)/N$, the number of exposed users who leave the state S are $\beta S(t)F(t)$ among which new forwarding and immuned users are $p\beta N(S(t)/N)F(t) = p\beta S(t)F(t)$ and $(1 - p)\beta S(t)F(t)$, respectively. As usual, $1/\alpha$ is the average duration a F-user remains active in forwarding.

An important distinction between a standard epidemic-SIR model and the Weibo-SFI model is the direct immunity of a susceptible user gained through exposure to the Weibo. The parameter $1 - p$ reflect the suitability of the Weibo for a susceptible user to trick the activity of forwarding. This SFI model

was used in [15, 16] as a basic building block of a more complicated framework. We adopt this for the Weibo information spread dynamics model in the hope this can be further expanded to discuss the dynamics of an ecosystem consisting of multiple Weibos sharing a same set of key words. The novelty of our approach here is to develop some analytic indices which will be used in our subsequent studies to consider Weibo spreading dynamics in a complex Weibo ecosystem, and explore the feasibility of using openly available Weibo data from the Chinese Sina-Microlog to estimate these indices for the purpose of nearcasting the propagation trend.

The official Chinese Sina-Microblog only provides limited information of the propagation, an important piece of information we can obtain from the Chinese Sina-Microblog directly is the number of cumulative forwarding users given by

$$E(t) = \int_0^t p\beta S(t)F(t)dt. \tag{2.4}$$

Observe that this is NOT a new compartment from the coupled system (2.1). Instead, from Eq (2.1), we can obtain

$$E'(t) = p\beta S(t)F(t). \tag{2.5}$$

It is easy to show that $S(t)$, $F(t)$, $I(t)$ and hence $E(t)$ all remain nonnegative. We consider the case when the Weibo is posted by a single user at the beginning, leading the initial condition: $E_0 = F_0 = 1$, $I_0 = 0$, $S_0 = N - F_0 = N - 1$.

From Eqs (2.1) and (2.5) it follows that $S(t)$ is decreasing since $S'(t) = -\beta S(t)F(t) < 0$, and that the function $E(t)$ is increasing since $E'(t) = p\beta S(t)F(t) > 0$. Therefore, $S(t)$ decreases to a limit $S_\infty := \lim_{t\to\infty} S(t) > 0$, $E(t)$ increases to a limit $E_\infty := \lim_{t\to\infty} E(t) < N$, $F(t)$ tends to 0 ($F_\infty = 0$), and $I_\infty = N - S_\infty$. These limits are shown to be relevant to the so-called final size of the Weibo spreading.

2.2. The Weibo propagation indices

Weibo reproduction ratio $\mathfrak{R}_o$: The initial growth of the F-population is given by $r = p\beta S_0 - \alpha$, and the propagation of the Weibo will never take off if $F'(t) > 0$. Since $F'(t) \leq (p\beta S_0 - \alpha)F(t)$ due to the decreasing property of $S(t)$ ($S(t) < S_0$). We then define

$$\mathfrak{R}_o := \frac{p\beta S_0}{\alpha} \tag{2.6}$$

as the Weibo reproduction ratio. Then $\mathfrak{R}_o < 1$ implies a rapid decline of F-user which results in information propagation never taking off. However, when $\mathfrak{R}_o > 1$ the F-population used grow exponentially initially.

Index $\mathfrak{R}_o$ denotes the number of F-users generated by one active forwarding user during an active period. A typical curve for F-population with a Weibo with $\mathfrak{R}_o > 1$ has the bell shape as shown in Figure 2.
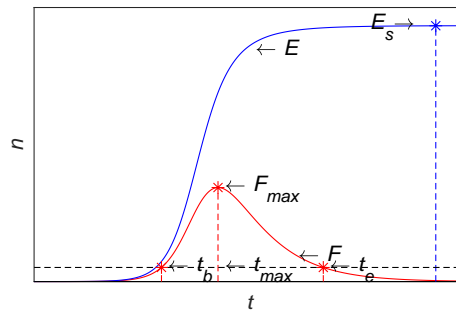
Figure 2. The temporal variation of the actively forwarding and cumulative forwarding users, $F(t)$ and $E(t)$, expressed as functions of the time variable t (horizontal axis), throughout the paper, the time unit is day and the variables for the numbers of the users in different compartments are real nonnegative numbers.

Maximal Weibo forwarding users F_{max}: The maximum of the F-curve is achieved when $F'(t) = 0$. At this point, we have $S(t) = \alpha/p\beta$. We also use Eqs (2.1) and (2.2) to get

$$F'(t) = \left(\frac{\alpha}{\beta S(t)} - p\right) S'(t). \tag{2.7}$$

Integrating then yields

$$F(t) = F_0 + p\left[S_0 - S(t)\right] + \frac{\alpha}{\beta} \log \frac{S(t)}{S_0}. \tag{2.8}$$

If we substitute $F_0 = 1$, we get (notify $F_{max} = F(t)$ for t when $S(t) = \alpha/p\beta$) that

$$F_{max} = 1 + pS_0 - \frac{\alpha}{\beta}\left(1 - \log \frac{\alpha}{p\beta S_0}\right) = 1 + pS_0 - \frac{pS_0}{\mathcal{R}_o}\left(1 - \log \frac{1}{\mathcal{R}_o}\right). \tag{2.9}$$

Maximum Weibo cumulative forwarding users E_s: The maximum Weibo cumulative forwarding users E_s, which represents the number of users that encountered the Weibo, is also a factor we are interested in. To establish an analytic formula, we integrate from 0 to $+\infty$ on both sides of Eq (2.1) to get

$$\int_0^{+\infty} \frac{1}{S(t)} dS(t) = -\beta \int_0^{+\infty} F(t)dt. \tag{2.10}$$

Adding Eqs (2.1) and (2.2) and integration yields also

$$p \int_0^{+\infty} dS(t) + \int_0^{+\infty} dF(t) = -\alpha \int_0^{+\infty} F(t)dt. \tag{2.11}$$

From Eqs (2.11) and (2.12), we get

$$\log\left(\frac{S_\infty}{S_0}\right) = \frac{\beta}{\alpha}\left[p\left(S_\infty - S_0\right) + (F_\infty - F_0)\right]. \tag{2.12}$$

In order to analyze the relation between E and S, we integrate from Eqs (2.1) and (2.5) to obtain

$$S_\infty = S_0 - \frac{E_\infty - E_0}{p}. \tag{2.13}$$

Similarly, $E_0 = F_0 = 1$, $F_\infty = 0$ and $\mathfrak{R}_o = p\beta S_0/\alpha$, from Eqs (2.12) and (2.13) we get (notify $E_s := E_\infty$) that

$$\log\left(1 - \frac{1}{pS_0}(E_s - 1)\right) = -\frac{\beta}{\alpha}E_s = -\frac{\mathfrak{R}_o}{pS_0}E_s \tag{2.14}$$

and then

$$E_s = 1 + pS_0\left(1 - e^{-\frac{\beta}{\alpha}E_s}\right) = 1 + pS_0\left(1 - e^{-\frac{\mathfrak{R}_o}{pS_0}E_s}\right). \tag{2.15}$$

We observe from Eqs (2.9) and (2.15) that, with a fixed $\mathfrak{R}_o$, the final outcome of the Weibo spread is determined by parameter S_0. Note that, differently from an epidemic study, the initial susceptible Weibo users are usually unknown.

Weibo public opinion times and velocities: To understand the global performance of the Weibo ecosystem, we need to know the take-off and extinctive time of a particular Weibo, denoted by t_b and t_e, respectively. The definitions depend on the threshold F^* which is set in advance such that $F(t_b) = F^* = F(t_e)$. In the experiment we design, we let $F^* = 0.1 \times F_{max}$. The difference, $t_e - t_b$, is the duration during which the Weibo remains active in the ecosystem. We denote this by $t_i := t_e - t_b$. Relevant to the timings and the spread duration, we can define the (initial) outbreak velocity V_o, the propagation decline velocity V_d, and the average spreading velocity V_a as follows:

$$V_o = \frac{F_{max} - F^*}{t_{max} - t_b}, \tag{2.16}$$

$$V_d = \frac{F_{max} - F^*}{t_e - t_{max}}, \tag{2.17}$$

$$V_a = \frac{F_{max} - F^*}{t_e - t_b}, \tag{2.18}$$

where t_{max} is given when $F(t) = F_{max}$. Note that t_b, t_e, t_{max}, V_o, V_d and V_a can be calculated when model parameters are estimated although we do not have the closed form of these quantities. A schematic picture is given in Figure 2.

2.3. An example:

The number of cumulative forwarding users can be collected through the Weibo' Application Programming Interface (API). Table 1 lists such a data set, studied in paper [22], for an actual event lasted 16 days.

Table 1. The cumulative forwarding numbers of a particular Weibo lasted for 16 days.

t(day)	1	2	3	4	5	6	7	8
number	10	198	737	2674	9644	33943	110751	550102
t(day)	9	10	11	12	13	14	15	16
number	739491	835032	880337	913164	919055	924084	925093	925664

The LS method [23] can be used to estimate the parameters of the SFI model (2.1), (2.2), (2.3) and (2.5). The parameter vector is set as $\Theta = (\beta, \alpha, p, S_0)$, and the corresponding numerical result for $E(t)$

is denoted by $f_E(t, \Theta)$. The LS error function

$$LS = \sum_{k=0}^{T} |f_E(t_k, \Theta) - ID_k|^2 \tag{2.19}$$

can be used, where ID denotes the real number of cumulative forwarding users given in Table 1, $t_k = k$ is the sampling time, $k = 1, 2, ..., 16$.

We use the standard package (fmincon) of MATLAB to solve a nonlinear LS problem, with parameter initial values and ranges given as follows: $\beta \in [2 \times 10^{-6}, 8 \times 10^{-6}]$, $p \in [0.1, 0.5]$, $\alpha \in [0.5, 1.5]$, $S_0 \in [7 \times 10^6, 2 \times 10^7]$, $\Theta_0 = (4.0 \times 10^{-6}, 0.3, 1.0, 1.0 \times 10^7)$.
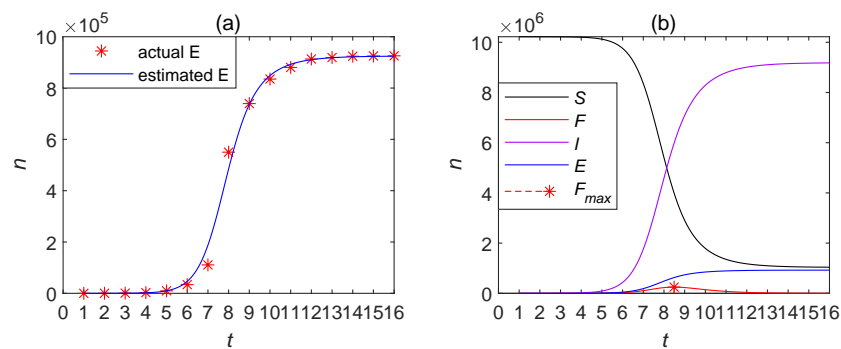

Figure 3. Numerical results of the Weibo given in Table 4, where the horizontal axis is the time in days and the vertical axis is the number of users. (a). data fitting to the actual number of E population. (b). four different populations: S-population, F-population (F_{max} is indicated with a red asterisk), I-population and E-population. The numerical results are as follows: $\beta = 2.5651 \times 10^{-6}$, $\alpha = 1.0365$, $p = 0.1006$, $S_0 = 1.0221 \times 10^7$. The relevant indices are $\mathfrak{R}_o = 2.5460$, $F_{max} = 2.4724 \times 10^5$, $E_s = 0.9243 \times 10^6$, $t_b = 5.92$(day), $t_e = 12.57$(day), $t_i = 6.65$(day), $t_{max} = 8.49$(day), $V_o = 0.8658 \times 10^5$/day, $V_d = 5.4538 \times 10^4$/day, and $V_a = 3.3461 \times 10^4$/day.

Figure 3 reports our data fitting and simulation results. Figure 3a shows the fitting result between the numerically calculated and the actual numbers of cumulative forwarding users. From the data fitting, we can estimate the model parameters and then inform other information not available from the API. Namely, we can obtain the temporal variations of the numbers of susceptible, forwarding, immunized and cumulative forwarding users, $S(t)$, $F(t)$, $I(t)$ and $E(t)$, after obtaining the model parameters.

Note that for the Weibo data listed in Table 1,the estimated basic reproductive ratio $\mathfrak{R}_o = 2.5460$, and since it is much greater than one, we should expect a rapid information spread at the beginning of the propagation. Figure 3a does confirm this expectation. We should also expect, from surveillance data of infectious disease outbreak such as influenza, a large number of F-users and a large number of E-users for the Weibo. The numerical simulations reported in Figure 3b, however, shows otherwise: despite a large number of susceptibles ($S_0 = 1.0221 \times 10^7$), the cumulative E-users during the entire propagation is only $E_s = 9.2430 \times 10^5$ and maximum F-users at the propagation peak is only $F_{max} = 2.4724 \times 10^5$. A reason is that p is only 0.1006 and hence a large number of individuals exposed to the Weibo became immunied immediately upon exposure.

3. Nearcasting capacity of SFI-modelling

3.1. Method 1 (all parameters estimated)

The numerical illustration in the last section indicates a good data fitting results using the numbers of cumulative forwarding users during the entire outbreak period. These numbers can be collected from the API, and therefore a Weibo can now be characterized by the SFI-model parameters (β, α, p, S_0), using the information from API retroactively.

An important question in Weibo information management is how many units (days) of data about the cumulative forwarding users we need to be able to estimate the model parameter and thus make the prediction for the near future about the propagation trend and calculate key Weibo indices. Figure 4 reports our numerical experiments in which parameters (β, α, p, S_0) are estimated from the past days, and then the cumulative F-users are predicted and compared with the actual data from API data.
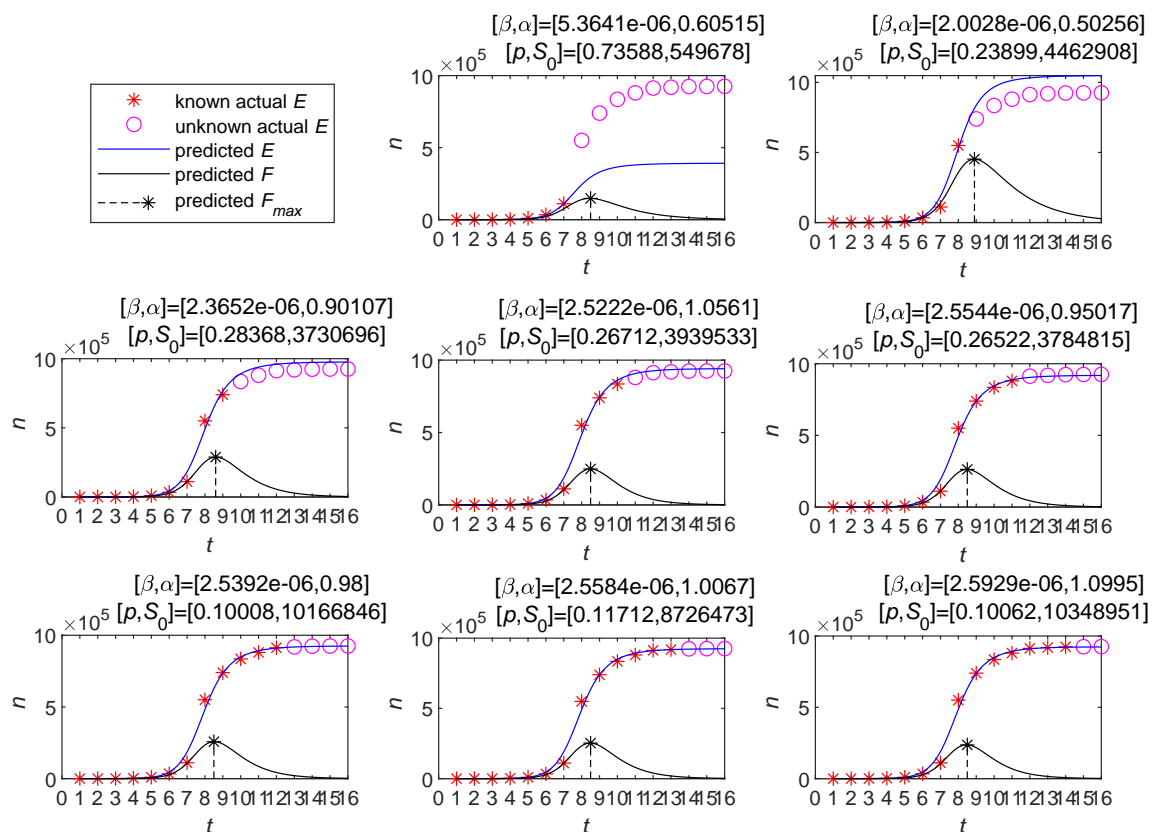

Figure 4. Numerical experiments with all parameter estimated based on days since the outbreak started: the numbers of the E-users which we used for prediction are marked with a red asterisk; the numbers of the E-users which we predict are marked as pink circles; the predicted E-population and F-population are shown using a blue line and black line, respectively; and the maximal F-population is marked as black asterisk. In the figure, the horizontal axis is the time (as unit time by day) and the vertical axis is the number of users.

Comparing with the estimated parameters $\beta = 2.5651 \times 10^{-6}$, $\alpha = 1.0365$, $p = 0.1006$, and $S_0 = 1.0221 \times 10^7$ from the entire outbreak duration, we observe that the estimated parameter and the resulted SFI-model prediction does not fit the actual data until 7 days have pasted, when the peaking time is 8.49 days.

3.2. Method 2 (a-prior estimation)

Therefore, the use of an SFI-model for nearcasting is not promising if we need to use the historical data to estimate all the model parameters altogether. On the other hand, in the Chinese Sina-Microblog, parameter β is determined by the compactness of the network structure and parameter α is usually user-specific rather than Weibo-specific. Thus, it is feasible to estimate parameters β and α before the outbreak of an event (for example from other events). Figure 5 gives the numerical results with fixed β and α that are given *a-priori*.

This experiment indicates the remaining parameters (p, S_0) can be estimated from the data in the first 3-days, much ahead of the outbreak peak time. Therefore, the nearcasting capacity of an SFI-model can be significantly enhanced should the network-specific parameter β and user-specific α be estimated in advance.

3.3. Comparison

The numerical experiments with all parameter estimated (Method 1) and some parameters *a-priori* estimated (Method 2) clearly show that Method 2 has a much better performance in nearcasting with limited historical data well before the peaking time of the Weibo information outbreak.

Tables 2,3,4 give the comparison of indices $\mathfrak{R}_o$, F_{max}, E_s, indices t_b, t_s, t_i, t_{max}, and indices V_o, V_e, V_a, separately. For Method 1, Tables 2,3,4 show that at the beginning of the Weibo propagation, our predictions have low accuracy when using the historical data until Day 7. Using the historical data until after the propagation peak (Day 8), the predicted maximum F-users (F_{max}), cumulative F-users (E_s), reproduction ratio ($\mathfrak{R}_o$) and the outbreak velocity (V_o) are larger than the actual values, indicating over-estimation of the propagation potential. However, as one gains more and more historical data, the prediction converges to the actual values. For the prediction of key times during the Weibo information outbreak, Method 1 can produce predictions of key instants, after Day 7, such as take-off, extinction, duration and propagation peak time within 1-day error. The predicted maximal forwarding users (F_{max}), the outbreak velocity (V_o), the propagation decline velocity (V_e) and the average spreading velocity (V_a) are all consistent with corresponding actual instants. So, the time after which a good nearcasting prediction is possible is Day 7, and the prediction quality is very high if historical data until Day 9 is used.

Method 2 is more effective in nearcasting both in terms of estimating the peaking time at the 3-day, but also estimating all relevant indices. Therefore, it is advised that one should perform the nearcasting by estimating parameters β and α before the outbreak of the event. This proposed *a-priori* estimation method for nearcasting emphasizes the importance of understanding the propagation of other previously happened Weibos propagation events in advance, in order to better monitor and respond to public opinion dynamics at a real time.
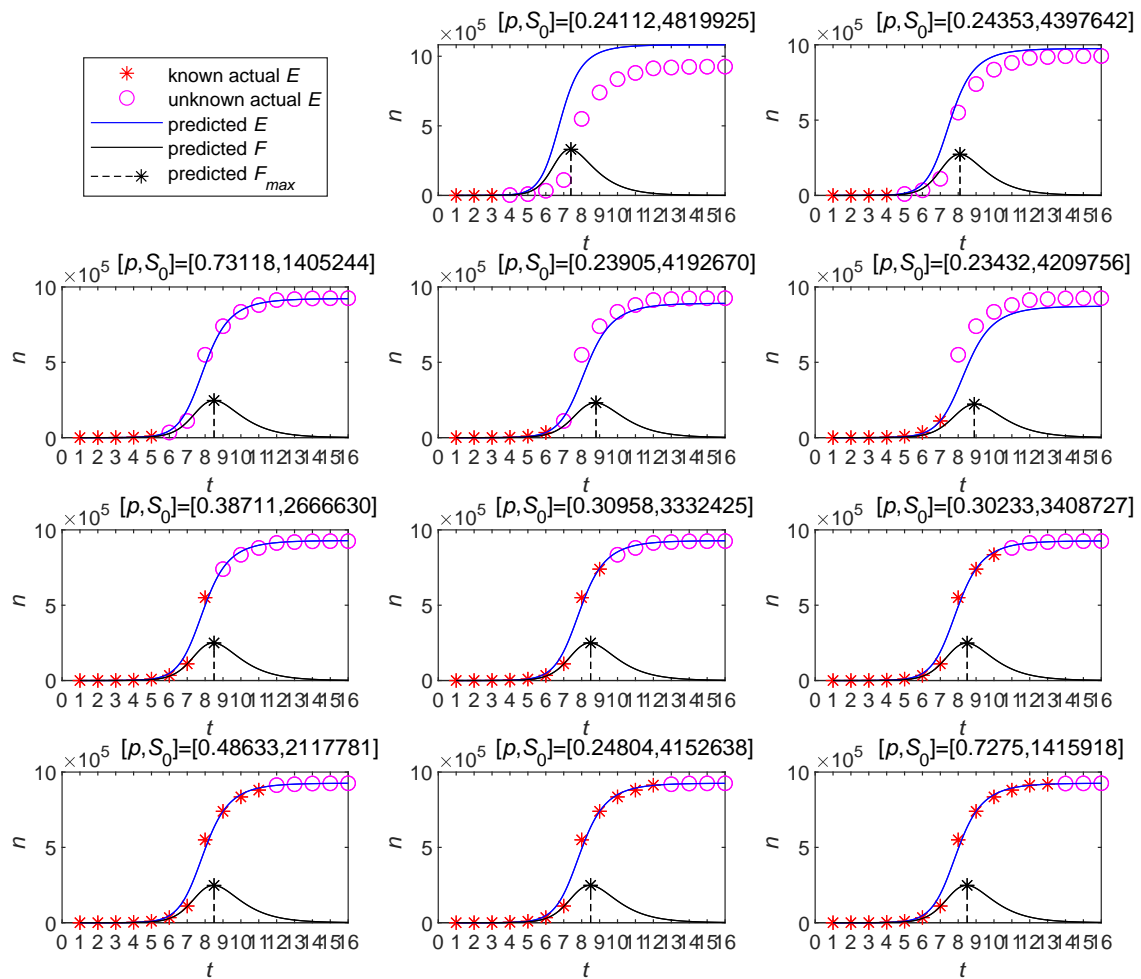
Figure 5. Numerical experiments with *a-priori* estimated (and thus) fixed β (2.5651×10^{-6}) and α (1.0365) based on days since outbreak started: the numbers of the E-users which we used for prediction are marked with a red asterisk; the numbers of the E-users which we predict are represented by pink circles; the predicted E-population and F-population are shown using a blue line and black line, respectively; and the maximal F-population is marked as black asterisk. In the figure, the horizontal axis is the time (as unit time by day) and the vertical axis is the number of users.

Table 2. Comparison of indices $\mathfrak{R}_o$, F_{max}, E_s for Method 1 and Method 2 based on historical data since the outbreak started: Method 1: all parameter estimated; Method 2: *a-priori* estimation with fixed β (2.5651×10^{-6}) and α (1.0365).

	$\mathfrak{R}_o$(actual : 2.5460)		F_{max}(actual : 2.4724×10^5)		E_s(actual : 0.9243×10^6)	
	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
Day 3	–	2.8762	–	3.3083	–	1.0824
Day 4	–	2.6504	–	2.7307	–	0.9751
Day 5	–	2.5428	–	2.4641	–	0.9228
Day 6	–	2.4803	–	2.3126	–	0.8920
Day 7	3.5855	2.4412	1.4758	2.2194	0.3920	0.8726
Day 8	4.2506	2.5546	4.5234	2.4942	1.0504	0.9286
Day 9	2.7779	2.5531	2.8837	2.4930	0.9768	0.9278
Day 10	2.5132	2.5504	2.4809	2.4866	0.9411	0.9265
Day 11	2.6986	2.5489	2.6283	2.4759	0.9189	0.9258
Day 12	2.6363	2.5490	2.5757	2.4798	0.9248	0.9258
Day 13	2.5974	2.5492	2.5285	2.4789	0.9245	0.9259
Day 14	2.4557	2.5496	2.3632	2.4796	0.9233	0.9261
Day 15	2.4561	2.5496	2.3738	2.4788	0.9272	0.9261

Table 3. Comparison of indices t_b, t_s, t_i, t_{max} for Method 1 and Method 2 based on historical data since the outbreak started: Method 1: all parameter estimated; Method 2: *a-priori* estimation with fixed β (2.5651×10^{-6}) and α (1.0365).

	t_b(actual : 5.92day)		t_e(actual : 12.57day)		t_i(actual : 6.65day)		t_{max}(actual : 8.49day)	
	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
Day 3	–	5.20	–	11.05	–	5.85	–	7.35
Day 4	–	5.67	–	12.03	–	5.85	–	8.08
Day 5	–	5.93	–	12.59	–	6.66	–	8.49
Day 6	–	6.09	–	12.95	–	6.86	–	8.76
Day 7	5.71	6.20	13.93	13.19	8.22	6.99	8.47	8.94
Day 8	6.20	5.90	14.94	12.52	9.74	6.62	8.91	8.46
Day 9	6.02	5.90	12.98	12.53	6.96	6.63	8.61	8.46
Day 10	5.93	5.91	12.56	12.55	6.63	6.64	8.50	8.47
Day 11	5.92	5.91	12.73	12.56	6.81	6.65	8.48	8.47
Day 12	5.94	5.91	12.70	12.56	6.76	6.65	8.53	8.47
Day 13	5.91	5.91	12.61	12.56	6.70	6.65	8.48	8.47
Day 14	5.90	5.91	12.44	12.56	6.54	6.65	8.46	8.47
Day 15	5.90	5.91	12.43	12.56	6.53	6.65	8.46	8.47

Table 4. Comparison of indices V_o, V_e, V_a for Method 1 and Method 2 based on historical data since the outbreak started: Method1: all parameter estimated; Method 2: *a-priori* estimation with fixed β (2.5651×10^{-6}) and α (1.0365).

	V_o(actual : 0.8658×10^5/day)		V_d(actual : 5.4538×10^4/day)		V_a(actual : 3.3461×10^4/day)	
	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
Day 3	–	1.3863	–	8.0554	–	5.0948
Day 4	–	1.0199	–	6.2227	–	3.8647
Day 5	–	0.8663	–	5.4091	–	3.3299
Day 6	–	0.7800	–	4.9701	–	3.0357
Day 7	0.4813	0.7293	2.4331	4.7019	1.6161	2.8588
Day 8	1.5023	0.8773	6.7514	5.5317	4.6580	3.3926
Day 9	1.0021	0.8769	5.9390	5.5159	3.7290	3.3861
Day 10	0.8688	0.8745	5.4995	5.4869	3.3677	3.3715
Day 11	0.9241	0.8706	5.5661	5.4494	3.4737	3.3516
Day 12	0.8952	0.8721	5.5603	5.4586	3.4299	3.3573
Day 13	0.8855	0.8718	5.5105	5.4568	3.3967	3.3561
Day 14	0.8312	0.8720	5.3464	5.4582	3.2536	3.3570
Day 15	0.8349	0.8717	5.3839	5.4562	3.2732	3.3558

4. Discussions

This paper concerned with nearcasting for forwarding behaviors in the Chinese Sina-Microblog based on the use of a simple compartmental SFI model for a Weibo to be forwarded by users exposed to the Weibo. A significant difference of Weibo information propagation from the pathogen of an infectious disease is that an exposed user may gain immunity to the Weibo so the user becomes completely uninterested in forwarding the Weibo.

The relative simplicity of the model permitted the construction of various indices and their calculations/estimations: the Weibo reproduction ratio, the maximal forwarding users, the maximum cumulative forwarding users, and critical propagation peak time, taking off and extinction times, and propagation velocities during different phases of the information outbreak. An important issue, to use this simple model to predict these critical indices and thus contribute to nearcasting propagation trend based on historical Application Programming Interface (API) data, was addressed through some numerical experiments based on a real Weibo event data. We considered two cases, where we can accurately estimate all model parameters with the historical data passing the peaking time, or we can rapidly and accurately estimate two critical parameters relevant to the Weibo when two other model parameters relevant to the network characteristic and forwarding users' waning interest rate are estimated *a-priori*. This second approach clearly shows the nearcasting capacity of a simple compartmental SFI-model as long as we have some information about the network characteristics and the anticipated public interest in a class of Weibo events.

The simple model provides two important pieces of information about the propagation characteristics of a particular Weibo: the intrinsic growth rate (the basic Weibo reproduction ratio),

and the maximal capacity of the network for the particular Weibo event (the maximal number of cumulative forwarding users). As anticipated from long time intensive modeling study of biological ecosystems, the intrinsic growth rate and the maximal capacity gives the logistic growth dynamics [24] of a single Weibo propagation dynamics. A full understanding of this propagation dynamics of a single Weibo event is the first critical step towards examining the propagation dynamics of a group of interactive Weibo events in a complex network.

There are many different model frameworks, as briefly reviewed in the introduction. We have listed existing studies which are directly relevant to our mechanistic approach towards developing a simple compartmental model for characterizing the Weibo information spread in the Sina-Microblog. We refer to [25–27] and references therein for further studies at the intersection of mathematics, epidemics, information diffusion and control. Interestingly, the work [25] that introduces a class of infants with maternal antibodies giving passive temporary immunity for the considered infectious disease may find an application for our modeling Weibo spread within the Sina-Microblog as some immediate followers of opinion leaders may inherit a certain immunity for a Weibo. Obviously, the Sina-Microblog possesses all kinds of features of complex networks [26], the sub-network structure and the network topology require much further investigations that our simple mechanistic model can serve as a building block towards a more complicated network framework. A primary goal of nearcasting the propagation trend is to inform optimal design of interventions, so the methodologies designed for the control of infection dynamics [27] may also be adopted in our simple model and its variations.

We mention also that a variant of the epidemiological SIR model was used in [28] to accurately describe the diffusion of online content over the online social network Digg.com. The work examined also qualitative properties of our viral information propagation model, demonstrate the model's applications to nearcasting social media spread in online social networks. Alternatively, a linear diffusive model was proposed and considered in Feng et al. [29, 30] using a temporal-spatial partial differential equation model to explain these rates of spread in the DOSN. The PDE and network analogue of the SFI model should be developed to reflect the spatial spread and network heterogeneity.

Acknowledgments

Conflict of interest

The authors declare that they have no competing interests.

References

1. R. Xu, How to improve the impact of traditional media microblog: based on the content analysis of People's Daily's high transmitted microblog, *Practical J.*, **6** (2013), 15–18.

2. Q. Gao, F. Abel, G. J. Houben, et al., A Comparative Study of Users Microblogging Behavior on Sina Weibo and Twitter, *2012 UMAP,* Springer, Berlin, Heidelberg, (2012), 88–101.

3. X. Shuai, X. Z. Liu, T. Xia, et al., Comparing the pulses of categorical hot events Twitter and Weibo, *Proceedings of the 25th ACM Hypertext,* ACM, (2014), 126–135.

4. X. Li, S. Cheng, W. Chen, et al., Novel user influence measurement based on user interaction in microblog, *2013 IEEE/ACM ASONAM*, 615–619.

5. N. Zhang, Y. Chai, Y. Li, et al., Modeling micro-blog network structure based on combination of online communities, *CDC*, IEEE, (2013), 3419–3424.

6. H. Chen, Z. Xiao and C. Xin, Survey on information diffusion in microblog, *Appl. Res. Comput.*, **31** (2014), 333–338.

7. H. Huo and X. Zhang, Modeling the influence of twitter in reducing and increasing the spread of influenza epidemics, *SpringerPlus*, **5** (2016), 88.

8. Y. Mei, W. Zhao and J. Yang, Influence maximization on twitter: a mechanism for effective marketing campaign, *IEEE ICC*, (2017), 1–6.

9. E. F. Can, H. Oktay and R. Manmatha, Predicting retweet count using visual cues, *2013 CIKM*, ACM, (2013), 1481–1484.

10. F. Xiong, Y. Liu, Z. Zhang, et al., An information diffusion model based on retweeting mechanism for online social media, *Phys. Lett. A*, **376** (2012), 2103–2108.

11. D. J. Daley and D. G. Kendall, Epidemics and rumours, *Nature*, **4963** (1964), 1118.

12. J. Huang and Q. Su, A rumor spreading model based on user browsing behavior analysis in microblog, *10th ICSSSM*, IEEE, (2013), 170–173.

13. J. Li, K. Niu, Z. He, et al., Analysis of rumor spreading in communities based on modified SIR model in microblog, *18th ICAIMSA*, Springer, Cham, **8722** (2014), 69–79.

14. Q. Su, J. Huang and X. Zhao, An information propagation model considering incomplete reading behavior in microblog, *Physica A*, **419** (2015), 55–63.

15. J. Borge-Holthoefer, S. Meloni, B. Goncalves, et al., Emergence of influential spreaders in modified rumor models, *J. Stat. Phys.*, **151** (2013), 383–393.

16. B. Wang, J. Zhang, H. Guo, et al., Model study of information dissemination in microblog community networks, *Discrete Dyn. Nat. Soc.*, **2016** (2016), 8393016.

17. D. Li, X. Chen, Y. Zhan, et al., Propagation regularity of hot topics in Sina weibo based on SIR model-a simulation research, *2013 IEEE Conference Commun. Appl.*, IEEE, (2015), 310–315.

18. Y. Liu, B. Wang, B. Wu, et al., Characterizing super-spreading in microblog: an epidemic-based information propagation model, *Physica A*, **463** (2016), 202–218.

19. Y. Zhang and C. Tang, Information propagation model based on the dynamics of complex networks in mircoblogging, *J. Comput. Informat. Syst.*, **10** (2014), 443–451.

20. H. Wang, Y. Li, Z. Feng, et al., Retweeting analysis and prediction in microblogs: an epidemic inspired approach, *China Commun.*, **10** (2013), 13–24.

21. M. Tanaka, Y. Sakumoto, M. Aida, et al., Study on the growth and decline of SNSs by using the infectious recovery SIR model, *2015 10th APSITT*, IEEE, (2015), 1–3.

22. D. Liu, Y.Yin and M. Song, Simulation analysis of Weibo information diffusion rule based on SIR model, *JBUPT*, **16** (2014), 28–33.

23. Y. Xiao, Y. Zhou, and S. Tan, Biological mathematics theory, *in: Xi'an Jiaotong University Press*, (2012).

24. X. Wang, J. Wu and Y. Yang, Richards model revisited: validation by and application to infection dynamics, *J. Theor. Biol.*, **313** (2012), 12–19.

25. H. W. Hethcote, The mathematics of infectious diseases, *SIAM Rev.*, **42** (2000), 599–653.

26. V. Karyotis and A. Khouzani, Malware diffusion models for modern complex networks: theory and applications, *Morgan Kaufmann*, (2015).

27. C. Nowzari, V. M. Preciado and G. J. Pappas, Analysis and control of epidemics: a survey of spreading processes on complex networks, *IEEE Contr. Syst. Mag.*, **36** (2016), 26–46.

28. M. Freeman, J. McVittie, I. Sivak, et al., Viral information propagation in the digg online social network, *Physica A*, **415** (2014), 87–94.

29. F. Wang, H. Wang and K. Xu, Diffusive logistic model towards predicting information diffusion in online social networks, *2012 32nd ICDCS Workshops*, IEEE, (2013), 133–139.

30. F. Wang, H. Wang, K. Xu, et al., Characterizing information diffusion in online social networks with linear diffusive model, *2013 IEEE 33rd ICDCS*, IEEE, (2013), 307–316.

AIMS Press