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.
Characterizing super-spreading in microblog: An epidemic-based information propagation model

Yu Liu, Bai Wang, Bin Wu, Suiming Shang, Yunlei Zhang, Chuan Shi

Abstract

As the microblogging services are becoming more prosperous in everyday life for users on Online Social Networks (OSNs), it is more favorable for hot topics and breaking news to gain more attraction very soon than ever before, which are so-called “super-spreading events”. In the information diffusion process of these super-spreading events, messages are passed on from one user to another and numerous individuals are influenced by a relatively small portion of users, a.k.a. super-spreaders. Acquiring an awareness of super-spreading phenomena and an understanding of patterns of wide-ranged information propagations benefits several social media data mining tasks, such as hot topic detection, predictions of information propagation, harmful information monitoring and intervention. Taking into account that super-spreading in both information diffusion and spread of a contagious disease are analogous, in this study, we build a parameterized model, the SAIR model, based on well-known epidemic models to characterize super-spreading phenomenon in tweet information propagation accompanied with super-spreaders. For the purpose of modeling information diffusion, empirical observations on a real-world Weibo dataset are statistically carried out. Both the steady-state analysis on the equilibrium and the validation on real-world Weibo dataset of the proposed model are conducted. The case study that validates the proposed model shows that the SAIR model is much more promising than the conventional SIR model in characterizing a super-spreading event of information propagation. In addition, numerical simulations are carried out and discussed to discover how sensitively the parameters affect the information propagation process.

© 2016 Elsevier B.V. All rights reserved.
1. Introduction

Weibo\(^1\), launched by Sina in August 2009, is a Twitter-like microblogging service. On this online social network (OSN), users can obtain breaking news and other trending information, dig their interests and activities, share personal opinions and argue with other users, update statuses and establish relationships, etc. During the following years since the foundation of Weibo, there has been an explosive growth of microblog users and the contents (tweets and retweets) generated by users on this microblog platform. Tweets are posted and retweeted by Weibo users, forming cascades [1]. In the way that one microblogger posts a tweet and others repost it, information spreads in a considerable extent on online social networks.

Despite a significant amount of work on models, explanations, and patterns of information propagation on online social media, the dynamic underlying super-spreading events on OSNs remains uncertain. Towards developing a better understanding of super-spreading on OSN, in this paper, we aim to build a mathematical model to discover how to describe and characterize super-spreading of information propagation on OSNs.

According to studies on information propagation on online social networks [2–4], firsthand news reports, bizarre anecdotes as well as rumors awaiting clearance mostly disseminate like super-spreading epidemic events. In the circumstance of follower–followeree relationship between microblog users, one user transmits a message to his followers simply by reposting this message, and then the information is distributed instantaneously point-to-surface. After the awareness of some non-ordinary stories, various behaviors of participants' on OSNs lead to many kinds of information diffusion patterns: a small group of users remain silent and just being onlookers; the majority will pass the news on, in the same way that they are influenced by each other and this behavior is mocked, recognized as Herd Behavior [5]; meanwhile, only a very few individuals will transmit the news and influence a very vast coverage of crowds of users than normal users would do, and this minority should be referred to as super-spreaders, and this kind of information propagation should be viewed as a super-spreading event. In most cases, super-spreaders are generally opinion leaders, or those who have quantities of followers on microblog platforms, though sometimes not vice versa.

The diffusion of information on OSNs bears some similarity to an infectious disease spread [6,7]. Although the probability that one individual user would retweet a piece of message relates to his interests and intention, crowds of users would not resist telling others about uncommon stories from a macroscopic perspective. This pattern corresponds to an individual's infection caused by a contagion, who contacted by an infectious person will get infected, even though some are immune to such a contagion. The emerging super-spreaders and the consequent super-spreading phenomenon in a social media scheme are also both analogous to the respective connotations in the field of epidemiology. In epidemiology, the term super-spreader [8,9] refers to a host that infects disproportionately more secondary contacts than others who have also been infected with the same virus or bacteria. Correspondingly in information propagation, a super-spreading story will spread in a notably large range on OSNs through a minor portion of users. Cases in disease spreading mostly conform the 20/80 rule [9], where approximately 20% of infected individuals are responsible for 80% of disease transmissions. Comparably, situations in information diffusion commonly comply the rule of 1% [10] in Internet culture, in which 1% of users contribute to the distribution mostly, while the other 99% spread little.

Online social network, especially microblog, is a hot research focus. Model, analysis and prediction of information propagation on microblog occupies an important part. Till now, there have been some studies on information propagation based on epidemic models.

Yang et al. [11] modeled information diffusion in implicit networks through a simple linear model, which can be utilized to quantify and predict the influence of users, and inspired us. But their model does not fit quantitative analysis of characterizing super-spreading phenomena. Lerman et al. [6] are the first to characterize the spread of news on OSNs by using epidemic models. Then, Abdullah et al. [7] formulated a clear epidemic-based model for news spreading on Twitter, following whom, there have been several studies based on epidemic models. The simple but operational and adoptable epidemic method of modeling is exploited to track trending topics, rumors and generic information on OSNs. These information propagation models are modified based on conventional epidemic models, such as the SIR model, the SEIR model, etc. Based on the SIR model, Li et al. [12] considered various kinds of user behaviors on Weibo and divide users into four groups with different transition rates between groups. Meng et al. [13] added a new compartment into the conventional SIR model in order to catch the phenomenon of rumor spreading in real life. Xia et al. [14] introduced a new class representing authorities that can diffuse authoritative information to prevent rumors from spreading. Zhao et al. [15,16] modified the SIR model to integrate forgetting and remembering mechanisms with variable forgetting rate to catch rumor dissemination characteristics. Zhou et al. [4] studied the characteristics of “information bomb” by leveraging a modified SIR model, which include influence breadth and strength of dynamic of information diffusion. Work by Xu et al. [17] and Zhao et al. [18] took into account complex network features to model information dissemination in mobile social networks and rumor spreading in the new media age, respectively. Zhang et al. [19] introduced a latent group of users who would be infected by an infectious individual, become latent and later then infectious. Although none of them addressed the super-spreading phenomena that benefits the information diffusion cascades on OSNs, previous work has shed light on the modeling of super-spreading.

In the field of disease spreading analysis, super-spreading events have been of great interest among researchers since the global outbreaks of severe acute respiratory syndrome (SARS) in 2003 [20]. Moreover, many infectious diseases were

\(^1\) http://weibo.com/.
reported as super-spreading events [21], i.e., observed infections unveiled evident heterogeneities in pathogen transmission, with some individuals exhibiting a higher ability to infect others. An epidemic compartmental model, SIPR, was proposed to model the super-spreading events during SARS [22]. By splitting the infected people into two classes with different “removal” rates, a normal class \( I \) for normal infectious individuals and a new class \( P \) that stands for super-spreaders, the SIPR model can capture the effect of super-spreaders during the SARS outbreaks in Hong Kong. But due to its focus on particular infectious disease spread, this epidemic model cannot be easily modified to model information propagation. After all, we will build a new model based on the conventional SIR model to accomplish this task.

In this study, through the empirical statistical results of Weibo posts data and a primitive investigation of the ingredients of users participated in diffusions of information on Weibo, we figure out a special class of active spreaders, who are referred to as super-spreaders. Intuitively, this special class thus is added into the well-known deterministic epidemic model, the SIR model, aiming to capture and characterize the effect of super-spreaders and consequential super-spreading events that emerged on Weibo. The adopted deterministic model is simple enough to be operationally beneficial, which can be represented by a sequence of mean-field equations. By leveraging this modified model, primitive trial experiments on the effectiveness of characterizing super-spreading phenomenon on real-world Weibo dataset are conducted.

The remaining part of this paper is organized as follows. In Section 2, we introduce the Weibo dataset and show the empirical statistics of the data, which leads us to the new model. We briefly review some conventional epidemic models in Section 3, then present the detailed depiction of our proposed SAIR model based on the popular SIR model and make some analyses. In Section 4, we validate the proposed model through a case study. In addition, a series of numerical simulations are carried out and the results are discussed in this section. We conclude our work in Section 5 and provide some directions for future study.

2. Empirical observations

In order to discover the patterns of super-spreading of information diffusion, three main criteria are considered when selecting dataset:

- Scale: the chosen dataset needs to be large enough to be statistically significant.
- Completeness: all retweeting actions should be observed and recorded.
- Effectiveness: every retweet can be tracked through a path originated from source tweet so that each retweeter can be tracked, too.

To comply the scale and the completeness criteria, we use a 4.7 million Weibo posts (6190 tweets and their retweets) dataset through the year of 2012 provided in Ref. [23]. All 6190 tweets were posted by 4999 users, including 2111 ordinary users and 2888 verified users. The retweeted counts of each tweet range from 100 to nearly 400,000. A verified user means that the user’s status is administratively verified by Weibo, i.e., their self-reported identifications are true. For every post, either tweet or retweet in this dataset, information includes unique ID, posting time, posting content, poster ID, poster verified status, etc.

For the effectiveness criterion, we will see if we can infer the diffusion network of each tweet from the data. It is worth noting that the components and text content of a retweet on Weibo differs from that on Twitter. Unlike Twitter, a retweeter can add some opinions when he reposts the retweet on Weibo, and he will usually keep any text content that already added by previous retweeters. Trails implying the message propagation path from previous retweeters will be generated automatically by Weibo system whenever a post is retweeted, in the form of “@userC: Repost // @userB: {content_added_by_userB} // @userA: {content_added_by_userA}”. Through this structural form, we know that a tweet posted by an initial user has been reposted by userA, then userB reposts the retweet of userA, and finally userC is the last one to retweet. In this way, we can state that the initial user influenced userA, userA influenced userB, and finally userB influenced userC. Thus, the networks of information diffusion can be extracted from the dataset via all trails like this. In a diffusion network, nodes are users who repost (except the author of the initial tweet), and every directed edge indicates which user reposts whose retweet.

**Fig. 1(a)** demonstrates an example of the diffusion network of a tweet information propagation. This tweet\(^2\) leads to 1003 retweets influencing 945 users. From this figure, it clearly demonstrates that most of the nodes have no child or at most only one child except the initial node (the author of the initial tweet). However, several nodes emit more edges than the others, and we call them super-spreaders. As for the initial node, the author user of the original tweet, he influences a vast range of users, obviously.

**Fig. 1(b)** shows the distribution of out-degrees of nodes (the initial user excluded) in the example diffusion network shown in **Fig. 1(a)**. In the diffusion network, nodes are users who repost the tweet, and these nodes have different out-degrees, i.e., the number of users influenced by each user is varied. According to records of this tweet, there are 42 nodes (users), each of which emits greater than or equal to 5 edges, totally influenced 253 nodes; and 18 nodes, every of which emits more than 10 edges, influenced 168 nodes in this diffusion network.

\(^2\) This tweet was posted by a verified user on Dec. 24, 2012, which is a letter of apology for the poster’s plagiarism.
(a) The diffusion network of an example tweet, where nodes represent influenced users while edges represent retweet actions. The larger red nodes depict users who have influenced at least 5 other users.

(b) The distribution of out-degrees (i.e., numbers of users influenced by each user) of nodes in the diffusion network shown in Fig. 1(a), except for the original tweet's author.

Fig. 1. An example of diffusion network of a tweet information. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The observed data points of this tweet approximate a power-law distribution as shown in Fig. 1(b). For the entire dataset, the distribution of user influenced counts is depicted in Fig. 2, in which the distribution shows a long-tail shape that is common in real-world social networks. This distribution suggests that the majority of users influence relatively small portion of other users, and relatively large amounts of users are influenced by a minority of users. In more detailed investigations, we find that the majority of highly retweeted posts share the same trait as demonstrated in the example.

The empirical statistical result of all posts in the dataset is shown in Table 1. Here, we define “influence range” differently from the corresponding definition in influence maximization problem: in the result, the influence range means the out-degree of one node in the diffusion network. Symbols used in the result table are defined as follows:

- $C_{\text{infl}}$ — The number of users one user influenced, excluding the initial author. In a diffusion network, $C_{\text{infl}}$ means the out-degree of a node. User category $C_{\text{infl}} = n$ means each of the users has influenced $n$ other users.
- $C_{\text{infl ave}}$ — The average number of users influenced by a particular category of users.
- $R_{u}$ — The percentage ratio of users in a particular category.
- $R_{\text{infl}}$ — The percentage ratio of users (excluding those influenced by any original author users) influenced by a particular category of users in a diffusion of tweet information.
### Table 1

Empirical statistical result of involving users' influence.

| User category | $C_{\text{infl}} = 0$ | $C_{\text{infl}} = 1$ | $C_{\text{infl}} = 2$ | $C_{\text{infl}} = 3$ | $C_{\text{infl}} = 4$ | $C_{\text{infl}} \geq 5$ |
|---------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| $C_{\text{infl}}$ | 0                      | 1                      | 2                      | 3                      | 4                      | $>73$                  |
| $R_u$         | 92.62%                 | 4.31%                  | 0.94%                  | 0.32%                  | 0.15%                  | 0.41%                  |
| $R_{u\text{infl}}$ | 0                      | 36.14%                 | 11.58%                 | 5.03%                  | 2.81%                  | 44.44%                 |

All fields are in average for every tweet in dataset, so the sum of each row may not be equal to one.

---

According to this empirical statistical result, it is obvious that most users are not so influential, and in the meantime, a very few individuals influence a vast range of users. Through diffusions of 6190 tweets with their retweets in the dataset, we observe that $92.62\% \pm 6.23\%$ of users in the dataset are "pure receivers", i.e., they have influenced no one. The ratio of users who have influenced only one follower (nodes with out-degree equals to 1, $C_{\text{infl}} = 1$) is respectable ($R_u = 4.31\%$) and the ratio of their influenced followers approaches approximately $2/5$ ($R_{u\text{infl}} = 36.14\%$). The users who have induced at least 5 other users to post a retweet is quite a few, the ratio of which is under $1\%$ ($R_u = 0.41\%$). However, these users induce more others to post a retweet and the influenced users counts 44.44%. The average number of users influenced by them is greater than 73.

To look deeper into the entire data, we demonstrate the statistical result of user distribution with different $C_{\text{infl}}$'s (influence range) in Fig. 3. The scatter plots (a)–(f) approximately imply the ratio of users of each $C_{\text{infl}}$ (i.e., $R_u$), depicted by the slope of the thin red line in each sub-figure (through a different means that a linear fit is performed on each category, yielding a slightly different value from $R_u$). Meanwhile, the corresponding distributions of users in each category are illustrated in sub-figures (g)–(l), respectively.

Furthermore, Fig. 4 illustrates the influence range of users with distinct $C_{\text{infl}}$. In every sub-figure, each point represents a diffusion. It is especially evident that users of $C_{\text{infl}} = 0$ have always influenced no one, as shown in Fig. 4(a). Other than this, there always exists situations where the information propagation is achieved by only one type of users, as denoted by the points with value 1 in y-axis in other sub-figures. As $C_{\text{infl}}$ grows from 1 to 4, the influence range of corresponding user categories become less. Conversely, the influence ranges are much broader when $C_{\text{infl}}$ exceeds 5, as depicted in Fig. 4(f).

In studies of diseases, super-spreaders have been characterized as infected individuals who excrete a higher than normal number of pathogens during the time they are infectious. This causes their contacts to be exposed to higher viral/bacterial loads than would be seen in the contacts of non-super-spreaders within the same duration of exposure [24]. As a result of aforementioned observations in this Weibo dataset, it is found that in information diffusion, there also exists super-spreading event and super-spreaders (users with $C_{\text{infl}} \geq 5$). To discover the role behavior of super-spreaders, the traditional epidemic-based models may be modified to address the super-spreading phenomena.
3.3. Information propagation model based on SIR

As introduced in Section 1, the information propagation dynamic of a tweet message is analogous to a contagious disease spread: an initiator posts an original tweet; numerous users (followers of this initiator and/or followers of followers) who are interested in this piece of information, will be influenced (infected) to post a retweet; these influenced users will be “infectious” in turn to influence other users (to spread this message, “contagion”, to others); after a period of time, they are disinterested in this message and halt being infectious (recovered from such a disease). Thus, in line with the characteristics of each compartment, social network users involved in an information propagation process can be recognized as either susceptible users, spreaders, or recovered users.

In the circumstance of information propagation on online social networks, a recent study [30] considered the diffusions on complex networks in which information carriers have varied topological features, e.g., node degree. A number of SIR-based information propagation models [31,18] that take node degree into account make the spreading process more realistic and apparent. According to the transition pattern detailed in Section 3.1, the compartment transition of the SIR model is
Fig. 5. The transfer diagram for the SIR model.

demonstrated in Fig. 5, and the SIR model of information propagation on complex network can be represented by mean-field equations as follows,

$$\frac{dS(t)}{dt} = -\frac{k\beta S(t)I(t)}{N(t)},$$  

$$\frac{dI(t)}{dt} = \frac{k\beta S(t)I(t)}{N(t)} - \gamma I(t),$$  

$$\frac{dR(t)}{dt} = \gamma I(t),$$  

$$N(t) = S(t) + I(t) + R(t) \equiv K,$$

where $S(t)$, $I(t)$ and $R(t)$ represent the numbers of susceptible users, spreaders and recovered users at time $t$, respectively. Moreover, all initial values are nonnegative, and the number of total population $N(t)$ is a constant. All parameters are nonnegative and the meanings of them are as follows:

$k$ denotes the average out-degree of nodes in the I compartment, i.e., the number of followers or latent susceptible users who read the post of an infectious user, and then they will get a chance to post a retweet;

$\beta$ is the infection rate per-user per time; and

$\gamma$ is the removal (recovered) rate.

3.3. The proposed SAIR model

Phenomena of super-spreading exist in both epidemic spreads and information diffusions. However, super-spreaders act similarly but with a little bit variation under the two circumstances. A super-spreader [8,9], once a susceptible one, but infected by an infectious user, will infect more secondary contacts from the susceptible class than other infectious individuals. On the contrary, the majority of individuals infect relatively fewer secondary contacts during epidemic spreading. Similarly, a super-spreader on OSNs would influence more users than others, causing users to be influential to influence others. Correspondingly, most users are less influential. One simple heuristic is to develop a new population class portraying the dissimilar behavior of super-spreaders.

Despite of the similarity, the characteristic of super-spreader differs a little in these two different scenarios. In a real epidemic, a super-spreader will continue being severely infectious before he is quarantined or gets cured, as described in the SIRP model [22]. On OSNs, however, a super-spreader may halt being highly influential due to the talking topic being aging (i.e., topic not being a recency) or this super-spreader being tired of mentioning such a story, and then he turns into a normal infectious user who still have chance to influence others with a common infection rate.

As previously introduced, the SIR model is used to model tweet messages diffusion dynamic and depicts the information propagation well, which is the basis we progress on. Furthermore, with the effect of super-spreaders, it is essential to add a new compartment, Active spreaders, to address the functionality of super-spreaders for the different behaviors between super-spreaders and ordinary users (aforementioned facts observed from real-world Weibo data in Section 2). In addition, the establishment of the new compartment diversifies the corresponding transition rates between classes. The mechanisms of information propagation in microblog considering super-spreading can be modeled as shown in Fig. 6.

Under the real-world Weibo scenario, all users involved in an information propagation process (the whole population) are divided into four classes: susceptible users ($S$), super-spreaders ($A$), normal spreaders ($I$) and recovered users ($R$). The information propagation rules can be summarized as follows.

1. An initiator posts the original tweet that starts the process of information propagation. If the initiator is a verified user, then he/she belongs to the A compartment, otherwise the I compartment. The initiator is infectious and has the ability to influence other users in the S compartment to retweet the message.

2. The followers of users in either the A compartment or the I compartment belong to the S compartment, and they are susceptible to be infected to repost after they have read the original tweet or its retweet.

3. If a susceptible user is a verified user, he/she will become a super-spreaders with probability $\alpha$, namely super-spreaders emerging rate, when the user has the message.

4. If a susceptible user is a normal user, he/she will become a normal spreader with probability $\beta_1$, after the user reads the message posted by a super-spreaders; or with probability $\beta_2$, when the user gets the information from a normal spreader. $\beta_1$ and $\beta_2$ are the infection rates of a super-spreaders and a normal spreader, respectively.
5. After a specific period of time $1/\varepsilon$, a super-spreader will become less infectious due to topic not being a recency, i.e., become a member of the I compartment. $\varepsilon$ is the retrograde rate.
6. After a specific period of time $1/\gamma$, a normal spreader will lose interest in the message and get recovered. $\gamma$ is the recovery rate.

Let $N$ be the total population of users on the diffusion network of a tweet message, and we assume that the total population retains unchangeable. $S$ represents the number of users that are susceptible to a tweet and will be infected by the tweet initiator or other retweeters, i.e., they are followers of users of compartments A and I. $A$ refers to the number of active spreaders, namely super-spreaders, who are highly infectious, and $I$ refers to the number of normal infectious users, namely normal spreaders. $R$ represents the number of recovered users who are not interested in a certain piece of information any more. Based on the above depiction of the dynamics, the SAIR model can be described by the following mean-field equations,

$$\frac{dS(t)}{dt} = \frac{\alpha [k_1 A(t)S(t) + k_2 I(t)S(t)]}{N(t)} - \frac{\beta_1 k_1 A(t)S(t)}{N(t)} - \frac{\beta_2 k_2 I(t)S(t)}{N(t)},$$

$$\frac{dA(t)}{dt} = \frac{\alpha [k_1 A(t)S(t) + k_2 I(t)S(t)]}{N(t)} - \varepsilon A(t),$$

$$\frac{dI(t)}{dt} = \frac{\beta_1 k_1 A(t)S(t)}{N(t)} + \frac{\beta_2 k_2 I(t)S(t)}{N(t)} + \varepsilon A(t) - \gamma I(t),$$

$$\frac{dR(t)}{dt} = \gamma I(t),$$

$$N(t) = S(t) + A(t) + I(t) + R(t) \equiv K,$$

where $S(t)$, $A(t)$, $I(t)$ and $R(t)$ represent the number of susceptible users, super-spreaders, normal spreaders and recovered users at time $t$, respectively. Similar to the SIR model, their initial values are essentially nonnegative to model a diffusion event. The number of total population $N(t)$ is a constant. $k_1$ and $k_2$ are average out-degrees of nodes of users in the A compartment and the I compartment, respectively, i.e., the numbers of followers or latent susceptible users who read the post of a super-spreader and a normal spreader, respectively.

According to the modeling of super-spreaders, parameters $\alpha$, $\beta_1$, and $\varepsilon$ are introduced into the SAIR model. If we set these three parameters to zero, the SAIR model will degenerate into the SIR model with the consideration of information diffusion on complex network which is described in Eqs. (1)–(3). Furthermore, we leverage complex network features to distinguish only two groups of users to infer their distinct characteristics through $k_1$ and $k_2$, rather than fine-grained classes used in modeling dynamic in complex socio-technical systems [32].

It appears that a super-spreader has numerous followers, and thus can influence a large quantity of users in most cases. Followers of a super-spreader or an influential user, are more willing to transmit a message posted by this influential user than those of an ordinary user will do. As stated above, verified users usually act as super-spreaders on Sina Weibo. Therefore, we consider verified users super-spreaders in the model validations and simulations section.

3.4. Steady-state analysis

Since the total population $N(t)$ is a constant, dividing Eqs. (5)–(8) by this constant $N(t)$ yields

$$\frac{ds}{dt} = -\alpha [k_1 s + k_2 i s] - \beta_1 k_1 a s - \beta_2 k_2 i s,$$

$$\frac{da}{dt} = \alpha [k_1 a s + k_2 i s] - \varepsilon a,$$
\[
\frac{di}{dt} = \beta_1 k_1 as + \beta_2 k_2 is + \varepsilon a - \gamma i, \tag{12}
\]
\[
\frac{dr}{dt} = \gamma i, \tag{13}
\]

with \(s(t) + a(t) + i(t) + r(t) = 1\), where \(s(t), a(t), i(t)\) and \(r(t)\) are the fractions of corresponding compartments. From now on, we use the corresponding fraction of each compartment. From Eqs. (11) and (12), we see that \(\lim_{t \to 0} a(t) = 0\) and \(\lim_{t \to 0} i(t) = 0\). Since the total population has a fixed number, \(r\) increases proportionally to \(i\).

Let Eqs. (11)–(13) be 0, solutions of equilibrium will be obtained. Thus it follows that the equilibria are \((s, a, i, r)_e = (1, 0, 0, 0)\) that is a non-diffusible equilibrium (disease-free equilibrium in epidemiology) we will not put analysis on, and \((s, a, i, r)_e = (s_e, 0, 0, r_e)\) for any \(s_e \geq 0, r_e > 0\) with \(s_e + r_e = 1\) after a super-spreading event. One equilibrium shows that a certain number of individuals remain susceptible \((s_e > 0)\), which greatly depends on initial conditions and adjustments of parameters [33]. Although this scenario does exist in the modeling of a disease spread, the number of susceptible individuals in a real event is difficult to be precisely quantified. Furthermore, let us consider a tweet message diffusion on OSNs, the entire population involved in a super-spreading event is a very large quantity of users. But relatively very few users remain unaware of that information after a vast extent of diffusion (remaining users in the \(S\) compartment), therefore the fraction of this class can approximately be 0. Furthermore, analyzing zero solution is sufficient, since problems having non-zero solutions can be converted to a modified problem where zero solution is used to analyze the corresponding stability problem [34]. Herein, we assume the equilibrium after a super-spreading event is \((s, a, i, r)_e \approx (0, 0, 0, 1)\).

Notice that \(r(t)\) in the differential equation (13) is uncoupled from the other three equations. With the equilibrium and the first three equations above, we can compute the Jacobian matrix with equilibrium \((s, a, i, r)_e \approx (0, 0, 0, 1)\):

\[
J(s, a, i)|_{(0,0,0)} = \begin{pmatrix}
-a\kappa_1 a - a\kappa_2 i & -a\kappa_1 s - \beta_1 k_1 s & -a\kappa_2 s - \beta_2 k_2 s \\
-\beta_1 k_1 a - \beta a_2 k_2 i & a\kappa_1 i + a\kappa_2 i & a\kappa_2 s \\
\alpha k_1 i + \alpha k_2 i & a\kappa_1 s - \varepsilon & a\kappa_2 s \\
\beta_1 k_1 i + \beta_2 k_2 i & \beta_1 k_1 s + \varepsilon & \beta_2 k_2 s - \gamma
\end{pmatrix}_{(0,0,0)}
\]

It follows that the characteristic equation is

\[
\lambda(\lambda + \varepsilon)(\lambda + \gamma) = 0, \tag{15}
\]

and the eigenvalues are \(\lambda_1 = 0, \lambda_2 = -\varepsilon\) and \(\lambda_3 = -\gamma\). Eigenvalues \(\lambda_2\) and \(\lambda_3\) are negative real numbers, and eigenvalue \(\lambda_1\) is 0 which has algebraic multiplicity 1 and corresponding geometric multiplicity 1. Thus, the criterion for a stable linear system are satisfied [35] and the proposed SAIR model is neutrally stable when it reach the equilibrium after a super-spreading event occurred, i.e., \((s, a, i, r)_e = (0, 0, 0, 1)\).

4. Model validation, simulation and discussion

4.1. Model validation

As a case study, we select a tweet and all its retweets from the dataset described in Section 2 to validate the proposed SAIR model. This tweet3 influenced a huge amount of users and 53,220 retweets were produced. The empirical cumulative distribution function of involved retweeter numbers is shown in Fig. 7(a). As you can see in the figure, the information diffusion progress mainly took place in the first 24 h (see Fig. 7(b) for the retweeter numbers in the first 24 h), which caused 48,174 distinct users to participate in and generate 49,637 retweets. After the extensive propagating period, the rest diffusion of 3583 retweets was contributed by another 3500 users in the remaining hours, accounting only for less than 7% of the entire distribution of all retweets. Thus, the model validation is based on the data of the first 24 h.

Given the posting time of each retweet following this tweet, we can obtain the number of retweets in every 60 min (here, we set the time interval to 60 min for simplicity) in the first 24 h. The count of normal users who have retweeted corresponds to the number of normal infectious spreaders in the \(I\) compartment of the SAIR model, and the count of verified users corresponds to the number of super-spreaders in the \(A\) compartment. To establish the connection between the real-world scenario and the mathematic model introduced in Section 3.3, we roughly recognize normal retweeters as elements of the \(I\) compartment, verified users as components of the \(A\) compartment, assuming that verified users act as super-spreaders.

---

3 This microblog was released by a common user on the early morning of Jan. 13, 2012, which was attached three photos showing scenes of numerous buyers who crowded around the Apple Retail Store at Sanlitun in Beijing when iPhone 4s came to China.
In order to find out the most suitable parameters in the SAIR model to characterize the real-world information propagation of this tweet, and compare the performance of it with the conventional SIR model reviewed in Section 3.2, an objective function is proposed as follows,

$$J = \| X - \hat{X} \|_F^2, \quad (16)$$

where $\| \cdot \|_F$ denotes the Frobenius norm, and $\hat{X}$ is the fitted result of observed data $X$. For the SAIR and SIR models, the respective objective functions are materialized as follows,

$$J_{\text{SAIR}} = \sum_{t=1}^{n} | I(t) - \hat{I}(t) |^2 + \sum_{t=1}^{n} | A(t) - \hat{A}(t) |^2, \quad (17)$$

$$J_{\text{SIR}} = \sum_{t=1}^{n} | I(t) - \hat{I}(t) |^2, \quad (18)$$

where $\hat{I}(t)$ and $\hat{A}(t)$ are fitted values of $I(t)$ and $A(t)$ (observed values) at time $t$, respectively. $n$ is the number of observations, i.e., the number of time epochs. The best fitted parameters can be obtained by minimizing the objective functions presented in Eqs. (17) and (18).

In order to determine the best fitted parameters in the SAIR model and the conventional SIR model, we perform pattern search, a multi-parameter nonlinear optimizing/fitting technique, to minimize the optimization functions in MATLAB with the time-divided data and the parameterized models. Required by the optimization tool, the constraints of parameters are given in Table 2. Parameters of degrees ($k_1$, $k_2$ and $k$) are referenced from a Weibo study [36]. The ranges of super-spreader emerging rate ($\alpha$) and various infection rates ($\beta_1$, $\beta_2$, and $\beta$) conform to the picking principle in epidemiology. However, the super-spreader retrograde rate $\varepsilon$ and the recovery rate ($\gamma$) can be greater than 1, since the time interval is much smaller in information diffusions than that in epidemics.

The metric root-mean-square error (RMSE)$^3$ is used to evaluate the performance of the fitted results by both models. Since only SAIR models the A compartment, which SIR does not, RMSE is measured between the fitted values and the observed

---

3 Although RMSE is oftentimes used as an evaluation metric for predictions, it is also fair and efficient to measure how close one model’s computed values are to the observed data points.
data points of the I compartment that both models can estimate. Thus, the metric RMSE is defined as follows,

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} |I(t) - \hat{I}(t)|^2}.$$  \hspace{1cm} (19)

As indicated by the definition, a smaller RMSE implies a better fitting result.

The fitted results of both the proposed SAIR model and the conventional SIR model are demonstrated in Figs. 8 and 9, respectively. For the SAIR model, the lowest error (RMSE$_{SAIR} = 744.6416$) can be reached with parameter set: $k_1 = 473.7$, $k_2 = 49.9$, $\alpha = 0.0029$, $\beta_1 = 0.1817$, $\beta_2 = 0.0109$, $\epsilon = 2.5609$, $\gamma = 1.1311$. While for the SIR model, the lowest error (RMSE$_{SIR} = 792.5963$) is obtained with parameter set: $\tilde{k} = 80.2$, $\beta = 0.0470$, $\gamma = 1.2180$. By comparing the fitted results, one can easily conclude that the new proposed model that considers the effect of super-spreaders is much better than the conventional one on characterizing information propagation process with super-spreading phenomenon in microblog. For the observed data points, it is observed in the figure that the peak of the A compartment and the peak of the I compartment come at the same time (at $time = 4$). However, for the fitted result, the peaks of the I compartment and the A compartment come at $time = 5$ and $time = 4$, respectively, which implies that super-spreaders in the A compartment lead the trend of information diffusion.

4.2. Numerical simulation and discussion

Given the mean-field equations of the SAIR model, numerical simulations are carried out using the Runge–Kutta method in MATLAB to solve the ordinary differential equations and analyses of the effects on the information propagation process by the new compartment A are conducted. Based on the model description in Section 3.3 and parameters obtained in the previous case study, we assume $N = 1000$, and set parameters that $\tilde{k}_1 = 473.7$, $\tilde{k}_2 = 49.9$, $\alpha = 0.0029$, $\beta_1 = 0.1817$,
Density $t$

Fig. 10. Density of each compartment in the SAIR model over time, with parameters setting: $k_1 = 473.7$, $k_2 = 49.9$, $\alpha = 0.0029$, $\beta_1 = 0.1817$, $\beta_2 = 0.0109$, $\gamma = 1.1311$, $\varepsilon = 2.5609$.

$\beta_2 = 0.0109$, $\varepsilon = 2.5609$, $\gamma = 1.1311$ to conduct numerical simulations. In the initial stage of the information propagation process, there is only one normal infected user, i.e., $S(0) = 999$, $A(0) = 0$, $I(0) = 1$, and $R(0) = 0$, and thus the corresponding fractions are $s(0) = 0.999$, $a(0) = 0$, $i(0) = 0.001$, and $r(0) = 0$, respectively. Without loss of generality, we use fractions in the following simulations.

Fig. 10 shows the overall trends of all four compartments of the SAIR model. From the simulation, we can find that there is a sharp increase in the number of users in the $I$ compartment in the early stage of the information propagation, which indicates that the SAIR model captures the swift eruption of massive emergence of retweeters (influenced users). As information propagates, the number of influenced individuals reaches a peak and then declines. Finally, the density of compartment $I$ comes down to zero, when this tweet stops spreading on Weibo. During the entire process, it should be noted that the density of the $S$ compartment decreases quickly as the fraction of the $I$ compartment increases. This is because the susceptible users in compartment $S$ quickly transit into either super-spreaders in compartment $A$ or normal infectious spreader in compartment $I$. As for the $A$ compartment, super-spreaders take up only a small fraction of the whole population but lead the information propagation which can be found in the case study in Section 4.1.

The simulation shows a case where a residue exists in the $S$ compartment after an information super-spreading event. A small portion of susceptible users are not infected during the information propagation process. In the parameter setting, the recovery rate $\gamma$ and super-spreader retrograde rate $\varepsilon$ are both greater than 1, which indicates that super-spreaders and normal spreaders get recovered very fast (lose interest to participate in future discussion of current topic very soon). Thus the simulation represents a trending topic that will get faded out soon after a short period of time. Topics on breaking news lacking of sustainable motivation to be continuously discussed are of this type, such as the tweet used in Section 4.1. However, with a smaller $\gamma$ and $\varepsilon$, users’ involving in the discussion and information diffusion will last longer. The following sensitivity analysis will cover this case.

With the introduction of the $A$ compartment, new parameters are brought in, i.e., $\alpha$, $\beta_1$, $\varepsilon$, and we will analyze the sensitivities of them in the following part. In actual simulations, we use finer granularity in each parameter than that shown in the results. For example, we run the sensitivity analysis of $\alpha$ with step size 0.0001, but show only the result with step size 0.001 which already provides adequate information.

4.2.1. Sensitivity analysis of super-spreaders emerging rate $\alpha$  

$\alpha$ is the emerging rate of super-spreaders. This rate controls the amount and the speed how super-spreaders involves in the information propagation. As experimented, the percentage of each class in the SAIR model will not change dramatically during information propagation process if $\alpha > 0.04$, i.e., the curve of each compartment follows a similar pattern. Moreover, it is evidently true that scenarios under $\alpha > 0.1$ can hardly be seen in the real world, for the reason that super-spreaders are always a minority. Thus, we set the range of $\alpha$ from 0.001 to 0.04 with step size 0.001.

Fig. 11 illustrates the sensitivity analysis of the super-spreaders emerging rate $\alpha$:

- $S$—susceptible users. As we can see in Fig. 11(a), with a greater super-spreaders emerging rate, the percentage of compartment $S$ drops much faster than that with a smaller one. This result shows that much more super-spreaders appear so that they will induce susceptible users to transit into either super-spreaders or normal infected individuals.
- $A$—super-spreaders. For the $A$ compartment, apparently seen from Fig. 11(b), a larger super-spreaders emerging rate will lead to a sharper rise and a higher peak of the $A$ compartment, which is in agreement with reality. Also, a larger $\alpha$ leads to early peaking in compartment $A$. 
Fig. 11. Sensitivity analysis of $\alpha$, the super-spreader emerging rate. $\alpha$ ranges from 0.001 to 0.04, step size 0.001, other parameters have the same values as described in Fig. 10.

- **I**—normal spreaders. As for the I compartment, if this rate gets bigger, the peak of the A compartment will get higher, which leads the peak of the I compartment to come much earlier and higher, as shown in Fig. 11(c). Furthermore, one should keep in mind that users of the A compartment will transit into the I compartment eventually.
- **R**—recovered users. Fig. 11(d) shows the simulation result of compartment R with varied $\alpha$, demonstrating that the super-spreader emerging rate does not affect the fading process in information propagation.

It is worth noting that for a very small $\alpha$, a large quantity of residual users exist in the S compartment. The scenario of small $\alpha$ together with large $\varepsilon$ and $\gamma$ combinatorially leads to this particular phenomenon.

4.2.2. Sensitivity analysis of super-spreader infection rate $\beta_1$

Unlike the super-spreader emerging rate $\alpha$ controlling the accumulating amount and speed of super-spreaders, the super-spreader infection rate dominates the spread of information. Fig. 12 shows the percentage change of each compartment with varied $\beta_1$. The range of $\beta_1$ is from 0.01 to 0.4 with step size 0.01.

On Weibo, a much larger infection rate of $\beta_1$ leads to much faster diffusion of tweets among users. Thus, the number of susceptible users of compartment S decreases a bit faster/slower with a larger/smaller $\beta_1$, while users in the I compartment get recovered much faster/slower to compartment R, as shown in Fig. 12(a) and (d), respectively. It differs from the monotonous trends of compartments S and R that varied $\beta_1$ causes diverse changes in compartments A and I:

- **A**—super-spreaders. As shown in Fig. 12(b) and (c), the rate of $\beta_1$ indirectly affects the growing and descending process of the A compartment by the I compartment. The fraction of compartment A is much small, and this is the evidence in support that the super-spreaders are a minority group but very "contagious", i.e., have a higher chance of being succeeded in influencing users than any other ordinary users. If the super-spreader infection rate gets larger, the peak of the A compartment will get higher and come much earlier.
- **I**—normal spreaders. For the I compartment in Fig. 12(c), the rate of $\beta_1$ directly control the growing process of the I compartment. If $\beta_1$ gets larger, the peak of compartment I has the same pattern as that of the A compartment, which means that the bigger $\beta_1$ is, the stronger the influence of super-spreaders are, and the higher chance of being influenced the susceptible users have, therefore the wider range the information diffuses.

Furthermore, we have examined and compared how the two infection rates $\beta_1$ and $\beta_2$ affect the fraction of each compartment during a tweet diffusion. Result shows that $\beta_2$, the normal spreader infection rate, has less impact on the information diffusion than $\beta_1$, the super-spreader infection rate, due to the fact that a super-spreading event is generally governed by the influence caused by super-spreaders.
4.2.3. Sensitivity analysis of super-spreader retrograde rate $\varepsilon$

$\varepsilon$ is the super-spreader retrograde rate that weighs the speed (or time) a super-spreader transits into a normal infectious person. Fig. 13 demonstrates the percentage change of each compartment with varied $\varepsilon$. The range of $\varepsilon$ is from 0.01 to 3 with step size 0.06.

From the diagrams, variations of $\varepsilon$ show a less notable effect on the densities of compartments S, I and R. Only the curve of the A compartment containing super-spreaders will become more sharp after peaking with the growth of rate $\varepsilon$. These characteristics are obviously true according to the differential equations of the SAIR model, i.e., the super-spreader retrograde rate has direct impacts on changes of the number of super-spreaders. However, the effect of varied $\varepsilon$ on compartment I is trivial, as the proportion of compartment A is extremely too small (compared with compartment I) for a variable $\varepsilon$ to induce notable changes on the I compartment.

Other than measuring the speed of state transition, this rate also indicates that the statistically mean time for a super-spreader being highly infectious is $1/\varepsilon$ (time epoch). Here, the time epoch we set in simulations is 1 h that best fits the Weibo data. Thus, the curve style displayed in Fig. 13(b) is in accordance with the reality that a small/big $\varepsilon$ indicates a slow/quick transition speed from compartment A to I, meaning that a super-spreader is more/less interested in disseminating such information.

Remarkably, if $\varepsilon$ is so large that super-spreaders become normal spreaders much fast, the spread of information will be diluted. And the consequence is that a portion of susceptible users remain in the S compartment. The consequence alters at the critical value of $\varepsilon$ at 1. All susceptible users will be influenced if $\varepsilon \leq 1$.

4.2.4. Sensitivity analysis of recovery rate $\gamma$

Although the recovery rate $\gamma$ is not introduced by the new A compartment, i.e., this rate is inherited from the SIR model, it always fairly affects the information propagation. The pattern of information diffusion with different $\gamma$ is similar to the depictions in study [37]. However, the conclusion in this study drawn by Su has some flaws with the underlying truth we observed. Fig. 14 demonstrates the percentage change of each compartment with varied $\gamma$. The range of $\gamma$ is from 0 to 2 with step size 0.04.
As demonstrated in Fig. 14(a) and (b), compartments S and A have little change with a varied $\gamma$. It is consistent with the statement made by Su et al. in Ref. [37], where they argue that the recovery rate does not cause dramatic changes if it
varies, and this rate cannot significantly affect the information propagation process. However, our study draws a different conclusion. As shown in Fig. 14(c), a tuning $\gamma$ significantly affects the curve pattern of compartment A similarly to that of compartment B caused by varied $\epsilon$ (cf. Fig. 13(b)): the larger the rate $\gamma$ is, the lower the peak of the fraction of compartment I is, and the faster it drops after reaching the peak, which implies that if users recover at a high rate (lose interest of a tweet fast), the information diffusion terminates early. For the R compartment shown in Fig. 14(d), the number of it raises sluggishly if this rate is small, but eventually reaches the equilibrated state as long as $\gamma > 0$.

5. Conclusions

In this paper, the patterns of information propagation on microblog services are depicted briefly. The mechanism of super-spreading phenomena in information propagation is discussed, and the notion of super-spreader in information propagation is defined with the help of corresponding definition in epidemiology field. Based on the empirical observation result from real-world Weibo dataset of tweets diffusion, an A compartment characterizing super-spreaders is introduced into the conventional epidemic model, and a parameterized model SAIR is built to characterize super-spreading phenomena of tweet information propagation. A case study of fitting to real world dataset is conducted, and parameter settings are obtained for both the SAIR model and the SIR model. Through comparison between fitted results of both models, the SAIR model is validated to show much more promising to characterize a super-spreading event of information propagation than the conventional one. In addition, numerical simulations are carried out to discover the sensitivity of parameters.

For future work, we plan to model the super-spreading on OSNs as a deterministic compartmental model but analyze it using a stochastic method by exploiting the graph topology of online social networks, which considers the varied effectiveness of nodes (users) with different network attributes. Another potential extension to our work would be considering the temporal variations of parameters. Not only the recency of burst news is time dependent, but the behavior of users is also time concerning. It has been observed that users on Weibo are more active in the daylight time of China than in the nighttime, which would affect the information spreading pattern. Furthermore, large scale real-world data can be fitted with our proposed model in order to fulfill the task of information influence breadth and strength analysis, information propagation intervention and control, etc.

Acknowledgments

We are grateful to the anonymous reviewers for their careful reading and valuable suggestions that greatly improve the quality of this paper. We are also thankful to King-wa Fu for the Weibo data and suggestions on data analysis. This research is supported in part by the National Key Basic Research and Development (973) Program of China (No. 2013CB329606), and the National Natural Science Foundation of China (No. 71231002).

References

[1] M. Gomez Rodriguez, J. Leskovec, B. Schölkopf, Structure and dynamics of information pathways in online media, in: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM’13, ACM, New York, NY, USA, 2013, pp. 23–32. http://dx.doi.org/10.1145/2433396.2433402.
[2] K. Thompson, R. Estrada, D. Daugherty, A. Cintron-Arias, A deterministic approach to the spread of rumors. Cornell University, Department of Biological Statistics and Computational Biology, Technical Report BU-1642-M. URL http://hdl.handle.net/1813/32243.
[3] Y. Bao, C. Yi, Y. Xue, Y. Dong, A new rumor propagation model and control strategy on social networks, in: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM’13, ACM, New York, NY, USA, 2013, pp. 1472–1473. http://dx.doi.org/10.1109/ASONAM.2013.7249259.
[4] X. Zhou, Y. Hu, Y. Wu, X. Xiong, Influence analysis of information erupted on social networks based on sir model, Internat. J. Modern Phys. C 26 (02) (2015) 1550018. http://dx.doi.org/10.1142/S0129183115500187.
[5] Herd behavior. https://en.wikipedia.org/wiki/Herd_behavior (accessed 22.12.15).
[6] K. Lerman, R. Ghosh, Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks. ArXiv e-prints arXiv:1003.2664.
[7] S. Abdullah, X. Wu, An epidemic model for news spreading on twitter, in: 2011 23rd IEEE International Conference on Tools with Artificial Intelligence, ICTAI, 2011, pp. 163–169. http://dx.doi.org/10.1109/ICTAI.2011.33.
[8] J.O. Lloyd-Smith, S.J. Schreiber, P.E. Kopp, W.M. Getz, Superspreading and the effect of individual variation on disease emergence, Nature 438 (7066) (2005) 355–359. http://dx.doi.org/10.1038/nature04153.
[9] A.P. Galvani, R.M. May, Epidemiology: Dimensions of superspreading, Nature 438 (7066) (2005) 293–295. http://dx.doi.org/10.1038/438293a.
[10] E. Hargittai, G. Walejko, The participation divide: Content creation and sharing in the digital age, Inf. Commun. Soc. 11 (2) (2008) 239–256. http://dx.doi.org/10.1080/13691180801946150.
[11] J. Y. Jang, J. Leskovec, Modeling information diffusion in implicit networks, in: 2010 IEEE 10th International Conference on Data Mining, ICDM, 2010, pp. 599–608. http://dx.doi.org/10.1109/ICDM.2010.22.
[12] D. Li, Y. Zhang, X. Chen, L. Gao, Propagation regularity of hot topics in sina weibo based on sir model - a simulation research, in: Computing, Communications and IT Applications Conference, ComComAp 2014 IEEE, 2014, pp. 310–315. http://dx.doi.org/10.1109/ComComAp.2014.7017216.
[13] D. Meng, L. Wan, L. Zhang, A study of rumor spreading with epidemic model based on network topology, in: W.-C. Peng, H. Wang, J. Bailey, Y.S. Tseng, T.B. Ho, Z.-H. Zhou, A.L. Chen (Eds.), Trends and Applications in Knowledge Discovery and Data Mining, in: Lecture Notes in Computer Science, vol. 8443, Springer International Publishing, 2014, pp. 381–392. http://dx.doi.org/10.1007/978-3-319-1186-3_35.
[14] L. Xie, C. Jiang, Y. Song, B. Song, Modeling and analyzing the interaction between network rumors and authoritative information, Entropy 17 (2015) 471–482. http://dx.doi.org/10.3390/e17010471.
[15] L. Zhao, X. Qiu, X. Wang, J. Wang, Rumor spreading model considering forgetting and remembering mechanisms in inhomogeneous networks, Physica A 392 (4) (2013) 987–994. http://dx.doi.org/10.1016/j.physa.2012.10.031.
[16] L. Zhao, W. Xie, H.O. Gao, X. Qiu, X. Wang, S. Zhang, A rumor spreading model with variable forgetting rate, Physica A 392 (23) (2013) 6146–6154. http://dx.doi.org/10.1016/j.physa.2013.07.080.

[17] Q. Xu, Z. Su, K. Zhang, P. Ren, X.S. Shen, Epidemic information dissemination in mobile social networks with opportunistic links, IEEE Trans. Emerging Top. Comput. 3 (3) (2015) 399–409. http://dx.doi.org/10.1109/ETC.2015.2414792.

[18] L. Zhao, H. Cui, X. Qiu, X. Wang, S. Zhang, A rumor spreading model in the new media age, Physica A 392 (4) (2013) 995–1003. http://dx.doi.org/10.1016/j.physa.2012.09.030.

[19] W. Zhang, S.-Y. Bai, R. Jin, The model of microblog message diffusion based on complex social network, Internat. J. Modern Phys. B 28 (20) (2014) 1450136. http://dx.doi.org/10.1142/S0217979214501367.

[20] Z. Shen, F. Ning, W. Zhou, X. He, C. Lin, D.P. Chin, Z. Zhu, A. Schuchat, Superspreading sars events, Beijing, 2003, Emerg. Infect. Dis. 10 (2) (2004) 256–260. http://dx.doi.org/10.1016/S1059-1105(04)00272-0.

[21] R.A. Stein, Super-spreaders in infectious diseases, Int. J. Infect. Dis. 15 (8) (2011) e510–e513. http://dx.doi.org/10.1016/j.ijid.2010.06.020.

[22] T. Mikhatskhwa, A. Mummert, Modeling Super-spread- ing Events for Infectious Diseases: Case Study SARS, ArXiv e-prints arXiv:1007.0908.

[23] K. wa Fu, C. hong Choe, M. Chau, Assessing censorship on microblogs in China: Discriminatory keyword analysis and the real-name registration policy, IEEE Internet Comput. 17 (3) (2013) 42–50. http://dx.doi.org/10.1109/MIC.2013.28.

[24] K.J. Rothman, S. Greenland, T.L. Lash, Modern Epidemiology, third ed., Lippincott Williams & Wilkins, Philadelphia, 2008.

[25] W. Guan, H. Gao, M. Yang, Y. Li, H. Ma, W. Qian, Z. Gao, X. Yang, Analyzing user behavior of the micro-blogging website sina weibo during hot social events, Physica A 395 (2014) 340–351. http://dx.doi.org/10.1016/j.physa.2013.09.059.

[26] Epidemic model. https://en.wikipedia.org/wiki/Epidemic_model (accessed 13.01.16).

[27] R. Ross, The Prevention of Malaria, NaBu Press, Dutton, 1910.

[28] W. Kermack, A. McKendrick, Contributions to the mathematical theory of epidemics. I. The problem of endemicity, Bull. Math. Biol. 53 (1–2) (1991) 57–87. http://dx.doi.org/10.1007/BF02464424.

[29] H.W. Hethcote, A thousand and one epidemic models, in: S. Levin (Ed.), Frontiers in Mathematical Biology, in: Lecture Notes in Biomathematics, vol. 100, Springer, Berlin, Heidelberg, 1994, pp. 504–515. http://dx.doi.org/10.1007/978-3-642-50124-1_29.

[30] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, Eur. Phys. J. B 26 (4) (2002) 521–529. http://dx.doi.org/10.1140/epjb/e20020122.

[31] V. Isham, S. Harden, M. Nekovee, Stochastic epidemics and rumours on finite random networks, Physica A 389 (3) (2010) 561–576. http://dx.doi.org/10.1016/j.physa.2009.10.001.

[32] A. Vespignani, Modelling dynamical processes in complex socio-technical systems, Nat. Phys. 8 (1) (2012) 32–39. http://dx.doi.org/10.1038/nphys2160.

[33] H.W. Hethcote, The mathematics of infectious diseases, SIAM Rev. 42 (4) (2000) 599–653. http://dx.doi.org/10.1137/S0036144500371907.

[34] Stability theory. http://texas.math.ttu.edu/~gilliam/ttu/ode_pde_pdf/Ch4.pdf (accessed 14.05.16).

[35] G. Teschl, Ordinary Differential Equations and Dynamical Systems, Vol. 140, American Mathematical Soc., 2012.

[36] Z. Guo, Z. Li, H. Tu, Sina microblog: An information-driven online social network, in: 2011 International Conference on Cyberworlds, CW, 2011, pp. 160–167. http://dx.doi.org/10.1109/CW.2011.12.

[37] Q. Su, J. Huang, X. Zhao, An information propagation model considering incomplete reading behavior in microblog, Physica A 419 (2015) 55–63. http://dx.doi.org/10.1016/j.physa.2014.10.042.