Empirical paths to the spread of information in location-based social networks

Ming-Yang Zhou[1]**, Wen-Man Xiong**[1]**, Hao Liao**[1,3,4]**, Tong Wang**[2] **and Zong-Wen Wei**[1,2]

[1] Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, People's Republic of China
[2] Department of Electronic Science and Technology, University of Science and Technology of China, Hefei, 230027, People's Republic of China
[3] Physics Department, University of Fribourg, Chemin du Musée 3, 1700 Fribourg, Switzerland

E-mail: jamesliao520@gmail.com

Abstract.   Spreading phenomena in complex networks have attracted much attention in recent years. However, most of the previous works only concern the critical thresholds and final states of the spread. In this paper, we investigate the empirical spreading paths in real location-based networks and find an abnormal phenomenon that the transferring probability of an epidemic between users varies with time, which violates the classical spreading models with a constant transferring probability. Besides, we observe an interesting delay gap between the maximal spreading velocity and the maximal transferring probability, where the spreading velocity refers to the fraction of newly infected nodes, and transferring probability represents the probability that a susceptible individual gets infected by one of its infected neighbors. Then we propose an advanced SI (susceptible-infected) model to analyze the problem, which could analytically explain the delay gap between the spreading velocity and the transferring probability. Experiments in BA and ER model networks demonstrate the effectiveness of our model. Thus, our work provides a deep understanding of the dynamics of the spreading problems.

[4] Author to whom any correspondence should be addressed.

Contents

1. Introduction

Spreading phenomena are ubiquitous in natural systems, e.g. epidemics in biological networks and computer viruses in internet networks [1, 2], where the systems could be represented by complex networks consisting of nodes and edges [3, 4] and information spreads along the edges of the networks [5, 6]. This issue has particular relevance in the analysis of complex network topologies and epidemic threshold on the dynamical information spreading [7, 8]. For example, classical spreading theory showed the presence of a critical threshold in the epidemic spreading rate below which the disease would vanish at an exponential rate, while it pervades the whole networks if the spreading rate was greater than the threshold [9, 10]. Consequently, the spreading threshold is one of the most important factors of spreading problems.

Previous works mainly focus on the relationship between spreading threshold [11, 12] and complex network topologies [13]. Karsai *et al* [14] found that the bursty and activity patterns of individuals and weight-topology correlations would slow down the spread in communication networks. Lerman *et al* [15] explored the empirical characteristics of information contagion on Twitter and found that network structure affected dynamics of information flow. Newman *et al* [16] investigated empirically the network structure and studied some control strategies for the immunization of computer viruses in email networks. Pastor and Vespignani [17] found that BA networks were more fragile than ER networks against epidemics. Moreover, the threshold for susceptible-infected-susceptible (SIS) model always vanishes in the thermodynamic limit [18]. Kuperman and Abramson [19] explored that the epidemic infection proportion would have a wide amplitude oscillations at a large rewiring probability on SIRS epidemiological model. However, we still lack an in-depth investigation of the dynami-

Unlike previous works that mostly focus on the critical threshold of networks or the identification of influential nodes [20, 21], we aim to track the spreading paths and analyze the evolving patterns of epidemics. We find the spreading probability of information changes with time, where the spreading probability refers to the transferring probability. For example, a new product initially draws great attention and people are more likely to buy it, there the purchase probability could be characterized by spreading probability. However, people will lose interest in the product as time goes by. Intuitively, the sales volume of the product decreases with the diminished interest. However, we find a counterintuitive phenomenon in real location-based networks where a delay gap exists between the maximal transferring probability and the maximal velocity. Then we propose an improved SI spreading model and thoroughly analyze the delay gap. Our method gives the analytical relationship among the spreading velocity, dynamical transferring probability and the network structures. Finally, experiments in BA and ER networks illustrate the validity of our method.

The rest of the paper is organized as follows: in section 2, we introduce dataset description, classical SI model and our proposed method. In section 3, we apply the proposed method to BA and ER model networks and compare their differences. Finally, the conclusion is given in section 4.

2. Materials and methods

2.1. Dataset description

We utilize two different datasets to explore the empirical spreading paths, named Gowalla and Brightkite [22]. In the two datasets, users travel to a place, check in and share the location with their friends on the Gowalla (or Brightkite) websites. Based on the check-in information, their friends are likely to travel the same place, check in and follow the sharing process. We consider checking into the different locations as independent spreading events. Different spreading events share the same underlying social networks. The total number of check-ins for Gowalla is 6.4 million over the period of February 2009–October 2010 and 4.5 million between April 2008 to October 2010 for Brightkite. Table 1 displays an example of the checking-in information of Gowalla, which lists user-id, time and location information. The social network for Gowalla is undirected and unweighted, and for Brightkite is directed and unweighted. But we treat Brightkite as undirected for simplicity. Structural properties of the Gowalla and Brightkite friendship network are shown in table 2. We are interested in understanding the pattern of how users travel to the same place at a certain time, and how likely a person is going to a place under the conditions that they have a friend who has been there.

2.2. The standard SI model

We use a SI spreading model [23] to describe the check-in phenomenon. Supposing that the information spreads only along the friendships in the social networks. If their friends check in the same place, we treat that the information has successfully spread

Table 1. Checking-in information of network Gowalla.

User-id	Check-in time	Latitude	Longtitude	Location-id
196514	2010-07-24T13:45:06Z	53.364 8119	$-2.272\,346\,5833$	145064
196514	2010-07-24T13:44:58Z	53.360 511 233	$-2.276\,369\,017$	1275991
196514	2010-07-24T13:44:46Z	53.365 389 5945	$-2.275\,408\,7046$	376497
196514	2010-07-24T13:44:38Z	53.366 370 9833	$-2.270\,076\,4333$	98503
196514	2010-07-24T13:44:26Z	53.367 408 7524	$-2.278\,381\,3477$	1043431
196514	2010-07-24T13:44:08Z	53.367 566 3377	$-2.278\,631\,763$	881734
196514	2010-07-24T13:43:18Z	53.367 964 0626	$-2.279\,294\,3689$	207763
196514	2010-07-24T13:41:10Z	53.364 905	$-2.270\,824$	1042822

from the user to their friends. The spreading probability in the SI model characterizes the likelihood whether the information could be passed through an edge.

Based on the above explanation between the dataset and spreading model, we focus on the SI spreading model in which each individual is in one of two states, either susceptible (S) or infected (I). We consider a network composed of N nodes tied with E links. In the spreading process, a susceptible node will become infected with a certain probability if its infected neighbors transmit the epidemic information to it. In heterogeneous random networks [17], let $i_k(t)$ be the probability that nodes with degree k are infected. The classical SI model can be written as [24, 25],

$$\frac{\mathrm{d}i_k(t)}{\mathrm{d}t} = \beta k(1 - i_k(t))\Theta(i_k(t)) \tag{1}$$

where β is the spreading rate, $1 - i_k(t)$ is the density that nodes with degree k are not infected, and $\Theta(i_k(t))$ is the probability that any given edge is pointing to an infected node with degree k. Then we can obtain

$$\begin{aligned}
\Theta(t) &= \sum_{k'} k'P(k' \leftarrow k)i_{k'}(t) \\
&= \sum_{k'} \frac{k'P(k')i_{k'}(t)}{\sum_k P(k)} \\
&= \frac{1}{\langle k \rangle} \sum_{k'} k'P(k')i_{k'}(t),
\end{aligned} \tag{2}$$

where $\langle k \rangle = \sum_{k'} k'P(k')$ and $P(k' \leftarrow k)$ are the probability that nodes with degree k are pointing to nodes with degree k', $\langle k \rangle$ and $P(k)$ represent the average degree and degree distribution of the network, respectively. Based on equations (1) and (2), we can obtain

$$\frac{\mathrm{d}i_k(t)}{\mathrm{d}t} = \beta k\Theta(t) \tag{3}$$

$$\frac{\mathrm{d}\Theta(t)}{\mathrm{d}t} = \beta \frac{\langle k^2 \rangle}{\langle k \rangle}\Theta(t). \tag{4}$$

So we can get the probability of infected nodes with degree k is,

$$i_k(t) = \frac{\langle k \rangle k}{\langle k^2 \rangle}\mathrm{e}^{\frac{\langle k^2 \rangle \beta}{\langle k \rangle}t + C}, \tag{5}$$

Table 2. Structural properties of different real location-based networks and two artificial networks. Properties include network size (N), total edge number (E), degree heterogeneity ($H = \langle k^2 \rangle / \langle k \rangle^2$), degree assortativity (r), average clustering coefficient ($\langle CC \rangle$), and average shortest path length ($\langle d \rangle$).

Network	N	E	H	r	$\langle CC \rangle$	$\langle d \rangle$
Gowalla	196 591	950 327	31.7105	−0.0293	0.0235	4.43
Brightkite	58 228	214 078	8.6635	0.0108	0.111	4.86
BA	10 000	19 997	2.1187	0.0022	0.000 01	5.5006
ER	10 000	50 090	1.0992	−0.000 03	0.000 58	4.2572

where C is a constant. The infection proportion of the whole network is $i(t) = \sum_k P(k) i_k(t)$, then

$$i(t) = \frac{\langle k \rangle^2}{\langle k^2 \rangle} e^{\frac{\langle k^2 \rangle \beta}{\langle k \rangle} t + C} \tag{6}$$

where $\langle k^n \rangle = \sum_k k^n P(k)$. This readily implies that the spread velocity is determined by the heterogeneity $H = \langle k^2 \rangle / \langle k \rangle^2$ [26].

2.3. Our proposed method

To simulate the dynamics of the spreading probability and spreading velocity, figure 1 gives a vivid explanation in a small artificial network to show the calculation of the spreading probability and spreading velocity at different periods. We mark infected nodes as 'I' and susceptible nodes as 'S' in each period. Initially, only one node is infected and others are susceptible, while four links exist between infected nodes and susceptible nodes. At time t_1, two nodes are newly infected. Thus the current spreading probability is calculated by $\beta(t_1) = \frac{1}{2}$. The spreading velocity refers to the fraction of newly infected proportion in the new period. The total infected nodes of the whole networks is 22, then the current spreading velocity is $\delta(t_1) = \frac{1}{11}$. Suppose that information spreads as figure 1, we can easily obtain the spreading probability and spreading velocity at each period. Here we introduce the spreading probability and spreading velocity in a real location-based network Gowalla. As shown in table 1, there are many different independent spreading events. By taking the interference effect of information into consideration, we only concern the events that more than 1000 people participate in. Moreover, for the spread of each event, we divide the spreading time span into 12 intervals. In different periods, we consider users who have ever checked-in to the location as infected users s_t and who have not checked-in the location as susceptible users. Then we compute the spreading probability $\beta(t)$ and spreading velocity ($\delta(t)$) in different periods as the following,

$$\beta(t) = i/b \tag{7}$$

$$\delta(t) = s_t/s \tag{8}$$

where i is the number of newly infected users at time t who are neighbors of infected users before time t, while b is the number of susceptible users who are neighbors of

infected users before time t, and s refers to the total number of infection users at the final state.

Figure 2 displays the dynamics of spreading probability and spreading velocity of network Gowalla in different periods. Remarkably, the spreading probability varies as time and violates the classical SI spreading model that possesses a constant spreading probability. Meanwhile, both of the spreading probability and velocity have a peak value. Comparing figures 2(a) and (b), we find a delay gap Δt between the maximal spreading velocity and the maximal spreading velocity. Besides, the delay gap exists in different time spans. The spreading paths of other events in other networks also reveal the delay phenomenon.

Note that based on equation (6), we cannot observe the delay gap. Because equation (6) considers the spreading rate as a constant, while the spreading rate varies in real scenarios. Thus we need a better model to describe the phenomenon. Here we present an improved SI spreading method:

$$\frac{\mathrm{d}i_k(t)}{\mathrm{d}t} = \beta(t)k(1 - i_k(t))\Theta(i_k(t)). \tag{9}$$

Comparing with equation (1), the $\beta(t)$ varies as time. Other parameters are the same to equation (1). Based on equations (2) and (9),

$$i_k(t) = \frac{\langle k \rangle k}{\langle k^2 \rangle}\mathrm{e}^{\frac{\langle k^2 \rangle}{\langle k \rangle}\int \beta(t)\mathrm{d}t + C}, \tag{10}$$

where C is a constant. The infection proportion of the whole network $i(t) = \sum_k P(k)i_k(t)$, thus

$$i(t) = \frac{\langle k \rangle^2}{\langle k^2 \rangle}\mathrm{e}^{\frac{\langle k^2 \rangle}{\langle k \rangle}\int \beta(t)\mathrm{d}t + C}. \tag{11}$$

At the initial state of the spread, we set $i(t = 0) = i_0$, where i_0 refers to a small number of infected proportion.

$$i(t) = i_0\mathrm{e}^{\frac{\langle k^2 \rangle}{\langle k \rangle}\int \beta(t)\mathrm{d}t}. \tag{12}$$
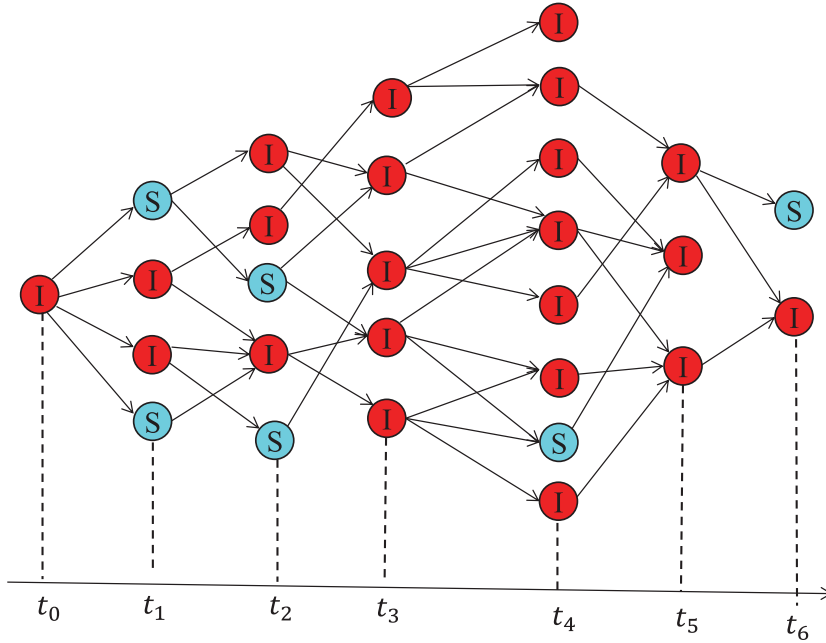
The spreading velocity follows,

$$\frac{\mathrm{d}i(t)}{\mathrm{d}t} = \frac{\langle k^2 \rangle}{\langle k \rangle}i_0\beta(t)\mathrm{e}^{\frac{\langle k^2 \rangle}{\langle k \rangle}\int \beta(t)\mathrm{d}t}. \tag{13}$$

To simplify the rate of change in spreading velocity, we denote $\gamma(t) = \frac{\langle k^2 \rangle}{\langle k \rangle}i_0\mathrm{e}^{\frac{\langle k^2 \rangle}{\langle k \rangle}\int \beta(t)\mathrm{d}t}$, then

$$\frac{\mathrm{d}i(t)^2}{\mathrm{d}^2t} = (\langle k \rangle\frac{\mathrm{d}\beta(t)}{\mathrm{d}t} + \langle k^2 \rangle\beta^2(t))\gamma(t). \tag{14}$$

In practical scenarios, $\beta(t)$ may follow Gaussian similar distribution, which can be written as,

$$\beta(t) = a\mathrm{e}^{-\frac{(t-\mu)^2}{2\delta^2}}. \tag{15}$$

Figure 1. A schematic illustration of spreading probability and spreading velocity of an artificial network. Nodes in red marked as 'I' stand for the infected nodes. While nodes in blue marked as 'S' represent the susceptible nodes. We can easily obtain the spreading probability and spreading velocity at each period.

The time corresponding to the maximal spreading probability and the maximal spreading velocity can be obtained by,

$$
\begin{cases}
\frac{\mathrm{d}\beta(t)}{\mathrm{d}t} = 0 \\
\frac{\mathrm{d}i(t)^2}{\mathrm{d}^2t} = 0.
\end{cases}
\tag{16}
$$

Clearly, there is no solution to equation (16). Therefore, we can say that the spreading probability $\beta(t)$ and velocity $\frac{dI}{dt}$ cannot reach the maximum at the same time. Denote the optimal time t_1 and t_2 corresponding to the maximal spreading probability and the maximal spreading velocity. Then the delay gap can be obtained by $\Delta t = t_2 - t_1$.

We now consider the reduced $\beta(t) = C_0$ that means the cases in equation (6), thus the spreading velocity follows

$$
\frac{\mathrm{d}i(t)}{\mathrm{d}t} = \frac{\langle k^2 \rangle}{\langle k \rangle} i_0 C_0 \mathrm{e}^{\frac{\langle k^2 \rangle}{\langle k \rangle} C_0 t}.
\tag{17}
$$

Obviously, the spreading velocity increases with time t at an exponential rate. There is no solution to equation (16) and no delay gap exists between the spreading probability and velocity in classical SI spreading model.

Apart from the two cases introduced above, the delay gap is also applicable for other $\beta(t)$, such as poisson distribution.

To verify the delay gap between the maximal spreading probability and velocity, figure 3 plots the rate of change in spreading velocity when the spreading probability follows different distributions in BA network. The average degree is $\langle k \rangle = 4$ in BA network and we give the initial infected proportion $i_0 = 10^{-3}$. Figure 3(a) shows the rate
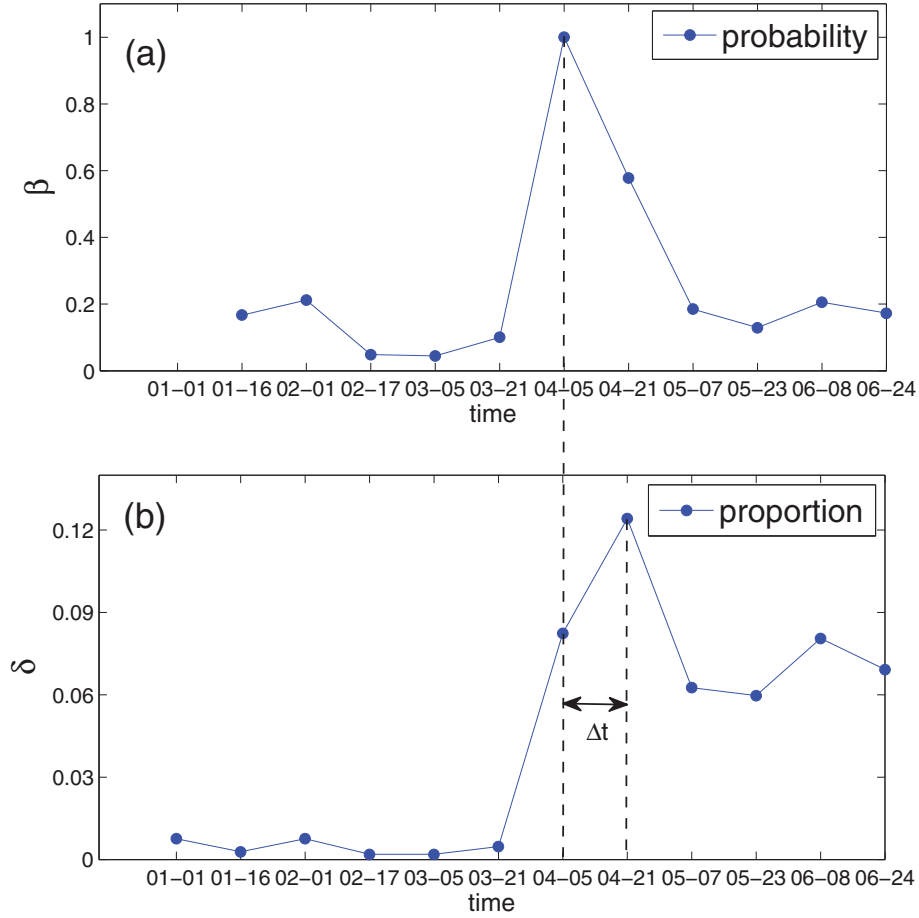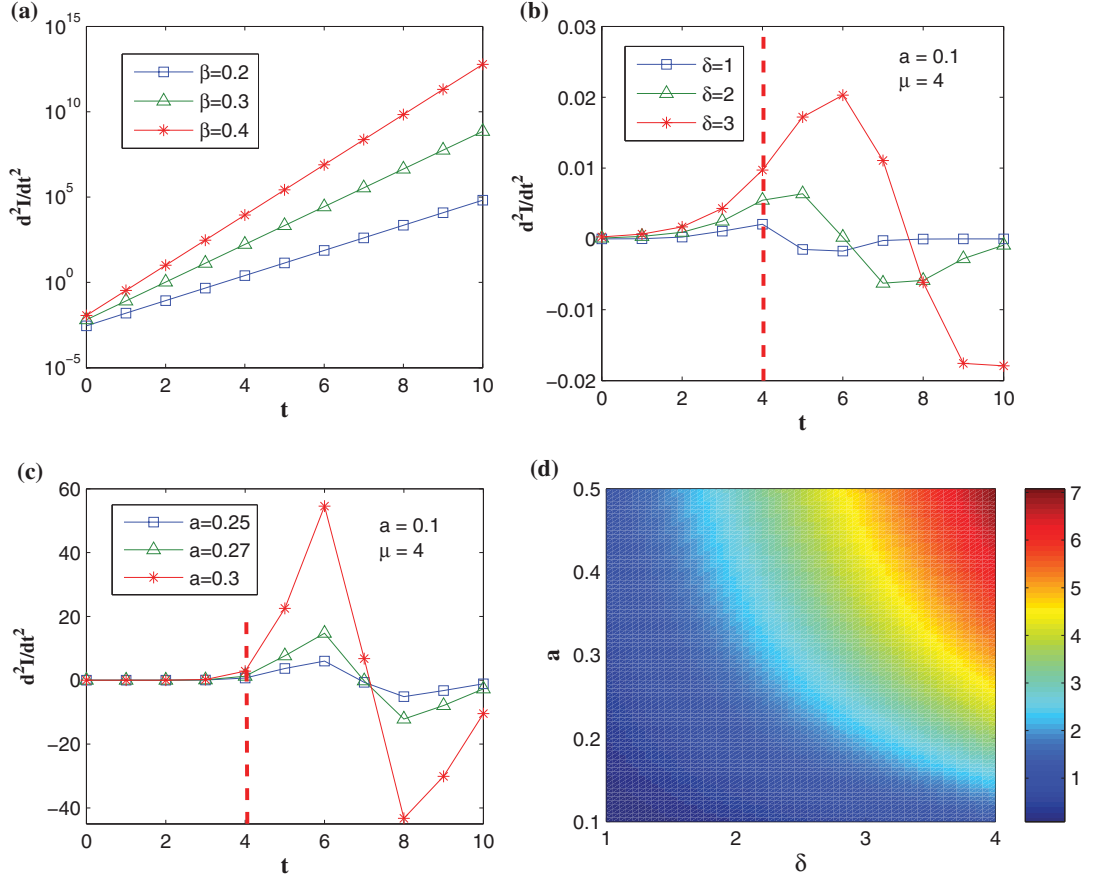
Figure 2. The spreading probability and spreading velocity as a function of time t. (a) and (b) Represent the spreading probability and spreading velocity of the event labeled 10 677 in Gowalla respectively. Δt is the delay gap between the maximal spreading probability and the maximal spreading velocity.

of change in spreading velocity increases with time t at an exponential speed when the spreading probability is given as a constant. With the increase of β, $\frac{\mathrm{d}^2 I}{\mathrm{d}t^2}$ varies more significantly. Figure 3(b) displays the rate of change in spreading velocity when the spreading probability follows Gaussian distribution. We consider the peak value of Gaussian distribution as $a = 0.1$ and the mean value as $\mu = 4$. It shows that the bigger δ is, the later the spreading velocity reaches the maximum ($\frac{\mathrm{d}I^2}{\mathrm{d}^2 t} = 0$). Obviously, there is a delay gap between the maximal spreading velocity and the maximal spreading probability ($t = \mu = 4$) (see the red dashed line in figure 3(b)). Figure 3(c) shows the rate of change in spreading velocity as a function of a. The spreading probability follows Gaussian distribution, and the standard degree deviation $\delta = 2$ and $\mu = 4$. When the spreading velocity reaches the maximum ($\frac{\mathrm{d}I^2}{\mathrm{d}^2 t} = 0$), $t \approx 7$. While the spreading probability reaches the maximum at $t = \mu = 4$ (see the red dashed line in figure 3(c)). Figure 3(d) indicates the delay gap between the maximal spreading velocity and the maximal spreading probability as a function of δ and a, where $\mu = 4$. The delay gap is positively correlated with both a and δ. Other settings are also applicable in the simulation and would influence the results little.

Figure 3. The rate of change in spreading velocity when the spreading rate follows different distributions in BA network. (a) The rate of change in spreading velocity as a function of time for different β. The spreading probability β is given as a constant, $0.2, 0.3, 0.4$, respectively. (b) The rate of change in spreading velocity as a function of time for different δ. The spreading rate β follows Gaussian distribution, where $a = 0.1$, $\mu = 4$. (c) The rate of change in spreading velocity as a function of time for different a. The spreading rate β follows Gaussian distribution, where $\delta = 2$, $\mu = 4$. The dashed line in (b) and (c) represent the time when the spreading probability reaches the maximum ($t = \mu = 4$). (d) The delay gap between the maximal spreading velocity and the maximal spreading probability as a function of different δ and different a when the spreading probability follows Gaussian distribution, where $\mu = 4$. The color depth indicates the delay gap.

3. Results

In this section, to verify the underlying factor of network structure on the delay gap, we test the performance of our proposed method both in Barabási–Albert (BA) and Erdö s–Rényi (ER) model networks [27, 28]. BA networks follow power-law degree distribution [29, 30] $P(k) = k^{-\gamma}$ where $2 < \gamma < 3$. The degree distributions of real location-based networks Gowalla and Brightkite are similar to BA network. ER network follows Poisson distribution. The statistical characteristics of BA and ER networks have been shown in table 2.

9
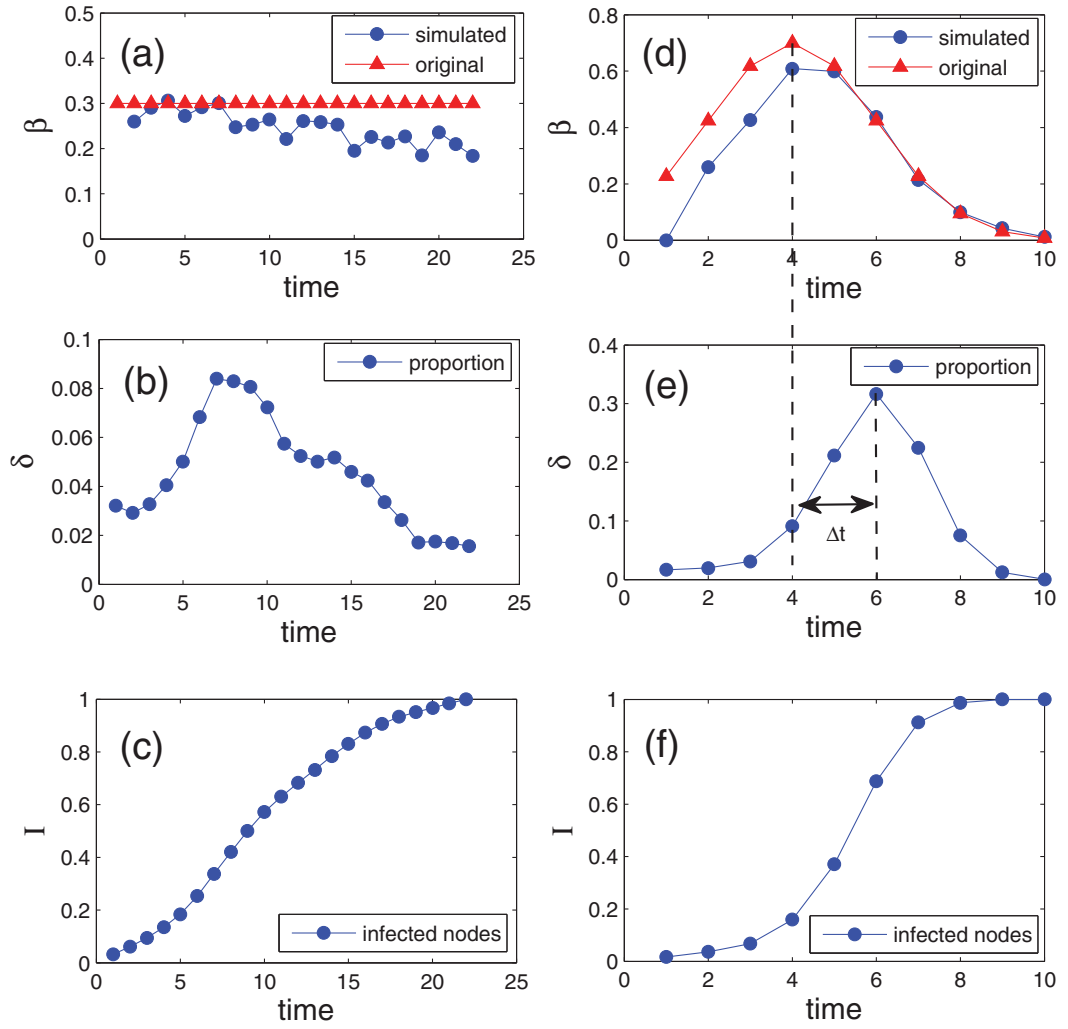
Figure 4. The spreading probability, spreading velocity and the total infected nodes as a function of time t respectively in BA network. Figures (a)–(c) are the spreading probability (blue circle), spreading velocity and total infected nodes at time t as a function of different periods respectively when the original spreading probability (red triangle) is set to be a constant 0.3 in figure (a). Because if the constant is too small, information will be difficult to spread throughout the whole system, if too big, information will spread throughout the whole system quickly with fewer steps. There is no sense to do research in the above two cases. As a consequence, we give the spreading probability as a moderate value 0.3. Figures (d)–(f) have the same meaning with figures (a)–(c) respectively when the original spreading probability is represented by a Gaussian function. The peak of the Gaussian function (red triangle) is set to be 0.7 in figure (d). Δt is the delay gap between the maximal spreading probability and the maximal spreading velocity. Spreading simulations will be introduced in the following: initially, 1% nodes are randomly chosen to be initial spreaders. At each step, every infected node infects all its neighbors with a given spreading probability until the number of susceptible nodes is less than 0.5%, which indicates the spreading process will be terminated. The results are obtained by averaging 100 independent simulations.
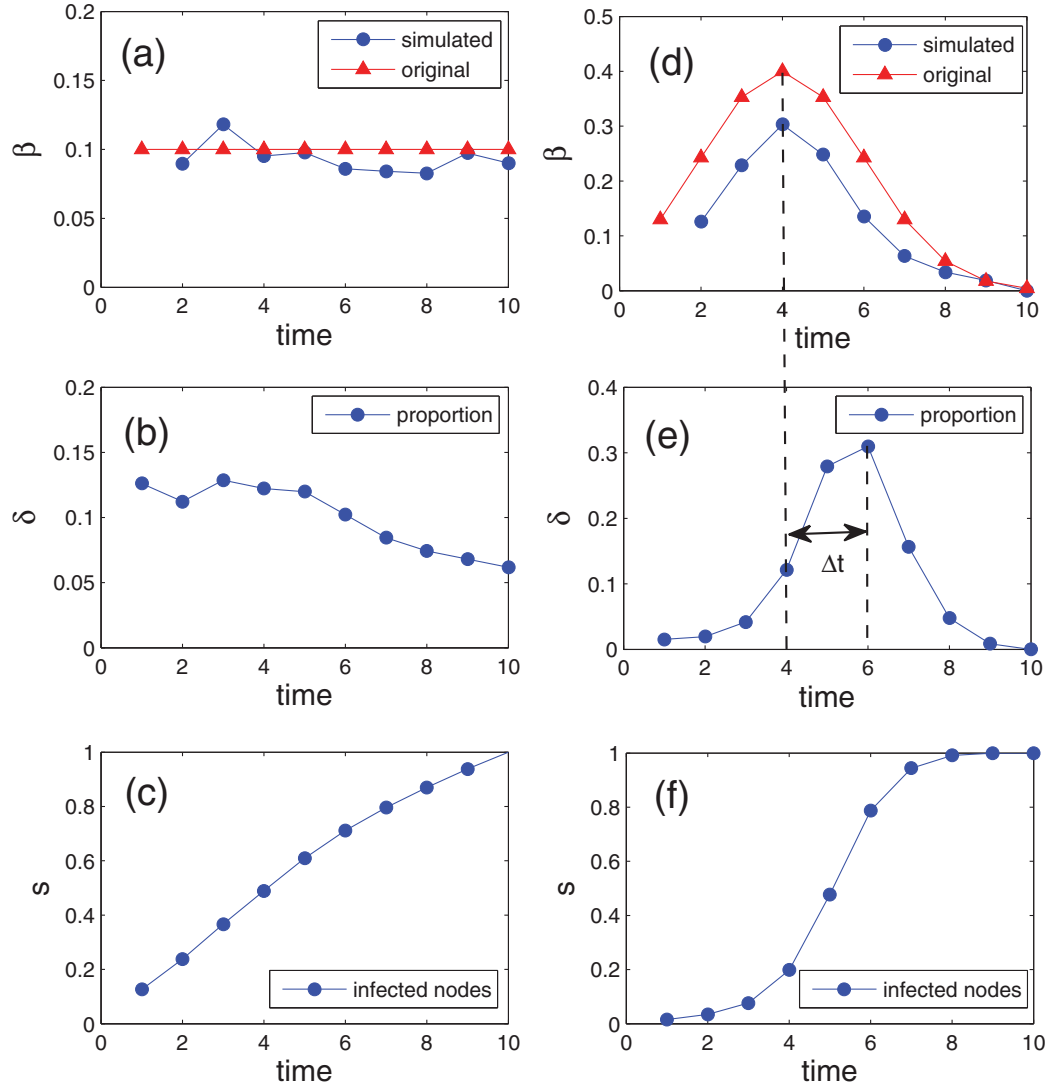
Figure 5. The spreading probability, spreading velocity and the total infected nodes as a function of time t respectively in ER network. Figures (a)–(c) are the spreading probability (blue circle), spreading velocity and total infected nodes at time t as a function of different periods respectively when the original spreading probability is given as a constant 0.1. Figures (d)–(f) are the results with a Gaussian function, which explains the delay gap in figure 2. Δt is the delay gap between the maximal spreading probability and the maximal spreading velocity. Spreading simulations are the same as figure 4. The results are obtained by averaging 100 independent simulations.

In the experiments, the BA network size is $N = 10\,000$ and the average degree is $\langle k \rangle = 4$. We consider the spreading probability as a constant ($\beta = 0.3$) or Gaussian distribution (see equation (15)) respectively to simulate the spreading path. We set $a = 0.7$, $\mu = 4$, and $\delta = 2$. At the beginning of the spreading process, 1% nodes are randomly chosen to be initial spreaders and other nodes are susceptible. Figure 4 shows the dynamics of simulated spreading probability (blue circle) and spreading velocity with different original spreading probabilities (red triangle). In figure 4(a), the simulated

spreading probability ranges from 0.2 to 0.3, approximating to the original spreading probability. It is due to this that we only count the newly infected nodes, but ignore whether the susceptible node is infected by one or more infected nodes. That is why the simulated probability has a little fluctuation. While in figure 4(d) the simulated spreading probability has similar performance to the original spreading probability. For the spreading velocity in figures 4(b) and (e), they both first increase and then decrease, but the latter has a greater fluctuation than the former. The proportion increases by 0.05 from $t = 1$ to $t = 7$ in figure 4(b), while it increases by 0.3 from $t = 1$ to $t = 6$ in figure 4(e). Figures 4(c) and (f) show the total infected nodes as a function of time t, which means the accumulation of the infection proportion in figures 2(b) and (e) respectively. Thus, the infected nodes reach the stable state in figure 4(f) earlier than figure 4(c). Moreover, the spreading velocity and spreading probability reaches the maximum at $t = 6$ and $t = 4$ respectively, which indicates the delay gap between them in figures 4(d) and (e). A similar phenomenon occurs in real location-based networks in figure 2.

To further investigate whether the delay gap is relevant to the network structure, figure 5 plots the dynamical spreading paths in ER network, where the size is $N = 10\,000$ and $\langle k \rangle = 10$. We consider the spreading probability as a constant ($\beta = 0.1$) or a Gaussian distribution, where $a = 0.4$, $\mu = 4$ and $\delta = 2$. In figure 5(a), the simulated spreading probability approximates to 0.1, almost equal to the original spreading probability. In figure 5(b), the simulated spreading probability has a similar trend with the original spreading probability. Comparing figures 5(d) with (e), there is a delay gap Δt. As a consequence, the delay gap is determined by both the spreading probability and the underlying network structures.

4. Conclusion

In conclusion, we first investigate the empirical spreading path in real location-based networks and show that the spreading probability varies with time, which is different from the constant spreading probability in classical SI model. Additionally, we show a delay gap exists between the maximal spreading probability and the maximal spreading velocity. By taking into account the dynamics of the spreading probability, we put forward an advanced SI spreading model to thoroughly analyze the delay gap. Finally, our method reproduces and characterizes the delay gap in random networks, especially in BA and ER networks, which illustrates the effectiveness of our method.

The exploration of empirical paths to the spread of information has great significance in improving information diffusion efficiency. When designing an advertising strategy for some activities (e.g. musical concert, film debut), we usually raise some advertising on TV and internet ahead of the time. Based on the delay gap in the paper, we can better determine when to put the advertising in the market to achieve maximal influence. Moreover, the proposed method can also help predict the trend and the coverage of the spread in a network. Thus, our paper has a deep understanding on the spreading problem and may help design better advertising strategies.

Acknowledgments

References

[1] Liljeros F, Edling C R, Amaral L A N, Stanley H E and Åberg Y 2001 The web of human sexual contacts *Nature* **411** 907–8

[2] Newman M E J 2001 The structure of scientific collaboration networks *Proc. Natl Acad. Sci.* **98** 404–9

[3] Sen P, Dasgupta S, Chatterjee A, Sreeram P A, Mukherjee G and Manna S S 2003 Small-world properties of the indian railway network *Phys. Rev.* E **67** 036106

[4] Stelling J, Klamt S, Bettenbrock K, Schuster S and Gilles E D 2002 Metabolic network structure determines key aspects of functionality and regulation *Nature* **420** 190–3

[5] Pastor-Satorras R, Castellano C, Van Mieghem P and Vespignani A 2015 Epidemic processes in complex networks *Rev. Mod. Phys.* **87** 925

[6] Castellano C, Fortunato S and Loreto V 2009 Statistical physics of social dynamics *Rev. Mod. Phys.* **81** 591

[7] Watts D J and Strogatz S H 1998 Collective dynamics of 'small-world' networks *Nature* **393** 440–2

[8] Boguñá M, Castellano C and Pastor-Satorras R 2013 Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks *Phys. Rev. Lett.* **111** 068701

[9] Nekovee M, Moreno Y, Bianconi G and Marsili M 2007 Theory of rumour spreading in complex social networks *Physica* A **374** 457–70

[10] Anderson R M, May R M and Anderson B 1992 *Infectious Diseases of Humans: Dynamics and Control* vol 28 (New York: Wiley)

[11] Chen Y, Paul G, Havlin S, Liljeros F and Stanley H E 2008 Finding a better immunization strategy *Phys. Rev. Lett.* **101** 058701

[12] Moreno Y, Pastor-Satorras R and Vespignani A 2002 Epidemic outbreaks in complex heterogeneous networks *Eur. Phys. J.* B **26** 521–9

[13] Yan G, Tsekenis G, Barzel B, Slotine J-J, Liu Y-Y and Barabási A-L 2015 Spectrum of controlling and observing complex networks *Nat. Phys.* **11** 779

[14] Karsai M, Kivelä M, Pan R K, Kaski K, Kertész J, Barabási A-L and Saramäki J 2011 Small but slow world: how network topology and burstiness slow down spreading *Phys. Rev.* E **83** 025102

[15] Lerman K and Ghosh R 2010 Information contagion: an empirical study of the spread of news on digg and twitter social networks *Proc. Fourth International AAAI Conf. on Weblogs and Social Media* **10** 90–7

[16] Newman M E J, Forrest S and Balthrop J 2002 Email networks and the spread of computer viruses *Phys. Rev.* E **66** 035101

[17] Pastor-Satorras R and Vespignani A 2001 Epidemic spreading in scale-free networks *Phys. Rev. Lett.* **86** 3200

[18] Castellano C and Pastor-Satorras R 2010 Thresholds for epidemic spreading in networks *Phys. Rev. Lett.* **105** 218701

[19] Kuperman M and Abramson G 2001 Small world effect in an epidemiological model *Phys. Rev. Lett.* **86** 2909

[20] Tao Z, Zhongqian Fu and Binghong W 2006 Epidemic dynamics on complex networks *Prog. Nat. Sci.* **16** 452–7

[21] Chakrabarti D, Wang Y, Wang C, Leskovec J and Faloutsos C 2008 Epidemic thresholds in real networks *ACM Trans. Inf. Syst. Secur.* **10** 1
[22] Cho E, Myers S A and Leskovec J 2011 Friendship and mobility: user movement in location-based social networks *Proc. of the 17th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (ACM) pp 1082–90
[23] Newman M E J 2002 Spread of epidemic disease on networks *Phys. Rev.* E **66** 016128
[24] Wang X, Li X and Chen G 2012 *Network Science: an Introduction* (Beijing: Higher Education Press)
[25] Shu P, Wang W, Tang M and Do Y 2015 Numerical identification of epidemic thresholds for susceptible-infected-recovered model on finite-size networks *Chaos* **25** 063104
[26] Barthélemy M, Barrat A, Pastor-Satorras R and Vespignani A 2004 Velocity and hierarchical spread of epidemic outbreaks in scale-free networks *Phys. Rev. Lett.* **92** 178701
[27] Barabási A-L 2013 Network science *Phil. Trans. R. Soc.* A **371** 20120375
[28] Lewis T G 2011 *Network Science: Theory and Applications* (New York: Wiley)
[29] Barabási A-L and Albert R 1999 Emergence of scaling in random networks *Science* **286** 509–12
[30] Burda Z, Correia J D and Krzywicki A 2001 Statistical ensemble of scale-free random graphs *Phys. Rev.* E **64** 046118