Privacy Spreading on Networks

Atuya OKUDAIRA
Faculty of Economics
The International University of Kagoshima
Kagosima 891-0197, Japan [*]

Yasushi NAKAO
Center for Fundamental Education
The University of Kitakyushu
Kitakyushu 802-8577, Japan [†]

Abstract

The spreading phenomenon of private information is investigated in the context of spreading phenomena on networks. We simulated spread of private information on scale-free networks and on lattice. The epidemic disease spreading models such as the SIS and the SIR are applied. In our simulation, creation of new information through the interaction between two packs of information are taken into consideration; such new information can propagate on networks. Information creation made the spreading of privacy easier to take place and at a wider range in networks in many calculations, including in the parameter region where the spreading is usually difficult. More strict rules against the leaking of privacy, and/or, technologies will be needed in order to inhibit the spread of private information, when the integration of various information is easy in the era of "Big Data". To set time-to-live (TTL) to the private data may be a good method.

[*] <okudaira(at)eco.iuk.ac.jp>
[†] <nakaoy(at)kitakyu-u.ac.jp>

概要

プライバシー情報が広がる現象をネットワーク上の拡散現象の文脈で調べた。我々はプライバシー情報の拡散をスケールフリーネットワークと格子の上で計算機実験した。感染症のモデルであるSISとSIRを使用した。計算機実験では、拡散する情報の塊の相互作用で新しい情報が生まれる。それがまた拡散する。情報の生成によって多くの計算で拡散がより簡単に起こりより広く広がった。通常は拡散が困難なパラメーター領域でもそうであった。いろいろな情報の集積が簡単なこの「ビッグデータ」の時代では、通常よりより厳しい法律や技術がプライバシー情報の広がりを防ぐために必要になる。プライバシーデータに寿命 (TTL) を設定して放棄することが一つのやり方かもしれない。

Key words: privacy, epidemic spreading, SIS, SIR, Big Data

1 Introduction

Privacy leakage is a big problem in an era of “Big Data”, although it is difficult to define “privacy”. [9]. There are some privacy protecting technologies (e.g. [10][6][2]). However, there are effective attacks on some of the technologies [6]. There are not only technologies but also some laws that protect privacy or let the government reveal private data. However, different nations have different laws [7], so the situation is not uniform. If the leakage of the data (or information) occurs, it is able to spread through the Internet or by word-of-mouth communication.

The spreading (propagating) processes of private information are similar to the epidemic spreading. The epidemic spreading is a generic phenomenon and applied not only to diseases but also rumors or information. The epidemic spreading in complex networks has been widely studied [11]. For instance, Pastor-Sattoras and Vespignani (2001) reported the absence of the epidemic threshold on a wide range of scale-free networks [12]. The Internet has the scale-free network structure [14], so an understanding of spreading processes on networks are important for the privacy research.

Although privacy preserving mechanisms are presented (e.g. [8]), "Big Data" may enable us to uncover some private information of persons [2]. For example, using well-known information or data which is easy to get, private information will have a chance of being uncovered, if someone relates those information together. The new information is, so to speak, created using well-known or less-known information. Moreover, the information used has the property of spreading. It is natural to assume that different data spread on different networks.

In this paper, we will investigate the effect of information creation on private information (or data) spreadings on networks. Our approach is as follows. The three layers of different types of networks, on which the same people belong, are assumed. The well-known information, "I_0" and "I_1", will spread on the network layer "0" and layer "1", respectively. If the two pieces of information are acquired by the same person at the same time, we assume the privacy information is uncovered by that person with some probability. Once the privacy information "I_2" is uncovered, that privacy information can spread on the network layer "2" independently, besides the continuous process of the uncovering privacy through the network 0 and 1.

In the following sections, we describe the models, the method of simulation, present the results of the simulation, then discuss them and show our conclusion.

2 Spreading models

The spreading models used in this paper are the SIS and SIR models. Here, we briefly describe these epidemic spreading models and also a model which has an interaction between diseases. We will use this model. A more detailed description of the SIS and SIR models can be found in [4] or [11].

2.1 The SIS model

The persons on the network are in the susceptible state, S, if they are not infected by the disease at some time. If the persons are infected, their states are changed to I. Infected persons will recover with no immunity to the disease at some time. That is, recovered persons are in the susceptible state, S, again. This transition process, $S \rightarrow I \rightarrow S$, is called the SIS model.

There are two parameters, β and μ, which describe the SIS transition. The parameter β is the infection probability. The susceptible persons will be infected with the disease with the probability β when there is an infected person in the neighborhood. On the other hand, the infected person recovers spontaneously and the parameter μ is the recovering probability. Thus, we express the SIS transition as

$$I + S \xrightarrow{\beta} 2I, \tag{1}$$

$$I \xrightarrow{\mu} S. \tag{2}$$

In the context of privacy spreading, S denotes the state of the persons who do not have the "information". If the persons get the "information" through interaction with their neighbors on the network, their state changes to the infected state I. After some time, infected persons will forget the "information" thoroughly, and return to the state S again.

2.2 The SIR model

The SIR model is a variation of the SIS when considering the immunization. Infected persons will recover with immunity to the disease at some time. The recovered persons no longer have the chance to be infected by the disease after recovering. This state is denoted by R. Therefore, the transition process in the SIR is as follows: $S \rightarrow I \rightarrow R$.

The SIR model also has two parameters, β and μ. The transition of states in the SIR is as follows:

$$I + S \xrightarrow{\beta} 2I, \tag{3}$$

$$S \xrightarrow{\mu} R. \tag{4}$$

Again, in the context of privacy spreading, R denotes the state of persons who had the information and deleted it intentionally. So, they will neglect the information immediately if the same information is transmitted by neighbors.

2.3 Interaction of informations/diseases

We assume that there are three information/diseases named 0, 1 and 2. They spread on different networks but the nodes on the networks are common. We write the state of a node as (X_0, X_1, X_2) where X_i is a state of the information/disease i $(i = 0, 1, 2)$, and the values of X_i is S and I in the SIS model, and S, I and R in the SIR model.

Information/disease 2 is created by the interaction of the information/disease 0 and information/disease 1 in a node.

$$(I, I, S) \rightarrow (I, I, I). \tag{5}$$

In the simulation, we assume the interaction shown in eq.(5) occurs with the probability p_{012} in each time step.

3 Computer Experiments

We performed computer experiments of the spreading phenomena on networks according to both the SIS and SIR models. The network making programs and spreading simulation programs are made using a high performance Scheme (a lisp dialect [1]) compiler Bigloo [13]. As a random number generator, Mersenne Twister is used.

3.1 Topology of the network

As described in the introduction, there are three layers of different types of networks, on which the same people belong. The network models used in this paper are the scale-free network, a random network and the lattice network.

3.1.1 Scale-free networks

The network described in Barabási and Albert (1999) [3] is a typical scale-free network. We constructed the Barabási and Albert network using the algorithm proposed by Bollobás and Riordan (2004) [5]. We call this network "BA-BR network", hereafter. The BA-BR algorithm is slightly different from the original Barabási and Albert's algorithm, since it allows the connection to the adding node itself.

Initially it has one node which points itself. The algorithm add nodes with m edges which stochastically points to old nodes. The pointing probability is proportional to the numbers of edges of the nodes. Figure 1 show the cumulative degree k distribution of a BA-BR network.

Our program for constructing the BA-BR network is named "BA-naive". An example of the output of BA-naive is as follows:

```
$ BA-naive -m 3 -a 1 -b 1 -t 10 -i 1
;;; 5 9 4 8 3 3 1 1 9 2
;;; BA-naive v2.0; t = 10; i = 1; a = 1; b = 1; dms = 0

((0 0 0 0) 9 8 7 7 6 5 5 4 3 3 3 2 2 2 1 1 0 0 0)
((1 0 1 0) 6 4 1)
((2 0 0 0) 4)
((3 0 0 0) 6)
((4 2 0 1) 8 5)
((5 0 4 0) 7)
((6 1 0 3))
((7 5 0 0) 9)
((8 4 0 8) 9 8)
((9 0 8 7))
```

The output is Scheme (a lisp dialect) expressions. Contents can

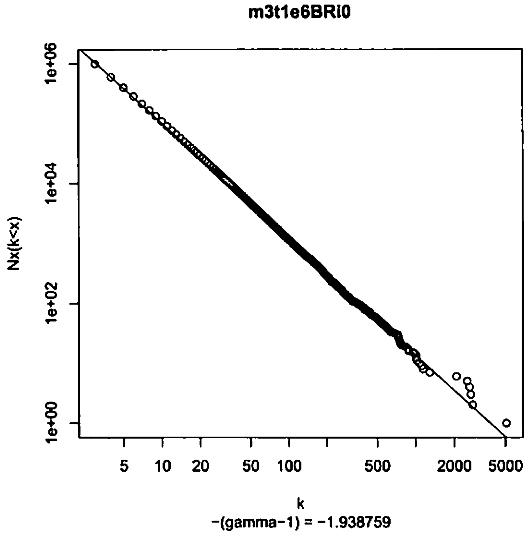


Figure 1: The log-log plot of the cumulative connectivity degree k distribution of a BA-BR network created by the program BA-naive. The line of a linear model fit is also shown.

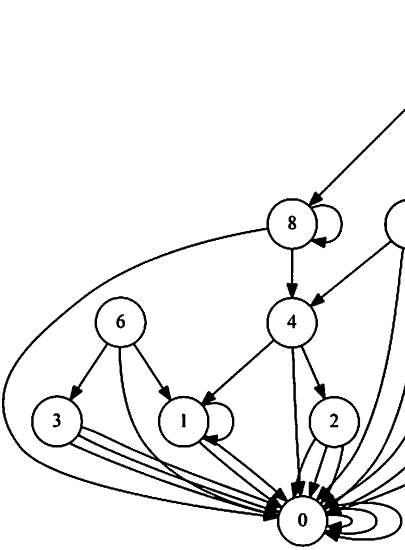


Figure 2: A BA-BR network of $m = 3, N = 10$ presented as a directed network. In the spreading simulation, we treated this network as a undirected graph.

be extracted using Scheme procedure `car` and `cdr` [1] . When the line of the output is x, `(car (car x))` is the label of the node, `(cdr (car x))` is the list of the labels of the nodes to which the node is directed. `(cdr x)` is the list of labels of the nodes which are directed to the node. For example, when x is `((4 2 0 1) 8 5)`, `(car (car x))` is 4, `(cdr (car x))` is `(2 0 1)` and `(cdr x)` is `(8 5)`. The structure of the above network is shown in figure 2 . The network graphs are made using Graphviz.

[1] When `x` is `(a b c d)`, `(car x)` is `a` and `(cdr x)` is `(b c d)`.

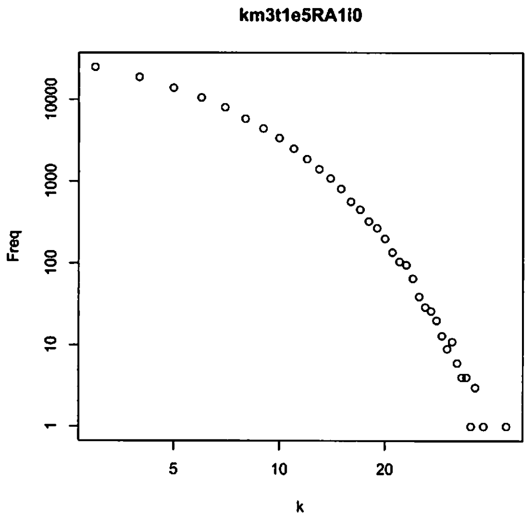


Figure 3: The degree k distribution of a RA network.

3.1.2 Random networks

By replacing the preferential attachment of nodes in BA-BR algorithms by the random attachment, we can construct Random networks. We call this type of network RA network hereafter. Figure 3 shows the distribution of connectivity degree k in a random network.

3.1.3 Lattices

We also use the 2-dimension square lattice with periodic boundary conditions i.e. a torus. An example output of the lattice-making program which makes a 4x4 lattice follows.

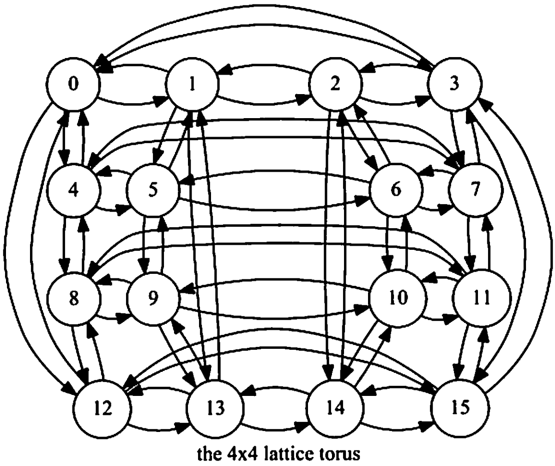


Figure 4: The 4x4 lattice presented as a directed graph. Each node has 4 neibours.

```
$ make-lattice 4 4
;;; make-lattice 1.1
((0 3 1 4 12))
((1 0 2 5 13))
((2 1 3 6 14))
((3 2 0 7 15))
((4 7 5 8 0))
((5 4 6 9 1))
((6 5 7 10 2))
((7 6 4 11 3))
((8 11 9 12 4))
((9 8 10 13 5))
((10 9 11 14 6))
((11 10 8 15 7))
((12 15 13 0 8))
((13 12 14 1 9))
((14 13 15 2 10))
((15 14 12 3 11))
```

Figure 4 is the graph of the 4x4 lattice presented as a directed graph.

3.2 Spreading Simulation of SIS and SIR

Spreading Simulation of SIS and SIR are performed as follows. There are N nodes in a network. Suppose a node labelled i ($0 \leq i < N$). It has a state of S, I or R. If the node labelled i is in the state of S, I or R, we write S_i, I_i and R_i respectively. In each time step, it changes the state following the model of SIS or SIR. As for the SIS model, for each time step, for all infected neibors (directly connected nodes) j,

$$I_j + S_i \xrightarrow{\beta} I_j + I_i \tag{6}$$

and for each time step,

$$I_i \xrightarrow{\mu} S_i \tag{7}$$

where and β and μ are constants.

As for the SIR model, the infection process is the same as that of the SIS model; for each time step, for all infected neighbors (directly connected nodes) j,

$$I_j + S_i \xrightarrow{\beta} I_j + I_i \tag{8}$$

and for each time step,

$$I_i \xrightarrow{\mu} R_i \tag{9}$$

where and β and μ are constants.

3.3 Spreading with Interaction

In our simulations, there are three informations which are named to x_0, x_1 and x_2, respectively. They spread on three different networks, but the nodes are common on those networks. We wrote the state of the node i as $(X_{i,0}, X_{i,1}, X_{i,2})$. $X_{i,s}$ ($s = 0, 1, 2$) having the value of S, I or R. When the node i has the value of S for x_0, R for x_1, and I for x_2, we wrote (S_i, R_i, I_i) and so on.

Information x_2 can be created by the interaction of the information x_0 and information x_1 in each node on the networks, if the node has both information x_0 and x_1. We assumed this interaction occurs with the probability p_{012} in each time step.

If a node has both information 0 and 1, then the node will get information 2 with the probability of p_{012} in a time-step,

$$(I_i, I_i, S_i) \overset{p_{012}}{\rightarrow} (I_i, I_i, I_i) \tag{10}$$

where $(X_{i,0}, X_{i,1}, X_{i,2})$ is the state of node i .

The informations spreads on the different networks according to the SIS model or SIR model. All informations have different parameters (β, μ). We sometimes referred to them as β_i and μ_i respectively for information-i.

We made the programs in order to simulate the spreading processes. The number of informations is not restricted to three. However, all informations propagate under the SIS model or all informations propagate under the SIR model.

4 Results

At first, we presented the results of SIS and SIR simulations on various networks and then presented the results of the interaction simulations. We also presented the interpretation of the results in the context of privacy information spreading.

4.1 Typical SIS time variation

We showed typical SIS time variations $\rho(t)$ on networks. Here, the density of the infected nodes $\rho(t)$ can be calculated by,

$$\rho(t) = \frac{\text{the number of the infected nodes}}{\text{the total number of the nodes}} \tag{11}$$

in each time step. In the context of the privacy spreading, $\rho(t)$ is the privacy information density.

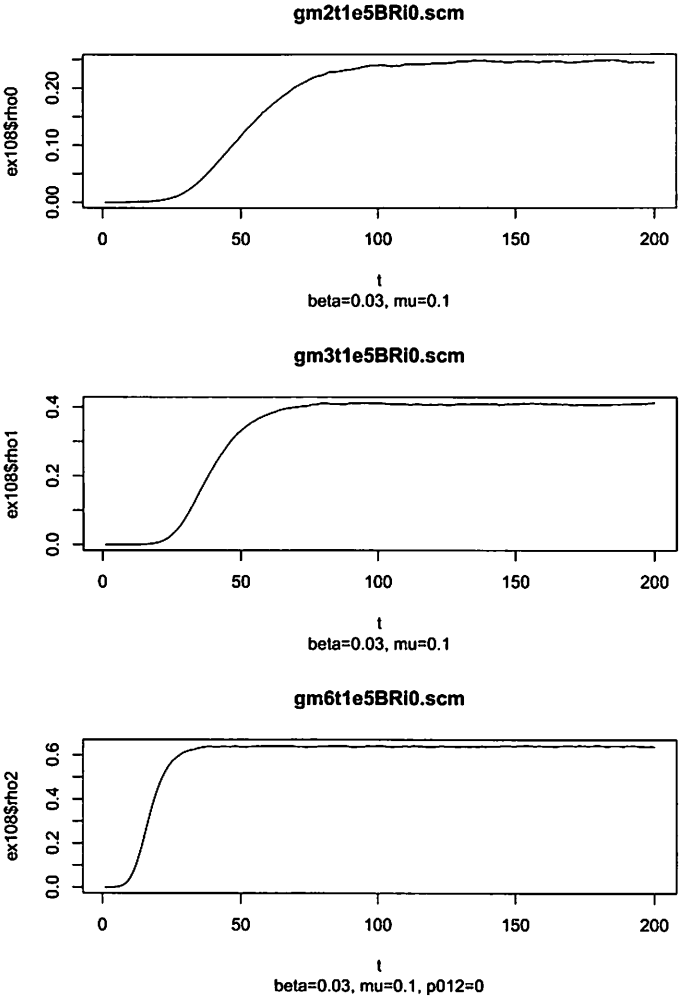


Figure 5: BA-BR networks with $m = 2, 3, 6$. For all simulations, parameters are $\beta = 0.03, \mu = 0.1$ and $p_{012} = 0$. For all panels, the horizontal axis is the simulation time-step and the vertical axis is the number density ρ_i of the nodes which have the information-i $(i = 0, 1, 2)$.

4.1.1 SIS on BA-BR networks with $m = 2, 3, 6$

In figure 5, we plotted the SIS information density time-variation on the BA-BR networks with $m = 2, 3, 6$ where m is a parameter of the network. (Each new node has m edges which points to an old node.) The horizontal axes are the simulation time-steps and the vertical axes are the number densities ρ_i of the nodes which have the information-i ($i = 0, 1, 2$). In this case, we set $p_{012} = 0$, so there was no interaction. All simulations have one initial infected node.

In many case, $\rho(t)$ showed exponential rise at the first stage of our simulations, and then reached metastable state. In some case, however, the epidemic ended when the state $\rho(t) = 0$ occurred at some time.

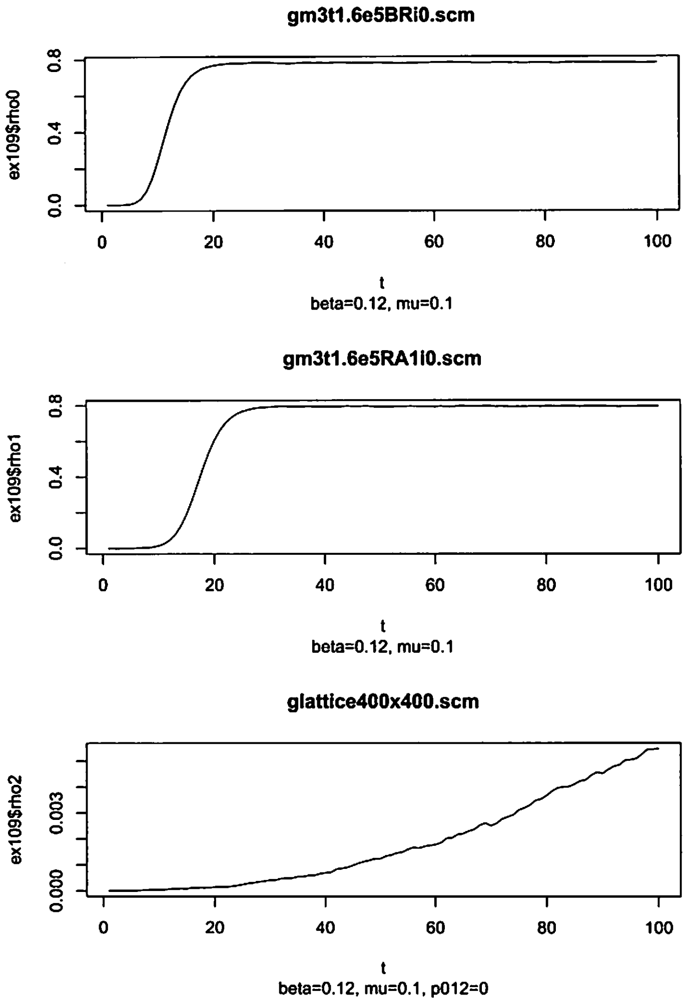
However, ρ in the metastable state increased as m increased and the rising time decreased as m increased. The privacy information spread more rapidly in the networks which has larger m.

4.1.2 SIS interaction on (a BR-BR, a RA1, a lattice)

In figure 6, we plotted the similar simulations with that in figure 5 on a BA-BR network, a RA network and a lattice. We plotted the same simulations in a longer time range in figure 7.

Although the spreading on the BA-BR network was more rapid than that on the RA network, the behaviors were similar to each other. The spreading on lattice was much slower than others. It reached metastable state at a time-step of about 1300 although the others reached it at a time-step of about 30.

The spreading of privacy information was most prominent when it was on the BA-BR and the lattice. Hereafter, we will concentrate on BA-BR networks and lattice; and as for the the BA-BR networks, we will use $m = 3$.



gm3t1.6e5BRi0.scm

gm3t1.6e5RA1i0.scm

glattice400x400.scm

Figure 6: Spreading on a BA-BR network, a RA network and a lattice. For all simulations, parameters are $\beta = 0.12, \mu = 0.1$ and $p_{012} = 0$.

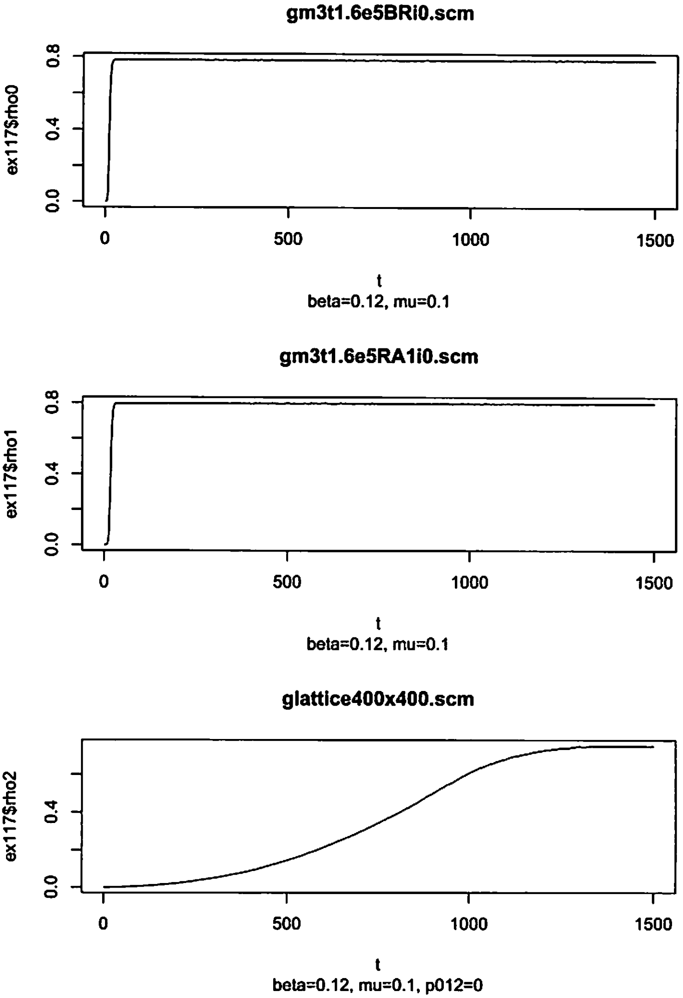


Figure 7: Same plot as figure 6 in longer time range. Spreading on BA-BR network, RA network and lattice. For all simulations, parameters are $\beta = 0.12, \mu = 0.1$ and $p_{012} = 0$.

4.2 Typical SIR time variation

4.2.1 SIR on a BA-BR, a RA1 and a lattice

Figures 8 and 9 show the same simulation with different time scales of the time variations of $\rho_i (i = 0, 1, 2)$ for the SIR spreading on a BA-BR network, a RA1 network and a lattice. The density of the infected nodes $\rho(t)$ showed rapid exponential rise and the following slower decay on both the BA-BR network and RA network. The spreading phenomenon on the lattice was much slower and more localized than those on the BA-BR network or the RA network. Here, the "localized" means

$$\rho_{\text{max, lattice}} << \rho_{\text{max, BA-BR}}, \tag{12}$$

where ρ_{max} is the maximum value of $\rho(t)$.

Thus the information spreads more rapidly on scale-free networks than on lattices. In order to spread on a lattice, large β is needed. Moreover, the time scale of the spreading on BA-BR networks are very weakly dependent on the size of the network, but those on the lattice are heavily dependent on the size.

4.2.2 SIR on lattices with various β

In figure 10, we plotted the SIR time variation of ρ_i with $\beta = 0.105, 0.1, 0.095$ on a lattice. The $\rho_i (i = 0, 1, 2)$ decayed soon under the condition of $\beta < 0.1$.

4.3 Interaction for SIS

We studied the SIS spreading with the interaction on various networks. The variations of ρ_0, ρ_1, and ρ_2 were simulated on different networks. At first, we treated all networks as scale-free.

4.3.1 Spreading on BA-BR networks (BA-BR, BA-BR, BA-BR)

In figure 11, we plotted the time variation of the density of the

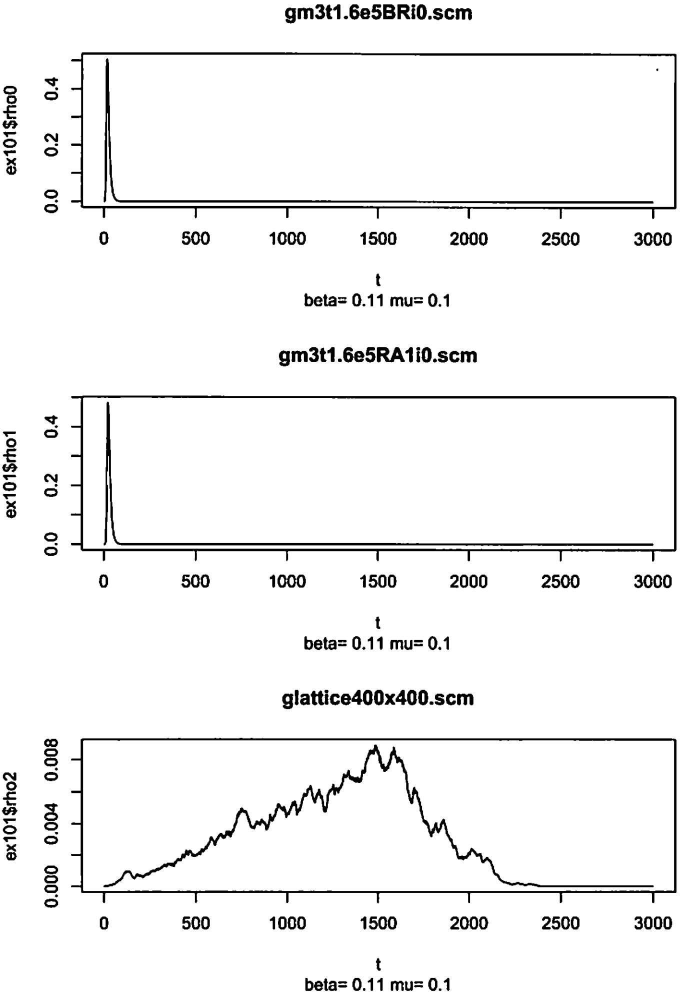


Figure 8: SIR Time variation of $\rho_i(i = 0, 1, 2)$ on various networks.

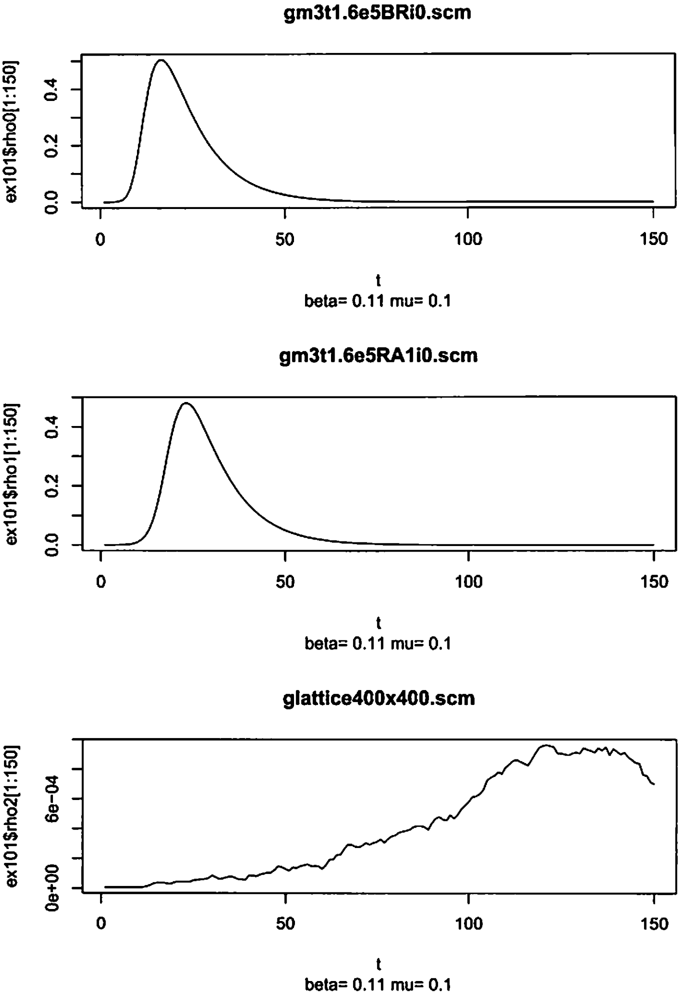


Figure 9: SIR Time variation of $\rho_i(i = 0, 1, 2)$ on various networks. This plot is the same as figure 8 but in a narrow time range.

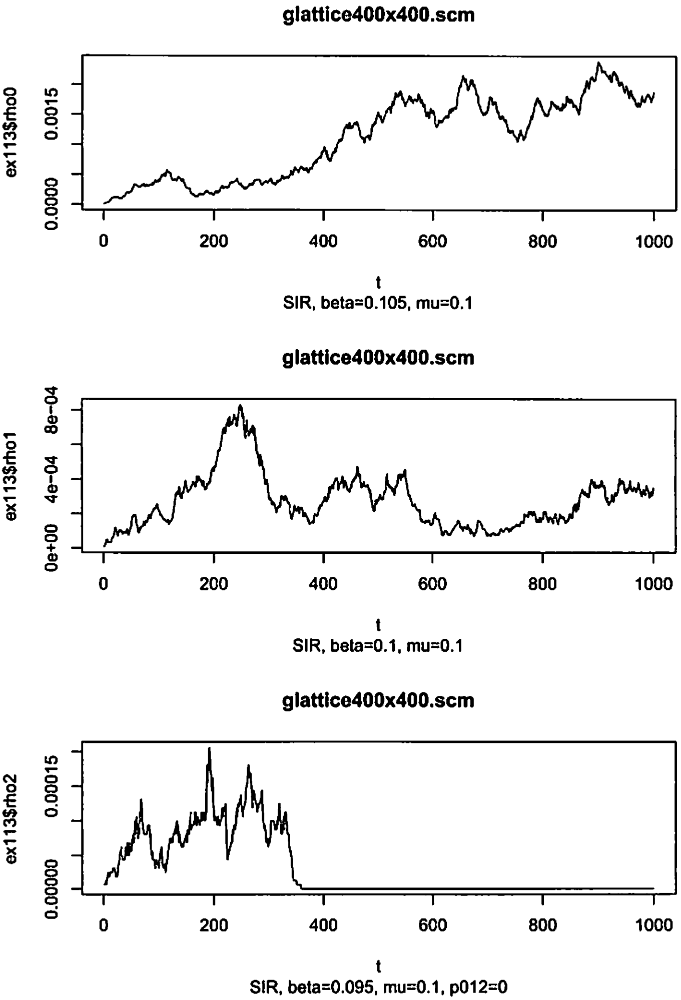


Figure 10: SIR time variation of $\rho_i(i = 0, 1, 2)$ on a lattice. The upper panel is ρ_0 with $\beta = 0.105$. The middle panel is ρ_1 with $\beta = 0.1$. The upper panel is ρ_2 with $\beta = 0.095$.

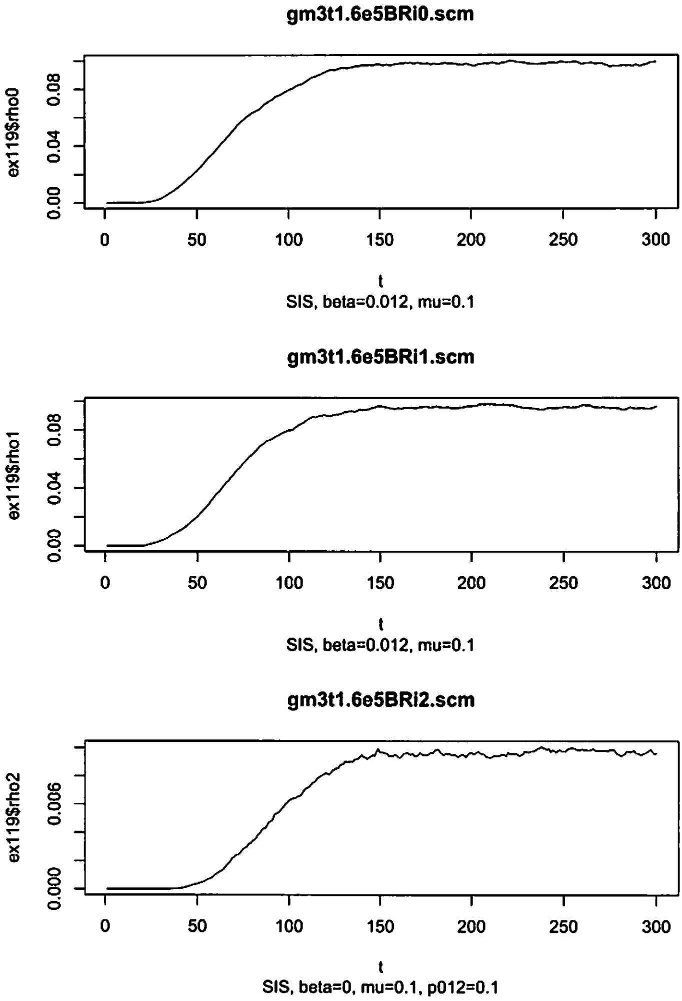


Figure 11: SIS spreading on BA-BR networks and the density time variation of the created information (upper and middle panels). The lower panel shows ρ_2 which is created by the interaction of information-0 and information-1. In the lower panel, $\beta = 0$, so the network is a dummy.

information-0 ρ_0 on the upper panel, the density of the information-1 ρ_1 on the middle panel and the created information-2 ρ_2 on the lower panel. ρ_0, ρ_1 and ρ_2 are on the BA-BR networks. The BA-BR networks were not same but different networks. As shown in the figure, SIS parameter $\mu = 0.1$ for all variation of $\rho_i (i = 0, 1, 2)$. $\beta = 0.012$ for ρ_0, $\beta = 0.012$ for ρ_1 and $\beta = 0$ for ρ_2. We set the information creation probability $p_{012} = 0.1$. ρ_0 and ρ_1 showed typical SIS time variation on the scale-free network. ρ_2 showed similar variation profile to that of ρ_0 and ρ_1. However, information-2 cannot spread by itself because $\beta = 0$.

In this case, the information-2 was created only by the interaction of the information-0 and information-1. The time variation of ρ_2 represented the time variation of the density of nodes which have both information-0 and information-1.

In figure 12, we plotted again the similar time variation to that of figure 11. However, in this case, β of information-2 was not zero. However, the β was very small. The ρ_2 of information-2 was larger than that of figure 6. It was difficult for the information-2 to survive long with the β. However, it survived in this case because the seeds created by the interaction were supplied continuously.

Thus if the information-2 is created by the interaction of information-0 and information-1, the spreading of it is more rapid and broader than without the creation. If β of information-2 is as large as that of information-0 or information-1 (we did not show the plot), the role of the interaction is only a supplier of some initial seeds of SIS spreading.

4.3.2 SIS interaction on (BA-BR, lattice, lattice)

In figure 13, we plotted the time variation of the density of the information-0 ρ_0 on the upper panel, the density of the information-1 ρ_1 on the middle panel and the created information-2 ρ_2 on the lower panel. ρ_0 is on the BA-BR network and ρ_1 and ρ_2 are on the lattice. As shown in the figure, SIS parameter $mu = 0.1$ for all variations of ρ. $\beta = 0.01$ for ρ_0, $\beta = 0.2$ for ρ_1 and $\beta = 0$ for ρ_2. We then set the information creation probability $p_{012} = 0.1$. ρ_0

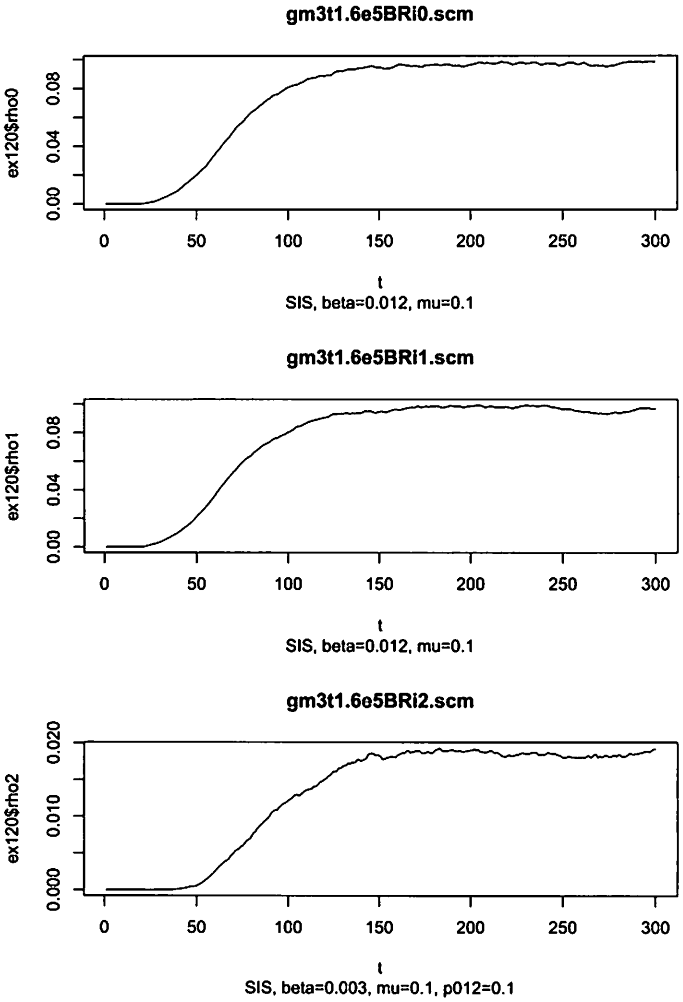


Figure 12: SIS spreading on BA-BR networks and the density time variation of the created information (upper and middle panels). The lower panel shows ρ_2 which is created by the interaction of information-0 and information-1.

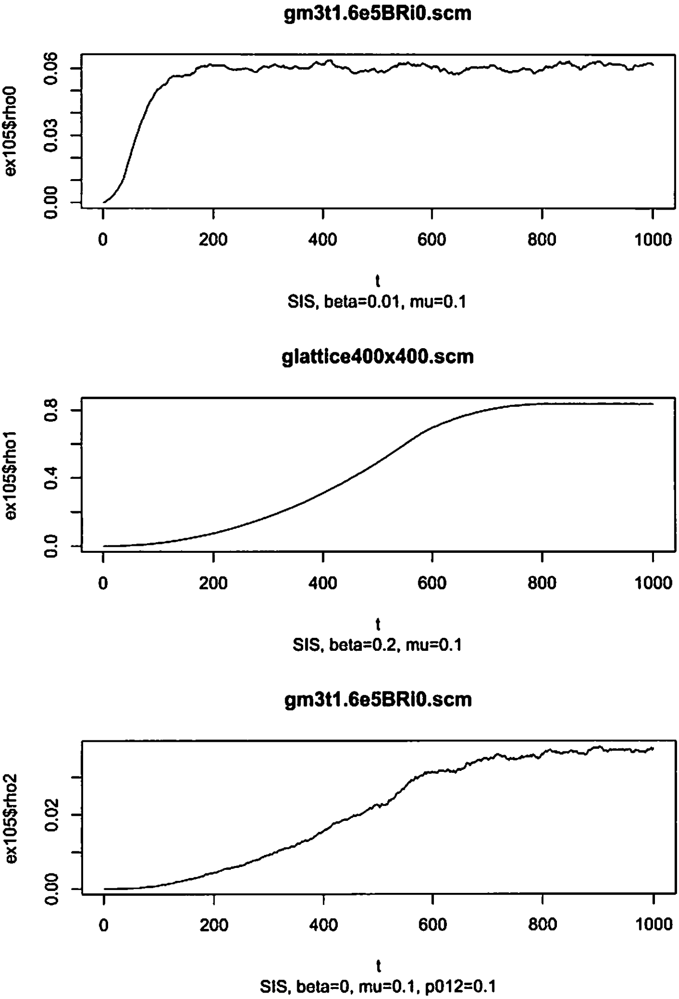


Figure 13: SIS spreading on networks and the density time variation of the created information. CAUTION: the network of the lower panel is meaningless because $\beta = 0$.

showed a typical SIS time variation on the scale-free network. ρ_1 showed a much slower rise. ρ_2 showed a similar variation profile to that of ρ_1.

Information spreading on the BA-BR network reached a metastable state much earlier than that on the lattice, so the profile of ρ_2 is similar to that of ρ_1.

In figure 14, the authors again plotted the similar time variations to that of figure 13. In this case, β of ρ_2 is not 0 but 0.1. The time variation of ρ_2 has shape which is a mixture of those of ρ_0 and ρ_1. At first time (time-step 50 – 100), ρ_2 showed a rapid increase and after that (time-step 150 – 250), it showed a slight increase and (time-step 250 – 400) then it increased again. The first rise was caused by the spreading on the BA-BR network after the creation of information/disease2 in a node on the network. It was followed by a metastable state of SIS. However, it was followed by the increase caused by the increase of ρ_1. At last (time-step 700−), it showed a metastable state.

ρ_2 of the metastable state was larger than ρ_0 and was larger than ρ_2 in figure 13. In this case, the interaction and the creation of the information made the spreading of the information broader.

4.3.3 SIS interaction on (BA-BR, BA-BR, lattice)

In figure 15, we plotted the time variation of the density of the information-0 ρ_0 on the upper panel, the density of the information-1 ρ_1 on the middle panel and the created information-2 ρ_2 on the lower panel. ρ_0 and ρ_1 are on the BA-BR networks. ρ_2 spreads with $\beta = 0.12$ on a lattice.

In this case, the spreading of information-2 on the lattice was much faster than that in figure 7. The spreading is supposed to be mainly caused by the interaction and creation. However, the metastable ρ_2 was similar in value to that in figure 7.

4.4 SIR cases

We also simulated SIR spreading with interactions.

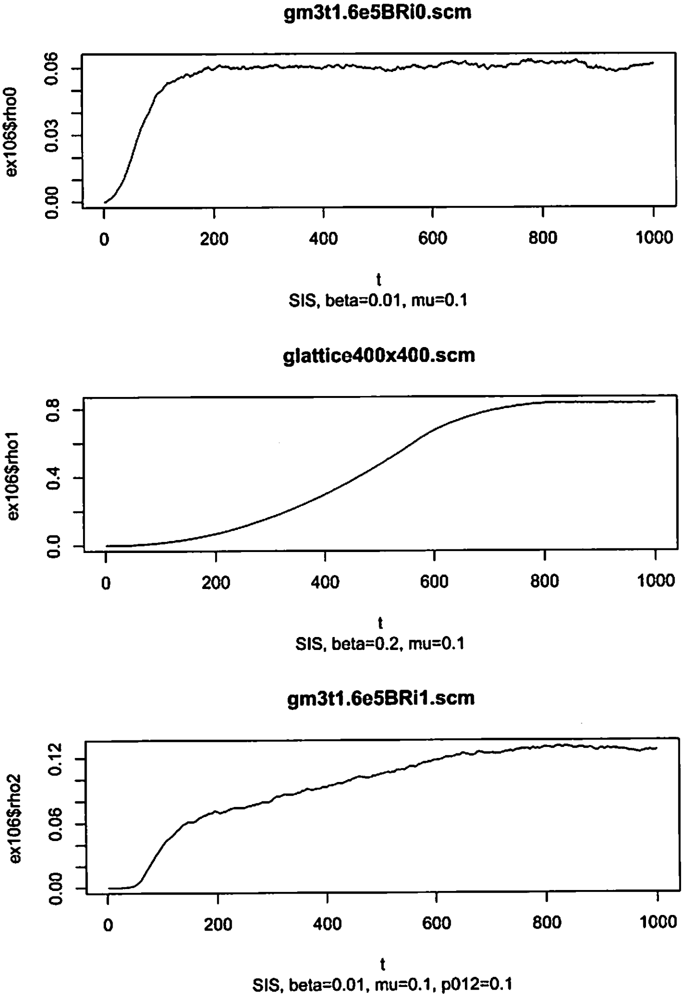


Figure 14: SIS spreading on networks and the density time variation of the created information.

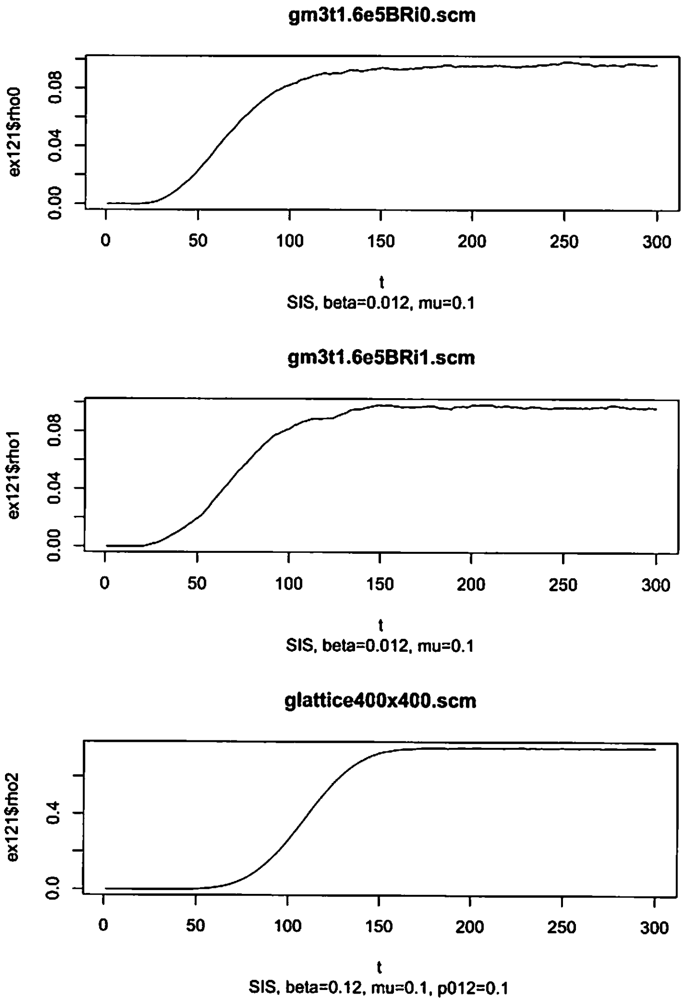


Figure 15: SIS spreading on networks and the density time variation of the created information.

4.4.1 SIR interaction on (BA-BR, BA-BR, BA-BR)

In figure 16, we plotted the time variations of ρ in a SIR case. The spreading on the BA-BR networks (the top and middle panels) showed a typical rapid rise and decay of ρ_0 or ρ_1. On the bottom panel, the time variation of ρ_2 is shown. In this case, the information 2 has little ability of spreading by itself ($\beta = 0.0012$. The profile of the time variation looks like a product of ρ_0 and ρ_1 but the maximum value was much smaller.

4.4.2 SIR interaction on (BA-BR, lattice, BA-BR)

In figure 17, we plotted the time variations of ρ in a SIR case. The spreading on the BA-BR network (the top panel) shows a typical rapid rise and decay of ρ_0 and on the lattice (the middle panel) shows the monotonic rise of ρ_1 in this time scale. On the bottom panel, the time variation of ρ_2 is shown. In this case, the information 2 had no ability of spreading by itself ($\beta = 0$. The profile of the time variation looks like a product of ρ_0 and ρ_1.

In figure 18, the authors plotted again the time variations of ρ in a SIR case. In this case, information-2 was created when both information-0 and information-1 exist, and it spread mainly by itself afterwards. It showed a typical SIR time variation. The maximum of ρ_2 in figure 18 was much larger than that of ρ_2 in figure 17. However, the maximum of ρ_2 was comparable to that of ρ_0, and the profile of ρ_2 was similar to that of ρ_0, so the spreading of information-2 should be the usual SIR spreading.

4.4.3 SIR interaction on (BA-BR, BA-BR, lattice)

In figure 19, we plotted time variations of ρ in different networks. In this case, information-0 spreads with $\beta = 0.02$ and $\mu = 0.1$ on a BA-BR network information-1 spread with $\beta = 0.01$ and $\mu = 0.1$ on another BA-BR network, and information-2 was created with the probability $p_{012} = 0.1$ and had the ability of spreading with $\beta = 0.01$ and $\mu = 0.1$. The information-2 spread on a lattice.

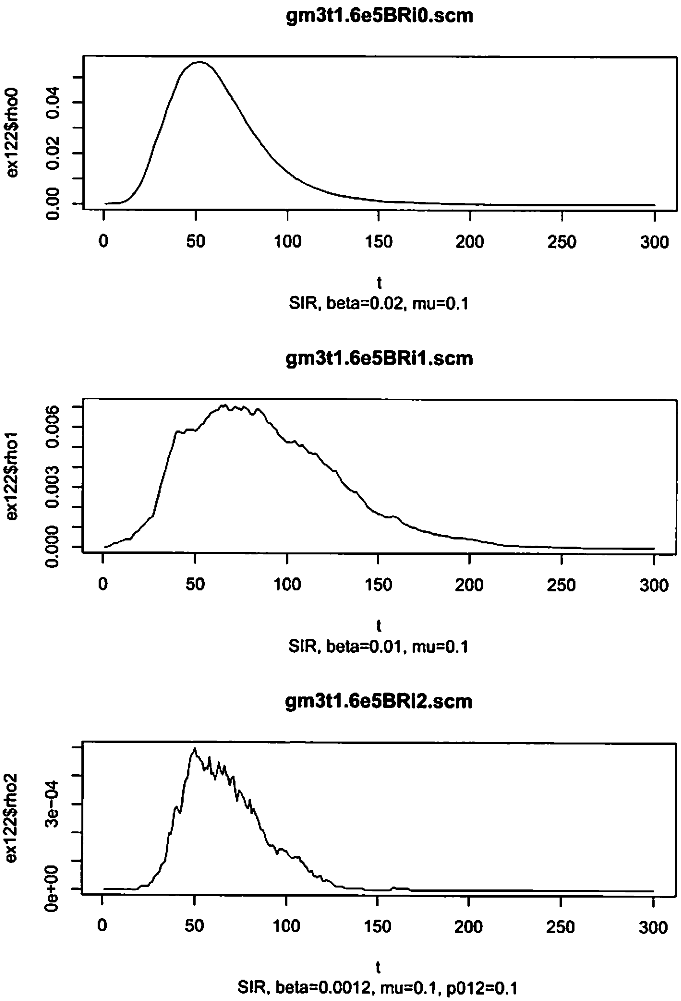


Figure 16: A case of 2 SIR-informations interaction.

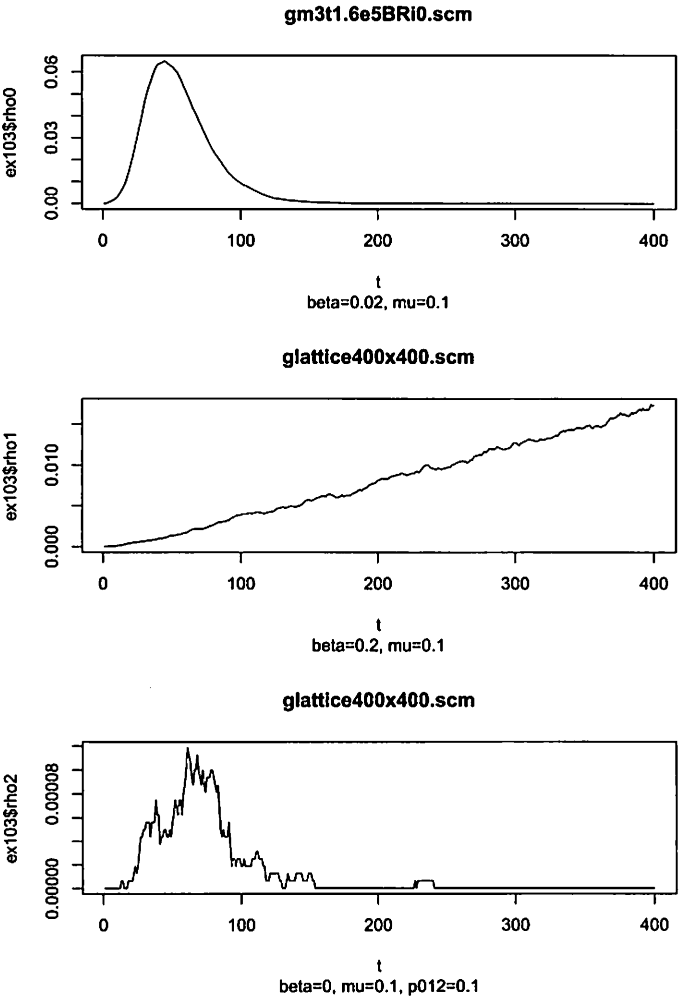


Figure 17: A case of 2 SIR-informations interaction.

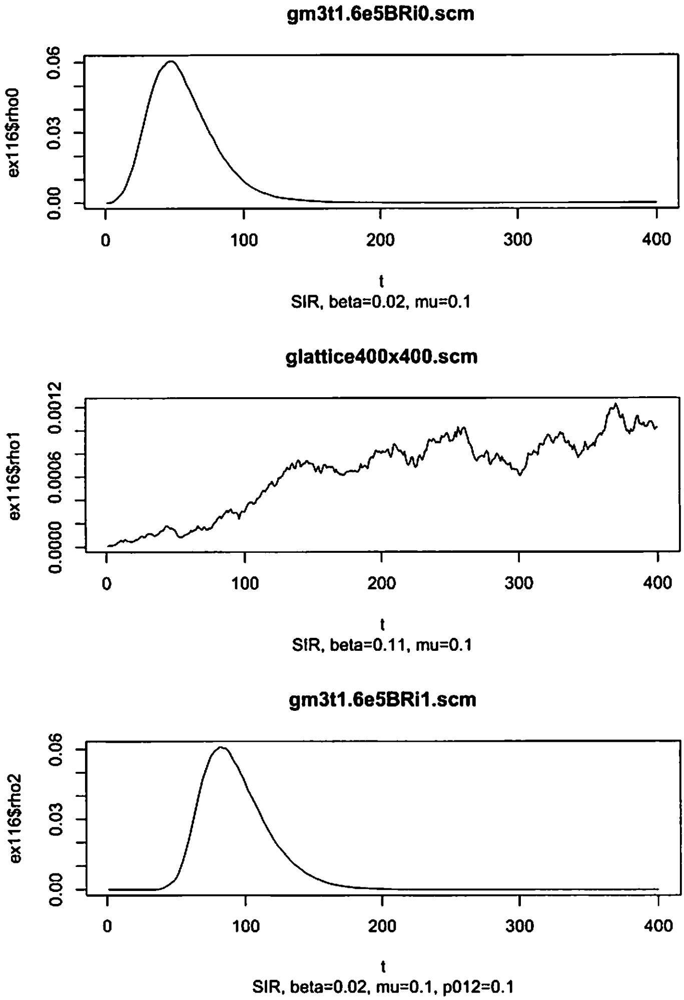


Figure 18: A case of 2 SIR-informations interaction. The ρ in the lower panel has typical SIR time variation.

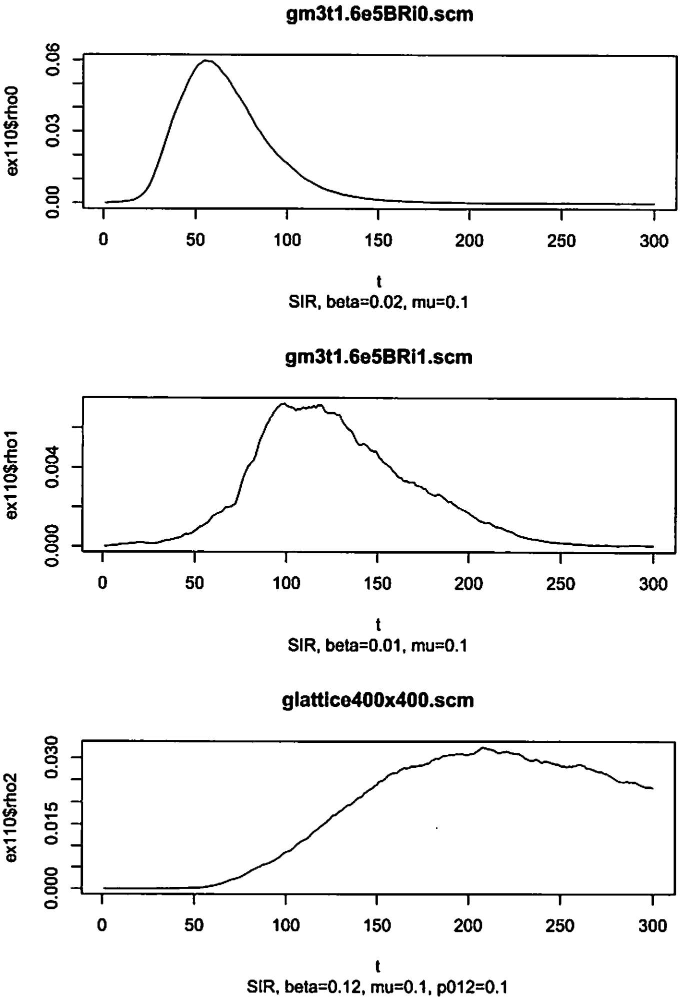


Figure 19: Another case of 2 SIR-informations interaction.

In this simulation, information-0 showed a rapid rise and followed decay. Information-1 had similar a profile to that of information-0 but the rise was delayed. Information-2 was created when both information-0 and information-1 existed, and it spread by itself afterwards.

In figure 20, we again plotted a similar case to figure 19. However, β of information-2 is low in this case. Therefore, the profile of ρ_2 was similar to that of ρ_2 in figure 19, but the value was lower.

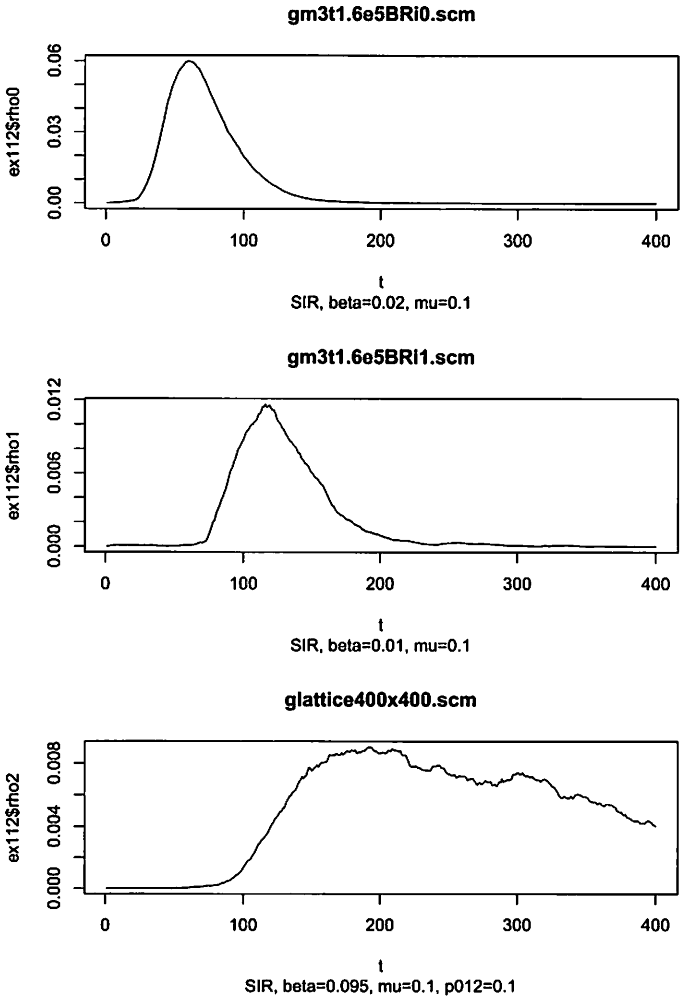
However, the peak value of the profile of ρ_2 was much larger than in cases without creation (figure 10) and the duration time was longer. In the cases of spreading on the lattices with $\beta < 0.1$, the spreading tends to decay in the early phase. In this case, the spreading must survive owing to the the interaction and the creation of the seeds of information and must have a larger expansion than in cases without creation.

5 Discussion and Conclusion

Privacy information is guarded in many ways. However, when it leaks, it spreads by itself in many ways. And if someone gathers the data (or information) of a person, one could create new information about the person using the data.

In order to apply our results to a real-world situation, we showed what the networks and the spreading models corresponded to. If the leak of the privacy information is inhibited by law or technologies, the spreading of it corresponds to low β. The value of β depends on the strictness of the law or the completeness of the technology. If the information spreads through the Internet, it spreads on the scale-free network. If the information spreads by word of mouth, it spreads on the lattice. If there is a law to erase improper information, the information propagates according to the SIR model, otherwise it propagates according to the SIS model.

The SIS spreading on the BA-BR networks with various ms shows that the spreading is more rapid in the network which has


gm3t1.6e5BRi0.scm

SIR, beta=0.02, mu=0.1

gm3t1.6e5BRi1.scm

SIR, beta=0.01, mu=0.1

glattice400x400.scm

SIR, beta=0.095, mu=0.1, p012=0.1

Figure 20: A 2 SIR-informations interaction. The β of information 2 is low.

a larger m. This means that the privacy spreading is more rapid on a network in which each node has more connections.

A node can be thought of as a person or an organization. A person forgets the information and an organization sometimes loses data or may have a rule to abandon the data, i.e. the information has a lifetime. In this case, the spreading is well modeled by the SIR model if the lifetime of the data is shorter than the time scale of the rising, i.e. the typical outbreak time.

The typical SIS time variations show that if the information is opened in the Internet, the spreading of it is very rapid and in order to make the effect lower, β should be lower, i.e. we need strict laws or technologies. The typical SIR time variations show that if the lifetime of the information is short, the effect of privacy leakage is limited.

The SIS interactions on BA-BR networks shows that if the seeds of privacy information spreads in the Internet or other scale-free networks, the privacy information is created somewhere and spreads widely. Even if the creation is difficult in our simulation (small p_{012}), it spread rapidly after the creation.

The SIS interactions on BA-BR networks and lattice shows the situation in which some of the information spreads through word-of-mouth communication. It may be the situation in which the leak of the information is forbidden, so the spreading is very slow. The case on the section 4.3.3 (BA-BR, BA-BR, lattice) shows that even if the leakage of the privacy is forbidden, if the seed information of privacy spreads on scale-free networks, the privacy information spreads rapidly.

The SIR interactions simulation shows the situation in which the person forgets the information or the organization makes the privacy data to have a time-to-live (TTL), i.e. the organization sets the life time to the data and deletes them at that time. The SIR interaction on the section 4.4.2 (BA-BR, lattice, BA-BR) shows that if β is large, the privacy information spreads rapidly and decays after the creation even if some of the seed data is difficult to spread. If β is small, the created privacy information is a product of seed ρ_0 and ρ_1. Moreover, if the rising times of ρ_0 and ρ_1 are different enough with each other, the privacy information creation is vary rare. The SIR interaction on the section 4.4.3 (BA-BR, BA-BR, lattice) shows that (comparing figure 20 and figure 10) if there is the creation of privacy information, the privacy information spreads more widely and for a longer time because many seeds are supplied.

The interactions for SIS and SIR are enabled using the "Big Data". The interactions for SIS and SIR show that if one is able to create the privacy information using known information, the spreading of it is easier than without the creation.

The ability to create information itself may be information. Sometimes the information may be hidden from public access. In this case, it spreads on a lattice. However, the interactions for SIR on the lattice show that it is not a complete method.

In conclusion, the creation of privacy information using well-known and less-known data is effective for privacy spreading. In the era of "Big Data", we need more strict laws or technologies in order to inhibit the spreading of the privacy information. To set Time-to-live (TTL) to the privacy data by a technology or a law may be effective. Or if we are able to forget, it may be good for the privacy.

Acknowledgments

This paper originates from the results of a sabbatical stay of the author A.O. at the Inria Sophia-Antipolis from August 2013 to August 2014. It was funded by the International University of Kagoshima; Tsumagari Gakuen. The author A.O. gratefully acknowledges this sabbatical stay and expresses his gratitude to Manual Serrano and members of the Indes team at Inria Sophia-Antipolis for their kind support during the sabbatical stay. This research has made use of NASA's Astrophysics Data System. The simulation programs are made using a Bigloo compiler. The authors thank Dave Timson for reading the manuscript.

この論文は、著者の一人、奥平敦也が、津曲学園鹿児島国際大学からの留学助成資金により、2013年8月から、2014年8月まで、フランス国ソフィア・アンティポリス市フランス国立情報学自動制御研究所ソフィア・アンティポリス (Inria Sophia-Antipolis) に留学した成果である。奥平は津曲学園および Inria Sophia-Antipolis と受け入れていただいたマニュエル・セラノ (Manuel Serrano) さん他 Indes チームのみなさんに感謝いたします。

References

[1] N. I. Adams IV, D. H. Bartley, G. Brooks, R. K. Dybvig, D. P. Friedman, R. Halstead, C. Hanson, C. T. Haynes, E. Kohlbecker, D. Oxley, et al. Revised^5 report on the algorithmic language scheme. *ACM Sigplan Notices*, 33(9):26–76, 1998.

[2] O. Angiuli, J. Blitzstein, and J. Waldo. How to de-identify your data. *Communications of the ACM*, 58(12):48–55, 2015.

[3] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, Oct. 1999.

[4] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Oct. 2008.

[5] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structure and Algorithms*, 18:279–290, 2004.

[6] J. Feigenbaum and B. Ford. Seeking anonymity in an internet panopticon. *Commun. ACM*, 58(10):58–69, Sept. 2015.

[7] T. Geller. In privacy law, it's the U.S. vs. the world. *Commun. ACM*, 59(2):21–23, Jan. 2016.

[8] F. Koufogiannis and G. Pappas. Diffusing Private Data over Networks. *ArXiv e-prints*, Nov. 2015.

[9] C. Landwehr. Privacy research directions. *Commun. ACM*, 59(2):29–31, Jan. 2016.

[10] K. Misata, R. A. Hansen, and B. Yang. A taxonomy of privacy-protecting tools to browse the world wide web. In *Proceedings of the 3rd Annual Conference on Research in Information Technology*, RIIT '14, pages 63–68, New York, NY, USA, 2014. ACM.

[11] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87:925–979, July 2015.

[12] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, June 2001.

[13] M. Serrano and P. Weis. Bigloo: A portable and optimizing compiler for strict functional languages. In *Proceedings of the Second International Symposium on Static Analysis*, pages 366–381. Springer-Verlag, 1995.

[14] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the Internet. *Physical Review E*, 65(6):066130, June 2002.

(received 30 March 2016)
(revised 1 September 2016)