PAPER

Information Distribution Analysis Based on Human's Behavior State Model and the Small-World Network

Pao SRIPRASERTSUK[†a)], *Nonmember and* Wataru KAMEYAMA[††], *Member*

SUMMARY In this paper, an information distribution model based on human's behavior is proposed. We also propose dynamic parameters to make the model more practical for real life social network. Subsequently, the simulations are conducted based on the small-world network and its characteristics, and the parameters in the model are analyzed to increase efficiently the power of information distribution. Our study and simulation results show that the proposed model can be used to analyze and predict the effectiveness of information distribution. Moreover, the study also shows how to use the model parameters to increase power of the distribution.
key words: *information distribution, primary and secondary information distribution, small-world network*

1. Introduction

In general, there are two kinds of information flows in an information distribution such as advertisement distribution: one is the primary information distribution, and the other is the secondary information distribution. The primary information distribution is the distribution done by providers or broadcasters to consumers through certain kinds of media such as television, newspaper, etc. The secondary information distribution is the distribution done by users to users such as word-of-mouth. Today, the advanced information technologies enable the secondary information distribution to perform by various methods and through media. This type of distribution has been playing an important role in our society and we cannot ignore its power anymore. Therefore, it is necessary to analyze the power of secondary information distribution.

We have investigated how the secondary information distribution affects the information circulation based on the statistical observation [1]. Our experiment in [1] shows that the model can present a part of the real world information distribution, and analyze its effectiveness. However, the proposed model does not consider the human network structure but only the statistical aspects of information distribution being consumed.

On the other hands, there are researches about information distribution taking into account the structure of the human network structure, that is the small world network. The small-world network have been very popular in the scientific community because of its ability to model many complex systems in our nature and man-made networks [2]. The World Wide Web [3], food web [4], scientific collaboration networks [5], electronic circuits [6] and even human languages [7] are the examples of these systems.

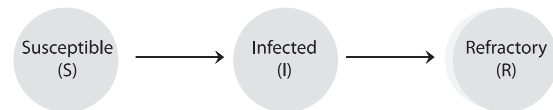


Fig. 1 The SIR model.

The small-world network model has been widely applied in various research domains. One of the popular domain is epidemics, particularly disease propagation in the human network. An epidemic model popularly used in such researches is the SIR(Susceptible, Infected and Refractory) model proposed by W.O. Kermack and A.G. Mckendrick [11]. The SIR model is composed of three states, Susceptible, Infected and Refractory states, and a node in the network can be in one of them as shown in Fig. 1.

There are researches [8]–[10] using the SIR model to analyze the effectiveness of information distribution. According to the nature of human's behavior, we consider the SIR model is not suitable to represent the information propagation cycle. Therefore, in order to analyze and control the power of secondary information distribution efficiently, we propose a new information distribution model and equations considering the human's behavior. Subsequently, the proposed model is simulated and analyzed by applying the small-world network and it characteristics.

This paper is organized as follows: in Sect. 2, some problems of the SIR model are described. We explain the proposed model and equations in Sect. 3. Sections 4 and 5 describe the simulation results, and finally Sect. 6 shows our conclusion and future works.

2. Problems of SIR Model

The SIR model has been widely used in many research areas such as computer viruses [12], information propagation [8]–[10] and so forth [13], [14]. For example, [8]–[10] use the SIR model for their proposals and simulation by considering "Susceptible" as individuals who have not heard the information and they are susceptible to be informed, "Infected" as individuals who are spreading the information, and "Re-

Manuscript received June 11, 2008.
Manuscript revised December 12, 2008.
[†]The author is with the Global Information and Telecommunication Institute, Waseda University, Tokyo, 169–0051 Japan.
[††]The author is with the Graduate School of Global Information and Telecommunication Studies, Waseda University, Tokyo, 169–0051 Japan.
a) E-mail: pao@akane.waseda.jp
DOI: 10.1587/transinf.E92.D.608

fractory" as individuals who know the information but no longer spreading it. However, using the SIR model is not suitable because it concerns but not fully the human behavior of information distribution. For instance, after individuals get the information, they may not spread the information or they may forget it and become susceptible again. Even they become information spreaders, they may forget the information, and they reverse to be susceptible in the future time. Moreover, infected nodes do not always spread the information to neighbors as mentioned in the SIR model in [8]–[10]. These points show the limitation and constraint of the SIR model to represent the real world information propagation by human.

Furthermore, the result from [8]–[10] applying the SIR model shows that all of information propagation reaches almost all individuals when time passes. In other words, any information is passed on to almost all nodes in the network. However, this kind of phenomenon is rare in the real world information distribution .

Another problem of [8]–[10] is that there are some cases in practical information distributions such as advertisement distribution where the primary information distribution is also done during the secondary information distribution. But their model and algorithm cannot explain and analyze this kind of distribution because they only concern the secondary information distribution.

Hence, in this paper we propose a model and equations that can be used for analyzing practicle information distribution as described in Sect. 3.

3. Proposed Information Distribution Model

3.1 Model Overview

With regard to the human's behavior state for information distribution, we consider that there are three states that are "Unknown", "Known" and "Distribute". The proposed model is shown in Fig. 2, and the definitions are described as below.

- Unknown State (U)
  Individuals in this state do not know information. They either do not receive information yet or they forget it. In the case of forgetting information, individuals transit from Known State to this state.
- Known State (K)
  In this state, individuals know information but do not have any action to the information distribution.

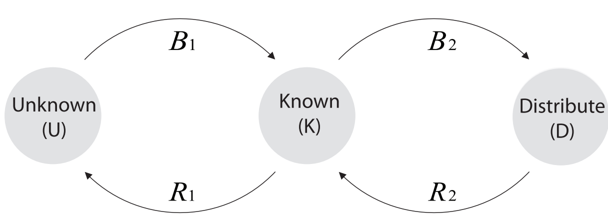


Fig. 2 The proposed information distribution model.

- Distribute State (D)
  Individuals in this state are active to distribute information. The distribution by their own intentions and other individuals requests are considered as the same.
- Probability of Becoming Known State(B_1)
  This parameter shows the probability of individuals in Unknown State to change into Known State. For example, some individuals in Unknown State are informed with the information by individuals in Distribute State.
- Probability of Becoming Distribute State (B_2)
  This parameter shows the probability of individuals in Known State to change into Distribute State. For example, if individuals in Know State start to distribute the information, they move from Known State to Distribute State.
- Probability of Returning to Unknown State (R_1)
  This parameter shows the probability of individuals in Known State to change into Unknown State. For example, they forget the information because they are not interested or the information becomes stale after time passes.
- Probability of Returning to Known State (R_2)
  This parameter shows the probability of individuals in Distribute State to change into Known State. For example, after individuals distributed the information, they may change their mind to stop the action. Thus, their state is changed to Known State.

With regards to Fig. 2 and the above definitions, the notations with time series are defined as below.

- N : the number of all individuals in the network
- $U(t)$: the number of Unknown State individuals in N at time t
- $K(t)$: the number of Known State individuals in N at time t
- $D(t)$: the number of Distribute State individuals in N at time t

Thus, $U(t) + K(t) + D(t) = N$.

3.2 The Process of Transiting to Known State from Unknown State

New individuals get information by being in contact with individuals in Distribute State. We assume that each individual in Distribute State contacts k neighbors in each period, and the probability of successful distribution of each individual in Distribute State is B_1. Therefore, the probability of successful distribution for individuals in Unknown State depends on numbers of neighbors in Distribute State. This probability is defined as Eq. (1), where G_i is successful distribution probability of individual i and n is the number of neighbors of it in Distribute State. The transition process from Unknown State to Known State with $B_1 = 0.5$ is illustrated in Fig. 3. In Fig. 3, the probability of successful distribution of "Unknown (11)" is 0.875 because there are three neighbors in Distribute State, while this probability of

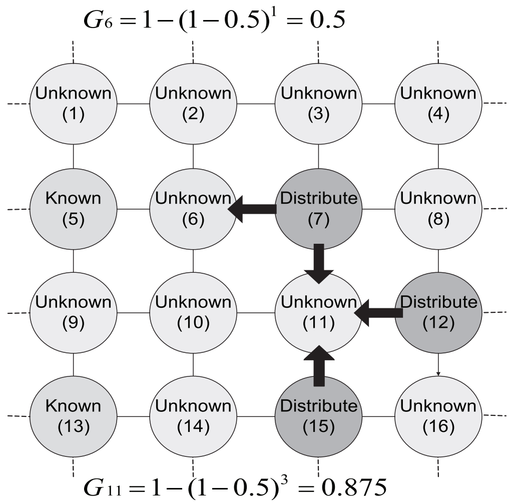


Fig. 3 The process of transiting to known state from unknown state with $B_1 = 0.5$.

"Unknown (6)" is 0.5 because there is only one neighbor in Distribute State. If the distribution is successful, Unknown State of the informed node changes to Known State. Note that this figure shows an illustration of the process on a regular lattice model. However, the actual simulation is done on a small-world network.

$$G_i = 1 - (1 - B_1)^n \tag{1}$$

3.3 Dynamism in R_1

R_1 shows the probability of forgetting information. We apply the forgetting curve theory of Hermann Ebbinghaus [15]. According to [15], the human forgetting formula is described by $R = e^{-(\frac{t}{S})}$ where R is memory retention, S is the relative strength of memory, and t is time. As discussed in Sect. 3.2, new individuals might transit from UnKnown and Distribute State to Known State, and individuals might leave Known State any time. We consider that, when those new individuals transit to Known State, they start to forget the information with the probability of R_1. Thus, $R_1(t)_i$ for individual i is defined in Eq. (2) where t_p is time when individual i transit to Know State. An example graph of R_1 is shown in Fig. 4 where $t_p = 0$ and $S = 100$.

$$R_1(t)_i = 1 - e^{-(\frac{t-t_p}{S})} \tag{2}$$

3.4 Dynamism in R_2 and B_2

We also consider B_2 and R_2 as dynamic parameters in this paper because their values depend on motivation and feeling of individuals to information itself. For instance, individuals might be motivated to distribute information in the beginning of the distribution but their intention might diminish when time passes. So, we consider the information distribution life cycle is similar to the Product Life Cycle

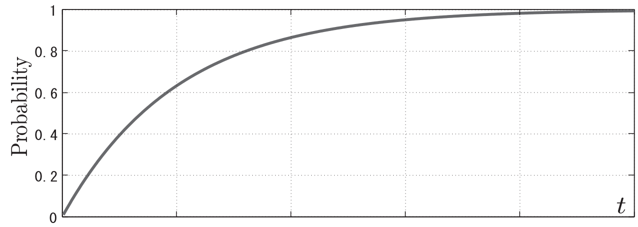


Fig. 4 An example graph of R_1 with $t_p = 0$ and $S = 100$.

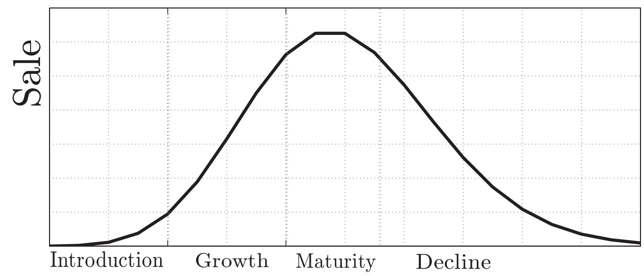


Fig. 5 The product life cycle (PLC) model [16].

(PLC) [16] in the marketing study area, because the PLC concept is based on consumers and manufacturers behaviors, and the social psychology. In addition, information is a product that is designed with a purpose in mind, while data serves as the ingredient in this product [17]. The model of the PLC is similar to the bell curve as shown in Fig. 5. In the PLC, there are 4 states that are Introduction, Growth, Maturity and Decline. The length of each state varies enormously because it depends on the product itself.

B_2 , i.e. motivation or incentive to distribute information, increases in the Introduction state where the motivation is created and the individuals in Distribute State start to distribute the information. In Growth state, the motivation and distribution are growing significantly along with public awareness. After Growth state, the motivation to distribute information is still growing for a short period of time and starts to decline because information becomes stale, and it happens in Maturity state. Finally, in Decline state, motivation decreases continuously or it becomes stagnant.

However, for R_2, i.e. when individuals have no motivation to re-distribute information after having distributed it, R_2 might not have Decline state, because lack of motivation of distributors is difficult to decline if there is no any stimulation.

Nevertheless, human intention, motivation and feeling of information distribution are so varied due to information itself. Such variations are able to be represented in many kinds of graph such as asymmetric bell curve. Hence, we propose Eq. (3) to represent $B_2(t)$ and $R_2(t)$. Eq. (3) is flexible to produce various kind of graphs by changing μ, λ_{up}, λ_{down} and β. Changing μ, λ_{up}, λ_{down} and β means changing time when the graph reach maximum value, changing graph scale when $t \leq \mu$, changing graph scale when $t > \mu$, and changing maximum value of y-axis, respectively. Some examples of graphs produced by Eq. (3) is shown in Fig. 6.

$$B_2(t), R_2(t) = \begin{cases} \dfrac{e^{-(t-\mu)^2 \times \lambda_{up}^2}}{\beta} & (t \leq \mu) \\ \dfrac{e^{-(t-\mu)^2 \times \lambda_{down}^2}}{\beta} & (t > \mu) \end{cases} \tag{3}$$

As a note, $R_1(t)_i$ of individual i varies depending on a point of time(t_p) when each individual transit from Unknown State or Distribute State to Known State. On the other hand, we assume that $B_2(t)$ and $R_2(t)$ depend on the information life cycle but not depend on individuals' state transition like $R_1(t)_i$. In other words, $B_2(t)$ and $R_2(t)$ have the fixed curves independent from individuals' state transition. It means an individual who get the information at early time tends to distribute it to others with high probability, but an individual who get the information at later time tends not to be interested information and to distribute it to others with low probability because of the oldness of the information. Therefore, $B_2(t)$ and $R_2(t)$ for all individuals at time t are same and fixed as given by Eq. (3).

3.5 Primary Information Distribution

Occasionally the primary information distribution is also done during the secondary information distribution. Therefore, we define the primary information distribution parameter $P(t)$ as shown in Eq. (4). In this equation, c and $tprimary$ are number of individuals informed by primary information distribution and time when primary information is distributed, respectively. For instance, if $c = 10,000$ and $tprimary = 200$, it means 10,000 individuals in Unknown state are randomly selected, and transit to Known state when $t = 200$ as they are deemed to be informed by primary information distribution power such as television, newspaper, etc.

In addition, the primary information distribution can be also done many times during the secondary information distribution. For example, if $c = 10,000$ and $tprimary = 200$ and 400, 10,000 individuals in Unknown state are randomly chosen and transit to Known state when $t = 200$ and $t = 400$.

$$P(t) = \begin{cases} c & (t = tprimary) \\ 0 & (\text{otherwise}) \end{cases} \tag{4}$$

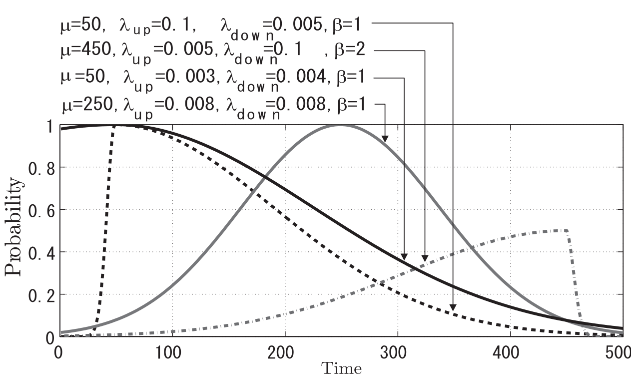


Fig. 6 The graphs from Eq. (3).

4. Simulation and Analysis of Proposed Model

4.1 Simulation Overview

In the present, the most popularly used small-world network models are WS model [2] and NW model [18]. In this paper, we use the NW model instead of the WS small-world network model, because there is probability for the WS model to be broken into unconnected cluster and the average distance between pairs of nodes on the graph is poorly defined due to the rewiring connection [18].

In our simulation, we apply our model and proposed equations to the one-dimension ring lattice small-world network. We deploy the following two steps to construct a small-world network which has the average degree equal to four because real social networks usually have average coordination numbers significantly higher than two [19]. First, we arrange and connect all nodes, then the network is formed as a regular one-dimension ring lattice model where each node has four connections to the nearest neighbor nodes. Second, we create shortcuts by repeatedly connecting two nodes chosen randomly according to the NW model. Furthermore, in social networks there are not only a few shortcuts in networks arising the small-world effect but also a few nodes in the network which have unusual high connections [20] called hubs. In order to make the simulation to be more practical for real-world social networks, we also randomly add a few hubs in our simulation. Table 1 is a list of parameters used in the simulation. An illustration of the NW model in our simulation is depicted in Fig. 7. In this

Table 1 Parameters used in the simulation.

Parameter	Explanation
$K(0)$	Number of initial Known State individuals at the beginning of information distribution. This parameter can be also considered as the result of the initial primary information distribution done in advance.
k	Average number of connections of each node in the network. In other words, k is average degree.
SC	Number of shortcuts.
Hub	Number of hubs.

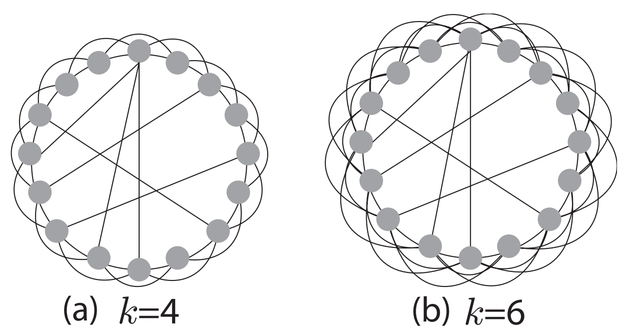


Fig. 7 An illustration of one-dimensional ring lattice small-world network (NW Model).

figure, (a) and (b) represent the NW model with $k = 4$ and $k = 6$, respectively. A node which has more shortcut than one is a hub.

In our simulation, the initial Known State nodes at the beginning of information distribution ($K(0)$), shortcuts and hubs are randomly chosen from all nodes in the network. During the process of simulation, all nodes that are in Distribute State distribute information to their neighbors. And, the probability of nodes in Unknown State becoming Known State depends on numbers of neighbor that are in Distribute State as discussed in Sect. 3.2. For the process of becoming Distribute State and returning to Unknown State, nodes in the Known State are checked to change their state with the probability of $B_2(t)$ and $R_1(t)_i$, respectively. In the process of returning to Known State, nodes in Distribute State are checked to change their state with the probability of $R_2(t)$.

In order to determine the number of shortcuts in our simulation, we consider the characteristic path length (L) in the network because this parameter shows the average distance between two nodes over all pairs of nodes on the network, and adding a few of shortcuts can reduce L dramatically [2]. M.E.J Newman et al. propose interesting method to calculate L using a mean-field-like approximation [21], and a complete solution for L is shown as Eq. (5). In this equation, ξ is characteristic length and $\xi = 1/(k'\phi)$ where k' and ϕ represent constant range($\frac{k}{2}$) and shortcut density($\frac{SC}{N}$), respectively.

$$Ł = \frac{\xi}{2k'\sqrt{1+2\xi/N}} \tanh^{-1} \frac{1}{\sqrt{1+2\xi/N}} \tag{5}$$

$$\overline{C} = \frac{1}{N}\sum_{i=1}^{N} C_i, C_i = \frac{2\,|E(\Gamma_i)|}{k_i(k_i-1)} \tag{6}$$

We use the above Eq. (5) to find appropriate number of shortcuts which gives small-world effect by conducting simulation of L with $N = 1,000,000$ and $k = 4$. Comparing with L of no shortcuts, L is reduced 10, 34 and 250 times when 22, 100 and 1000 shortcuts are added in the network respectively, while Clustering Coefficient of the whole network $\overline{C}$ is almost constant where $\overline{C}$ can be obtained by Eq. (6) [2]. In this equation, C_i is clustering coefficient of subgraph Γ_i, k_i is number of neighbor connections of the nodes i in subgraph Γ_i and $|E(\Gamma_i)|$ is the numbers of existing links in the neighborhood of node i. A graph of L of $N = 1,000,000$ and $k = 4$ is shown in Fig. 8 where y-axis and x-axis are L/N and number of shortcuts, respectively. As seen in this figure, L/N reduces quite considerably when shortcuts are added. But after 100 shortcuts are added, L/N reduces gradually. Therefore, we decide to use 100 shortcuts as a default value in our simulation.

For $B_2(t)$, we use the values shown in Table 2, and these values produce a graph in which the probability increases rapidly at the beginning of the distribution time and gradually decreases after it reaches the maximum value of probability. For $R_2(t)$, the values in Table 2 generate a graph in which the probability increases gradually. In other words, we assume that, the motivation to distribute this information by the nodes in Known State is rapidly created, but after the motivation reaches to a peak, it gradually decreases. And, the lack of motivation to distribute this information by the nodes in Distribute State gradually increases and becomes stagnant later because there is no stimulation. Graphs of $B_2(t)$ and $R_2(t)$ generated by values in Table 2 are shown in Fig. 9.

Other parameters for the simulation are also shown in Table 2. The parameter values in this table are a set as default values, and we change the values in the simulation to analyze their impacts afterward. Furthermore, we define the effectiveness of information distribution as the number of individuals in the network who know the information at each time t. Hence, in the simulation, we focus on the effectiveness of the distribution by observing $K(t)+D(t)$. The default parameter values in Table 2 generate a graph of $K(t) + D(t)$, $U(t)$, $K(t)$ and $D(t)$ shown in Fig. 10. As a note, all nodes and shortcuts are randomly created in the simulation as mentioned above. Therefore, in order to make the result data be reliable, every simulation is conducted 5 times. Subsequently, we calculate all result data to find the average values, that makes Fig. 10.

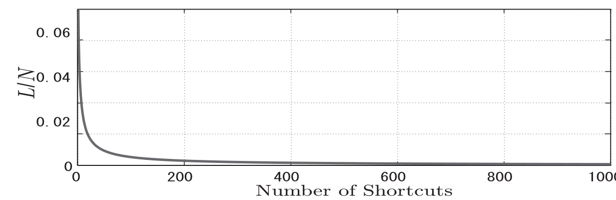


Fig. 8 Graph of L with $N = 1,000,000$ and $k = 4$.

Table 2 Default parameter values.

Parameter	Value
B_1	0.5
$R_1(t)$	Eq. (2) with $S = 100$
$B_2(t)$	Eq. (3) with $\mu = 5$, $\lambda_{up} = 0.1$, $\lambda_{down} = 0.005$, $\beta = 1$
$R_2(t)$	Eq. (3) with $\mu = 450$, $\lambda_{up} = 0.008$, $\lambda_{down} = 0$, $\beta = 1$
$K(0)$	1000
k	4
N	1,000,000
range of t	1 to 500
SC	100
Hub	100. Each hub has 10+k connections
$P(t)$	always 0

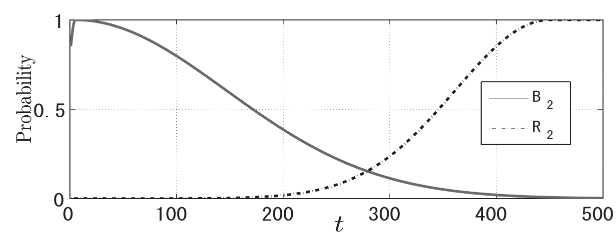


Fig. 9 Graph of $B_2(t)$ and $R_2(t)$ generated by values in Table 2.

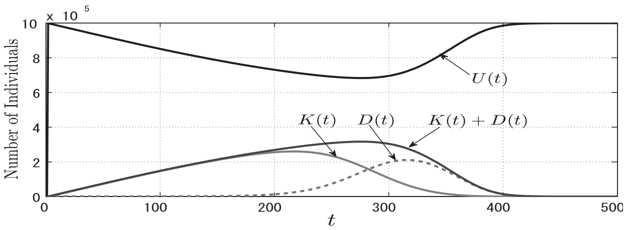


Fig. 10 Graph of $K(t) + D(t)$, $U(t)$, $K(t)$ and $D(t)$ generated by values in Table 2.

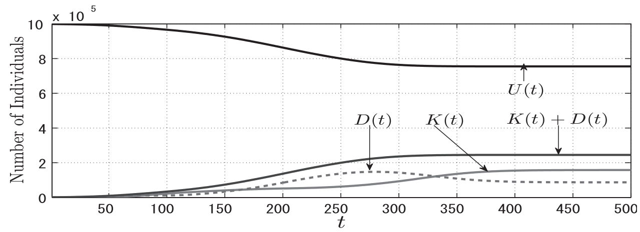


Fig. 11 The impact of $R_1(t)$, $R_2(t)$ and $B_2(t)$ in Pattern1.

Table 3 Groups of R_1, R_2, B_2 and B_1 values.

Pattern	Parameter	Value
Pattern1	$B_2(t)$	Eq. (3) with μ = 200, λ_{up} = 0.01, λ_{down} = 0.015, β = 1
	$R_2(t)$	Eq. (3) with μ = 120, λ_{up} = 0.03, λ_{down} = 0.009, β = 1
	$R_1(t)$	0
	B_1	0.5
Pattern2	$B_2(t)$	1
	$R_2(t)$	0.5
	$R_1(t)$	0.5
	B_1	0.5

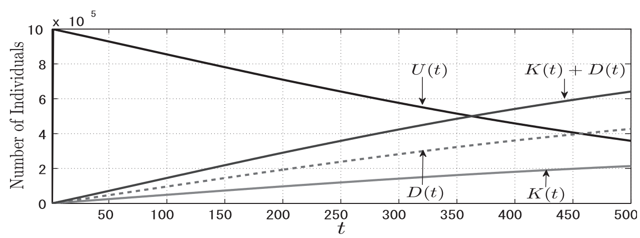


Fig. 12 The impact of $R_1(t)$, $R_2(t)$ and $B_2(t)$ in Pattern2.

4.2 Effect of R_1, R_2 and B_2 to Distribution Result

We consider that the effectiveness of general information distribution gradually increases at the beginning of the distribution time until they peak out and starts to gradually decrease to zero as shown in $K(t) + D(t)$ of Fig. 10. Therefore, the values of R_1, R_2 and B_2 in Table 2 are selected in order to represent this kind of typical phenomenon. Nevertheless, changing the values of R_1, R_2 and B_2 can change the phase transition and represent other kinds of phenomena as well. For example, the graphs of $U(t)$, $K(t)$, $D(t)$ and $K(t) + D(t)$ in Fig. 11 begin with gradual increase or decrease, then stabilize to constant values as time progresses. This graphs is generated by using the values of Pattern1 in Table 3. Furthermore, the changes of $U(t)$, $K(t)$, $D(t)$ and $K(t)+D(t)$ are linear shown in Fig. 12 when the values of Pattern2 in Table 3 are used. According to those results, changing the values of R_1, R_2 and B_2 can control the patterns of the speed and the effectiveness of the information distribution. In other words, those parameters can change phase transition of the distribution results.

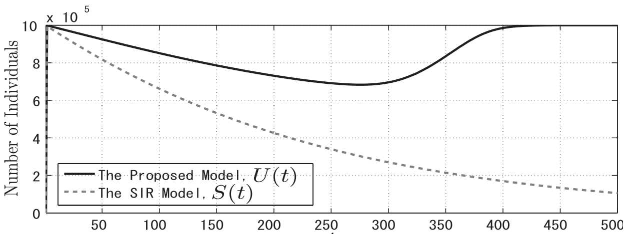
4.3 Comparison of Proposed Model and SIR Model

In this section, we compare the temporal variation of individual numbers in each state of the proposed model to that of the SIR model. First, we conduct the simulation of the SIR model on the small-world network with the similar process and the parameters in [8]–[10]. The process is as follows: at the beginning, the initial I State individuals are randomly chosen. Then, individuals in the I state distribute the information to their neighbors with the constant transmissibility rate δ whereby its value is 1. If the neighbors are in the S state, they become infected and change to the I state. However, if their neighbors are already in the I or R state, the original infected nodes become refractory and move to the R state. The process is repeated until simulation ends. However, in order to make the comparison more effective, we change the total number of individuals in the network, number of initial infected individuals and number of contacted neighbors in each time t, to the same numbers as the simulation of the proposed model. For the simulation result of proposed model, we use the result shown in Fig. 10.

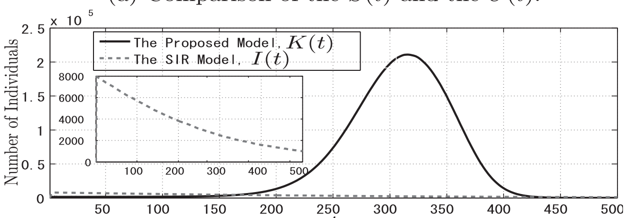
After conducting simulations of the SIR model, we compare the temporal variation of individual numbers in S, I, R states and $I(t) + R(t)$ of the SIR model to U, K, D states and $K(t) + D(t)$ of the proposed model, respectively. $S(t)$, $I(t)$, $R(t)$ are defined as number of individuals of each respective state. The result is shown in Fig. 13 where y-axis is the number of individuals and x-axis is time.

As seen in Fig. 13 (a), while the number of individuals in the S state of the SIR model gradually decreases, the number of individuals in the U state of the proposed model also decreases with slower rate but it later increases back slowly. According to Fig. 13 (b), in the beginning, the number of individuals in the K state of the proposed model starts to increase slowly until $t \approx 170$ where it rapidly increases up to $t \approx 315$, before decreasing back. On the other hand, the number of individuals of the I state of the SIR model increases up to almost 8000 individuals in the beginning and then gradually decreases as shown in upper inset of Fig. 13 (b).

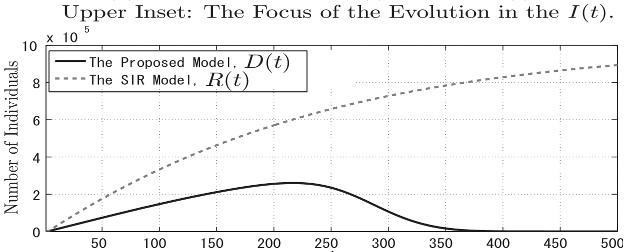
Furthermore, as shown in (c) and (d) of Fig. 13, the number of individuals in the R state of the SIR model and $I(t)+R(t)$ gradually increase and the number will reach to N in the future of time t. However, the number of individuals in the D state and $K(t) + D(t)$ of the proposed model gradually increase until they peak out and start to decrease. Note that we compare $I(t)+R(t)$ of the SIR model to $K(t)+D(t)$ of



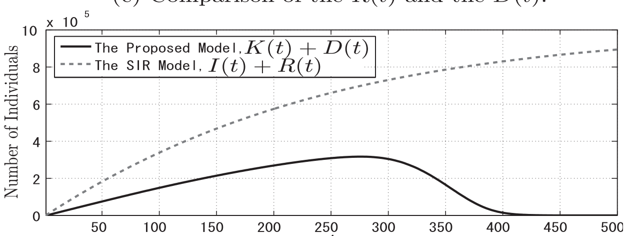
(a) Comparison of the $S(t)$ and the $U(t)$.



(b) Comparison of the $I(t)$ and the $K(t)$, Upper Inset: The Focus of the Evolution in the $I(t)$.



(c) Comparison of the $R(t)$ and the $D(t)$.



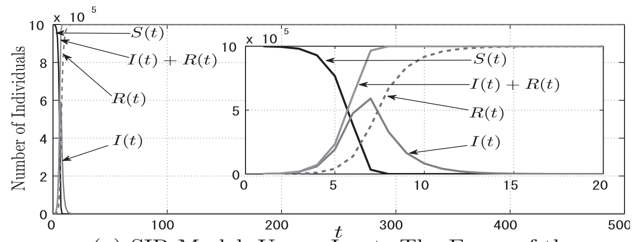
(d) Comparison of $I(t) + R(t)$ and $K(t) + D(t)$.

Fig. 13 Comparison of the SIR model and the proposed model.

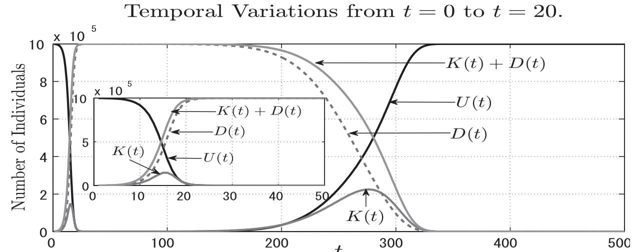
the proposed model because $I(t) + R(t)$ is considered as the information distribution effectiveness same as the $K(t)+D(t)$ in the proposed model.

The different results between the SIR and the proposed model are obtained because there are no returning states in the SIR model. Therefore, all individuals in the network would receive information when time passes and this kind of phenomenon is occasional in our society. In the proposed model when appropriate R_1 and R_2 are selected, individuals can return to Known and Unknown states again.

Moreover, we also compare the SIR model and the proposed model by the mean-field equation. According to [22], [23], the mean field differential equation of the SIR model in a population of N individuals is given by the below Eq. (7), where δ is transmissibility, and g is the rate of conversion from the I state to the R state(recovery rate). The mean-field differential equation of the proposed model is shown in Eq. (8). Subsequently, we calculate Eq. (7) with $\delta = 1$, $g = 0.5$ and initial infected node is 1000. For the Eq. (8), we use the B_1, B_2, R_1, R_1 and $K(0)$ values in Table 2. The



(a) SIR Model, Upper Inset: The Focus of the Temporal Variations from $t = 0$ to $t = 20$.



(b) Proposed Model, Upper Inset: The Focus of the Temporal Variations from $t = 0$ to $t = 50$.

Fig. 14 Comparison of the SIR model and the proposed model by the mean-field equation.

result is shown in Fig. 14 in which (a) and (b) are the result of Eq. (7) and Eq. (8), respectively.

$$
\begin{aligned}
\frac{dS(t)}{dt} &= -\delta S(t)I(t) \\
\frac{dI(t)}{dt} &= \delta S(t)I(t) - gI(t) \\
\frac{dR(t)}{dt} &= gI(t)
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
\frac{dU(t)}{dt} &= -B_1D(t) + R_1K(t) \\
\frac{dK(t)}{dt} &= (-R_1 - B_2)K(t) + B_1D(t) + R_2D(t) \\
\frac{dD(t)}{dt} &= -R_2D(t) + B_2K(t)
\end{aligned}
\tag{8}
$$

According to (a) of Fig. 14, the temporal variations of individual numbers in the S and R states of the SIR model show rapid decrease/increase, and once reaching a bottom or peak, they remain steady. However in the I state, there is a period where the number of individuals rapidly rises until it peaks out, and rapidly falls to the bottom and remains stable.

The temporal variations in the proposed model are different from the ones in the SIR model. In the proposed model, after the individual numbers reach a peak or bottom and remain steady, they decrease/increase back again as seen in (b) of Fig. 14. Since it is difficult to see the temporal variations in the beginning of the distribution time, the uppper insets that focus on the temporal variations from $t = 0$ to $t = 20$ and $t = 0$ to $t = 50$ are added to Fig. 14 (a) and (b), respectively.

Since there are no parameters in Eq. (7) that represent the potential to return individuals to their previous states, while there are such parameters instead in Eq. (8), the result of Fig. 14 gives the similar curve-shape patterns to the result of Fig. 13. However in Fig. 14, the changes are much more extreme because the network structure is neglected in Eq. (7) and Eq. (8).

5. Parameters' Impacts of Proposed Model

Figure 15, Fig. 16, Fig. 17 and Fig. 18 shows the impact of k, shortcuts, R_1 and B_1, respectively, where y-axis is the total number of individual who know the information($K(t)+D(t)$) and the x-axis is time(t). In the simulation of shortcuts impact (Fig. 13),we exclude hubs from the simulation because connections between hubs and other nodes can be considered as shortcuts, and, thus, they can create confusion (with normal shortcuts).

After conducted the simulation, the result from Fig. 15 shows that k gives distinguish impact of increasing maximum value of $K(t)+D(t)$ when $k \leq 8$. However, this impact is reducing when $k > 8$. Moreover, there is no much different results of maximum value of $K(t)+D(t)$ between $k = 12$ and $k = 14$. Those of results imply that the effectiveness of information distribution on the network which has high k is not much improved.

The impact of shortcuts(SC) is shown in Fig. 16. As seen in this figure, there is small but noticeable impact between $SC = 0$(the regular ring lattice model) and $SC = 100$. However, after adding 100 shortcuts there are hardly any impact by providing more several hundreds shortcuts, while there is distinguished impact by providing a huge number of shortcuts such as thousand and several ten thousands. However, as seen in the simulation results of $SC = 30000$ and $SC = 40000$ in Fig. 13, there is no much difference result between the two graphs. The results represent that if the network have very high number of shortcuts, it is not necessary to increase shortcuts in the networks in order to increase the power of the distribution.

The impact of R_1 is shown in Fig. 17. Increasing R_1 does not only increases the maximum value of $K(t) + D(t)$ but also increases the horizontal(x-axis) size of the graphs. Nevertheless, there is almost no impact for the maximum value of $K(t)+D(t)$ when $S > 3000$. This result implies that the maximum value of $K(t) + D(t)$ is not much improved when S is very high because only a few of individuals in the network forget the information.

Furthermore, according to Fig. 18, B_1 does not give impressive impact because $K(t) + D(t)$ is slightly changing when increased by 0.2.

5.1 The Impact of $P(t)$

In this subsection, we observe the impact of $P(t)$. In the simulation, we also use the data in Table 2 but change the value of $P(t)$ to $P(tprimary) = 10,000$. Subsequently, we conduct three simulations where $tprimary = 150$, $tprimary = 350$ and $tprimary = 450$ to compare the impacts of each value. In other words, when $t = 150$, $t = 350$ or $t = 450$, primary information distribution is done to the network and as a result 10,000 nodes in Unknown State are randomly selected and changed into Known State. The reason to select those of t is because the total number of individuals who know the information ($K(t) + D(t)$) is increasing when $t = 150$, decreasing when $t = 350$ and becoming nearly zero when $t = 450$, respectively, as seen in Fig. 10.

The simulation results are shown in Fig. 19, Fig. 20 and Fig. 21. As seen in those figures, using primary information distribution gives very impressive result if it is done when $K(t) + D(t)$ is increasing. In fact, the maximum value of $K(t) + D(t)$ with $P(150) = 10,000$ reaches approximately 2.5 times of the maximum value of $K(t)+D(t)$ with $P(t) = 0$.

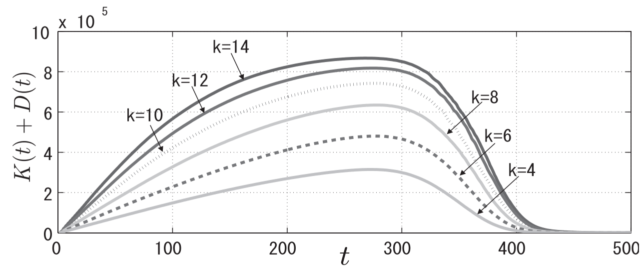


Fig. 15 The impact of k.

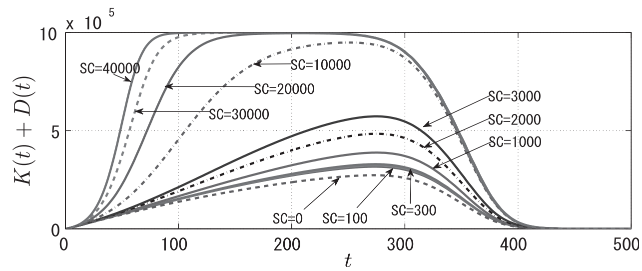


Fig. 16 The impact of shortcuts.

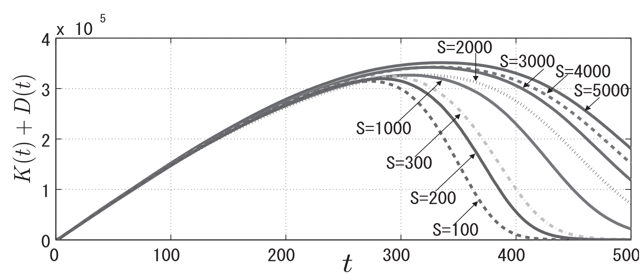


Fig. 17 The impact of R_1.

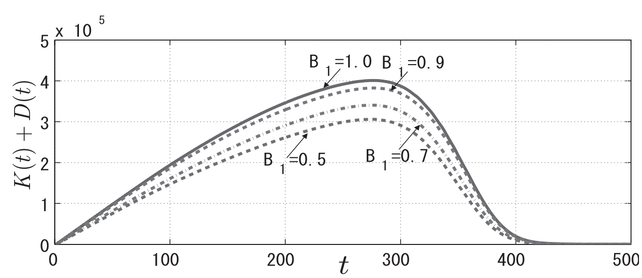


Fig. 18 The impact of B_1.

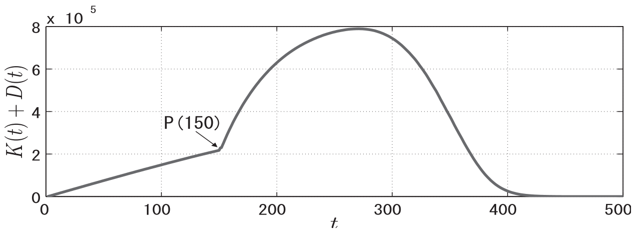


Fig. 19 The impact of $P(150) = 10,000$.

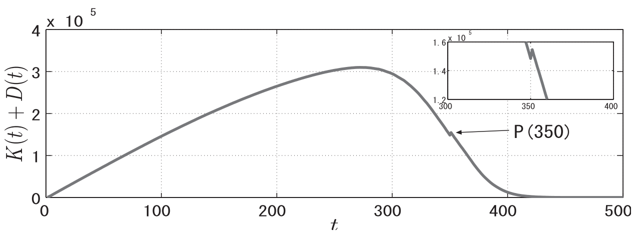


Fig. 20 The impact of $P(350) = 10,000$. Upper inset: The focus impact of $P(350) = 10,000$.

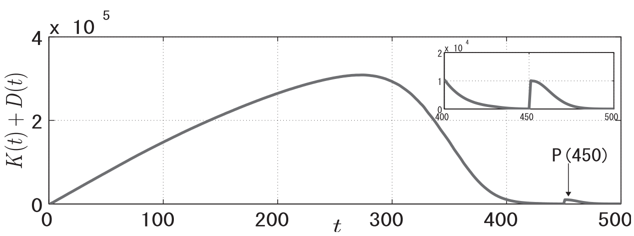


Fig. 21 The impact of $P(450) = 10,000$. Upper inset: The focus impact of $P(450) = 10,000$.

However, if this primary information distribution takes place during the period when $K(t) + D(t)$ is decreasing($P(350) = 10,000$) or becoming nearly zero($P(450) = 10,000$), there is almost no impact to $K(t) + D(t)$. The focus impacts of $P(350) = 10,000$ and $P(450) = 10,000$ are shown in upper insets of Fig. 20 and Fig. 21, respectively. As shown in upper insets of those figures, $K(t) + D(t)$ is slightly increased, then it is gradually decreasing. Hence, in order to increase the power of the distribution we conclude that $P(tprimary)$ should be applied when $K(t) + D(t)$ is increasing or $\frac{d(K(t)+D(t))}{dt} > 0$.

5.2 Consideration of Network Structure with Shortcuts

The number of shortcuts in the network is one of most important factors to the network structure. In order to observe the impact of network structure to information distribution result by the number of shortcuts, we define a new parameter calling Distribution Effectiveness Rate (*DER*) and its equation is shown in Eq. (9) whereas T is rang of the simulation time. This parameter represents the effectiveness of the distribution result, and the maximum value of *DER* is one. Subsequently, we conduct simulation by increasing number of shortcuts and observe impact to *DER*. The result is shown in Fig. 22 whereas y-axis is *DER* and x-axis is

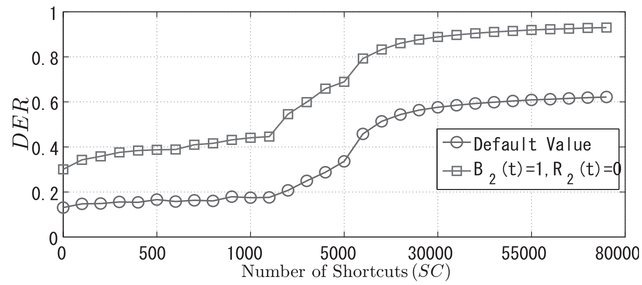
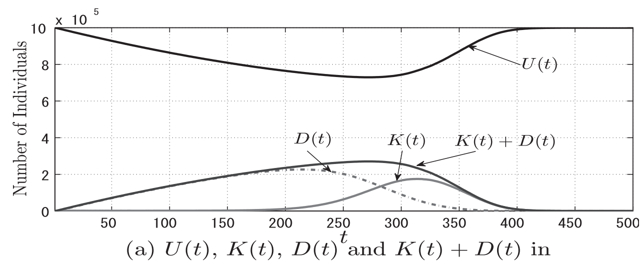
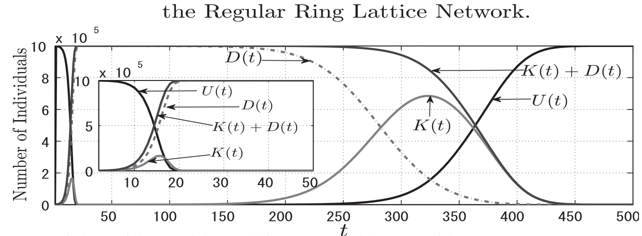


Fig. 22 A graph of shortcut's impact to *DER*.



(a) $U(t)$, $K(t)$, $D(t)$ and $K(t) + D(t)$ in the Regular Ring Lattice Network.



(b) $U(t)$, $K(t)$, $D(t)$ and $K(t) + D(t)$ in the Random Network, Upper Inset: The Focus of the Temporal Variations from $t = 0$ to $t = 50$.

Fig. 23 A comparison of the regular lattice network and the random network.

number of shortcuts, respectively.

$$DER = \frac{1}{T \cdot N} \sum_{t=1}^{T} K(t) + D(t) \tag{9}$$

As seen in this figure, the circle line graph is produced by using $B_2(t)$ and $R_2(t)$ from Table 2, while the square line graph is generated by using $B_2(t) = 1$ and $R_2(t) = 0$. Those results give the similar graphs pattern, and show that *DER* gradually increases but it slightly increases when number of shortcut is more than 15,000. In other words, when the network contains very high number of shortcuts, further increasing shortcuts does not have significant impact even though the motivation of nodes for the distribution is high, i.e. $B_2(t) = 1$ and $R_2(t) = 0$.

Furthermore, we conduct the simulation on the random network and the regular ring lattice network, and compare their results. In this random network, each node has four connections which are randomly connected to other nodes, i.e. both the networks have the same k ($k = 4$). For the other necessary parameters values, the values from Table 2 are used. The distribution results in regular ring lattice network and the random network are shown in (a) and (b) of Fig. 23, respectively. The upper inset of Fig. 23 (b) is focused on the temporal variations from $t = 0$ to $t = 50$.

According to the Fig. 23 (a) and (b), the temporal variation of individual numbers of the $U(t)$, $K(t)$, $D(t)$ and $K(t) + D(t)$ gradually change in the regular lattice model, while they show rapid changes in the random network. Moreover, the information is distributed to all individuals within very short-time in the random network, while it is distributed slowly and only less than 30 percent of all individuals can receive it in the regular ring lattice network. In other words, the information is distributed slowly in the regular ring lattice network because the network is highly ordered and the connections are localized, then the information can not reach to all individuals during the distribution time. However, the distribution result is much more extreme in the random network because the network is disordered and there is no effect of spatial correlations.

5.3 Consideration of $B_2(t)$

As described in Sect. 3.3, we consider and define $B_2(t)$ as a dynamic parameter. This parameter is also very important for the proposed model to achieve fruitful distribution results. The experimental study of the global social network by forwarding more than 60,000 e-mail messages shows a final result that if the individual incentive is insufficient during forwarding messages or searching for remote targets, the small-world effect does not occur even the network structure is well connected [24]. This experimental result supports the idea of $B_2(t)$ in our proposed model, because the proposed model can explain this kind of phenomenon by adjusting $B_2(t)$, in which the SIR model is not able to do so.

In order to prove that, we set a new value of $B_2(t)$ as a low value and generate a graph of $B_2(t)$ shown in Fig. 24 while $R_2(t)$ is still the same as shown in Table 2. In addition, we also increase 200,000 shortcuts in the network in order to produce well connected network structure. After conducting the simulation, our result is shown in Fig. 25. In this figure, the solid line and the dotted line show the simulation results when $B_2(t)$ of Fig. 24 and Table 2 are used, respectively. The result shows that even though the network has a lot of shortcuts and well connected, if $B_2(t)$, i.e. incentive to distribution is low, the distribution result would be poor(solid line) and the small-world effect does not occur. Nevertheless, if we use $B_2(t)$ of Table 2, it gives a very impressive distribution result(dotted line) as seen in Fig. 25. The result of this figure explains the conclusion of [24].

5.4 Discussion

There are some research studies [25] that give suggestions to concentrate on shortcuts to increase or accelerate the distribution. Nevertheless, it is difficult to find or locate shortcuts in real world networks due to the limitation of the whole network structure information. For example, [25] shows that the navigation process in the social network is an interplay between the network structure and the individual's decisions based on their limited information of the whole network structure. Since individuals are lacking of information of the whole network structure, therefore they find that it is very difficult to select the best routes when distributing information. Even recent research efforts [26]–[28] have studied algorithms to find such shortcuts in networks but those efforts are still far from complete when applying the algorithms in practical social networks. Therefore, not only the network structure but also incentive or motivation plays an important role in making fruitful information distribution result.

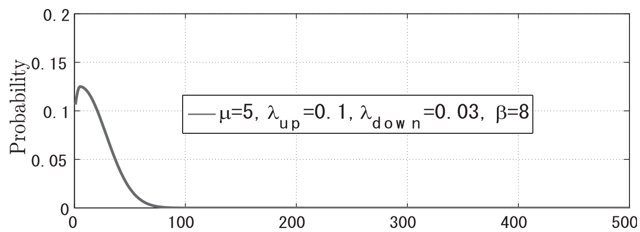


Fig. 24 A graph of $B_2(t)$ with low probability.

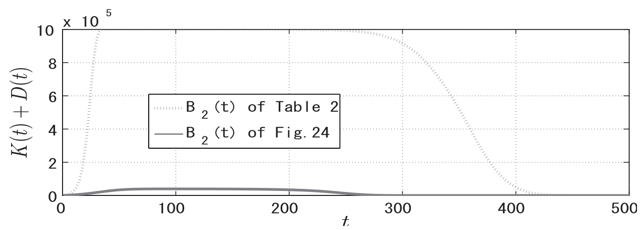


Fig. 25 The simulation results of high number of shortcuts with $B_2(t)$ of Fig. 24 and $B_2(t)$ of Table 2.

6. Conclusion and Future Works

In this paper, information distribution model based on human's behavior is proposed. We also introduce dynamic parameters to make the model more practical for the real life social network. Some limitations of the SIR model and related works are explained. One of the big differences in the proposed model and the SIR model is that there are no returning states in the SIR model. However, in the proposed model when appropriate R_1 and R_2 are selected, individuals can return to the previous states again. Subsequently, we apply small-world network and its characteristics to conduct the simulations.

Having conducted various simulations, the results show that the model can present, analyze and predict the effectiveness of the secondary information distribution. The simulation results also show how to increase the power of the distribution efficiently by using parameters in the proposed model, and how the model can be useful for optimizing and controlling information distribution on social networks. In summary, the impact of R_1 shows that increasing S in Eq. (2) increases both maximum value of $K(t)+D(t)$ and range of distribution effectiveness time. However, if value of S is big enough, further increasing its value does not have significant impact on the maximum value of $K(t)+D(t)$. Furthermore, using B_1 and R_1 to increase the maximum value of $K(t) + D(t)$ is not effective because they do not give impressive impacts. Our results also show that if the networks have very high shortcuts or high k, it is not necessary to increase shortcuts or k in the networks in order to increase the power of the distribution. The simulation results also show that changing the values of R_1, R_2 and B_2 can change the phase transition of the distribution results.

Moreover, in order to use the primary information distribution during the secondary distribution in effective way, the primary information distribution should be done when $\frac{d(K(t)+D(t))}{dt} > 0$.

According to the consideration of network structures and $B_2(t)$, network structure gives significant impact to the distribution result. Nevertheless, its impact is ineffective if motivation to distribute information is low. In other words, not only the network structure but also incentive or motivation plays an important role in making effective information distribution.

In this research, our ultimate goal is to analyze and observe the real-world phenomena by using the proposed model and its parameters. In order to achieve our goal, in this paper we propose a model with analytical results as the first step. We will find the relationship between the real-world phenomena and the parameters, and also how to map those parameters to such phenomena in our future works. There are also other remaining issues in the proposed model such as analysis of B_2 and R_2. Because these parameters can produce fruitful distribution results, B_2 and R_2 will be studied more in order to maximize the accuracy and the effectiveness of the proposed model. Furthermore, some influencing factors for real-world information distribution will be investigated to find the relation and impacts on the parameters in the proposed model.

References

[1] P. Sriprasertsuk, A. Seki, and W. Kameyama, "On content distribution model and analyzing distribution effectiveness," J. IPSJ, vol.48, no.8, pp.2878–2891, Aug. 2007.

[2] D.J. Watts and S.H. Strogatz, "Collective dynamics of small world networks," Nature, vol.393, pp.440–442, June 1998.

[3] A. Broder, "Graph structure in the web," Comput. Netw., vol.33, pp.309–320, 2000.

[4] J.M. Montoya and R.V. Sole, "Small-world patterns in food webs," J. Theoretical Biology, vol.214, pp.405–412, March 2002.

[5] M.E.J. Newman, "The structure of scientific collaboration networks," Proc. National Academy of Sciences of the United States of America, vol.98, pp.404–409, Jan. 2001.

[6] R.F.i Cancho, C. Janssen, and R.V. Sole, "Topology of technology graphs: Small world patterns in electronic circuits," Phys. Rev. E, vol.64, 046119, Sept. 2001.

[7] R.F.i Cancho and R.V. Sole, "The small-world of human language," Proc. Royal Society of London B, vol.268, pp.2261–2265, July 2001.

[8] H. Liang, P. Kwangho, and L. Ying-Chang, "Information propagation on modular networks," Phys. Rev. E, vol.73, 035103 (R), March 2006.

[9] F. Feng, L. Lianghuan, and W. Long, "Information propagation in a novel hierarchical network," Arxiv:math/0605293, May 2006.

[10] M. Yamir, N. Maziar, and F.P. Amalio, "Dynamics of rumor spreading in complex networks," Phys. Rev. E, vol.69, 066130, June 2004.

[11] W.O. Kermack and A.G. Mckendrick, "A contribution to the mathematical theory of epidemics," Proc. Royal Society of London A115, pp.700–721, Aug. 1927.

[12] C. Li-Chiou and K.M. Carley, "The impact of countermeasure propagation on the prevalence of computer viruses," IEEE Trans. Syst. Man Cybern. B, Cybern., vol.34, no.2, pp.823–833, April 2004.

[13] D.H. Zanette, "Critical behavior of propagation on small-world networks," Phys. Rev. E, vol.64, 050901, Oct. 2001.

[14] D. Dagon, C. Zou, and W. Lee, "Modeling botnet propagation using time zones," Proc. 13th Annual Network and Distributed System Security Symposium (NDSS'06), pp.235–249, Feb. 2006.

[15] H. Ebbinghaus, Memory: A Contribution to Experimental Psychology, Columbia University Press, New York (translated from Untersuchungen zur experimentellen Psychologie, Duncker and Humblot, Leipzig, Germany, 1885), 1913.

[16] T. Levitt, "Exploit the product life cycle," Harvard Business Review, vol.43, pp.81–94, Nov. 1965.

[17] M.K. Sparrow, "Information systems: A help or hindrance in the evolution of policing," The Police Chief, vol.54, no.4, pp.26–44, April 1991.

[18] M.E.J. Newman and D.J. Watts, "Renormalization group analysis of the small-world network model," Phys. Lett. A, vol.263, pp.341–346, Dec. 1999.

[19] M.E.J. Newman and D.J. Watts, "Scaling and percolation in the small-world network model," Phys. Rev. E, vol.60, pp.7332–7442, 1999.

[20] R. Kasturirangan, "Multiple scales in small-world graphs," Massachusetts Institute of Technology AI Lab Memo 1663 and cond-mat/9904055, April 1999.

[21] M.E.J. Newman, C, Moore, and D.J. Watts, "Mean-field solution of the small-world network model," Phys. Rev. Lett., vol.84, pp.3201–3204, April 2000.

[22] M.M. Telo da Gamaa and A. Nunes, "Epidemics in small world networks," European Physical J. B (Condensed Matter and Complex Systems), vol.50, pp.205–208, April 2004.

[23] A. Kleczkowski and B. Grenfell, "Mean-field-type equations for spread of epidemics: The 'small world' model," Physica A: Statistical Mechanics and its Applications, vol.274, pp.355–360, Dec. 1999.

[24] P.S. Dodds, R. Muhamad, and D.J. Watts, "An experimental study of search in global social networks," Science, vol.301, pp.827–829, Aug. 2003.

[25] S.A. Pandit and R.E. Amritkar, "Characterization and control of small-world networks," Phys. Rev. E, vol.60, pp.1119–1122, Aug. 1999.

[26] J.M. Kleinberg, "The small-world phenomenon: An algorithmic perspective," Proc. 32nd Annual ACM Symposium on Theory of Computing, pp.163–170, June 2000.

[27] Z. Han and Z.X. Huang, "Navigation in a small world with local information," Phys. Rev. E, vol.70, 036117, Sept. 2004.

[28] L. Adamic and E. Adar, "How to search a social network," Social Networks, vol.27, pp.187–203, July 2005.

Pao Sriprasertsuk received Bachelor degree from King Mongkut Institute of Technology, Ladkrabang (KMITL) in the major of Telecommunication Engineering in 1998. He received Master of Science from Graduate School of Global Information and Telecommunication Studies, Waseda Univ. in 2002. He has continued his doctoral study since then. Since 2007, he has been a visiting researcher at Global Information and Telecommunication Institute, Waseda Univ. His current research interests are metadata and content distribution system. He is a member of IEEE.

Wataru Kameyama received Bachelor, Master and Doctor of Engineering from School of Science and Engineering, Waseda Univ. in 1985, 1987 and 1990, respectively. He joined ASCII Corp. in 1992, and was transferred to France Telecom CCETT from 1994 to 1996 for his secondment. He has been appointed as a professor at Graduate School of Global Information and Telecommunication Studies, Waseda Univ. since 2002. He has been involved in MPEG, MHEG, DAVIC and the TV-Anytime Forum activities. He was a chairman of ISO/IEC JTC1/SC29/WG12, and a secretariat and a vice chairman of the TV-Anytime Forum. Since 2006, he has been a chairman of the SIG on Electronic Intellectual Property in IPSJ. He received the best paper award of Niwa-takayanagi in 2006 from the Institute of Image Information and Television Engineers. He is a member of IPSJ, ITE, IIEEJ, ACM and IEEE.