N-020

A Proposal of Contents Distribution Model and Analysis of Illegal Contents Distribution on the Small-World Network

Pao Sriprasertsuk† Wataru Kameyama††

1. Introduction

In general, there are two kinds of contents flows in an information 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. For examples, one student copies rental DVD and gives to his or her friends. Today, the advanced information technologies enable the secondary information distribution to be able to perform by various methods and media.

Nevertheless, the secondary information distribution has been also increasing the power of illegal distribution of contents, and contents industries have been suffering from it, particularly by peer-to-peer file sharing (p2p) networks. The study of [1] has predicted that global peer-to-peer networks will be effectively stopped by legal means.

In addition, there is another way to illegally distribute contents that is the Small-World Network(SWN)[2] or social network. With the advanced information communication technologies and availability of cheap media, storage devices and high bandwidth networks, distributing contents in the SWN has been becoming easier and more powerful than before. Therefore, it is necessary to analyze and control the power of secondary information distribution.

We have investigated how the secondary information distribution affects the information circulation based on the statistical observation[3]. However, the proposed model does not consider the SWN structure but only the statistical aspects of information distribution being consumed. There are researches[4][5][6] using the SIR model[7] to analyze the effectiveness of information distribution in the SWN. We consider the SIR model is not suitable to represent the real world information propagation cycle since it does not fully take into account the nature of human behavior for information distribution.

In this paper, we propose an information distribution model and equations considering the human behavior. Subsequently, the proposed model is simulated and analyzed by applying the SWN and it characteristics to analyze illegal contents distribution.

2. The Proposed Information Distribution Model

2.1 Model Overview

With regard to the human behavior state for information distribution, we consider that there are three states that are "Unknown", "Known" and "Distribute". The proposed model is shown in Figure 1, 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.
- 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 Figure 1 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 at time t
- $K(t)$: the number of Known State individuals at time t
- $D(t)$: the number of Distribute State individuals at time t

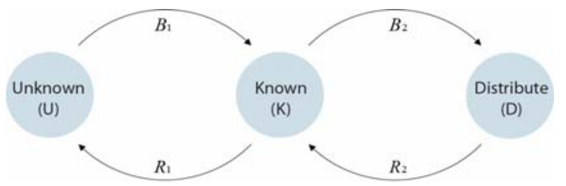

Figure 1: The Proposed Model

2.2 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 Equation 1, where G_i is successful distribution probability of individual i and n is the number of neighbors of it in Distribute State.

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

2.3 Dynamism in R_1

We apply the forgetting curve theory of Hermann Ebbinghaus[8] because it has clearly stood the test of time and has been validated in many numbers of dissertations and follow-up studies. As discussed in section 2.2, 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

† Global Information and Telecommunication Institute, Waseda University
†† Graduate School of Global Information and Telecommunication Studies, Waseda University

State, they start to forget the information with the probability of R_1. Thus, $R_1(t)_i$ for individual i is defined in Equation2 where t_p is time when individual i transit to Know State and S is the relative strength of memory.

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

2.4 Dynamism in B_2 and R_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. As we know, 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 Equation 3 to represent $B_2(t)$ and $R_2(t)$. Equation 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 $t \le \mu$, changing graph scale when $t > \mu$, and changing maximum value of y-axis, respectively.

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

4. Simulation and Analytical Results

4.1 Simulation Overview

In this paper, we use the Newman-Watts(NW) SWN[9] instead of the Watt-Strogatz(WS) SWN, 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. The parameters for the simulation are shown in Table 2 where $K(0)$ and k are number of initial Known State individuals at the beginning of information distribution and the average number of connections of each node in the network, respectively.

Table 2: A Default Parameter Values

Parameter	Value
B_1	0.5
$R_1(t)$	Equation 2 with S = 100
$B_2(t)$	Equation 3 with μ = 5, λ_{up}=0.1, λ_{down}=0.005, β = 1
$R_2(t)$	Equation 3 with μ = 450, λ_{up}=0.008, λ_{down}= 0, β = 1
$K(0)$	10
k	4
N	10,000
rang of t	1 to 500
shortcut	50

$$DER = \frac{1}{T \cdot N} \sum_{t=1}^{T} K(t) + D(t) \tag{4}$$

In order to observe the impact of distribution effectiveness, we define a parameter calling Distribution Effectiveness Rate(DER) and its equation is shown in Equation 4 whereas T is rang of t. We define the nodes in the network which have ability to copy and redistribute contents as Bad nodes, and nodes which don't have ability to redistribute contents are Constraint nodes. Firstly, we suppose that all nodes in the network are Bad nodes. Subsequently, we conduct simulations by changing Bad nodes to be Constraint nodes at constant rate and observing impact to DER.

4.2 Simulation Results and Considerations

Having conducted the simulations based on one-dimensional ring lattice SWN and two-dimensional square lattice SWN, the results are shown in Figure 2 and Figure 3 whereas y-axis and x-axis are DER and number of Constraint nodes, respectively.

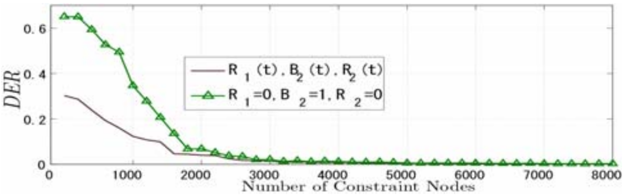

Figure 2: The Impact of Number of Constraint Nodes to DER in One-Dimensional Ring Lattice SWN.

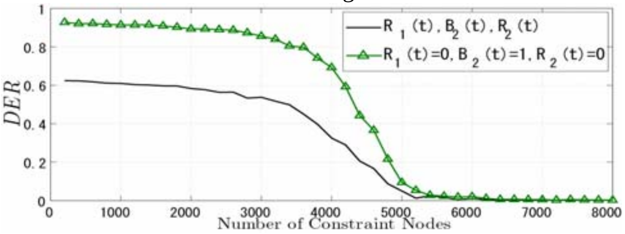

Figure 3: The Impact of Number of Constraint Nodes to DER in Two-Dimensional Square Lattice SWN.

As seen in Figure 2 and 3, the line graph shows the impact when we use $R_1(t)$, $B_2(t)$, $R_2(t)$ from Table 2, and triangle line graph shows the impact when we use $R_1(t) = 0$, $B_2 = 1$ and $R_2 = 0$ as static values. The results show DER is rapidly decreasing until Constraint nodes approximately reach 2,000 in Figure 2 and 5,000 in Figure 3. Surprisingly, the results imply that illegal contents distribution in the network which has 4 average degree distribution and a few of shortcuts can be controlled by suppressing the number of Bad nodes less than 80 percent for one-dimensional SWN and 50 percent for two-dimensional SWN, even though all nodes in the network have motivation to redistribute the contents(triangle ling graph).

The results also show that two-dimensional square lattice SWN is more appropriate to represent and analyze real world content distribution on SWN. Because only 20 percent of Constraint nodes in one-dimensional SWN can stop effectively illegal content distribution, and this kind of phenomenon is rare in the real world.

5. Future Works

There are remaining issues in our work. For examples, we will estimate the appropriate values of parameters in the proposed model for the real world social network. Furthermore, some influencing factors and their impacts for real-world contents distribution will be investigated and analyzed by using the proposed model to analyze and control the illegal contents distribution.

References

[1] P. Biddle, P. England, M. Peinado, and B.Willman: The Darknet and the Future of Content Distribution, In Proceedings of the 2002 ACM Workshop on Digital Rights Management, (2002)
[2] D.J.Watts and S.H. Strogatz: Collective dynamics of 'small world' networks, Nature, vol. 393, pp. 440-442,(1998).
[3] S. Pao, S. Akiko and K. Wataru: On Content Distribution Model and Analyzing Distribution Effectiveness, Journal of Information Processing Society Japan, vol.48 , pp. 2878-2891, No.8, (2007)
[4] H. Liang, P. Kwangho and L. Ying-Chang: Information propagation on modular networks, Physical Review E, vol. 73, 035103(R), (2006).
[5] F. Feng, L. Lianghuan, W. Long: Information propagation in a novel hierarchical network, Arxiv:math/0605293 , (2006).
[6] M. Yamir, Maziar, F.P. Amalio: Dynamics of rumor spreading in complex networks, Physical Review E, vol. 69, 066130, (2004).
[7] W.O Kermack and A.G. Mckendrick: A Contribution to the Mathematical Theory of Epidemics, Proceedings of the Royal Society of London A115, pp. 700-721, (1927).
[8] H. Ebbinghaus: Memory: A Contribution to Experimental Psychology, Columbia University Press, New York, (1913).
[9] M.E.J. Newman and D.J Watts: Renormalization group analysis of the SWN model, Physics Letters A, vol. 263, pp. 341-346, (1999).