A Consideration of Added Value Which Influences Information Diffusion

Yuya Ota

Soka University
Tokyo, Japan
Email: e18m5222@soka-u.jp

Norihiko Shinomiya

Soka University
Tokyo, Japan
Email: shinomi@soka.ac.jp

***Abstract*—In the case of existing information diffusion models, probability of information diffusion only follows parameters of neighbor nodes sending and receiving information. But, the momentum of diffusion gains in an exponential fashion whenever the information passes through influential nodes in a real network. For example, when information diffuses through social networking service (SNS), the speed of diffusion increases whenever information passes nodes that is influential users. This paper quantifies the degree of influence as information vitality, and proposes the model using it. In addition, the paper considers relationship between diffusion speed and information vitality in simulation of the model.**

Keywords–Information diffusion; Social network; Independent cascade model.

I. INTRODUCTION

Recent advances in communication technology have made it possible to increase individual information transmission capacity, thus social network was formed on the Internet. It enabled people to diffuse information wildly and rapidly without existing mass media. On the other hand, the anonymity of the network accelerates irresponsible information transmission; thus, social problems that are like fake news or hate speech are increasing. For example, fake news was used to attack political opponent and appeal the electorate in the American presidential election in 2016. There is large risk to change the result by wrong information. To solve the problem, we need to make mathematical model to predict information diffuse in the society, and find the factor to influence the diffusion.

Ample studies have proposed the information diffusion model, however, the model does not consider the change of added information value in the diffusion process because the probability depends completely on initial parameter that network nodes and edges have. One of the value is diffusion trend that is determined by the contents of the information.The paper by Vosoughi et al. [1] have concluded that wrong information spread 20 times faster than right one. This means that the speed of information diffusion depends on the content. Another study analyzes the big contents data that is collected from SNS to find the diffusion trend [2]. These studies elucidate how the content influences the diffusion process; however, it cannot lead to predict the information diffusion.

The purpose of our study is making the information diffusion model to explain how the change of added information value influences the diffusion process. To test the hypothesis that the information diffusion depends on the value change in the process, we have examined whether the centrality of network for information diffusion has correlation with range of the diffusion. To this end, we have performed a series of different simulation with stochastic model.

In Section 2, information diffusion model using added information value is explained. In Section 3, method of simulations and results are illustrated. Finally, considering and future work were suggested in Section 4.

II. MODEL REPRESENTATION

This model considers the added information value, for example the number of re-tweet on Twitter, as the factor which influences the diffusion process. As shown in Table 1, these kinds of parameter can be considered as the added information value. The value is defined by information engine theory as the name of ”information vitality (IV)” [3]. The first question which we have to answer in this paper is how the IV works in the information diffusion process.

Information engine theory has supposed that information has two parameters which are entropy and IV. Entropy is a parameter which is used in thermodynamics and statistical mechanics and considered as a physical quantity that expresses “ambiguity” in information theory. Information which has higher entropy is more likely to diffuse, so entropy and initial IV have a correlation.

In the IV model, we have used directed graph, and regarded the node receiving the information as active node, the ones not receiving node as inactive node, and the human relationship as edges. We have also considered the IV of the active nodes as diffusion probability $p_{u,v}$ $(0 \leq p_{u,v} \leq 1)$. When a node becomes active, the node tries to make neighbor nodes active only once in the next step. In addition, every node has the parameter named ”logical work” that change IV. When information passes through the nodes, each node gives logical work q $(q \in \mathbb{R})$ to the information, and change the IV. Therefore, when a node becomes active in the probability of $p_{u,v}$, the node tries to diffuse in the probability of $p_{u,v} + q$ in next step. Although the probability of information diffusion does not change in previous independent cascade (IC) model as figure 1 shows, in the case of Figure 2, initial IV is 0.2, and when it makes next node active, it becomes 0.21. This simulation keeps going until the state does not change. This model can explain spread of news or advertisement on SNS as information initiative model.

III. SIMULATION

We have performed diffusion simulation to assess the performance of proposed model using large network of news communication web site. For comparison, we have also performed simulation using IC model.

The network is a directed graph, and has 30398 nodes and 87627 arcs. We chose one node which has enough diffusing capacity as the starting node. In the case of IC model, we set 20% diffusion probability for every arc, whereas in the case

 ISBN: 978-1-61208-756-6

TABLE I. ADDITIONAL INFORMATION PROPERTY

	Added Value
Facebook	Number of Likes
Twitter	Number of Re-tweets
YouTube	Number of Views
Text	Language
General information	Security parameter

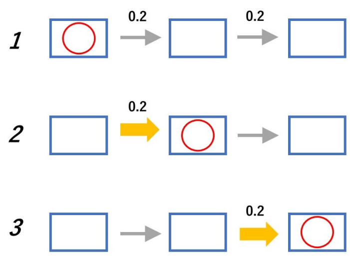

Figure 1. Concept of IC model

of IV model, the diffusion probability is increasing by logical work of every node which the information has passed.

Figure 3 shows that when the diffusion ends before it spreads widely, there is little difference between two models. The only difference is that threshold to cause explosive diffusion became high because of increasing IV. Next, Figure 4 shows the data of explosive diffusion. The simulation data demonstrates that IV changes the explosive diffusion wide and high frequency.

The second simulation has been designed to assess degree of the diffusion suppression per three types of centrality including degree centrality, page-rank, and personalized page-rank. Only personalized page-rank rates the closeness of the nodes to the starting node. We sorted all nodes according to each centrality, then decreased the probability of the nodes in order of the centrality. The simulation investigated how distance from the starting node influences to the centrality of the node in aspect of information diffusion. Results of the simulation, as illustrated in Figure 5, when we use personalized page-rank, information diffusion is well suppressed, and variation

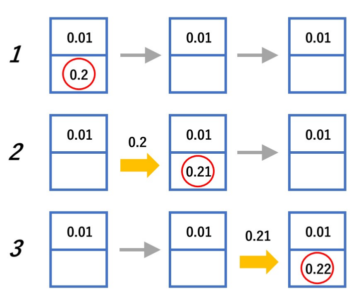

Figure 2. Concept of Information vitality model

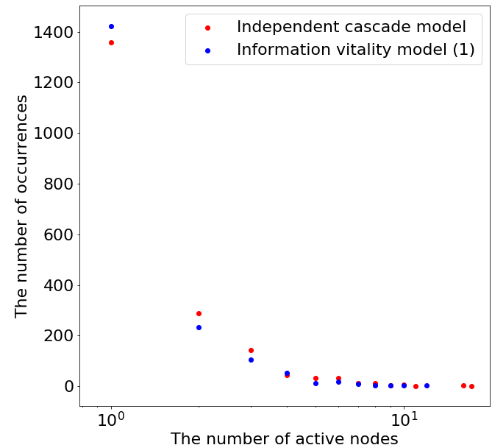

Figure 3. Scatter plot for small-scale diffusion (less than 1000 nodes)

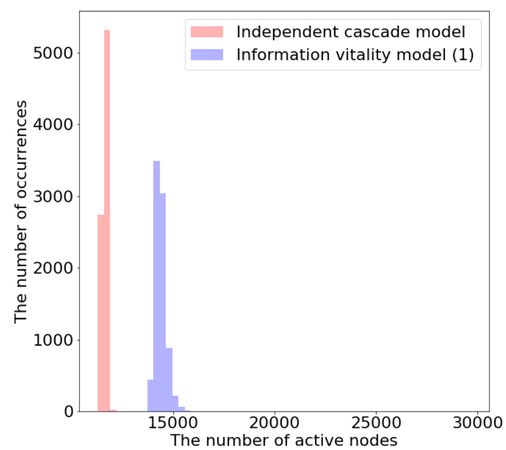

Figure 4. Histogram for Large-scale diffusion (more than 1000 nodes)

rate is the least in these types of centrality. It is reasonable to support that early stage of information diffusion has big chance to change conclusive wideness of diffusion, and it is more important to weaken the nodes which are close to the starting node than the nodes which have high degree.

IV. CONCLUSION

In order to make the model that explains information diffusion phenomena especially in social networks on the Internet, we quantified the added information value as IV. Then, we made the information diffusion model that the diffusion probability changes according to the diffusion process. As a result, it was found that in the case of small-scale diffusion with less than 1000 active nodes, the influence of added information value remained on the change of threshold leading to large-scale diffusion.

On the other hand, in the case of large-scale diffusion with 1000 or more active nodes, the range of diffusion increased more than the case that the added information value was not

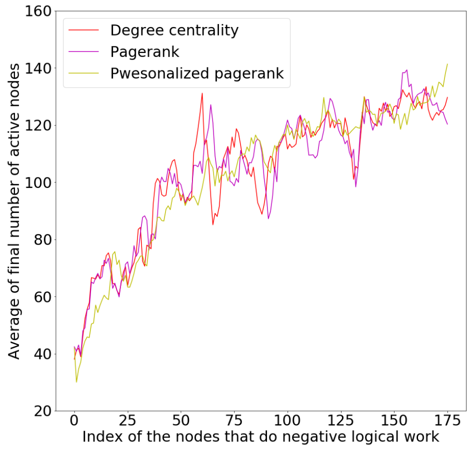

Figure 5. Diffusion suppression per each centrality

taken into account. Information such as the number of views on YouTube can be considered to have the effect to expand the range of spreads explosively after the video spread more than a certain range. In order to accurately model the actual information diffusion which is determined by human, more complicated algorithm will be necessary. We will examine and work on the construction of the more accurate information diffusion model.

REFERENCES

[1] S. Vosoughi, D. Roy, and S. Aral, The spread of true and false news online, Science: Vol. 359, Issue 6380, pp. 1146-1151, March 2018.

[2] K. Ohara, R. Odagiri, and S. Shirakawa, On a Topic-based Information Diffusion Model, SIG-KBS/SIG-KBS,B5(01) pp. 13-17, 2015.

[3] H. Watanabe, ”A Basic Theory of Information Network”, IEICE TRANS FUNDAMENTALS, VOL.E76-A, NO.3, pp.265-276, March 1993.