IEICE Proceeding Series

Analysis on network topology and dynamics of information diffusion

Akiyoshi Tanaka, Yutaka Shimada, Kantaro Fujiwara, Tohru Ikeguchi

Vol. 1 pp. 57-60
Publication Date: 2014/03/17
Online ISSN: 2188-5079

2012 International Symposium on Nonlinear Theory and its Applications
NOLTA2012, Palma, Majorca, Spain, October 22-26, 2012

Analysis on network topology and dynamics of information diffusion

Akiyoshi Tanaka†, Yutaka Shimada†, Kantaro Fujiwara†, and Tohru Ikeguchi†,‡

†Graduate School of Science and Engineering, Saitama University
255 Shimo-Ohkubo, Sakura-ku, Saitama-city, Saitama, 338-8570 Japan
‡Brain Science Institute, Saitama University
255 Shimo-Ohkubo, Sakura-ku, Saitama-city, Saitama, 338-8570 Japan
Email: {Tanaka, sima}@nls.ics.saitama-u.ac.jp, kantaro@mail.saitama-u.ac.jp, tohru@mail.saitama-u.ac.jp

Abstract—Information diffuses in real networks. In this paper, we investigated how information diffuses in complex networks and what are important factors in the information diffusion. To discuss this issue, we focused on two results: D. Watts and S. Strogatz showed that the information diffuse widely and quickly across a random network which contains many shortcuts (*Nature*, **343**, 440–442, 1998). On the other hand, D. Centola reported that the information diffuse widely and quickly across a lattice network which contains few shortcuts (*Science*, **329**, 1194–1197, 2010). We analyzed how the difference between the results is caused, introducing two hypotheses. First, assuming that network topology contributes greatly to the information diffusion, we analyzed the relations between network topology and information diffusion. Second, we focused on the dynamics of information diffusion, namely how to diffuse information in networks. We then proposed a simple model of the information diffusion. As a result, the network topology did not affect the information diffusion. However, our model can replicate Centola's result that the information diffuses widely and quickly in the lattice networks with few shortcuts. These results suggest that the dynamics of information diffusion affects the dynamics on networks more greatly than the network topology.

1. Introduction

In the real world, various phenomena occur by interactions between many components. If we regard the components as nodes and their interactions as links, we can describe many kinds of real systems as networks. For example, in a network of friendships, persons are nodes and their relationships are links. Because the number of nodes is large and their connections are complex in many real networks, these networks are called complex networks [1]. In the real complex networks, several kinds of information diffuse, for example, diseases, electric pluses in neural networks, electrical energy in power grids, and so on. Therefore, it is very important to clarify how the information diffuses in networks toward effective prevention against infectious diseases, understanding neural systems, and so on [2].

Here, we focused on two experimental results reported in Refs. [3] and [4]. In 1998, using a simple model of infectious diseases, D.Watts and S.Strogatz showed that information diffuses widely and quickly across random networks whose average path length between nodes are short. On the other hand, in 2010, D.Centola studied the information diffusion through social networking service on the Internet. As a result, the information diffused more widely and quickly across lattice networks whose average path length is longer than the random networks.

In this paper, we focused on the disagreement of these two results in Refs. [3] and [4], and we analyzed how these differences are originated. To solve this issue, we analyzed the information diffusion in terms of the network topology and the dynamics of information diffusion. We assume that the disagreement of two experimental results is caused by the difference in network topology or that in the dynamics of information diffusion. Then, we constructed a simple model of diffusion dynamics that is involved in Centola's experiments. Our results clearly show that the information diffusion does not depend on network topologies but on diffusion dynamics.

2. Network topology

In Ref. [3], Watts and Strogatz used a ring-lattice network (RLN) and generated the random networks by randomly rewiring links in the RLN. Because the degree of nodes changes after the random rewiring, we call this random rewiring degree-non-preserving-rewiring (DNPR). On the other hand, in Ref. [4], Centola used a hexagonal-lattice network (HLN) and generated the random networks by rewiring links in the HLN so that the degree of all nodes does not change after the random rewiring [5][6]. In this sense, we call this random rewiring degree-preserving-rewiring (DPR).

The RLN and the HLN have almost the same properties; namely nodes in the RLN and the HLN have the same degree, their average path length are long, and they have many clusters in which any three nodes are connected to each other. However, their network topology is slightly different (Fig. 1). In addition, because the processes of the rewiring are different in Refs.[3] and [4], random networks generated from the RLN and the HLN are different: a random network generated by DPR holds the degree of its original network but that by DPR does not. Then, we also investigated the clustering coefficients of the RLN, the HLN and random networks (The results are not shown). As a result, the clustering coefficients of the RLN and the HLN are high, and those of random networks are low. These results do not depend on whether the degree is held or not.

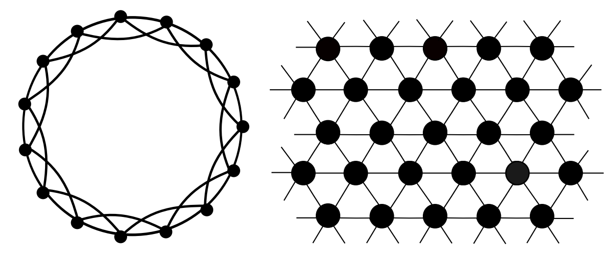


Figure 1: (a) ring-lattice-network [3]. (b) hexagonal-lattice-network [4].

3. Dynamics of information diffusion

3.1. Epidemic model

Using a simple epidemic model, Watts and Strogatz investigated how infectious diseases diffuse in the RLN and the random networks generated by DNPR [3]. At time $t = 0$, a node is selected from the RLN in which all nodes are healthy. The selected node is then infected. After that, infected nodes are permanently removed from the network after a period of sickness. During this time, they infect their healthy adjacent nodes with probability r. This process is repeated until all nodes are infected or the virus dies out.

3.2. Behavior in an online community

Centola studied the spread of a health behavior through a network-embedded population by creating an Internet-based health community website [4]. He gathered participants who were interested in health from other health-interest websites. The participants including Centola were randomly arranged to nodes in the HLN and a corresponding random network created from the HLN by RPD. At time $t = 0$, Centola send a message which contains information about health to his adjacent nodes. If the participants receive the message and have an interest, the message is automatically forwarded to their adjacent nodes. Once the message is forwarded, the corresponding participants never receive and send messages.

4. Analysis from view point of network topology

We first focus on topological differences between the networks used in Refs. [3] and [4]. Network topologies of both the RLN and the HLN are slightly different. The degree of nodes in the random networks generated by DPR is the same, but that by DNPR is not. We assume that these differences in the network topology affect information diffusion. In our experiment, we used the RLN, the HLN, and the random networks generated from the RLN and the HLN by DNPR and DPR. Thus, we used six networks in all. Applying the epidemic model to these networks, we investigated how the information diffusion depends on the difference of the network topology. The probability r is set to 0.5. The number of nodes N in the RLN and the HLN is 128, and the degree k of nodes is 6. We measured how many nodes are infected at time t by the diffusion rate R defined as follows:

$$R = \frac{1}{N}\sum_{i=1}^{N} a_i, \tag{1}$$

where N is the number of nodes in a network; if the ith node is infected, $a_i = 1$, otherwise $a_i = 0$.

We show the temporal changes of the diffusion rate in Fig. 2. The horizontal axis shows time t and the vertical axis shows the diffusion rate. The diffusion rate is averaged for one hundred initial states at each time. In each time, the disease typically diffuse across a wider range in the random networks (solid blue triangles) than in the RLN and the HLN (solid red circles). In addition, the disease diffuses more quickly in the random networks than the RLN and the HLN. Figure 2 indicates that the experimental results in Ref. [4] is not reproduced because the disease diffuses widely and quickly in the random networks. This suggests that the disagreement of two experimental results between Watts and Strogatz [3] and Centola [4] is not caused by the network topology.

5. Proposed model of information diffusion

Next, we focus on the diffusion dynamics, namely how the information is diffused dynamically. In the epidemic model of Watts and Strogatz and the experiments by Centola, when nodes receive information, the nodes send it to its adjacent nodes and thereby the information diffuses step by step. In this sense, their dynamics of information diffusion is similar. However, the nodes in the experiments by Centola have some typical features shown in the following subsections. Including these features into the epidemic model, we propose a new model and numerically reproduce the experiment conducted by Centola.

5.1. Intrinsic sending probability

When a participant receives a message, the probability that he/she forwards it to neighbors depends on every participant. The probability of each participant is defined as λ_i ($i = 1, \cdots, N$). In numerical simulations, we used three types of the probability λ_i: the fixed value, $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 0.5$, random numbers that obey the uniform distribution in $[0, 1]$, and those that obey the gamma distribution whose average μ and variance σ are

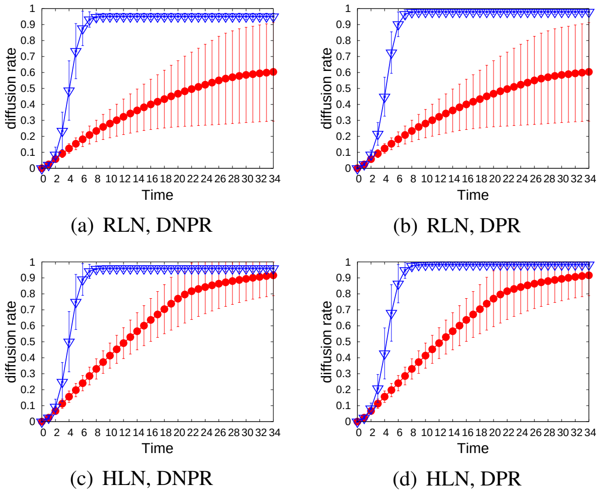


Figure 2: Temporal changes of the diffusion rate. The horizontal axis shows time and the vertical axis shows the diffusion rate (Eq. (1)). The solid red circles in (a) and (b) correspond to the RLN, and those in (c) and (d) correspond to the HLN. The solid blue triangles in (a) and (c) correspond to the random networks created by DNPR, and those in (b) and (d) correspond to the random networks by DPR. The vertical bars are the diffusion rate with the standard deviations of the diffusion rates.

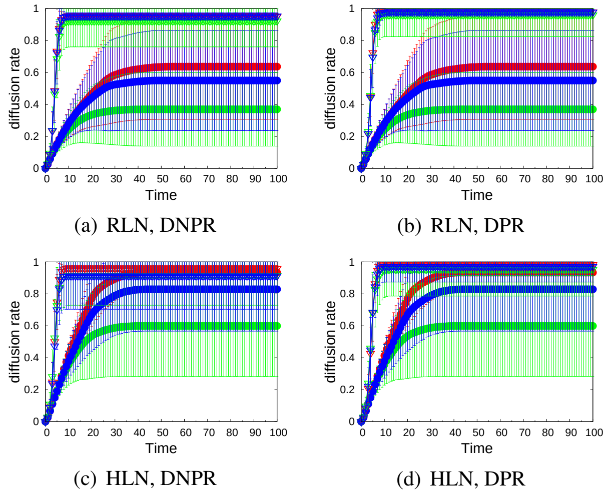


Figure 3: Temporal changes of the diffusion rate. The horizontal axis shows time, and the vertical axis shows the diffusion rate, where $\lambda = 0.5$ (red), the uniform distribution (blue), and the gamma distribution (green).

0.5 and 0.028. Then, we calculate the diffusion rate defined in Eq. (1), namely if the ith node has an interest to the message, $a_i = 1$, otherwise $a_i = 0$.

We show the temporal changes of diffusion rate in Fig. 3. From Fig. 3, we can see that depending on distribution, the diffusion rates in the RLN and the HLN are different, and that the messages diffuse less widely and less quickly in the RLN and the HLN than in the random networks for any case. In other words, the messages diffuse widely and quickly across the random networks regardless of whether λ is fixed or distributed. From these results, the distribution of λ does not affect the results that the information diffuses widely and quickly across the RLN and the HLN. In the following simulations, we fix λ for the sake of simplicity.

5.2. Dynamical sending probability

Centola reported that participants tend to have an interest in the received message if they receive two or more messages.To realize this, we introduce a new rule. If the node receives another message after receiving a message, its sending probability increases in $\lambda + \alpha$. In this way, we change the sending probability λ to $\lambda + \alpha$, so that we realize the above-mentioned dynamics.

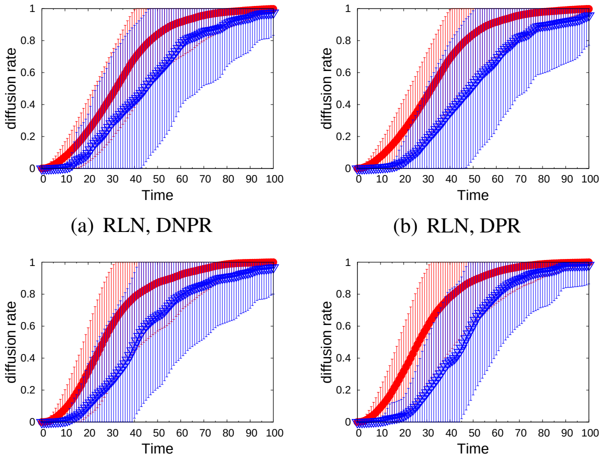
We set λ to 0.01 and α to 0.6 when we conducted the numerical experiments to realize significant difference between the case that a node receives only one message and that a node receives multiple messages. If α is too small, there is no advantage for receiving multiple messages, so that messages will diffuse widely and quickly across the random networks. We measured the rate of diffusion by Eq. (1).

The temporal changes of diffusion rates are shown in Fig. 4. The messages diffuse widely and quickly across the RLN and the HLN whose clustering coefficient is high. In particular, in Fig. 4(d), we used the HLN and its random network generated by RPD as well as in Centola's experiments. Fig. 4(d) indicates that the experimental results that Centola reported are replicated. However, the standard deviation of both the RLN and the HLN and the random networks take large values and the differences of the diffusion rate between the RLN and the HLN and the random networks are not statistically significant.

5.3. Elimination of nodes never sending messages

In our model, if a node receives messages, the node sends the messages to its adjacent nodes within a finite time period. However, it is natural to assume that nodes which have no interest in the messages never send them. To introduce this feature to our model, the nodes are forced not to send and receive the messages if they do not send the messages within a given time period T. The time period T depends on each node. We assign the time period T_i to the ith node. T_i is a normal random number whose average μ and variance σ are 15 and 1.6. If T_i is large, our models are essentially the same as the model that T is not introduced.

We show the temporal changes of diffusion rate in Fig. 5. The diffusion rate of the random networks and its standard deviation in Fig. 5 are lower than that in Fig. 4. Comparing with Fig. 4, the difference of diffusion rate between

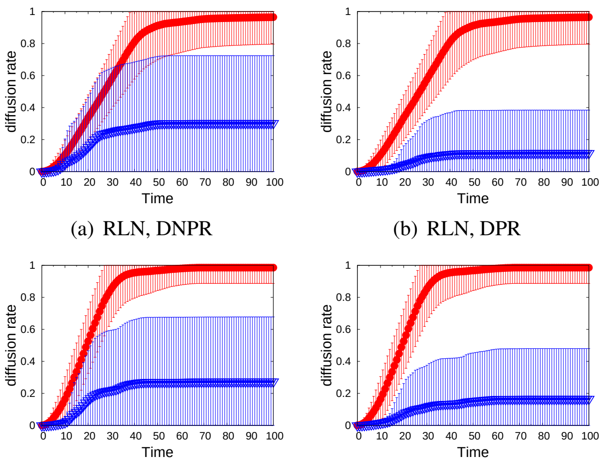


(a) RLN, DNPR (b) RLN, DPR

(c) HLN, DNPR (d) HLN, DPR

Figure 4: Temporal changes of the diffusion rate in the case that the nodes receiving two or more messages transmit more easily than those receiving only one message.

the RLN and the HLN and the random networks are statistically significant. Therefore, our model can reproduce Centola's results that the messages diffuse more widely and quickly across the RLN and the HLN than the random networks (Fig. 5(d)).



(a) RLN, DNPR (b) RLN, DPR

(c) HLN, DNPR (d) HLN, DPR

Figure 5: Temporal changes of the diffusion rate when the time period T is introduced.

6. Conclusion

In this paper, we studied the mechanism of information diffusion across the RLN and the HLN and corresponding random networks to investigate the disagreement between two results in Refs. [3] and [4]. In Refs. [3] and [4], the information diffusion in the RLN and the HLN and random networks is investigated. However, the networks are different in Refs. [3] and [4]. Based on the difference, first we applied the conventional epidemic model [3] to the RLN and the HLN and random networks. As a result, the results in Ref. [4] were not reproduced: the information diffused widely and quickly across the random network. Next, we proposed a new model of the information diffusion, introducing three properties into the conventional epidemic model: (1) the intrinsic sending probability of nodes that the nodes transmit information, (2) the dynamical sending probability that the nodes receiving two or more messages transmit more easily than those receiving only one message, and (3) the elimination of nodes which have not sent messages for a time period T. As a result, we showed that the intrinsic probability did not contribute to the results in Ref. [4]. However, by introducing the dynamical sending probability and the elimination of nodes, the information diffused widely and quickly across the RLN and the HLN. These results indicate that the disagreement between two results [3] and [4] is not caused by the network topology, but by the dynamics of information diffusion.

References

[1] R. Albert and A. L. Brabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, **74**, 47–94, 2002.

[2] D. Watts, "Small Worlds: The Dynamics of Networks Between Order and Randomness," *Princeton University Press*, 2003.

[3] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, **393**, 440–442, 1998.

[4] D. Centola, "The Spread of Behavior in an Online Social Network Experiment," *Science*, **329**, 1194–1197, 2010.

[5] S. Maslov and K. Sneppen, "Specificity and Stability in Topology of Protein Networks," *Science*, **296**, 910–913, 2002.

[6] S. Maslov and K. Sneppen and U. Alon, "Correlation profiles and motifs in complex networks," *Handbook of Graphs and Networks*, 168–198, 2003.