2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

Which vertices affect the spread of disease in temporal networks?

Koshi Abe†, Yutaka Shimada‡, Kantaro Fujiwara‡ and Tohru Ikeguchi‡

†Department of Management Science, Graduate School of Engineering, Tokyo University of Science
‡Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science
6-3-1 Niijuku, Katsushika, Tokyo 125-8585, Japan
Email: koshi.abe@hisenkei.net

Abstract—To prevent outbreaks of infectious diseases, it is one of the effective approaches to analyze mathematical models of the infectious diseases. To model the spread of the infectious diseases in realistic situations, contact networks of individuals have attracted much attention in recent years. In this paper, using a simple mathematical model of the infectious disease, we investigate the spread of the infectious disease on the real contact networks observed from person-to-person interactions recorded by radio frequency identifier at a hospital and a high school. We also investigate how removal of vertices with high centralities from networks affect the spread of the infectious diseases. As a result, we find that removal of the vertices with the highest number of contacts is the most effective method for suppressing the spread of the infectious diseases in the hospital. On the other hand, in the high school, the closeness centrality-based removal of vertices is more effective than the strategy based on the number of contacts.

1. Introduction

Several common features of real networks including the Internet, gene networks [1], economic systems, and face-to-face interactions between individuals [2] have been clarified. The small-world [3] and the scale-free [4] properties are known as the universal properties.

In previous research, information diffusion and spread of diseases have been discussed on static networks whose structures does not change with time [5]. However, the structures of real networks always change with time in the realistic situation [6].

We investigate how the infectious disease spreads on real temporal networks whose structures change with time, using the data of person-to-person interactions recorded by the radio frequency identifier (RFID) in a hospital and a high school [7]. In Ref. [8], we have reported that the infectious disease more widely spread on a real hospital network than a real high school network. This result in Ref. [8] is caused by the differences between the structure of the hospital and the high school networks. From this result, we can infer that an effective method for preventing undesirable outbreaks is different between the hospital network and the high school network. In this paper, we investigate how we can effectively slow down the spread of the infectious disease by removing vertices from the temporal networks. We also identify vertices which are most influential in the temporal networks.

As a result, removal of the vertices with the highest number of contacts can effectively suppress the spread of the infectious disease in the hospital network. On the other hand, removal of the vertices with the highest degrees and those with the highest closeness centralities is more effective method for preventing the infectious diseases from diffusing in the high school network than the removal of vertices with the highest number of contacts.

2. Data

In this paper, we used the data of face-to-face contacts of individuals recorded by SocioPatterns [7, 9, 10]. The data are classified into two types. The first data were observed at a hospital in Lyon, France, and the second were observed at a high school in the Lycée Thiers, Marseilles, France. The contacts between individuals were recorded every 20 seconds.

In the data of the hospital, subjects participating in the experiments were 29 patients and health care workers including 27 nurses and nurses' aides, 11 medical doctors, and 8 administrative staff members. The number of the subjects was 75. The data were collected over five days(see Ref. [9] for details). The number of contacts in the hospital was 32,424 during the five days.

In the data of the high school, the subjects were 126 high school students. The data were collected over four days (see Ref. [10] for details). The number of contacts in the high school was 28,561 during the four days.

3. Methods

We investigate which vertices are influential in the diffusion of the infectious disease on temporal networks by numerical simulations. We first constructed networks from the real data and then applied a mathematical model of the infectious disease to these networks. The edge between the vertex i and the vertex j at time t is described by $l_{ij}(t) \in \{0, 1\}$, where if a contact exists, $l_{ij}(t) = 1$, otherwise $l_{ij}(t) = 0$. A state of the vertex i at time t is described by $S_i(t) \in \{0, 1\}$, where if the vertex i is infected with the infectious disease, $S_i(t) = 1$, otherwise $S_i(t) = 0$.

In this experiment, we first choose the vertex i randomly and change its state into an infected state $S_i(t_1) = 1$, where t_1 is the time when the vertex i contacts the other vertices at the beginning in these data. We also define an infection rate r as the probability that a state of a susceptible vertex changes to an infected state. The number of infected subjects at time t is $I(t)$ $(t = 0, 20, \cdots, T)$, where T is the time when the infectious disease disappears from the networks. The ratio of the number of infected subjects at time t to the total number of subjects is $P(t) \equiv I(t)/N$, where N is the number of subjects. When the vertex i contacts with the vertex j under the situation where $S_i(t) = 1$ and $S_j(t) = 0$ at time t, the infectious disease is transmitted from the vertex i to the vertex j with the probability $rl_{ij}(t)$. When the vertex j is infected at time t ($S_j(t) = 1$), the vertex j has the infectious disease during a fixed period τ. After the period τ, the state of the vertex j changes to the recovered state. Recovered vertices are not infected again, and they do not transmit the infectious disease to other vertices.

We first investigate how the infectious disease spreads on the contact networks by $P(t)$. We next try to suppress the spread of the infectious disease by removing a few vertices from the networks. The number of removed vertices is defined by R. We removed R vertices based on their characteristic features: degrees, the number of contacts, the closeness centrality, the betweenness centrality, and the eigenvector centrality [11]. We calculate these centralities from the temporally aggregated static networks by using all contact data. The closeness centrality c_i of the vertex i is defined as follows:

$$c_i = \frac{N-1}{\sum_{j=1;j\neq i}^{N} d(v_i, v_j)}, \tag{1}$$

where $d(v_i, v_j)$ is the shortest path length between the vertices i and j. The betweenness centrality b_i of the vertex i is defined as follows:

$$b_i = \frac{\sum_{i_s=1;i_s\neq i}^{N} \sum_{i_t=1;i_t\neq i}^{N} \frac{g_i^{(i_s i_t)}}{N_{i_s i_t}}}{(N-1)(N-2)/2}, \tag{2}$$

where $N_{i_s i_t}$ is the total number of the shortest paths between the vertices i_s and i_t, and $g_i^{(i_s i_t)}$ is the number of the shortest paths passing through the vertex i out of all the shortest paths between the vertices i_s and i_t. The eigenvector centrality of the vertex i is described as u_i, and $\boldsymbol{u} \equiv (u_1, u_2, \cdots, u_N)^T$. Let A be an adjacency matrix of a network. Let λ_m be the mth eigenvalue of $A(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N)$. Then, the eigenvector centrality is given by

$$\lambda_1 \boldsymbol{u} = A\boldsymbol{u}, \tag{3}$$

that is, the vector of the eigenvector centralities is the eigenvector of A corresponding to the maximum eigenvalue λ_1.

4. Results

4.1. The relation between the recovery time τ and the normalized number of infected subjects $P(T)$

We first conducted experiments under the condition that $r = 0.01$. We calculated the final ratio of the number of infected subjects to the total number of subjects $P(T)$. Figure 1 shows how $P(T)$ changes when the recovery time τ changes. From Fig. 1, $P(T)$ in the case of the hospital is higher than that of the high school in all values of the recovery time τ. We supposed that there are two causes of the different tendency of the spread of the infectious disease between the hospital and the high school. One is the difference between the number of contacts in the hospital and that in the high school. The other is the difference between structural properties in the hospital and that in the high school. In Ref. [8], we have investigated how the infectious disease spreads on temporal networks. As a result, the infectious disease easily spreads when the number of contacts is large, and the structural properties affect the diffusion of the infectious disease. From these results, we infer that the properties of the influential vertices in the hospital network and in the high school network are different from each other.

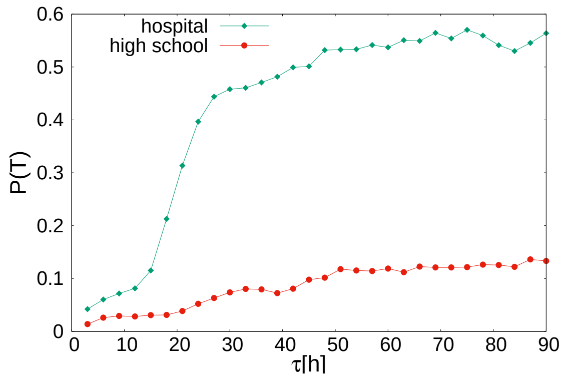
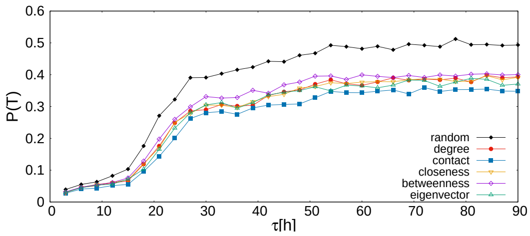


Figure 1: The relation between the recovery time τ and the ratio of the number of infected subjects $P(T)$ at time T. The range of τ is $3 \leq \tau \leq 90$[h]. The infection rate r is 0.01. The results are averaged over 1,000 trials.

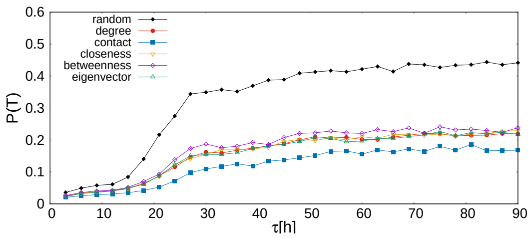
4.2. Preventing the spread of the infectious disease by the removal of vertices

We here try to prevent the infectious disease from widely spreading over networks by removing vertices based on the centrality measures. The degree, the number of contacts, the closeness centrality, the betweenness centrality, and the eigenvector centrality of each vertex are calculated. Then, we remove R vertices based on these centrality measures, that is, we remove R vertices with the highest values of these centrality measures.

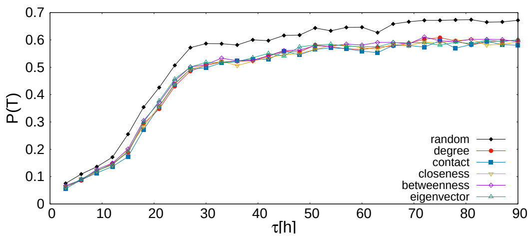
Figures 2 and 3 show how $P(T)$ changes when the recovery time τ changes at the hospital and the high school. In Figs. 2 and 3, we remove R vertices with the highest values of centralities which are the degree (red lines), the number of contacts (blue lines), the closeness centrality (yellow lines), the betweenness centrality (purple lines), and the eigenvector centrality (green lines). The black lines show the results for which vertices are randomly removed. From Fig. 2(a), (b), and (d), removal of the vertices with the highest number of contacts achieve the lowest values of $P(T)$ in all values of τ. On the other hand, from Fig. 2(c), the results are almost the same as each other. This is because the number of removed vertices is too small to slow down the spread of the infectious disease in spite of the high infection rate r. From these results, in the hospital, removing vertices with the highest number of contacts is the most effective method for slowing down the spread of the infectious disease.



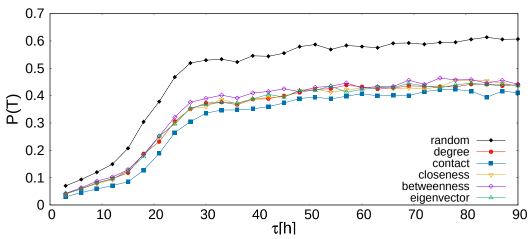
(a) $r = 0.01, R = 5$



(b) $r = 0.01, R = 10$

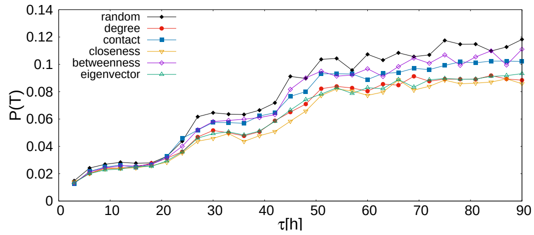


(c) $r = 0.02, R = 5$

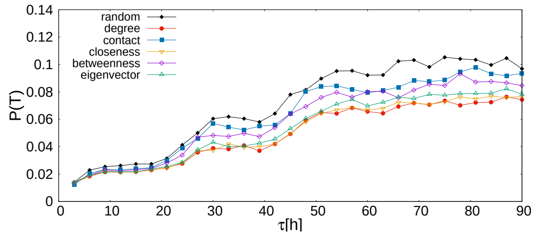


(d) $r = 0.02, R = 10$

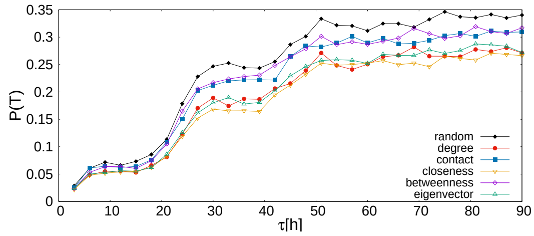
Figure 2: The relation between the recovery time τ and the ratio of the number of infected subjects $P(T)$ at time T, in the case of the hospital. The results are averaged over 1,000 trials.



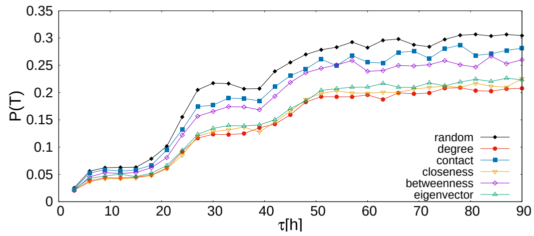
(a) $r = 0.01, R = 5$



(b) $r = 0.01, R = 10$



(c) $r = 0.02, R = 5$



(d) $r = 0.02, R = 10$

Figure 3: The relation between the recovery time τ and the ratio of the number of infected subjects $P(T)$ at time T, in the case of the high school. The results are averaged over 1,000 trials.

Figure 3 shows how $P(T)$ changes when the recovery time τ changes at the high school. Figure 3(a) and (c) is the result of removing five vertices ($R = 5$). Figure 3(b) and (d) is the result of removing ten vertices ($R = 10$). From Fig. 3, removing the vertices with the high degrees, with the high closeness centrality, and with the high eigenvector centrality leads to the good results that we can effectively slow down the diffusion of the infectious disease. In contrast to the hospital, removal of the vertices with the highest number of contacts is not so effective. Then, removing vertices with the high closeness centrality or those with the high degree are the effective method for slowing down the spread of the infectious disease in the high school.

5. Conclusion

We constructed temporal networks by using the data of person-to-person interactions recorded by RFID [2]. We investigated how the infectious disease spreads on the temporal networks by using a simple mathematical model of the infectious disease on the temporal networks. As a result, we found that the infectious disease spreads more widely in the hospital than in the high school. From this result, we conjectured that the effective methods for slowing down the spread of the infectious disease at the hospital are different from those at the high school.

We then investigated which vertices mainly affect the spread of the infectious disease. As a result, the vertices with the highest number of contacts significantly accelerate the spread of the infectious disease in the hospital, but does not in the high school. In contrast to the hospital, the vertices with the highest closeness centralities and with the highest degrees mainly contribute to the spread of infectious disease in the high school.

Acknowledgment

The authors would like to thank Mr. S. Ishii and AGS Corp. for their kind encouragements on this research. The research of Y.S. is supported by Grant-in-Aid for Research Activity Start-up (No. 26880020) and by Grant-in-Aid for Young Scientist B (16K16126) from JSPS. The research of K.F. was partially supported by Grant-in-Aid for Young Scientist B (16K16138) and by Grant-in-Aid for Challenging Exploratory Research (No 15K12137) from JSPS. The research of T.I. was partially supported by Grant-in-Aid for Exploratory Research (No. 24650116) from JSPS.

References

[1] U. Alon, "An introduction to systems biology - design principles of biological circuits," CRC Press, 2006.

[2] A. Barrat, C. Cattuto, V. Colizza, J-S. Pinton, W. V. Broeck, and A. Vespignani, *PLOS ONE*, **5**, e11596, 2010.

[3] D. J. Watts and S. H. Strogatz, *Nature*, **393**, 440–442, 1998.

[4] A-L. Barabási and R. Albert, *Science*, **286**, 509–512, 1999.

[5] D. Centola, *Science*, **329**, 1194–1197, 2010.

[6] P. Holme and J. Saramäki, *Physics Reports*, **519**, 97–125, 2012.

[7] http://www.sociopatterns.org/

[8] K. Abe, Y. Shimada, K. Fujiwara, and T. Ikeguchi, *Proceedings of International Symposium on Nonlinear Theory and its Applications (NOLTA2015)*, 337–340, 2015.

[9] P. Vanhems, A. Barrat, C. Cattuto, J-S. Pinton, N. Khanafer, C. Régis, B. Kim, B. Comte, and N. Voirin, *PLOS ONE*, **8**, e73970, 2013.

[10] J. Fournet and A. Barrat, *PLOS ONE*, **9**, e107878, 2014.

[11] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, *Advances in Physics*, **56**, 167–242, 2007.