Analysis on temporal effects in time-varying social networks

Yutaka Shimada* and Tohru Ikeguchi

Graduate School of Engineering Department of Management Science, Tokyo University of Science
* yshimada@ms.kagu.tus.ac.jp

Abstract—Recent researches on real networks greatly focus on the temporal properties of real evolving networks, that is, how real networks change their structural properties with time. We here analyzed structural and temporal properties in real social networks using a real e-mail network, and observed that the e-mail network has some characteristic properties simultaneously. Focusing on these properties, we propose a model which can generate networks with these characteristic features.

1. Introduction

A network, or a graph, is one of effective tools to describe various real systems. To analyze real networks, several measures that quantify their structural properties have been proposed, and thereby one can obtain structural and dynamical properties of the real systems (see for example [1]). Recent researches on real networks greatly focus on the temporal properties of real evolving networks [2], such as human mobility patterns [3, 4], e-mail activity patterns [5, 6], and citation patterns in scientific publications [7, 8]. In these researches, temporal properties in many real systems have been successfully revealed, and various models have also been proposed based on the empirical results towards effective prevention of the spread of infective diseases and e-mail warms to humans and computers, and effective transmission of various types of information over networks.

In this paper, we analyzed structural and temporal properties of a real e-mail network and observed that the e-mail network has some characteristic properties simultaneously: a bursting property in the interevent time between two e-mails of an e-mail user [5, 6], scale-free distributions of the number of sent and received e-mails [9, 10], and a moderate correlation between the number of sent e-mails and that of received e-mails. Focusing on these properties, we propose a model which can generate networks with these three characteristic features.

2. E-Mail network and its properties

We here use an e-mail network which is obtained from e-mail exchange recodes collected in Keil university over a observation period of 112 days [10]. Nodes in the e-mail network are users who have intrinsic e-mail addresses. We only use e-mail addresses of users who have sent messages at least once during the observation period. Links in the e-mail networks are communications between users: when node $v(i)$ send a message to other node $v(j)$, $v(i)$ and $v(j)$ are connected by a directed link. The e-mail network consists of $n = 59812$ e-mail addresses which are classified into internal and external accounts. The internal accounts correspond to e-mail addresses of local students in Keil university, and the external accounts correspond to all other observed e-mail addresses. In this paper, we only use the internal accounts because the number of sent and received messages by the internal accounts are exactly observed, whereas those by the external accounts are partially observed and thus its data is incomplete. In this case, the number of nodes n is 1296.

Let $s_i(t)$ be the number of massages sent by a node $v(i)$ and $r_i(t)$ be the number of massages received by the node $v(i)$ until time t. The e-mail networks have two well-known properties. The first one is that the distributions of $s_i(t)$ and $r_i(t)$ show a power-law behavior. Both distributions of $s_i(t)$ and $r_i(t)$ are shown in Figs.1 and 2. The e-mail network that we used here also shows the power-law behavior. The second one is a bursting property in an interevent time τ between two e-mails sent by the same e-mail user. Figure 2(a) shows the distributions of the interevent time τ_s between two e-mails sent by the same e-mail user, and Fig. 2(b) shows the distributions of the interevent time τ_r between two e-mails received by the same e-mail user. Both distributions of τ_s and τ_r show power-law decay in the middle and high values of τ_s and τ_r, which implies the existence of the bursting in sending and receiving messages.

Although the above features are important and well-known properties of the e-mail activity patterns, the correlation between the number of sent and received messages is also important to characterize the structural property of real networks. We here focus on the correlation between the number of sent messages $s_i(t)$ and the received messages $r_i(t)$ of the same user. The correlation coefficient between $s_i(t)$ and $r_i(t)$ is 0.43 in the case of the e-mail network, and this result indicates that the e-mail network has a moderate correlation between the send and received number of messages.

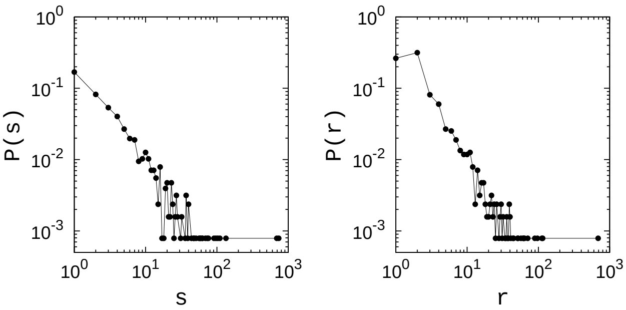


Figure 1: Distributions of the number of sent e-mails (left) and received e-mails (right) in the double logarithmic scale.

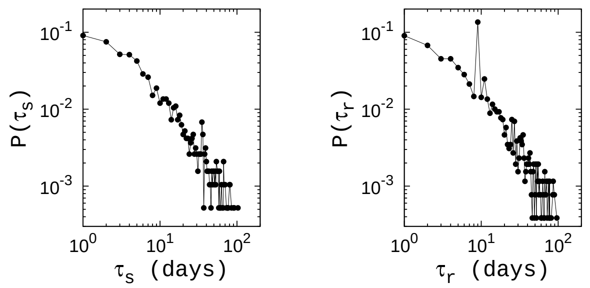


Figure 2: (a) The distribution $P(\tau_s)$ of the interevent time between two e-mails sent by a user, and (b) the distribution $P(\tau_r)$ of the interevent time between two e-mails received by a user in the double logarithmic scale.

3. Model

Even though these properties are typical characteristic features, a model realizing these properties simultaneously has not been proposed. Based on the empirical results shown in the previous section, we propose a model which realizes both structural and temporal properties of the e-mail network. At time $t = 0$, all n nodes do not have sent messages and received messages, namely $s_i(0) = 0$ and $r_i(0) = 0$, $(i = 1, \ldots, n)$. At each time step t, one message is sent from one node to other node. A sender node $v(i)$ is chosen with a probability proportional to the following probability:

$$S(i,t) = \rho \frac{s_i(t) + \kappa}{\sum_{l=0}^{n}[s_l(t) + \kappa]} + (1-\rho)\frac{r_i(t) + \kappa}{\sum_{l=0}^{n}[r_l(t) + \kappa]}, \tag{1}$$

where κ is an initial activity that enables nodes to send and receive a message at time $t = 1$. The nodes cannot send and receive any messages at time $t = 1$ without κ, because both $s_i(0)$ and $r_i(0)$ are zero. The parameter ρ determines the significance of $s_i(t)$ and $r_i(t)$ in the probability $S(i,t)$. The first term and the second term in the right-hand side of Eq. (1) are the same as the preferential attachment rule [9] with respect to the number of sent and received messages.

A receiver node $v(j)$ which is different from the sender node is chosen with a probability proportional to the following probability:

$$R(i,t) = \rho \frac{r_i(t) + \kappa}{\sum_{l=0}^{n}[r_l(t) + \kappa]} + (1-\rho)\frac{s_i(t) + \kappa}{\sum_{l=0}^{n}[s_l(t) + \kappa]}. \tag{2}$$

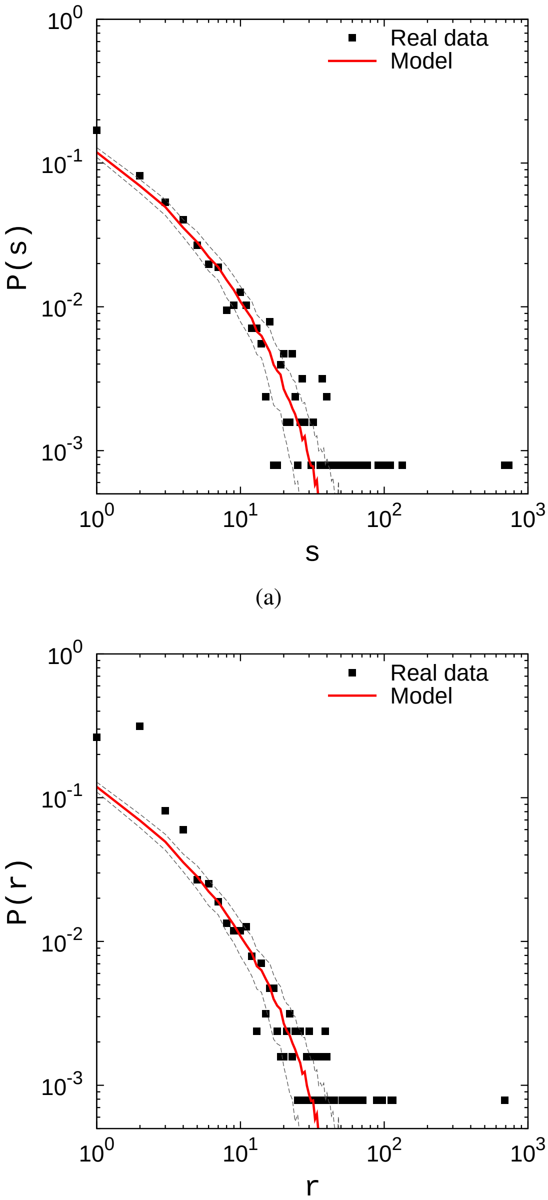
Equations (1) and (2) indicate that we can change the correlation between the number of sent messages and that of received messages, varying the value of ρ. For example, if $\rho = 1$, the sender node is determined only by the number of sent messages of each node, while the receiver node is determined by the number of received messages of each node. In this case, the correlation between $s_i(t)$ and $r_i(t)$ disappears. On the other hand, if $\rho = 0$, the probability of sending message $S(i,t)$ is determined only by $r_i(t)$ and that of receiving probability $R(i,t)$ is determined only by $s_i(t)$. Then the resultant correlation between $s_i(t)$ and $r_i(t)$ becomes high. The reason is as follows: when the node $v(i)$ is chosen as a sender, $s_i(t)$ increases and the sender node $v(i)$ acquires a better position to receive messages. In the same manner, when the node $v(i)$ is chosen as a receiver, $r_i(t)$ increases and the receiver node $v(i)$ acquires a better position to send messages. These processes are repeated, and thus the correlation between $s_i(t)$ and $r_i(t)$ increases.

In numerical simulations, we first determined the parameter κ and ρ such that the correlation coefficient between $s_i(t)$ and $r_i(t)$ of the model takes the similar value to that of the real e-mail network, namely 0.43, as much as possible. We set κ to 0.15 and ρ to 0.85 in which the correlation coefficient is 0.44 that is the average over 100 networks generated from our model. We set the number of nodes n to 1296 and increased the time step t until the number of sent (received) messages reaches to the similar number of sent messages in the real e-mail networks.

4. Results and discussions

Figure 3 shows the distributions of the final number of sent messages $s_i(t)$ [Fig.3(a)] and that of received messages $r_i(t)$ [Fig.3(b)] obtained from real data (black rectangles) and from our model (red lines). The networks generated from our model agree with the real networks in the small and middle number of sent messages s ($1 \leq s \leq 60$) in Fig.3(a), and in the middle number of received messages r ($3 \leq r \leq 60$) in Fig.3(b). Although our model cannot generate hub nodes which have the great number of sent and received messages from Fig.3, we find that these nodes show strange behavior and it might be a rare event to observe such nodes. For example, the activity of the node with the highest number of sent messages is very low and the number of days when the node is active is about 2 days. However, amazingly, the node sent 669 messages during just two days. Our model cannot describe this behavior as shown in Fig.3, but such behavior might be a rare event even in a realistic situation.

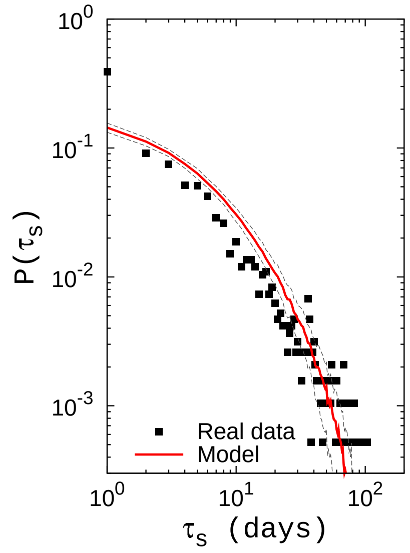
Figure 4 shows the distributions of τ_s [Fig.4(a)] and those of τ_r [Fig.4(b)] obtained from real data (black rectangles) and from our model (red lines). The networks generated from our model agree with the real data, especially in the tail of the distributions well. Although it is known that the distribution of the interevent times in the e-mail networks is well approximated by a fat tailed distribution $P(\tau) \sim 1/\tau$, in our cases, the distributions of the interevent times obtained from the model might not obey the power law, but the exponential distribution. In this sense, our model is incomplete, and we expect that the e-mail network still has important properties hidden from our model.



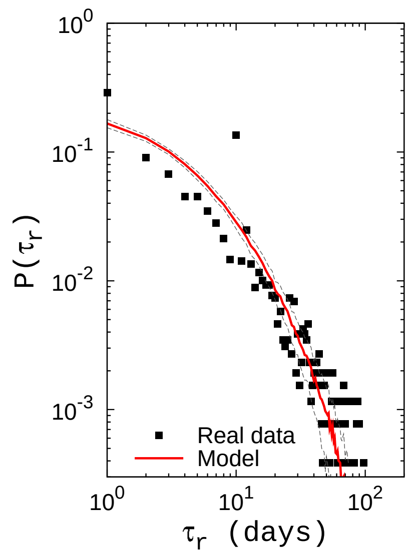
(a)

(b)

Figure 3: Distributions of the number of sent and received e-mails obtained from the real e-mail network (black rectangles) and our model (red line). (a) The distribution of the number of sent e-mails. (b) The distribution of the number of received e-mails. We generated 100 networks from our model and calculated the average of the distributions of the number of sent and received e-mails over these 100 networks, and the gray dashed lines show $\pm\sigma$ where σ is the standard deviation.



(a)



(b)

Figure 4: (a) Distribution $P(\tau_s)$ of the interevent time between two e-mails sent by a user. (b) Distribution $P(\tau_r)$ of the interevent time between two e-mails sent by a user. The black rectangles are obtained from the real e-mail network, and the red lines are obtained from our model. We generated 100 networks from our model and calculated the average of the distributions of the interevent time over these 100 networks, and the gray dashed lines show $\pm\sigma$ where σ is its standard deviation.

5. Conclusions

In this paper, we investigated both structural and temporal properties of real e-mail networks and showed that the e-mail network has three characteristic properties simultaneously, that is, the bursting property in the interevent time between two e-mails, scale-free distributions of the number of sent and received e-mails, and the moderate correlation between the number of sent e-mails and that of received e-mails. We then proposed a model to satisfies these properties simultaneously and showed that our model agrees with the real e-mail network well.

The research of T.I. is supported by Grant-in-Aid for Exploratory Research (No. 24650116) from JSPS.

References

[1] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, (2010).

[2] P. Holme and J. Saramäki, *Temporal Networks*, Understanding Complex Systems, Springer-Verlag Berlin Heidelberg (2013).

[3] L. Isella et al., Journal of Theoretical Biology, **271**, 166-180, (2011).

[4] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, Physical Review Letters, **110**, 168701 (2013).

[5] A. Vazquez et al., Physical Review E, **73**, 036127 (2006).

[6] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási, Physical Review Letters, **98**, 158702 (2007).

[7] M. Medo, G. Cimini, and S. Gualdi, Physical Review Letters, **107**, 238701 (2011).

[8] M. Golosovsky and S. Solomon, Physical Review Letters, **109**, 098701 (2012).

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

[10] H. Ebel and L.-I. Mielsch, and S. Bornholdt, Physical Review E, **66**, 035103 (2002).