Title	A Study on the Dynamics of Friendship Network Formation Using a Directed Network Model
Author(s)	Nakada, Toyohisa; Kato, Yoshihiko; Kunifuji, Susumu
Citation	
Issue Date	2007-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/4080
Rights	
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, KICSS 2007 : The Second International Conference on Knowledge, Information and Creativity Support Systems : PROCEEDINGS OF THE CONFERENCE, November 5-7, 2007, [Ishikawa High-Tech Conference Center, Nomi, Ishikawa, JAPAN]

A Study on the Dynamics of Friendship Network Formation Using a Directed Network Model

Toyohisa Nakada **Yoshihiko Kato**[1] **Susumu Kunifuji**
School of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST)
{t-nakada,y-katou,kuni}@jaist.ac.jp

Abstract

In this paper, we discuss the dynamics of friendship network formation by the analysis of real networks and simulations. First, we survey real networks that comprise students in a graduate school and developed within two months of admission. From the results, we find that reciprocity and sibling bias have a considerable effect on the creation of a friendship network. Second, we perform computer simulations to construct a friendship network that is similar to a real network by using these biases. Because we can construct a similar friendship network only by using these biases, we believe that the simulation can be used as a model for representing real friendship networks.

Keywords: Friendship network, Dynamics, Directed Network

1 Introduction

Many researches for enhancing real world communities (human networks) have been proposed in recent years. For example, Kautz et al. [1] proposed a visualization system to show the relations between researchers based on published documents. Sumi et al. [2] have developed a support system to enhance real human networks in an academic society by enabling interactions between the real world and cyber space with a handheld device and/or a public display. These researches are based on an assumption that we can benefit from the enhancement of human networks that we are a part of and/or those we intend to join. However, the important point is that it is not clear how a human network grows naturally. A human network will increase or decrease the relations between people, which are normally called links of the network, even if support systems for enhancing interactions between people do not exist. We believe that it is necessary to understand the natural growth of a human network for distinguishing the effects of a support system from a network without a support system.

The purpose of this study is to survey human network growth. First, we attempt to determine the characteristics of growth from observed data, which are obtained from questionnaires on the friendships of graduate students. Second, since the obtained characteristics may not be irrelevant to the growth of human networks, computer simulation is performed to construct a similar network from only some of the obtained characteristics. If a similar network can be constructed from only some of the obtained characteristics, it is considered that they are the most important properties required for the growth of the network.

This paper is organized as follows. Section 2 describes related works that comprise the history of human network analysis and some measurements to understand networks such as ”small world” and ”scale free.” The analysis of a human network is described in section 3 and the simulation to construct a similar network is described in section 4. Finally, we conclude this study in section 5.

2 Related Work

In this section, we discuss the history of human network analysis and recent significant researches on networks such as ”small world” and ”scale free.”

2.1 Psychology and sociometry

In psychology, many researchers have focused on obtaining an insight into human characteristics such as personality and profile. These properties are sometimes used in simulations to construct human networks, for example, [3].

On the other hand, there are many researches for understanding human networks that employ a visualization technique called sociometry [4; 5]. The represented network is called a sociograph. Since the motive behind using this technique is

[1] presently at Alpine Information System, Inc.

to help understand human networks using the visualization, it does not appear to be reliable for the analysis of the structure.

2.2 Network analysis

In the field of graph theory research, the random graph theory [6; 7] was proposed in the 1950s. In the random graph theory, links are connected to other nodes randomly. Since the theory is not adaptable to friendship networks, another model called the biased network model [6; 8; 9] was also proposed simultaneously. The word ”bias” means that a link is connected based on the network structure instead of being connected randomly. Some types of biases were proposed; for example, ”reciprocity bias” in directed networks is a link having a tendency such that a node A often connects back to node B that is already connected to A. In ”sibling bias,” two nodes are often connected when they are already connected to another node. When these two biases appear simultaneously, the resulting bias is called ”sibling reciprocity bias.” Furthermore, in ”inbreeding bias,” two nodes that are in a group are often connected to each other.

In the 1960s, Milgram’s experiment [10] was performed. The results showed a connection with people in six steps in the U.S. In the 1970s, Granovetter [11] stated that weak ties enable reaching populations and audiences who are not accessible via strong ties. Based on the principle ”the rich get richer,” [12] Price [13] also showed a relation model for references in research papers. The model has become a base for the ”scale free” model. In the 1980s, an exponential random model [14] was proposed. The generalization presented in [15] enables us to quantitatively estimate the characteristics of a network such as triad.

In recent years, Watts et al. [16] presented the WS model to construct a small world phenomenon. Barabashi et al. [17] have developed the BA model to construct a network having a scale-free feature. Moreover, a model representing a network having scale-free and high-clustering coefficients was proposed [18]. Many other models have been proposed [19] in recent years.

2.3 Analysis of friendship network based on recent research on network analysis

After the WS and BA models were proposed, many measurement parameters of network structures were developed. For example, distance, clustering coefficient, link distribution, etc., were proposed. Based on the measurements, new models representing friendship networks have been proposed. Jin et al. [20] have developed a model in which two people who have a mutual friend often have a chance encounter with each other. The model proposed by Davidson et al. [21] that represents the recommendation of one friend to another and the lifetime of humans shows a small-world and a scale-free phenomenon on the network. The Vertex fitness model [22] handles a symmetric function that estimates the probability of two friends contacting each other. The scale-free phenomenon appears by the manipulation of the probability. Newman [23] developed a model in which two people in a group often connect to each other. The connection using the nearest-neighbor model [24] involves two people who are nearest neighbors often connecting with each other. All the models described here are based on undirected networks.

In contrast, based on the directed network, Garlaschelli et al. [25] developed a measurement that enables us to compare the ratio between directed and undirected links. Caldarelli et al. [26] represented an e-mail send/receive network as a weighted directed network. The network showed a scale-free feature.

2.4 Differences between our study and related works

We investigate friendship networks from the viewpoint of recently developed measurements such as distance, clustering coefficient, link distribution, and link reciprocity and biases proposed by Fararo [8]. Moreover, these biases are used to construct a network similar to real networks. The simulation is like a biased model in which each bias has a probability of emerging. From another viewpoint, the simulation is interpreted as an exponential random graph model with dynamics. A model representing network dynamics using biases has not been reported to date.

Table 1. Observed data

Organization	School of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST).
Participants	Graduate students who have just entered the shool.
Period	Every week from April 2005 to May 2005. 8 weeks for group A and 7 weeks for group B.
Group	People who are attending a lecture form a group. Two groups (two lectures) are investigated. The number of members in group A and group B are 26 and 16, respectively. The groups do not overlap.
Method	Questionnaire that enquires about the friendship among the people in a particular group.

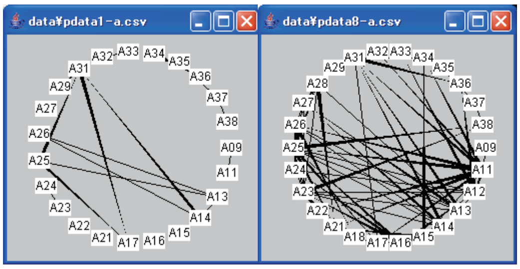

Figure 1. Start-up network (left) and the network at the end of the survey (right) in group A.

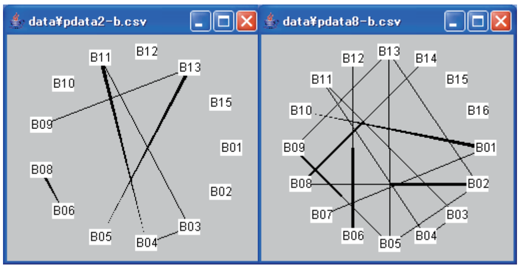

Figure 2. Start-up network (left) and the network at the end of the survey (right) in group B.

3 Analysis of real network data

In this section, we describe our data that is to be analyzed and its characteristics.

3.1 Data Overview

Table 1 shows an overview of the surveyed data. Figures 1 and 2 show the start-up network and the network at the end of the survey. A person and his/her friendship are represented as a node and a link, respectively. The network is a directed network. A link with a uniform thickness is a bi-directed link, which means that friendship exists between the two connected people. On the other hand, a link-like arrow indicates a single-directed link, which means a person who is positioned at the source of the link has friendship but the other person has not acknowledged. Moreover, the statistics are shown in figures 3 and 4.

3.2 Measurement

In this study, we first considered three measurements developed on the basis of researches on the small-world and scale-free networks. These measurements are the mean distance between two nodes, the clustering coefficient, and the link distribution. After the WS and BA models were proposed, these measurements have often been used to understand network structures. With regard to the distance, when there is an unlinked node in a network, the distance becomes infinity. This measurement does not appear to be useful because we cannot compare networks that often have unlinked nodes. Thus, we use the inversed distance instead of the distance in this study, i.e., the distance is replaced with the inversed distance because the former becomes infinity if there is an unlinked node. This measurement deals with unlinked nodes [19]. With regard to the clustering coefficient, some researchers have proposed that it deals with directed networks. For example, Park et al. [27] have proposed the measurement in order to investigate the relationship of influences among nodes. In this study, since we focus on friendship networks, we assume that single-linked people are not friends. Thus, the clustering coefficient is calculated on a network in which all single links are removed.

Furthermore, since our observed data is for a directed network, link reciprocity [25] measurement is also employed. This measurement enables us to estimate the ratio of single and bi-links.

The equations for the abovementioned mea-

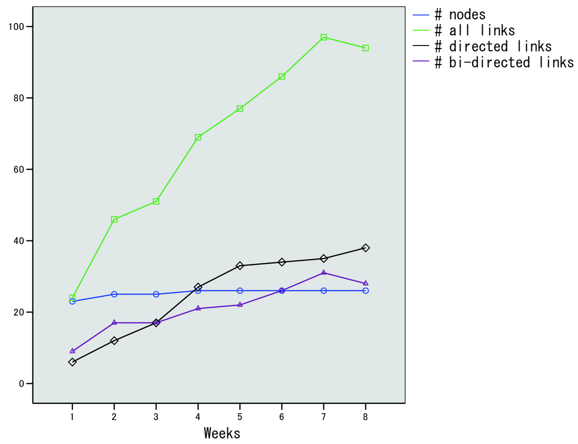

Figure 3. Number of nodes and links in group A

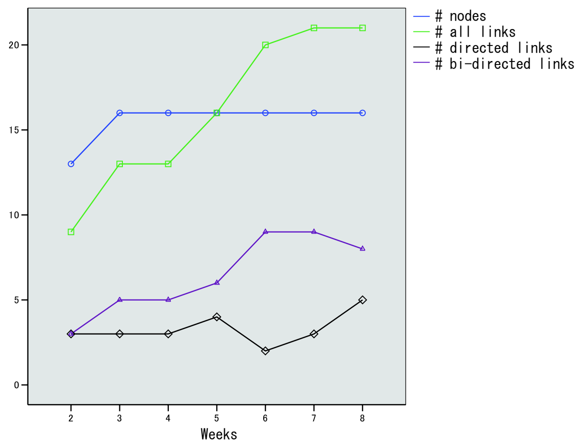

Figure 4. Number of nodes and links in group B

surements are given by equations (1) to (4).

$$L^{-1} = \frac{1}{n(n-1)} \sum_{i,j \in V, i \neq j} d_{ij}^{-1} \quad (1)$$

$$C = \frac{1}{n} \sum_{i \in V} C_i,$$

$$C_i = \begin{cases} \frac{2B_i}{k_i(k_i-1)} & k_i \geq 2 \\ 0 & k_i \leq 1 \end{cases} \quad (2)$$

$$p(i) \sim i^{-\gamma} \quad (3)$$

$$LR = \frac{B/E - D}{1 - D},$$

$$D = \frac{E}{n(n-1)} \quad (4)$$

Here, L^{-1} denotes the inversed distance; C, the clustering coefficient; n, the number of

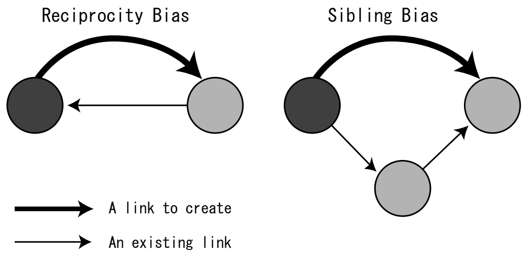

Figure 5. Reciprocity and sibling bias

nodes; V, a set of nodes; d_{ij}^{-1}, the smallest distance from node i to j; k_i, the number of nodes that are mutually connected to node i (called nearest-neighbor node); and B_i, the number of bi-links among node i's nearest-neighbor nodes. $p(i)$ is the link distribution; the network is called a scale-free network if and only if the distribution follows the power law. LR denotes the link reciprocity measurement value; B, the number of bi-links; E, the number of all links (a bi-link is counted twice); and D, the density of the network.

3.3 Reciprocity and sibling bias

Although we introduced some measurements proposed in recent researches in the previous subsection, we consider other viewpoints in this study. These are the two biases shown in figure 5. When we consider that a network structure is one of the causes for adding a new link, the nearest-neighbor nodes may be the most relevant to the structure. From a node that will be added to a new link to a node that will add a new link, the closest nearest-neighbor is a case of reciprocity bias, as shown on the left-hand side of the figure. In the reverse direction, the nearest neighbor becomes a case of sibling bias, as shown on the right-hand side of the figure. It is reasonable to suppose that these two biases are the simplest biases in the viewpoint that a new link is created based on the structure of the network.

3.4 Statistics of networks

Figures 6 and 7 show the inversed distances, clustering coefficients, and link reciprocities. Since the high clustering coefficient and high inversed distance are observed in two groups, it seems that the networks are a small-world phenomenon.

Red line in figure 11 shows the link distribu-

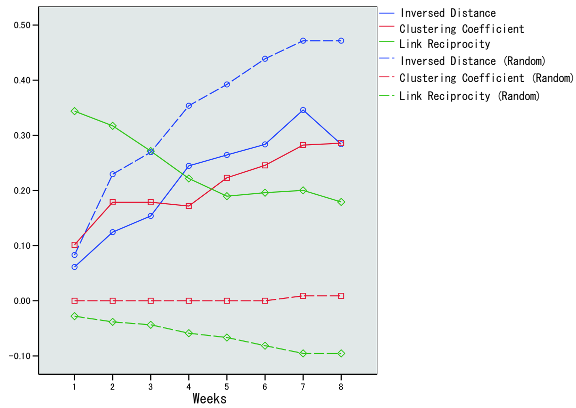

Figure 6. Inversed distance, clustering coefficient, and link reciprocity in group A

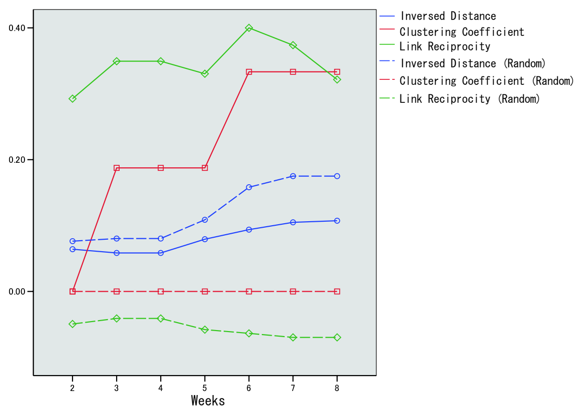

Figure 7. Inversed distance, clustering coefficient, and link reciprocity in group B

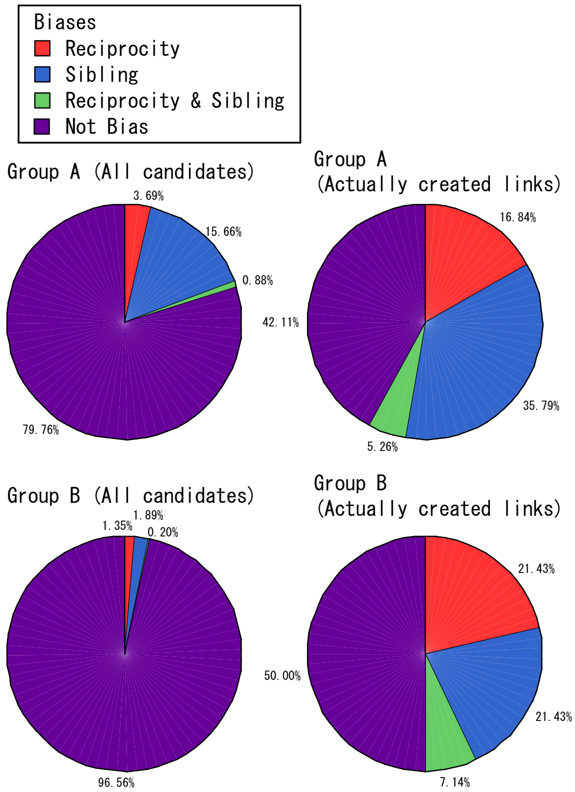

Figure 8. Ratio of types of biased links

tions. Although the power law distribution is drawn as a line in a graph on the log scale, the graphs are not on the log scale. Thus, the network cannot be called "scale free."

Furthermore, figure 8 shows the distribution of the biases. The left-hand sides of the graphs show all the candidates for new links, and the right-hand sides show the actually created links. It appears that all the biases are often selected to create a new link.

4 Simulation for creating similar networks

In the previous section, we showed some characteristics of real human network dynamics. In this section, we would like to clarify causal relations, i.e., which characteristics are the causes and which characteristics are the effects. We used five measurements that include the inversed distance, clustering coefficient, link distribution, link reciprocity, and biases. It is possible that the biases become causes because the other measurements cannot be operated directly. Thus, in this section, we perform a simulation to construct a network having characteristics similar to real data networks by using only biases. To clarify the causal relations, a support system for activating human networks must be designed.

4.1 A model to represent network dynamics

Our simulation is performed according to the following steps. The values of n_0, n_t, and m are equal to those of real networks.

1. Add n_0 nodes to the network.

2. Add m links to the network based on the probability calculated by equation 6.

3. Add n_t nodes having no links to the network if necessarily.

4. Repeat steps 2 to 3 an arbitrary number of times.

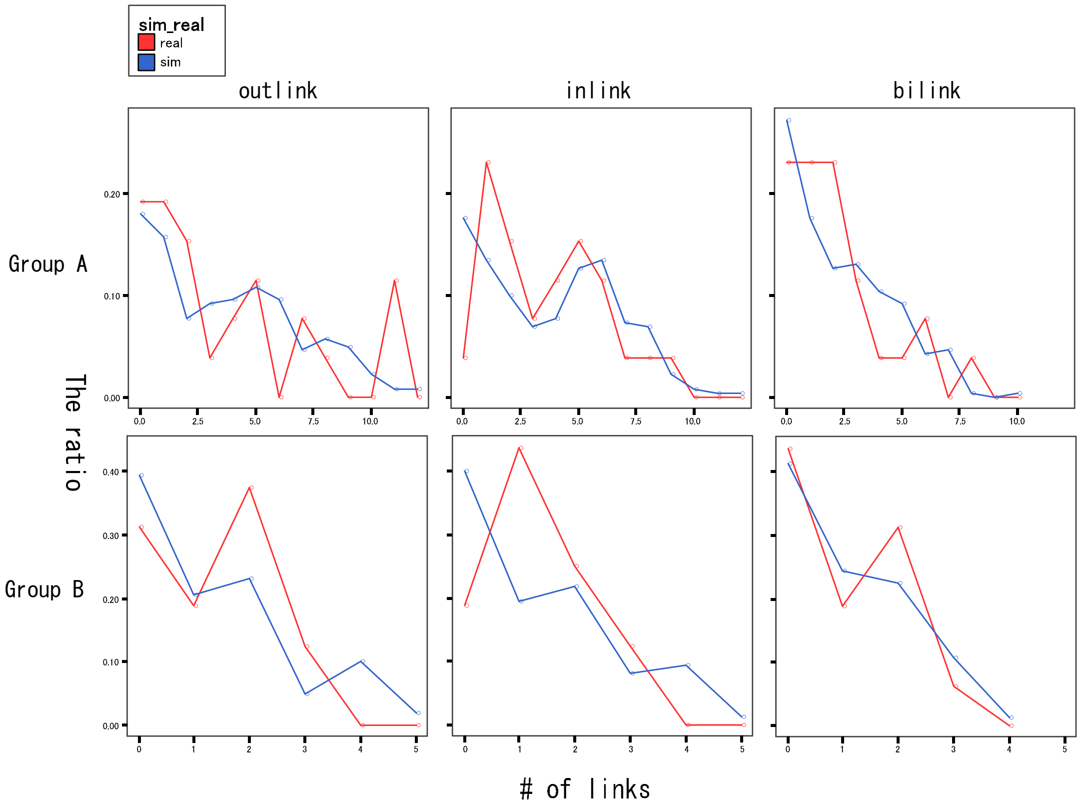

Figure 11. Link distributions of the real and simulated networks at the end of the survey

$$
\begin{aligned}
cad(x \rightarrow y) &= \theta_{rep} * rep(x \rightarrow y) + \\
&\quad \theta_{sib} * sib(x \rightarrow y) + \\
&\quad \theta_{nob} * nobias(x \rightarrow y), \\
&\quad \theta_{rep} \geq 0, \theta_{sib} \geq 0, \\
&\quad \theta_{nob} \geq 0, \\
&\quad \theta_{rep} + \theta_{sib} + \theta_{nob} = 1 \quad (5) \\
P(x \rightarrow y) &= cad(x \rightarrow y) / \\
&\quad \sum_{i \in X} \sum_{j \in Y} cad(i \rightarrow j) \quad (6)
\end{aligned}
$$

where $rep(x \rightarrow y)$, $sib(x \rightarrow y)$, and $nobias(x \rightarrow y)$ denote the reciprocity bias, sibling bias, and absence of bias from nodes x to y, respectively. θ_{rep}, θ_{sib}, θ_{nob} are the simulation parameters.

4.2 Comparison between the simulated and real networks

Figures 9 and 10 show the characteristics of the real and simulated networks and their parameters. Since the simulation was performed ten times while changing the seed of the random function, the characteristics of the simulated networks have means and standard deviations that are presented as lines through the mean points and which the values are in case arc. The figures show that the values of the three measurements are similar for both the networks. The values of real networks are almost within the standard deviations. On the other hand, The link distributions are shown in figure 11. The figure shows these distributions are similar relative to a diagonally right-down line. Since the characteristics of simulated networks are similar to those of real networks, it is reasonable to suppose that the simulation can re-create a real network by selecting appropriate parameters.

4.3 Discussion of the simulation parameters

With regard to θ_{nob}, it is considered that the bias is a link caused by a factor other than the network structure. Such a link is called a "shortcut" in the WS model. It appears that the shortcut causes a network to possess a small-world. Moreover, the shortcut is called a weak tie in Granovetter's paper [11]. The weak tie plays important role for reaching new job opportunity. Thus, it is likely

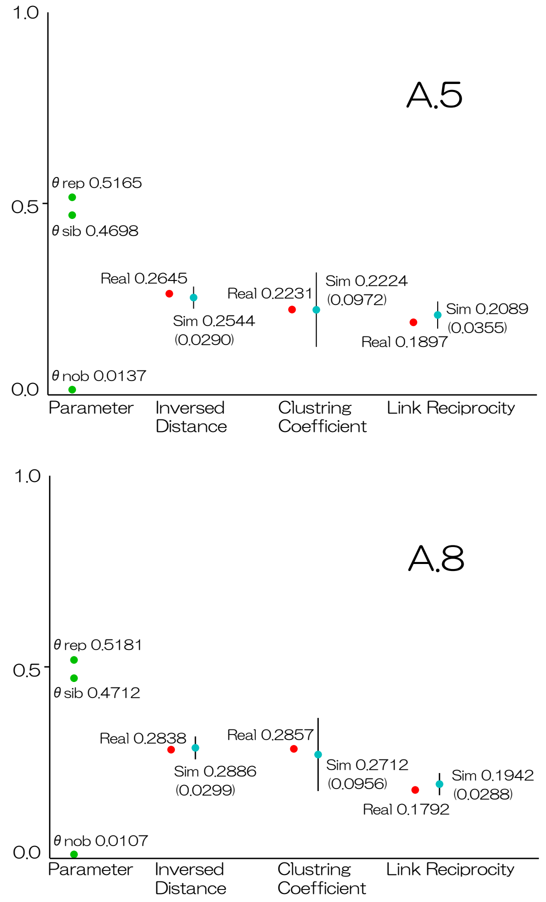

Figure 9. Simulation result in group A

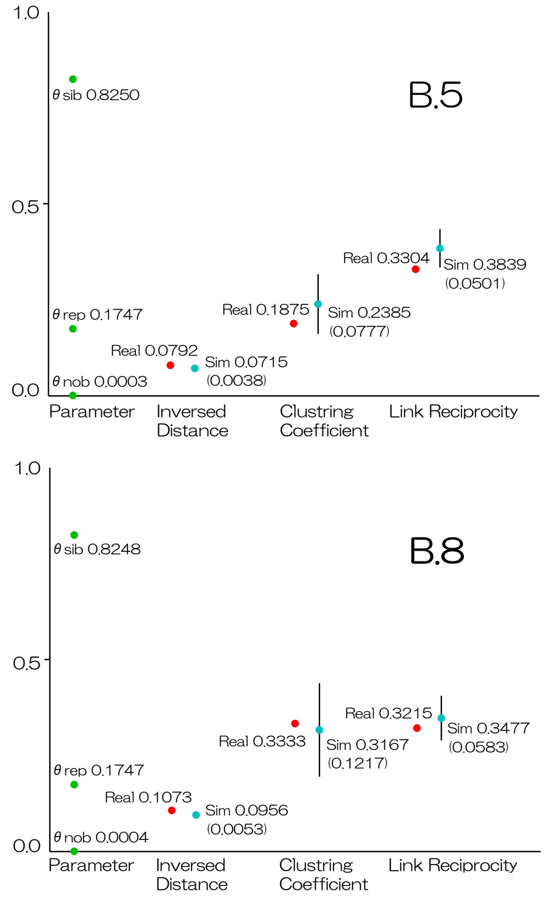

Figure 10. Simulation result in group B

that the bias creates new interactions.

As discussed above, it may be presumed that when θ_{nob} is large, the network will be activated in the future. In fact, θ_{nob} of group A, which was activated more often than group B, is considerably high. Moreover, θ_{nob} of group A 5 weeks later is greater than that of the same group 8 weeks later. It may represent that human community often gets down its activity while time passing.

For designing a support system for activating human networks, θ_{nob} may be important. We believe that it is possible to effectively activate human networks when the system attempts to create biased links.

5 Conclusion

We analyzed the growth of human networks and simulated the networks using a two biased links and a non-biased link. Since the resultant networks are similar to the real networks on the basis of recently developed measurements, we believe that the bias composition is important for the growth of a human network. For example, if we observe the biases, it may be possible to estimate the network growth. The biases may become one of the design concepts to develop a support system for activating human networks. We have a plan to gather more real data and improve the sophistication of our model.

Acknowledgment

Our research is partly supported by the fund from Ministry of Education, Culture, Sports, Science and Technology, JAPAN, under the name of Cluster for Promotion of Science and Technology in Regional Areas.

References

[1] Kautz, H. A., Selman, B., and Shah, M.: The Hidden Web, AI Magazine, Vol. 18, No. 2, pp.27-36, (1997)

[2] Sumi, Y. and Mase, K.: Digital Assistant for Supporting Conference Partici-

pants: An Attempt to Combine Mobile, Ubiquitous and Web Computing, Ubicomp 2001, LNCS 2201, pp.156-175, Springer, (2001)

[3] Zeggelink, Evelien, P. H., Stokman, F. N., Van De Bunt, and Gerhard, G.: The Emergence of Groups in the Evolution of Friendship Networks, Journal of Mathematical Sociology, Vol. 21(1-2), pp. 29-55, (1996)

[4] Rapoport, A. and Horvath, W. J.: A Study of a Large Sociogram, Behavioral Science, Vol. 6, pp.279-291, (1961)

[5] Guldner, C. E., Stone-Winestock, and Penny: The Use of Sociometry in Teaching at the University Level, Journal of Group Pshchotherapy, Psychodrama & Sociometry, Winter 95, Vol. 47, Issue 4, pp.177-185, (1995)

[6] Rapoport, A.: Contribution to the Theory of Random and Biased Nets, Bulletin of Mathematical Biophysics, Vol. 19, pp.257-277, (1957)

[7] Erdös, P., and Rényi, A.: On Random Graphs, Publicationes Mathematicae, Vol. 6, 290-297, (1959)

[8] Fararo, T. J. and Sunshine, M.: A Study of a Biased Friendship Network, Syracuse University Press, Syracuse, (1964)

[9] John, S., Thomas, J. F., and Filip, A.: Advances in Biased Net Theory: Definitions, Derivations, and Estimations, Social Networks, Vol. 26, pp.113-139, (2004)

[10] Milgram, S.: The Small World Problem, Psychology Today, Vol. 2, pp.60-67, (1967)

[11] Granovetter, M. S.: The Strength of Weak Ties, American Journal of Sociology, Vol. 78, pp.1360-1380, (1973)

[12] Herbert, A. S.: On a Class of Skew Distribution Functions, Biometrika, Vol. 42, No. 3/4, pp.425-440, (1955)

[13] Price, D. J. de S.: A General Theory of Bibliometric and Other Cumulative Advantage Processes, Journal of the American Society for Information Science, Vol. 27, pp.292-306, (1976)

[14] Paul, W. H. and Samuel L.: An Exponential Family of Probability Distributions for Directed Graphs, Journal of the American Statistical Association, Vol. 76, No. 373, pp.33-50, (1981)

[15] Frank, O. and Strauss, D.: Markov Graphs, Journal of the American Statistical Association, Vol. 81, 832-842, (1986)

[16] Watts, D. J.: Networks, Dynamics, and the Small World Phenomenon, American Journal of Sociology, Vol. 105, pp.493-592, (1999)

[17] Barabási, A.-L., Albert. R., and Jeong, H.: Mean-field Theory for Scale-free Random Networks, Physica A, Vol. 272, pp.173-187, (1999)

[18] Petter, H. and Beom, J. K.: Growing Scale-free Networks with Tunable Clustering, Phys. Rev. E, Vol. 65, 026107, (2002)

[19] Newman, M. E. J.: The structure and Function of Complex Networks. SIAM Review, Vol. 45, No. 2, pp.167-256, (2003)

[20] Emily, M. J., Michelle, G., and Newman, M. E. J.: Structure of Growing Social Networks, Phys. Rev. E, Vol. 64, 046132, (2001)

[21] Davidsen, J., Ebel, H., and Bornholdt, S.: Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks, Phys. Rev. Lett., Vol. 88 128701, (2002)

[22] Caldarelli, G., Capocci, A., De Los Rios, P., and Muñoz, M. A.: Scale-Free Networks from Varying Vertex Intrinsic Fitness, Phys. Rev. Lett., Vol. 89, No. 25, (2002)

[23] Newman, M. E. J.: Properties of Highly Clustered Networks, Phys. Rev. E, Vol. 68, 026121, (2003)

[24] Vazquez, A.: Growing Networks with Local Rules: Preferential Attachment, Clustering Hierarchy and Degree Correlations, Phys. Rev. E, Vol. 67, 056104, (2003)

[25] Diego, G. and Maria, I. L.: Patterns of Link Reciprocity in Directed Networks, Phys. Rev. Lett., Vol. 93, 268701, (2004)

[26] Caldarelli, G., Coccetti, F., and De Los Rios P.: Preferential Exchange: Strengthening Connections in Complex Networks, Phys. Rev. E, Vol. 70, 027102, (2004)

[27] Sung, M. P. and Beom, J. K.: Dynamic Behaviors in Directed Networks, Phys. Rev. E, Vol. 74, 026114, (2006)