ISSN 1346-5597

NII National Institute of Informatics

NII Technical Report Oct 2002

Proposal of the extension network model for personal human network by using neighborhood matchmaker

Masahiro Hamasaki and Hideaki Takeda

NII-2002-005E
Oct 2002

URL: http://www.nii.ac.jp/ Tokyo 101-8430, JAPAN +81-3-4212-2000 FAX: +81-3-3556-1916

Proposal of the extension networks model for personal human networks by using neighborhood matchmaker

Masahiro HAMASAKI †, Hideaki TAKEDA ‡†

† *The Department of Informatics, The Graduate University for Advanced Studies,*
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN
‡ *National Institute of Informatics,*
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, JAPAN
†*hamasaki@grad.nii.ac.jp* ‡*takeda@nii.ac.jp*

Abstract

In this paper, we propose the neighborhood matchmaker model as a network extension model for personal human network. Personal human network is useful for information gathering, but in order to adapt various needs for information gathering, it is necessary to extend and optimize it. Using the neighborhood matchmaker model, we can increase a new friend who is expected to share interests via all own neighborhoods on the personal human network. Iteration of matchmaking is used to optimize personal human networks. We simulate the neighborhood matchmaker model with the real data and the virtual data and compare the results by our model with those by the central server model. The neighborhood matchmaker model can reach almost the same results obtained by the sever model with each type of data.

Keywords Internet Agents Technology, Collaboration Technology

1 Introduction

Information exchanging among people is one of powerful and practical ways to solve information flood because we can act intelligent agents to collect, filter and associate necessary information. We exchange information with other people who are connected to with our personal human network. If we need variable information to exchange, we must have a good human network. In this paper, we propose a network extension model called "Neighborhood Matchmaker Model". It can form optimized networks from the arbitrarily given networks.

This paper is organized as follows. In Chapter 2, we show how to make relationship for information exchanging. In Chapter 3, we show how a personal human network can be optimized using the proposed model. Explanation of the experiments is shown in Chapter 4. The results of the experiments are shown in Chapter 5, and this paper concluded in Chapter 6.

2 How to make relationship for information exchanging

There are two different types in information exchange, i.e., mail is peer-to-peer type, and Mailing List and BBS are client/server type. When using mails, we exchange information with persons who are known already or known by homepage or others. So, we can select easily who is appropriate to exchange information from relationship with us. In case of Bulletin Board System (BBS) and Mailing List, we exchange information via shared space. So we can get easily new persons to exchange information with us.

The weak point of the former is difficulty to get a new partner. In this type, we just exchange information with the known people. However, there is a limit of a number of the known people, so that it is difficult to make a best environment for information exchanging. And the weak point of the latter type is difficulty to select a new partner. We have a chance to know a good person for information exchanging among the participants. But we should select persons from possibly a huge number of participants. Since it is very intelligent and time-consuming task, it is difficult for us.

There are some systems to support to form communities and exchange information. Kautz et al. [1] emphasized importance of human relations for WWW and showed done primary work for finding human relations, i.e., their system called ReferralWeb can find people by analyzing bibliography database. Sumi et al. [2][3] supported people to meet persons who have same interests and share information using mobile computers and web applications. Kamei et al. supported to form communities using visualization relationship among participants[4].

In these systems, they assume a group as a target either explicitly or implicitly. If the group is bigger, it is more likely to contain valuable persons to exchange information. However, we have to make more efforts with these systems in order to select such persons from a lot of candidates in the group.

And it is difficult for us to organize and manage such the large group. If some persons who have same interests are in the group, we might be able to be supported by these systems. However, it is difficult to organize a good group for that purpose.

ReferralWeb uses high quality information such as bibliography database, and has realized high precision support. Sumi et al. applies their system to groups such as a meeting for the study and visitors of a museum, and reported that it has good enough effect. However, anyone cannot use both systems. Community Organizer supports to visualize relationship between participants and do not consider making a group with good members.

If we can need better relationship for information exchanging, we must meet and select partners more and more. So, we need information exchanging systems to have a method that realizes the above two requirements i.e. how to meet and select new partners. In our real life, we often get new friends via our friends. Since we do not know new friends before meeting them, we have no ways to select them. But one of our friends selects a new person for us, because s/he knows both. Friends work as matchmaker for new friends. We propose a new model to expand network by using this idea.

3 Neighborhood Matchmaker Model

In this chapter, we introduce a model that can optimize a personal human network by adapting using the method in our real life. We call that method "Neighborhood Matchmaker Model (NMM)" after this. Before explaining NMM, we define the network model for this problem. At first we define person as node, and connection for information exchanging between people as path. Here we assume that we can measure a degree of connection between two nodes (and we call it "connection value" after here). Then, we can define that making a good environment for information exchanging is optimizing this network. In NMM, the network is optimized by matchmaking of neighbor nodes.

If we use this model automatically on an information exchanging system, there are some conditions. These conditions needed as follows.

- All nodes can connect to each other
- All nodes can calculate relationship between nodes connected to them

Under this condition, each node can change connections to others in order to connect better nodes autonomously. The behavior of a node is as follows.

1. Each node calculates connection values between its neighbor nodes.

2. it finds some pairs of nodes which have good enough connection values, it recommends both nodes of the selected pairs to others.

3. A node that receives recommendation decides whether it accepts the recommendation or not. If it accepts the recommendation, it adds the recommended node to its neighbor nodes.

We can get personal human network by iteration of these behaviors. Figure 1 shows these behaviors. In the next chapter, we test this model with simulations.

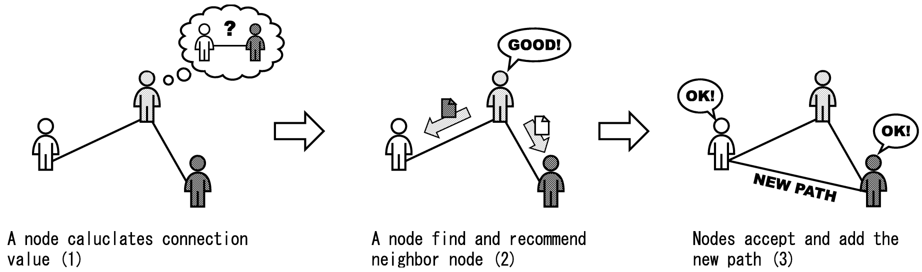


Figure 1: behavior of nodes

4 Experiments

4.1 The Procedure of the simulation

The main objective of the experiment is how our proposed model is useful to optimize a human network. We simulate optimization with NMM using the real data and the virtual data. All nodes of the network can calculate connection values between their neighbor nodes, select some of pairs as recommendation, and recommend these pairs to nodes of these pairs. A node can judge whether connect to recommended node or not. In this simulation, a node takes a simple tactics. Each node wants to connect to other nodes that have better connection values i.e., if a new node is better in connection than the worst existing node, the former replaces the latter.

Figure 2 shows the flow chart of this simulation. At first, we create nodes each of which has some data to represent a person. In this experiment, the data is a 10-dimensional vector or WWW bookmark data. Next, we put paths between nodes randomly. However, we fix the number of paths for simulation. The number of initial paths is a control parameter in this experiment. All paths have connection values between nodes. A new path must be better than the least of current paths. If all nodes cannot get a new path using matchmakers, the network is converged. Thus the simulation ends. In this simulation, a single turn consists of selection of a node and selection of paths at that node. At first, all nodes calculate connection values between own neighbor nodes. Next, a node that may exchange path is selected randomly. Only this node can exchange paths in this turn. The selected node selects the best path from recommended nodes and the worst path it has. The selected node replaces the worst one to the best one if the connection value of former is higher than the latter. All nodes that have more than two paths works as matchmaker, i.e., they calculate connections between their neighbor nodes whether the node can exchange paths. If all nodes cannot exchange paths, the simulation ends.

Figure 3 shows an example of exchanging path. In figure 3, node A has currently four paths to node B, node C, node D, and node F. The paths recommended to A can are node E and G, since neighborhood matchmakers i.e., node B and F, recommend these nodes. Note that connection value of recommended paths can be calculated by these matchmakers, not by node A, because A has no information node B and G before recommended. The neighbor nodes know each node, so they can calculate these connection values.

Node A judges a new node with that information. At first, node A chooses the worst path among nodes that it has. In this example, the worst one is "0.2". Next, node A chooses the best path recommended by matchmakers. In this case, the path from node G is best. So node A exchanges these paths, i.e., node A deletes a path from B and creates a path from G. It should be noted that a single node determines deletion and addition of paths. It may cause bad effects to nodes of the other end of paths that are deleted or added. For example, such a node can be isolated. So this

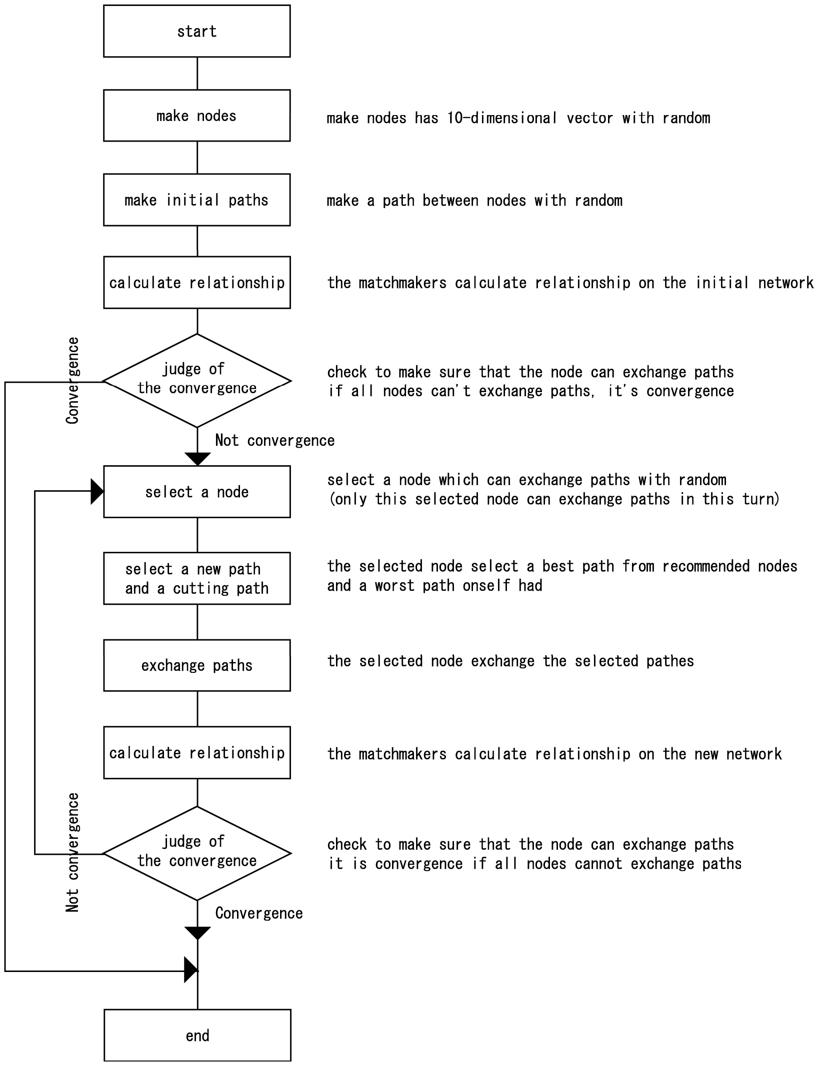


Figure 2: Flow chart of simulation

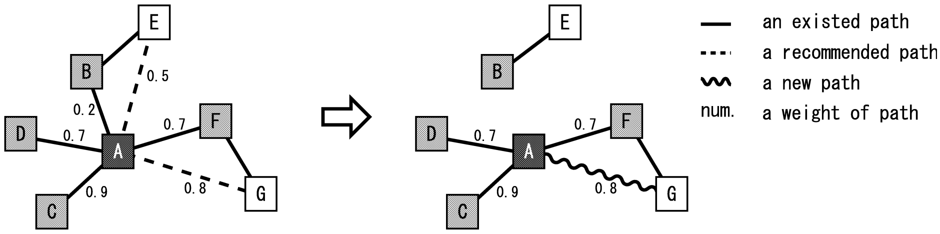
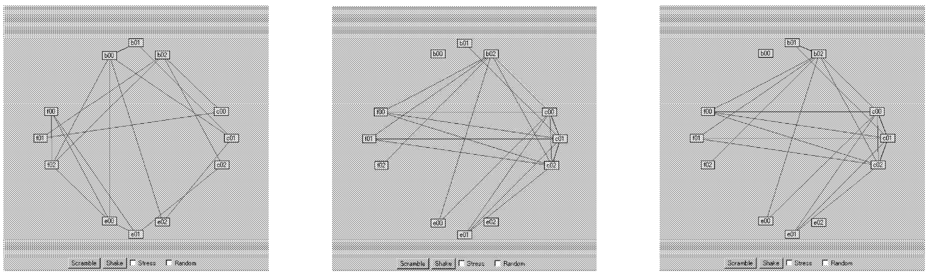


Figure 3: How to make a new path

algorithm cannot ensure the convergence.

We implement this simulation as a Java application. The simulator is implemented by Java. Figure 4 shows the output interface of the simulator. These are the examples of nodes and paths at the initial state (a), the converged state (b), and the ideal state (c) by the simulator. We can see that the network at Figure 4-b(converged) is more similar to one at Figure 4-c(ideal) than one at Figure 4-a(initial).



a) initial state b) converged state c) ideal state

Figure 4: the networks at the some states

4.2 The Measurement of Evaluation

If it is possible to calculate connection values between all nodes at a place, we can create the best network definitely. We can measure how to our method works by comparing the best network and the network generated by our method. We compare two networks in the following two ways. One is cover rate that is how much paths in the best network is found in the generated network. It means how much similar in structure two networks are. The other is reach rate that is comparison of the average of connection values between the best and generated networks. It indicates how much similar in effectiveness two networks are. These parameters are defined as the following formulas:

$$cover - rate = \frac{\mid \{P_{current} \cap P_{best}\} \mid}{N}$$

$$reach - rate = \frac{\sum_{l=1}^{N} f(p_l | p_l \in \{P_{current}\})}{\sum_{m=1}^{N} f(p_m | p_m \in \{P_{best}\})}$$

p	:	a path
N	:	the size of paths
$\{P\}$	:	a set of paths
$\{P_{best}\}$	:	the best set of paths
$\{P_{current}\}$	:	the current set of paths
$f(p)$	:	a value of path

5 Simulation

5.1 Use the virtual data

In this simulation, we should provide data that can be used to calculate connection values between nodes. In the virtual data, we define that each node has a 10-dimensional vector generated randomly, and the connection value between nodes is product of two vectors. An element of a vector is a random number from 0.0 to 1.0. The probably that zero is set on the element is 0.5. The reason is to make heterogeneous relationship among nodes.

There are two parameters to control experiments. One is the number of nodes and the other is the number of paths. In this experiment, we set the size of nodes from 10 to 100 and the size of paths from the 1 to 5 times the number of nodes. The parameters are shown in Table 1. The simulation is performed 10 times for each set of parameters, and we use the average as the results.

The graphs in Figure 5 and Figure 6 plot the average of cover-rate against

Table 1: The set of parameters

node	10-dimensional vector
connection value	product of the pair of vectors
nodes	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
paths	$\times 1, \times 2, \times 3, \times 4, \times 5$

turn. Figure 5 only shows the results from the size of paths is thrice and Figure 6 only shows the resultis from the size of paths is 60 for ease-to-see.

In this simulation, we cannot know whether the network will convergence. However, we can see all graphs became horizontal and find all networks convergence using matchmaking. And we can find the average of measurements and the turn of convergence are effective the size of nodes and paths.

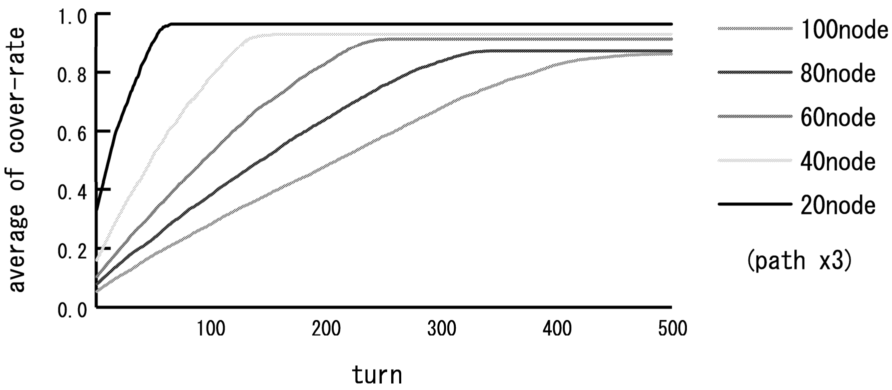


Figure 5: Cover-Rate and nodes in the virtual data

The graphs in Figure 7 and Figure 8 plot the average of reach-rate against turn. In this case, we can find the effect of the size of paths and nodes is weak

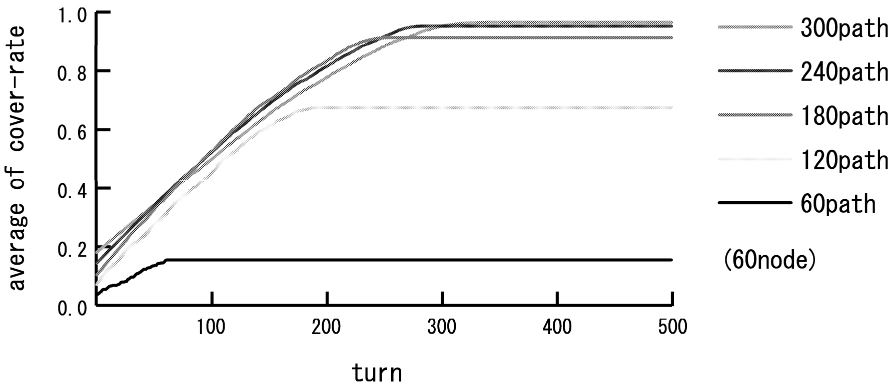


Figure 6: Cover-Rate and paths in the virtual data

more than with reach-rate. However, there is a same tendency with cover-rate. So, we examine the relevance among the average of measurements and the size of paths and nodes and the turn of convergence.

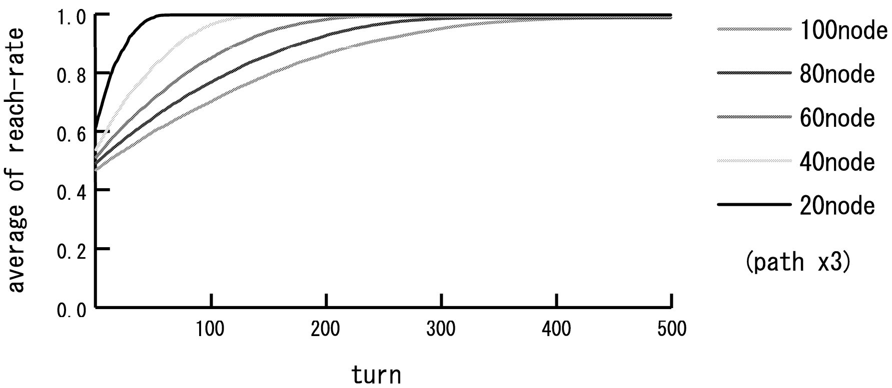


Figure 7: Reach-Rate and nodes in the virtual data

At first, we examine the relevance between the average of measurements and the size of nodes and paths. The left graph in Figure 9 plots cover-rate at

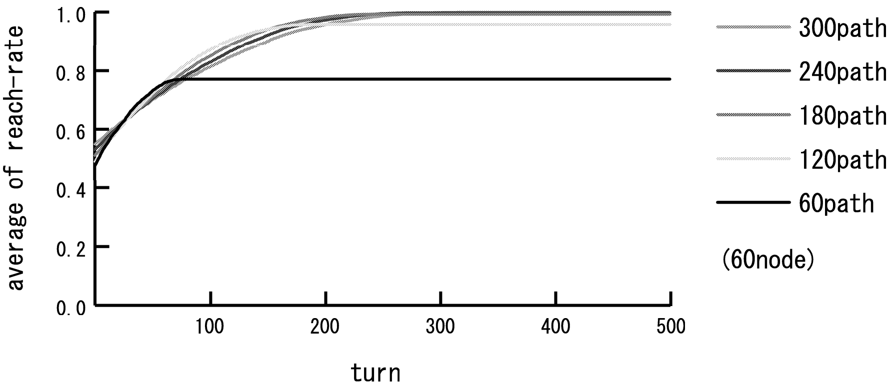


Figure 8: Reach-Rate and paths in the virtual data

the initial state against the size of nodes, and the right plots cover-rate at the converged state against the size of nodes. When the size of nodes increases, the average of cover-rate decreases when the size of paths increases, the average of cover-rate decreases. It means that larger size of nodes is more difficult which larger size of paths is easier for optimization using this model.

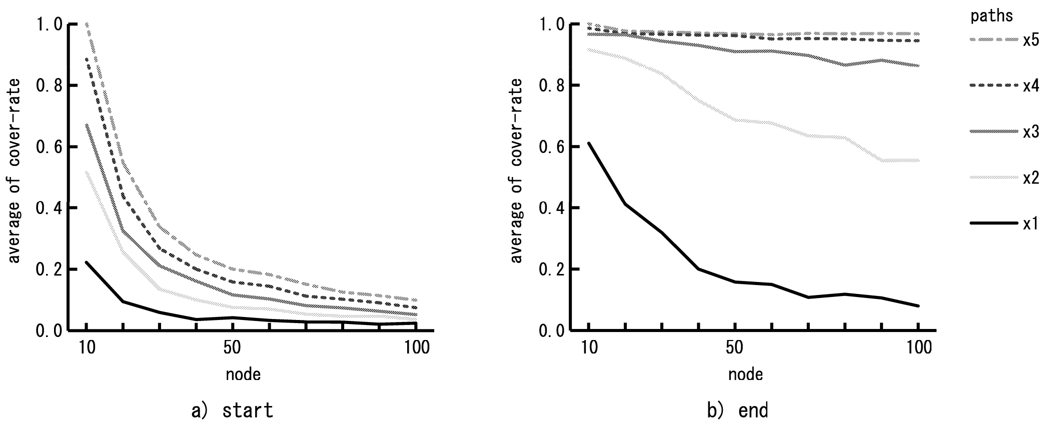


Figure 9: The average of cover-rate

The left graph in Figure 10 plots reach-rate at the initial state against the size of nodes, and the right plots reach-rate at the converged state against the size of nodes. There is a similar tendency to the average of cover-rate (Figure 9). These graphs indicated that the optimization using this method is sufficiently effective if the size of paths is more than twice of the size of nodes.

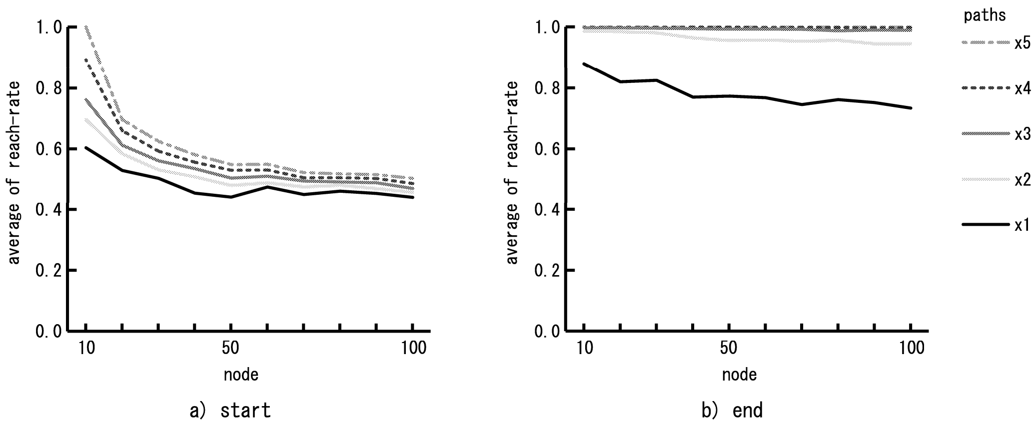


Figure 10: The average of reach-rate

Next, we examine the relevance between the average of measurements and the turn of convergence. The graph in Figure 11 plots the average of convergence turns against the size of nodes. A ratio of convergence is 1.0. Almost this model does not ensure network convergence, we can conclude that the model of converges practically from this result. This graph indicated that the turn of convergence increase linearly when the size of nodes increases. In this simulation, only a single one node can exchange paths in a turn, so the times of exchanging per node do not became so large.

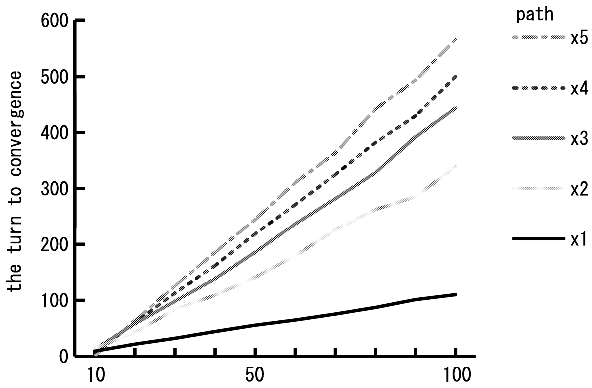


Figure 11: Average of Convergence Turn

5.2 Use the real data

We use the virtual data generated randomly in the previous experiment. In next experiment, we use the real data generated by people. We use WWW bookmarks to measure connection values among people. Users always add a web page in which she/he is interested and organize topics as folder in WWW bookmark. So it can be said that WWW bookmark represents the user profile. In this simulation, we need to calculate relationship between nodes. We use a parameter called "category resemblance" such as a value of relationship between nodes [5]. This parameter is based on resemblance of folder structure of WWW Bookmark.

We asked twelve persons to submit their WWW bookmark files to the system that can calculate relationship between WWW bookmarks. We set up the size of nodes as 12 and the size of paths as 12, 16, 20, and 24. The parameters are shown in Table 2. The simulation is performed 10 times for each set of parameters, and the average as the results.

Table 2: The set of parameters (2)

node	WWW Bookmark
connection value	categorize resemblance
nodes	12
paths	12, 16, 20, 24

The graph in Figure 12 plots the average of cover-rate against turn, and in Figure 13 plots the average of reach-rate. A vertical line on a line in the graph indicates convergence. There is the same tendency with the virtual data. These results indicate that the network could be optimized in the real data.

6 Conclusion

In this paper, we propose the way to get a new person who is a partner for exchanging information and proposed a method called "Neighborhood Matchmaker Model (NMM)". Our method only use collaborative and autonomous matchmaking and do not need any central servers. Nevertheless,

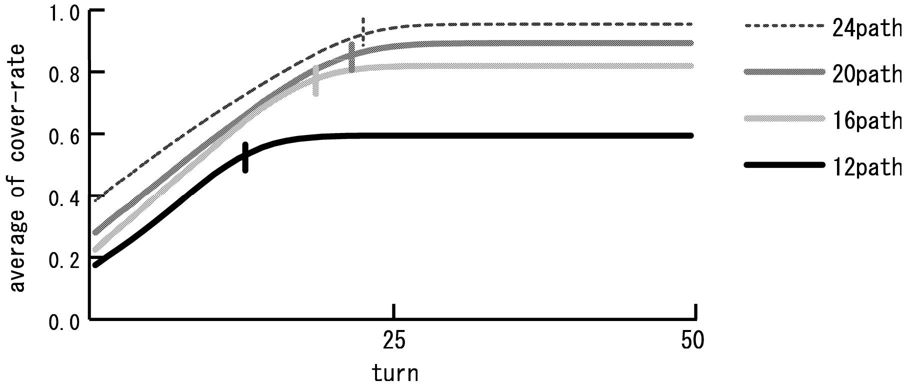


Figure 12: the average of cover-rate in the real data

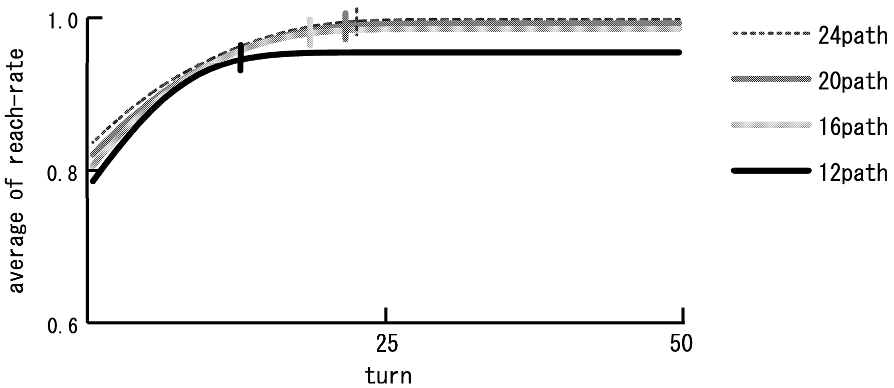


Figure 13: the average of reach-rate in the real data

by examining our experiment results, the optimal personal human network can be obtained. In this simulation we need the number of paths that is 2 or 3 times of the number of nodes and the number of turns that is 1.5 to 2 times the number of nodes in order to optimize the network sufficiently.

This model does not need the centered server. This feature has two advantages for information exchanging. One is, we can use this model easily and quickly, so that we can assist to form dynamic and emergent communities that are typical in the Internet. And the other is, we can deal with any size of groups, because it calculates relationship among people without collecting all data at he server. It is possible to assist bigger groups that are more likely to contain valuable persons to exchange information.

A further direction of this study will be to develop a system using this proposed model and investigate effectiveness for it in real world.

References

[1] Henry Kautz, Bart Selman, Mehul Shah. ReferraWeb: Combining Social Networks and Collaborative Filtering. In *the Communications of the ACM*, Vol. 40, No. 3, 1997.

[2] Yasuyuki. Sumi, Kenji. Mase. Collecting, Visualizing, and Exchanging Personal Interests and Experiences in Communities. *the 2001 Interna-*

tional Conference on Web Intelligence (WI-01), 2001.

[3] Yasuyuki. Sumi, Kenji. Mase. Digital Assistant for Supporting Conference Participants: An Attempt to Combine Mobile, Ubiquitous and Web Computing, *Ubicomp 2001*, in G.D.Abowd, B.Brumitt, S.A.N.Shafer (Eds.), Ubicomp 2001, LNCS 2201, pp.156-175, Springer, Sep. 2001.

[4] Koji Kamei, Eva Jettmar, Kunihiko Fujita, Sen Yoshida, Kazuhiro Kuwabara. Community Organizer: Suppoting the Formation of Network Communities throuh Spatial Representation. In Proceedings of *the 2001 Symposium on Applications and the Internet (SAINTŐ1)*, 2001.

[5] Masahiro Hamasaki, Hideaki Takeda. Experimental Results for a Method to Discover Human Relationship Based on WWW Bookmarks. In Proceedings of *the Knowledge-Based Intelligent Information & Engineering Systems (KESŐ1)*, 2001.