Mining Social Networks in Message Boards

Naohiro Matsumura, **David E. Goldberg** and **Xavier Llorà**

IlliGAL Report No. 2005001
January, 2005

Illinois Genetic Algorithms Laboratory
University of Illinois at Urbana-Champaign
117 Transportation Building
104 S. Mathews Avenue Urbana, IL 61801
Office: (217) 333-2346
Fax: (217) 244-5705

Mining Social Networks in Message Boards

Naohiro Matsumura[*], **David E. Goldberg** and **Xavier Llorà**
Illinois Genetic Algorithms Laboratory (IlliGAL)
Department of General Engineering
University of Illinois at Urbana-Champaign
104 S. Mathews Ave, Urbana, IL 61801
`matumura@econ.osaka-u.ac.jp`, `{deg, xllora}@uiuc.edu`

Abstract

In the paper, we first present an approach to extract social networks from message boards on the Internet. Then we show structural features of 3,000 social networks extracted from 3,000 message boards from 15 categories in Yahoo!Japan Message Boards to prove the relationships between the features and the categories. After we classify social networks into three types (interactive communication, distributed expertise communication and soapbox communication), we suggest an approach for mining social networks to identify the types of communication, the roles of individuals, and important ties, all of which can be used to redesign the means communication as well as understand the state of communication.

1 Introduction

With the advent of popular social networking services on the Internet such as Friendster [1] and Orkut [2], social networks are again coming into the limelight with respect to this new communication platform. A social network shows the relationships between individuals in a group or organization where we can observe their social activities. For example, individuals who share a serious problem might all join a discussion on ways to solve the problem, while in another case only a few knowledgeable individuals might give information when individuals seek to find out more about specific topics.

Social networks have been studied for decades. Milgram discovered what we now call 'small-world phenomena' which show that everyone is connected to each other through a short chain in their social networks (Milgram 1967). Rogers classified each person into five types according to their stage regarding the adoption of new ideas (Rogers 1995). Granovetter insisted that weak ties spanning local relationship boundaries contribute to the diffusion of information (Granovetter 1973). Freeman proposed centrality measures to identify the importance of individuals and ties in a social network (Freeman 1978). Scott used social networks to identify gaps in information flow within an organization to find ways to get work done more effectively (Scott 1992). Krackhardt studied the importance of informal networks in organizations and revealed the effect of these networks on the accomplishment of tasks (Krackhardt and Hanson 93).

---

[*]Visiting scholar from Graduate School of Economics, Osaka University, Osaka 560-0043 Japan
[1]http://www.friendster.com/
[2]http://www.orkut.com/

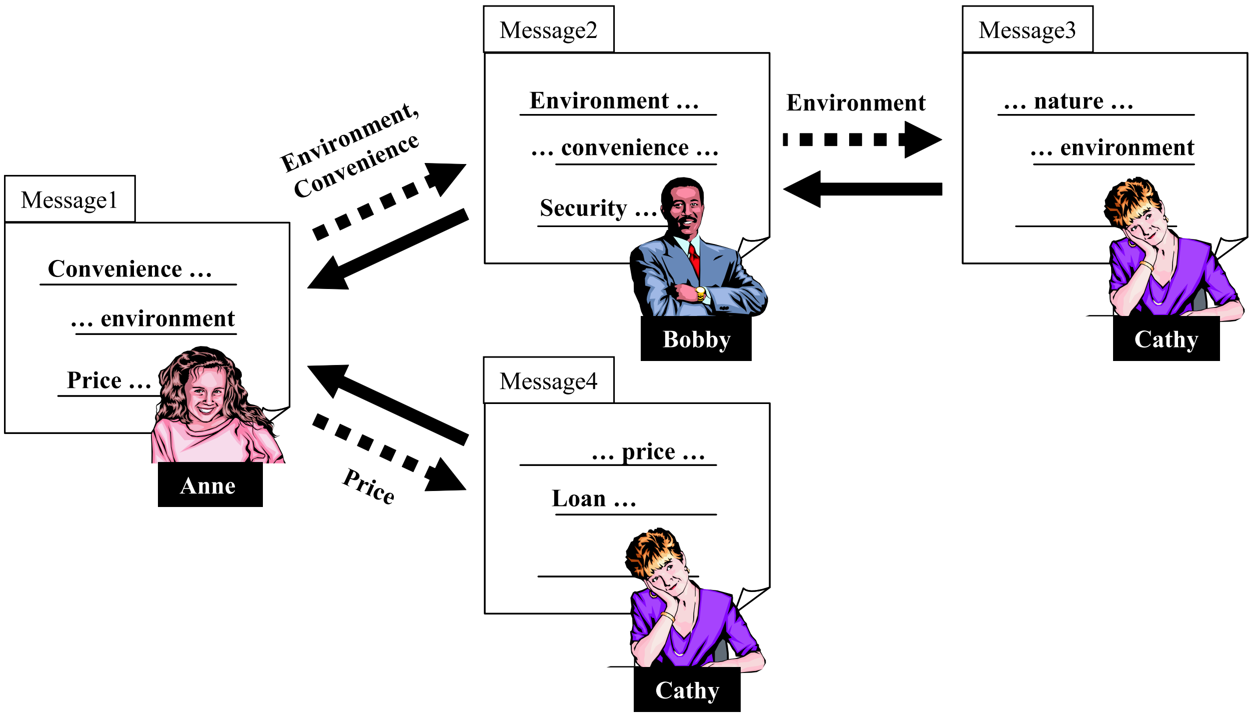

Figure 1: A message chain of five messages sent by three individuals.

Social networks in online communication have been studied as well. Ohsawa et al. classified communities into six types according to the structural features of the word co-occurrence structure of communications (Ohsawa et al. 2002). Tyler et al. analyzed e-mail logs within an organization to identify communities of practice – informal collaborative networks – and leaders within the communities (Tyler et al. 2003). Matsumura et al. revealed the effect of anonymity and ASCII art on communication through message boards (Matsumura et al. 2004).

In addition, much research on social networks has been done in past decades and many properties of these networks are now well known. However, the causality between social networks and communication types are still veiled in mystery. In this paper, we aim at revealing some of the mystery by mining social networks in message boards to understand the state of communication and obtain new ideas regarding communication redesign.

The remainder of this paper proceeds as follows. In Section 2, we present an approach to extract social networks from message boards on the Internet. Then we introduce five indices that can be used to measure the structural features of social networks in Section 3. We describe three types of communication based on the classification of 3,000 social networks in Section 4. In Section 5, we suggest an approach for mining social networks which enables us to identify the types of communication, the roles of individuals, and important ties. Our conclusions and directions for future work are given in Section 6.

2 Extracting Social Networks

In a social network based upon online communication, the distance between individuals does not mean ‘geographical distance’ because each person lives in a virtual world. Instead, distance can be considered ‘psychological distance’ and this can be measured by the “influence” wielded among the members of the network.

Consider the situation where an individual p has a great deal of influence on an individual q. In this case, we can consider three types of relationship.

Case 1. p is close to q.

Case 2. q is close to p.

Case 3. p and q are close to each other.

Cases 1 and 2 show uni-directional relationships, and Case 3 shows a bi-directional relationship. Most previous studies of social networks employed undirectional social networks, i.e., Case 3, because of the simplicity of analysis. However, the relationship between individuals is not symmetric because of their activities and social situations (Wallace 1999). In addition, the difference of the distances between two individuals could be a key to understanding the relationships since it shows the communication gap between them. For this reason, we treat a social network as an asymmetric network where vertices denote individuals and directed links denote the flows of influence. In this paper, we do not need to distinguish Case 1 from Case 2 (e.g., the direction of the relationship distance) because our approach to measuring the communication gap, described later, produces the same results regardless of the direction.

We measure the influnce by using the IDM (Influence Diffusion Model) algorithm in which the influence between a pair of individuals is measured as the sum of propagating terms among them via messages (Matsumura 2003). Here, let a message chain be a series of messages connected by post-reply relationships, and the influence of a message x on a message y (x precedes y) in the same message chain be $i_{x\to y}$. Then, $i_{x\to y}$ is defined as

$$i_{x\to y} = |w_x \cap \cdots \cap w_y|, \tag{1}$$

where w_x and w_y are the set of terms in x and y, respectively, and $|w_x \cap \cdots \cap w_y|$ is the number of terms propagating from x to y via other messages. If x and y are not in the same message chain, we define $i_{x\to y}$ as 0 because the terms in x and y are used in a different context and there is no influence between them.

Based on the influence between messages, we next measure the influence of an individual p on an individual q as the total influence of p's messages on other's messages through q's messages replying to p's messages. Let the set of p's messages be α, the set of q's messages replying to any of α be β, and the message chains starting from a message z be ξ_z. The influence from p onto q, $j_{p\to q}$, is then defined as

$$j_{p\to q} = \sum_{x\in\alpha}\sum_{z\in\beta}\sum_{y\in\xi_z} i_{x\to y}. \tag{2}$$

Here we see the influence of p on q as q's contribution toward the spread of p's messages. The influence of each individual is also measurable using $j_{p\to q}$. Let the influence of p be k_p, and all other individuals be γ. Then, k_p is defined as

$$k_p = \sum_{q\in\gamma} j_{p\to q}. \tag{3}$$

As an example of measuring the influence, let us use the simple message chain shown in Figure 1 where Anne posted Message 1, Bobby posted Message 2 as a reply to Message 1, and Cathy posted Message 3 and Message 4 as replies to Message 2 and Message 1, respectively. In the figure,

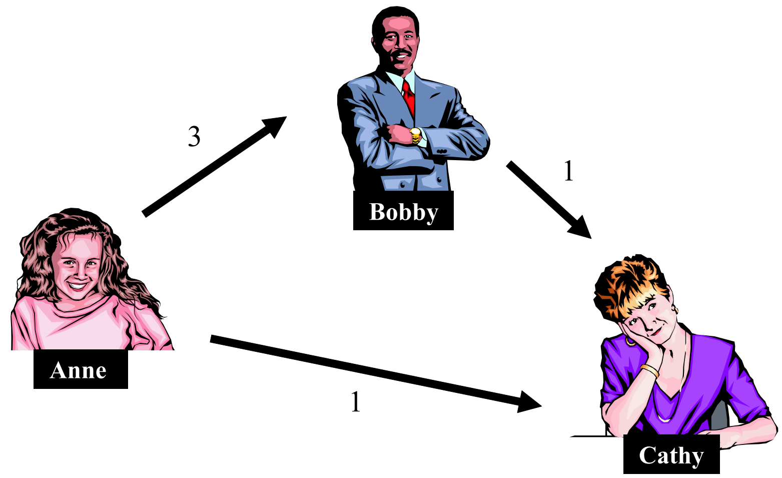

Figure 2: A social network showing the influence from Figure 1.

solid arrows show the replies to previous messages, and dotted arrows show the flows of influence. Here, the influence between a pair of individuals is as follows.

- The influence of Anne on Bobby is 3 (i.e., $j_{Anne \rightarrow Bobby} = 3$), because two terms ("Environment" and "Convenience") were propagated from Anne to Bobby, and one term ("Environment") was propagated from Anne to Cathy via Bobby.

- The influence of Anne on Cathy is 1 (i.e., $i_{Anne \rightarrow Cathy} = 1$), because one term ("Price") was propagated from Anne to Cathy.

- The influence of Bobby on Cathy is 1 (i.e., $i_{Bobby \rightarrow Cathy} = 1$), because one term ("Environment") was propagated from Bobby to Cathy.

- The influence of Bobby on Anne and of Cathy on Anne is 0 (i.e., $i_{Bobby \rightarrow Anne} = 0$ and $i_{Cathy \rightarrow Anne} = 0$), because no term was propagated to Anne from either Bobby or Cathy.

Note that we ignore the influence of Anne on Cathy, even though a term "Environment" was propagated from Anne to Cathy via Bobby, because we want to measure direct influence between individuals. Instead, we consider the indirect influence of Anne on Cathy via Bobby as the contribution of Bobby, and add it to the influence of Anne on Bobby.

By mapping the influence between individuals, we can obtain a social network showing influence as in Figure 2 where their relationships are shown as directional links and the influence between them. From the figure, we can understand the influential relationships between individuals, and guess their roles in the communication — e.g., as opinion leaders and followers (Rogers 1995) — from the influence. For example, Anne would be an opinion leader because she was the source of the most influence on others. Bobby would be a mediator because he was a recipient of influence and transmitted some of it to Cathy. Cathy would be a follower because she only received influence from others.

The influence between individuals also shows the distance between them with respect to contextual similarity since the influence indicates the degree of their shared interest represented as terms. The influence and contextual distance between individuals are inversely related; i.e., the greater the influence, the shorter the distance. Here, let us define the length of a link (i.e., distance) as follows.

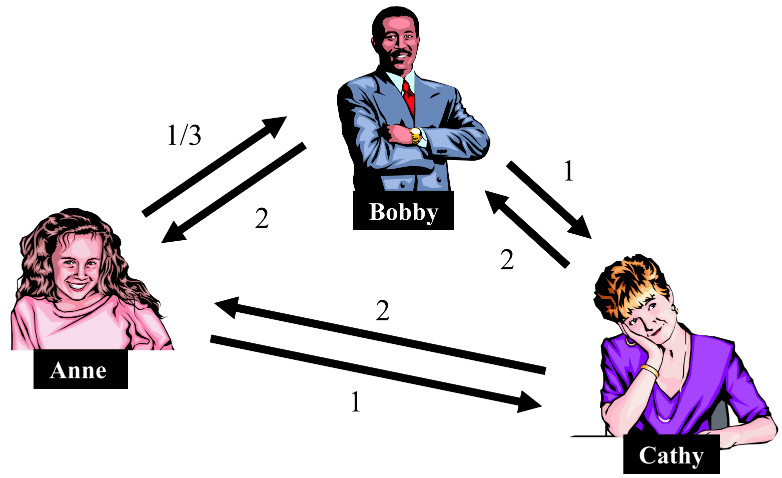

Figure 3: A social network showing the distance from Figure 1.

Definition 1 (The distance of a link) *The distance from an individual p to an individual q, $d_{p \to q}$, is defined as the value inversely proportionate to the influence from p to q; i.e., $d_{p \to q} = 1/j_{p \to q}$.*

The distance is between 0 and 1 when the influence is more than 0. However, the distance cannot be measured by the above definition if the influence is 0. In that case, we define the distance $n-1$ (n is the number of individuals participating in communication) as the case of the weakest relationships; i.e., the diameter of a social network where all individuals are connected linearly with maximum distance. In this way, an asymmetric social network with distance is extracted from message chains as shown in Figure 3.

3 Communication Gaps in Social Networks

Based on the forward and backward distances between two individuals in a social network, we can consider three types of relationship between them.

Type 1. Two-sided communication: Two individuals actively exchange their ideas with each other. In this case, they are equivalent in communication, and their closeness can be identified by the short distances between them.

Type 2. One-sided communication: One individual's idea is received by an individual, while another individual's idea is not. In this case, the two senders are not equivalent in communication, as can be identified from the distance between them since the forward distances and backward distances will differ from each other.

Type 3. Sparse communication: Two individuals rarely exchange ideas with each other. In this case, they are not involved in communication, and their relationship can be identified from the long distances between them.

These three types, in other words, correspond to various "communication gaps" between two individuals. By measuring the communication gaps for all pairs of individuals in a social network, we can understand the state of their communication. For example, a small gap indicates active communication, while a large gap suggests inactive communication.

The distances can also be measured for two individuals who are not directly connected, but are indirectly connected via other individuals because they can exchange their ideas through others. Interestingly, the shortest path between individuals is often not an existing direct path, but an indirect path. In this paper, we consider the distance of the shortest path as the distance between two individuals since it reflects the real channel of their communication.

Here, let the distance of the shortest path from an individual p to another individual q be $d_{p\to q}$ ($d_{p\to q} = 0$ if p is equal to q) and the set of all individuals in a social network be γ. We then propose three indices, G_{diff}, G_{max} and G_{min}, that measure the collective communication gap in the social network from different points of view.

- G_{diff} measures the communication gap by accumulating the differences in the distances of the shortest paths between all pairs of individuals. G_{diff} is defined as

$$G_{diff} = \frac{1}{2}\sum_{p\in\gamma}\sum_{q\in\gamma}|d_{p\to q} - d_{q\to p}|, \tag{4}$$

  where $|d_{p\to q} - d_{q\to p}|$ is the absolute value of $(d_{p\to q} - d_{q\to p})$.

- G_{max} measures the communication gap by accumulating the longer distances of the shortest paths between all pairs of individuals as a bottleneck hindering communication. G_{max} is defined as

$$G_{max} = \frac{1}{2}\sum_{p\in\gamma}\sum_{q\in\gamma}\max(d_{p\to q}, d_{q\to p}), \tag{5}$$

  where $\max(d_{p\to q}, d_{q\to p})$ returns the maximum value from $\{d_{p\to q}, d_{q\to p}\}$.

- G_{min} measures the communication gap by accumulating the shorter distances of the shortest paths between all pairs of individuals. G_{min} is defined as

$$G_{min} = \frac{1}{2}\sum_{p\in\gamma}\sum_{q\in\gamma}\min(d_{p\to q}, d_{q\to p}), \tag{6}$$

  where $\min(d_{p\to q}, d_{q\to p})$ returns the minimum value from $\{d_{p\to q}, d_{q\to p}\}$.

We can also consider another approach to identifying the state of communication by using only the distances between individuals instead of the differences between distances since the distance itself shows another aspect of communication gaps. The approach is known as "closeness centrality" where individuals nearby are more like to give/receive information more quickly than others (Freeman 1978). Based on closeness centrality, we propose two more indices for measuring the collective communication gap, C_{diff} and C_{dist}, as follows.

- C_{diff} measures the communication gap by accumulating the differences in the closeness centralities of individuals. C_{diff} is defined as

$$C_{diff} = \sum_{p\in\gamma}|c_p^{in} - c_p^{out}|, \tag{7}$$

  where c_p^{in} means the inward closeness centrality that shows the sum of distances of the shortest paths from all other individuals to p, and c_p^{out} means the outward closeness centrality that shows the sum of distances of the shortest paths from p to all other individuals.

Table 1: The average of normalized five indices for 15 categories measured from 3,000 message boards in Yahoo!Japan Message Boards.

Categories	C'_{dist}	C'_{diff}	G'_{diff}	G'_{max}	G'_{min}
Family & Home	0.032	0.028	0.032	0.048	0.017
Health & Wellness	0.065	0.065	0.073	0.100	0.029
Arts	0.068	0.070	0.081	0.108	0.029
Science	0.072	0.068	0.078	0.110	0.034
Cultures & Community	0.085	0.061	0.071	0.120	0.051
Romance & Relationships	0.081	0.079	0.089	0.125	0.038
Hobbies & Crafts	0.095	0.092	0.106	0.146	0.043
Regional	0.161	0.120	0.142	0.230	0.093
Entertainment	0.151	0.129	0.157	0.228	0.075
Government & Politics	0.217	0.167	0.197	0.313	0.120
Business & Finance	0.241	0.160	0.184	0.331	0.150
Schools & Education	0.253	0.161	0.195	0.349	0.158
Recreation & Sports	0.239	0.208	0.253	0.362	0.116
Computers & Internet	0.447	0.221	0.272	0.579	0.315
Current Events	0.455	0.220	0.271	0.588	0.322

- C_{dist} measures the communication gap by accumulating the closeness centralities of individuals. C_{dist} is defined as

$$C_{dist} = \sum_{p \in \gamma} c_p^{out}. \tag{8}$$

Note that C_{dist} doesn't change even if we use c_p^{in} instead of c_p^{out}.

4 Three Types of Communication

To determine the features of the five indices proposed in Section 3, we analyzed 3,000 social networks. The analysis procedure was as follows.

Step 1. We downloaded 3,000 message boards from 15 categories of Yahoo!Japan Message Boards. To equalize the number of messages for each message board, we selected message boards having more than 300 messages and downloaded the first 300 messages. Then, we removed stop words (words except for noun and verb words) from all the messages to accurately measure content-derived influence. In this way, we prepared 3,000 message boards with each having 300 messages.

Step 2. We extracted a social network from each message board using the approach described in Section 2. To equalize the number of individuals in a social network, we constructed a social network with the 10 most influential individuals identified by Equation (3). We thus obtained 3,000 social networks, each consisting of 10 individuals.

Step 3. We measured G_{diff}, G_{max}, G_{min}, C_{dist}, and C_{diff} for the 3,000 extracted social networks. To equalize the range of the indices, we normalized each index by dividing it by its theoretical

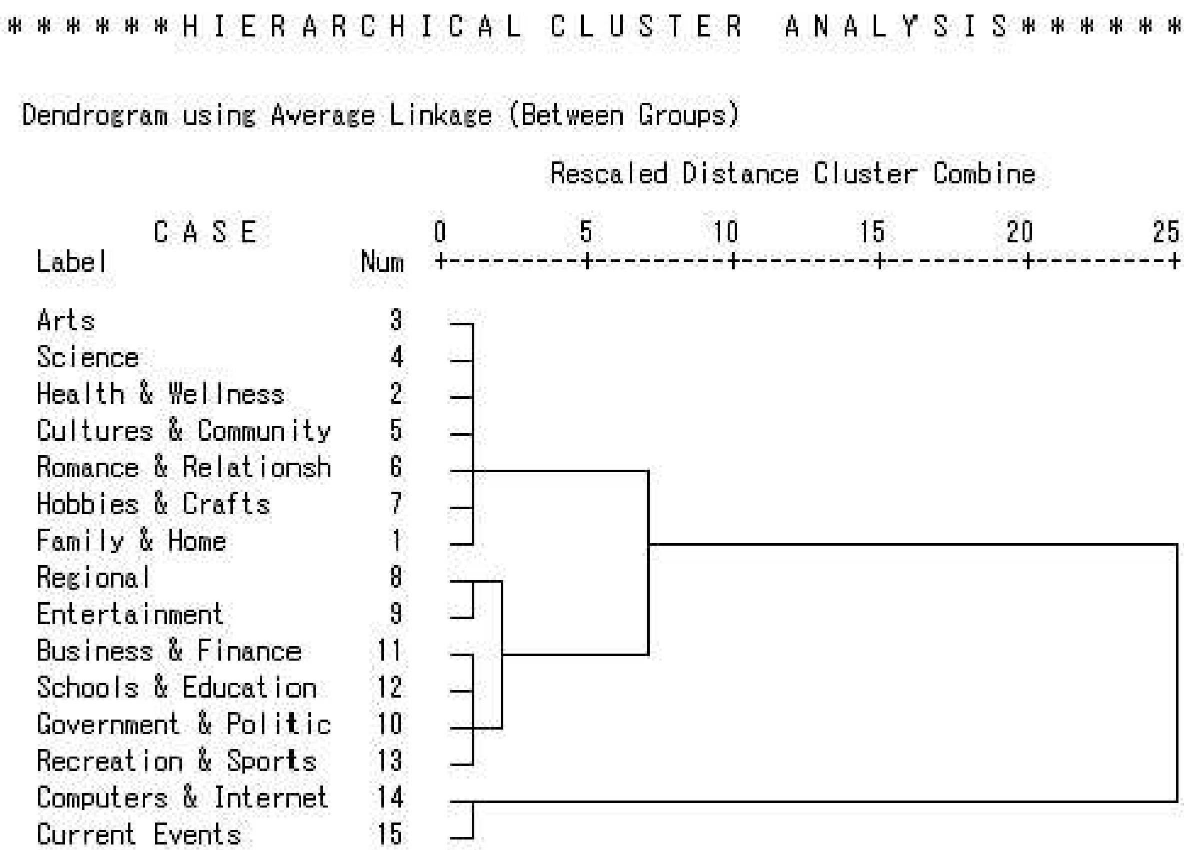

Figure 4: A dendrogram produced through hierarchical cluster analysis.

maximum. The normalized indices, G'_{diff}, G'_{max}, G'_{min}, C'_{diff}, and C'_{dist}, were

$$G'_{diff} = \frac{G_{diff}}{{}_nC_2(n-1)} \tag{9}$$

$$G'_{max} = \frac{G_{max}}{{}_nC_2(n-1)} \tag{10}$$

$$G'_{min} = \frac{G_{min}}{{}_nC_2(n-1)} \tag{11}$$

$$C'_{diff} = \frac{C_{diff}}{n(n-1)^2} \tag{12}$$

$$C'_{dist} = \frac{C_{dist}}{n(n-1)^2} \tag{13}$$

Step 4. The average of each index was calculated for each category.

The values of the five indices for the fifteen categories are shown in Table 1. Here, to investigate the relationships between the indices and categories, we applied hierarchical cluster analysis to the data. This analysis merges clusters based on the mean Euclidean distance between the elements of each cluster. A tree-like diagram, called a dendrogram, is then constructed as shown in Figure 4. From this figure, we can find three major clusters, each corresponding to a type of communication.

We named the clusters as follows.

Interactive Communication: This cluster includes seven categories ("Arts", "Sciences", "Health & Wellness", "Culture & Community", "Romance & Relationships", "Hobbies & Crafts", and "Family & Home"), the indices of which are considerably smaller than those of other categories. The topics in these categories are common and familiar to many individuals who share these interests. As a consequence, individuals are naturally involved in the communication, and actively exchange their ideas with others.

Distributed Expertise Communication: This cluster includes six categories ("Regional", "Entertainment", "Business & Finance", "Schools & Education", "Government & Politics", and "Recreation & Sports"), the indices of which are generally higher than those of the interactive communication categories. As the topics in these categories are somewhat specific and disputable, experienced or knowledgeable individuals contribute most to the communication.

Soapbox Communication: This cluster includes two categories ("Computers & Internet" and "Current Events"), the indices of which are higher than those of the above two clusters. The topics in these categories are mainly current affairs or topical news, and the communication is one-way from informers to audiences (or lurkers).

From the above results, we can say that there are roughly three types of communication in Yahoo!Japan Message Boards. If these types are common properties in other social networks, it will be possible to identify the state of communication by measuring the five indices.

We expected that the five indices would reveal different aspects of communication, however the Pearson correlation coefficients between them were over 0.9. This meant that these indices were not statistically distinct. In other words, we should be able to identify the types of communication by using only one index instead of all five. In the following, we show some approaches to mining social networks based on G'_{max} because G'_{max} proved to be the most discriminative index for identifying clusters.

5 Mining Social Networks

Mining of social networks is done to identify the types of communication, understand the roles of individuals, and explore remedies for communication gaps in social networks. In this section, we present case studies of mining social networks based on communication types and G_{max} explained before.

5.1 Communication Types

We prepared two types of message log, log 1 and log 2, each of which was extracted from a message board on DISCUS[3] (Goldberg 2003). Log 1 consisted of 72 messages posted by five individuals (one Japanese researcher, one Spanish researcher, and three Japanese businesspersons) who discussed "cell phones and women" in English. Log 2 consisted of 102 messages posted by six individuals (two Japanese researchers, one Spanish researcher, one Japanese businessperson, one Chinese student, and one Taiwanese student) who discussed "purchase of a house" in English. Note that the real names of individuals have been replaced with fictional names to protect their privacy.

[3]http://www-discus.ge.uiuc.edu/

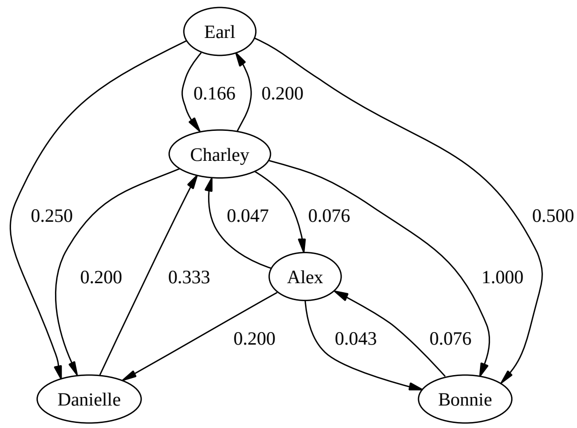

Figure 5: The social network with distance extracted from log1. $G_{max} = 0.069$.

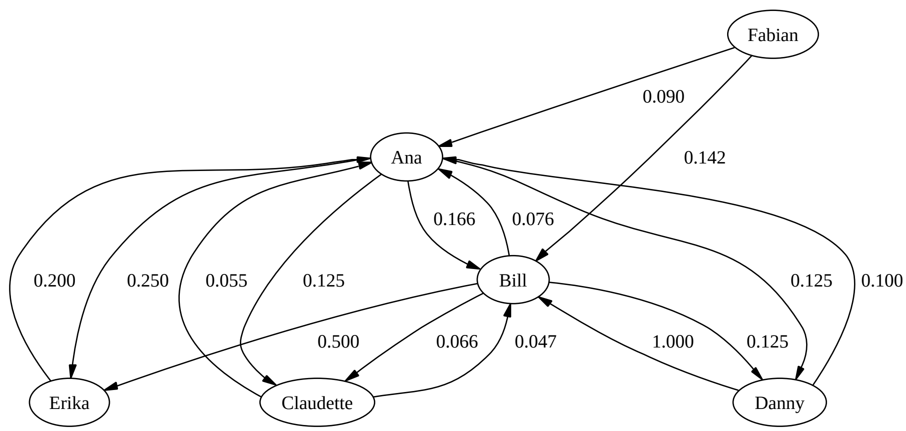

Figure 6: The social network with distance extracted from log 2. $G_{max} = 0.364$.

Figures 5 and 6 are social networks showing the distance extracted from log 1 and log 2, respectively, using the approach described in Section 2. Once we obtained the social networks, G'_{max} was measured as 0.069 from Figure 5 and 0.364 from of Figure 6. Comparing G'_{max} with the clustering results in Section 4, we can identify the types of communication of log 1 and log 2 as "interactive communication" and "distributed expertise", respectively.

The topic in log 1 was familiar to the individuals because they use cell phones everyday. On the other hand, the purchase of a house is a big event in life, and many individuals have no experience of making such a purchase. Therefore, the communication in log 2 was controlled by a few experienced individuals. Thus, the communication types of log 1 and log 2 seem to have properly reflected the types of real communication.

5.2 Roles of Individuals

As shown by the definition of G_{max} in Equation (5), G_{max} is measured by summing each individual's communication gaps with respect to other individuals. That is, G_{max} for each individual is easily measurable by translating the definition of G_{max}. Let the communication gap of an individual p

Table 2: Three most important links in Figure 6.

Link	l_{max}	G''_{max}	G'_{max}
Erika → Ana	0.574	0.938	0.364
Danny → Ana	0.270	0.634	0.364
Ana → Erika	0.226	0.590	0.364

Table 3: g'_{max} of each individual in Figure 6.

	Ana	Bill	Claudette	Danny	Erika	Fabian
g'_{max}	0.038	0.039	0.038	0.040	0.042	0.167

be g_{max}. We can then define g_{max} as

$$g_{max} = \frac{1}{2} \sum_{q \in \gamma} \max(d_{p \to q}, d_{q \to p}). \tag{14}$$

Normalized g_{max} is measured by dividing by the theoretical maximum. Let the normalized g_{max} be g'_{max}. Then, g'_{max} is defined as

$$g'_{max} = \frac{g_{max}}{\frac{1}{2}n(n-1)}. \tag{15}$$

As the G_{max} of Figure 6 is high, let us measure each individual's g'_{max} to reveal the source of the communication gap. As shown in Table 3, Fabian has the highest g'_{max}. From this, we can understand that there are communication gaps around Fabian.

The roles of individuals are also identified from their relationships with others. If removing a link raises G'_{max}, the link is considered to help reduce the communication gap and is therefore important. That is, we can measure the importance of each link by comparing the G'_{max} of the original social network with that link to G'_{max} of a social network without the link. Let the importance of a link be l_{max}. We then define l_{max} as

$$l_{max} = G''_{max} - G'_{max}, \tag{16}$$

where G''_{max} is the G'_{max} measured from a social network without the link. From the top three l_{max} values in Figure 6 listed in Table 2, we can see that a link from Erika to Anna is the most important.

During interviews with the individuals in Figure 6, we found that they considered Fabian creative, but strong-willed to the degree that nobody could counter his ideas. As a result, the communication around him did not go well. They also agreed with the results regarding the important links in Table 2 because these were the links actively used to exchange ideas during the period under study.

5.3 Remedies for Communication Gaps

Once we can obtain information about who are the causes of communication gaps and which links are most important for communication, we can prepare a remedy to improve communication. For example, the following three approaches could be the candidates of the remedies, apart from the

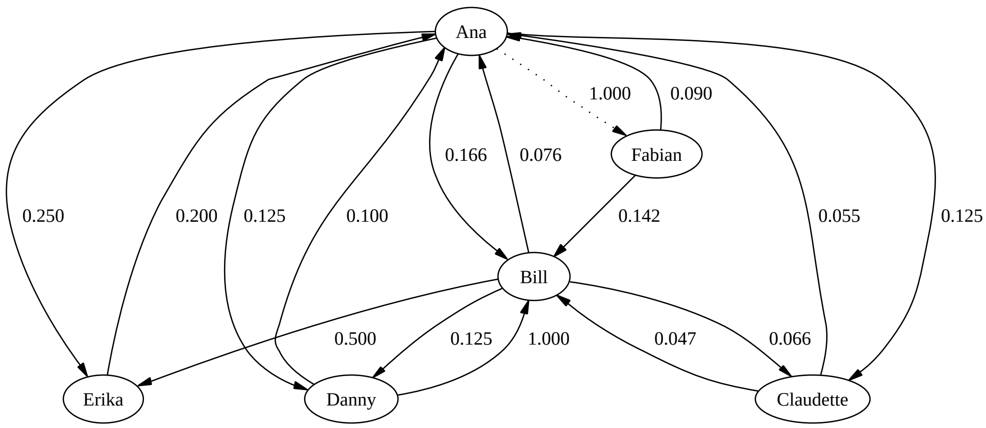

Figure 7: A link from Ana to Fabian causes G'_{max} to dynamically drop from 0.364 to 0.069.

feasibility.

- Persuade inactive individuals to contribute to the communication, or persuade strong individuals to listen to others' ideas. While this approach is straightforward, the effectiveness of such persuasion depends on many factors which are beyond the scope of this paper.
- Remove inactive or strong individuals from the communication. This approach is easy and has an immediate effect, but it is not constructive since we would lose the potential contributions of the excluded individuals.
- Add another individual who can communicate with inactive or strong individuals to bridge communication gaps. To find such individuals, the link importance can be used.

We can also simulate the importance of virtual links which make the communication gap smaller. For example, if there was a link from Anne to Fabian, G'_{max} would dynamically drop from 0.364 to 0.069 as shown in Figure 7.

Needless to say, there approaches above should be planned carefully before putting into practice. Having an inverview with individuals for surveying the effect of a remedy would further enhance the possibility of the success of mining social networks.

6 Conclusion

We have described an approach to extracting social networks from a message board. We then revealed three types of communication from point of communication gaps, and suggested some approaches to mining social networks.

Human beings are social creatures, and we could not survive without cooperating with others. It also suggests that understanding how relationships are created and functioned is essential to make our lives happier and richer. The range of our social networks is rapidly expanding than before as the Internet accelerates our communication via E-mail, Chat, Video conference system etc. Reflecting such a situation, managing and utilizing social networks will be an increasingly important part of our lives. We hope this study will contribute to the realization of a better way of life through human relationships.

Acknowledgements

We thank the Chance Discovery Consortium for their financial support. We are also extremely grateful to Prof. Yukio Ohsawa, Hiroshi Tamura, Yuichi Washida, and Masataka Yoshikawa, and all the members of IlliGAL for their insightful discussions.

References

Freeman, L.C. (1978) Centrality in Social Networks, *Social Networks*, **1**, 215–39.

David E. Goldberg, Michael Welge, and Xavier Llorà (2003) DISCUS: Distributed Innovation and Scalable Collaboration In Uncertaion Settings, IlliGAL Report No. 2003017

Granovetter, M. (1973) The Strength of Weak Ties, *American Journal of Sociology*, **78**, 1360–1380.

David Krackhardt and Jeffery R. Hanson (1993) Informal Networks: The Company behind the Chart, *Harvard Business Review*, July-August 1993, 104–111.

Naohiro Matsumura (2003) Topic Diffusion in a Community, Yukio Ohsawa and Peter McBurney (Eds.), *Chance Discovery*, Springer Verlag, 84–97.

Naohiro Matsumura, Asako Miura, Yasuyuki Shibanai, Yukio Ohsawa, and Toyoaki Nishida (2004) The Dynamism of Nichannel, *Journal of AI & Society*, Springer (in press)

Milgram, S. (1967) The Small World Problem, *Psychol. Today*, **1**, 61–67.

Yukio Ohsawa, Hirotaka Soma, Yutaka Matsuo, Naohiro Matsumura, and Masaki Usui (2002) Featuring Web Communities Based on Word Co-occurrence Structure of Communications, Proc. of WWW2002, 736–742.

Rogers, E.M. (1995) *Diffusion of Innovations* (Fourth Edition), New York, Free Press.

Scott, W.R. (1992) *Organizations: Rational, Natural, and Open Systems*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Tyler, J.R., Wilkinson, D.M. and Huberman, B.A. (2003) Email as spectroscopy: Automated Discovery of Community Structure within Organizations. (Preprint http://www.lanl.gov/arXiv:cond-mat/0303264)

Wallace, Patricia (1999) The Psychology of the Internet, Cambridge University Press.