Fractal Structure of Small World in a Friendship Network

Masaki Tomochi[1]
Atsushi Tanaka[2]
Tatsuhiro Shichijo[3]

Abstract

The effects of spatial and social distance on a friendship network are analyzed. We used the data obtained from "Tomocom" (tomocom.jp) which is a social network service where approximately 350 undergraduate students living in several areas in Japan are participating. From the data, we have found that spatial and social distance between individuals causes stratification in the friendship network and brings about fractal structure of small world. Based on what have been found in the data analysis, a model has been built and it successfully replicated the fractal structure of small world in the simulation.

Keywords

Social network, Small world, Fractal Structure, SNS

1. Introduction

When it comes to a friendship network, a series of Milgram's researches on "small-world phenomenon" are well known (Milgram 1967, Travers & Milgram 1969, Korte & Milgram 1970). Milgram et al. invented an experimental method called "small-world method" and repeatedly obtained the experimental results that showed there were only six degrees of separation in average between two arbitrarily sampled subjects. This was the beginning of the famous legend known as "six degrees of separation". An idea of "small world", all the people are connected by much shorter steps than we normally imagine, gives us some sense of security and has been accepted by many people.

1 College of Economics and Environmental Policy, Okinawa International University
2 Graduate School of Science and Engineering, Yamagata University
3 College of Economics, Osaka Prefecture University

Kleinfeld (2002a, 2002b), however, has raised questions about validity of the small-world experiments by Milgram et al. (Milgram 1967, Travers and Milgram 1969, Korte and Milgram 1970). According to her, Milgram's experiments have two defects, one is about sampling bias and the other is related to low frequency of completion of the chain letters. And then, suggesting that there exist effects of spatial and social distance on a friendship network, she proposed an idea, that is, "[t] he 'lumpy oatmeal' theory, that we live in a world with many small worlds possibly but not necessarily connected, might be viewed as the "weak" form of the small world phenomenon, for which we do have evidence" (Kleinfeld 2002a, p.65).

In this paper, the effects of spatial and social distance on a friendship network are analyzed. We used the data obtained from "Tomocom" (http://tomocom.jp/) which was opened and has been operated as a social network service (SNS) for undergraduate students. Detailed information on "Tomocom" is given in the next section. Based on what have been found in the data analysis in the section 2, modeling and simulations are conducted in the section 3. Discussion is given in the section 4.

2. Analysis on the Network Data

"Tomocom" is designed for assisting and promoting interchange between undergraduate students over the internet. It is established in January, 2009 and began to be operated practically in April, 2009. The data analyzed in this paper has been obtained in August, 2010. There are 351 undergraduates students, who register themselves at "Tomocom" by August, 2010. These students live in several separated prefectures such as Tokyo, Osaka, Okinawa, and so on, and go to 10 different universities. Among these 351 undergraduate students, 247 students are active users, that is, they have one or more friends in "Tomocom". In this paper, these 247 active users (senior: 43, junior: 128, and sophomore: 76) are the main subjects to be analyzed.

Figure 1 describes the friendship network in "Tomocom" which is drawn in Kamada-Kawai method (1989) where the nodes are configured tightly if the edges between them are dense. The edges in dark gray show friendship within the same university. On the other hand, the edges in light yellow show friendship over the universities. The differences of the colors in the nodes represent the differences of the university to which the active users belong. In Figure 1, the differences of the shapes in the nodes show the differences of the years of the active users (○: senior, △: junior, and □ : sophomore). We can see that the networks are cut into some pieces. The largest component consists of 162 active users and 645 edges between them. We can also see that the active users who belong to the same university gather together and that those who are in the same university are gathered furthermore according to the years.

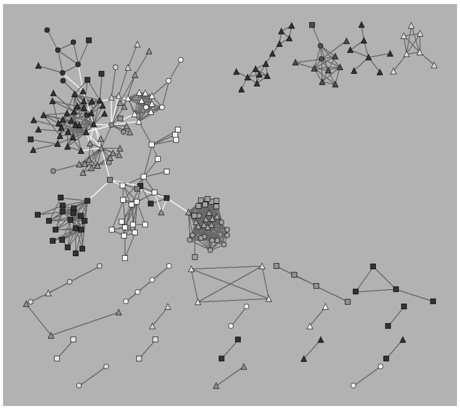

Figure 1

Table 1 shows the average degree in each 10 universities. The “within degree” in Table 1 (colored in dark gray in Figure 1) means the average number of edges between the active users who are within the same university. The “within degree” is divided into the “same-year degree” and the “different-year degree”. The “same-year degree” stands for the average number of edges between the active users who are in the same year in the same university. This is segmented into seniors, juniors, and sophomores. The “different-year degree” corresponds to the average number of edges between the active users who belong to the same university but are in the different years. On the other hand, the “beyond degree” in Table 1 (colored in light yellow in figure 1) stands for the edges between agents who have a link or edges beyond the universities.

	Average Degree	Within Degree	Same-year Degree				Different-year Degree	Beyond Degree
				Seniors	Juniors	Sophomores		
Mean	2.69	2.33	1.50	0.21	0.82	0.47	0.82	0.37
Median	2.38	1.66	1.28	0.08	0.83	0.33	0.22	0.10
C.V.	0.92	1.11	0.66	1.95	0.68	1.02	2.09	1.98

Table 1

The important fact in Table 1 is that the following inequalities hold;

the mean “within degree” 〉 the mean “beyond degree” …… (1)

and

the mean “same-year degree” 〉 the mean “different-year degree”,

…… (2)

that is, there is stratification by the differences of the universities and years. In other words, the most edges are likely to be created between those who are in the same university and the same year, and then, the second most edges tend to be made between those who are in the same university but the different years. The most difficult edges to be stretched are between those who are in the different universities. It is not hard to imagine what is mentioned above, though, the inequalities (1) and (2) show that the data is suggesting the spatial distance (difference in universities) and social distance (difference in universities and/or years) between individuals give stratification in the friendship network.

Table 2 shows clustering coefficient (C), average path length (L), the number of nodes (N), the total degree (K), and the average degree ($\langle k \rangle = 2K/N$) of the whole and partial networks in "Tomocom". The values for universities 1 to 5 in Table 3 are obtained for the active users who are in the largest component.

	C	L	N	K	〈k〉
Activer Users	0.448	–	247	745	6.032
Largest Component	0.512	5.069	162	645	7.963
University 1	0.608	3.136	33	98	5.939
University 2	0.322	2.562	24	34	2.833
University 3	0.423	2.308	36	108	6.000
University 4	0.624	1.784	19	63	6.632
University 5	0.816	1.372	31	302	19.484
Mean	0.558	2.232	29	121	8.178
Median	0.608	2.308	31	98	6.000

Table 2

	C / C_rand	L / L_rand
Activer Users	18.2604	–
Largest Component	10.3568	2.0671
University 1	3.2749	1.5981
University 2	2.6115	0.8394
University 3	2.4656	1.1540
University 4	1.6933	1.1460
University 5	1.2569	1.1865
Mean	2.260	1.185
Median	2.466	1.154

Table 3

Let us examine if the largest component and the partial networks of each university in "Tomocom" has the properties of small-world network. In order to do so, we compare C and L in Table 3 and those of random graphs ($C\_rand = \langle k \rangle/(N-1)$ and $L\_rand = \ln N / \ln \langle k \rangle$) whose size are the same as the cases in Table 3. If a network contains the features of small-world, then it will hold $C/C\_rand \gg 1$ and $L/L\_rand \approx 1$ that are considered in Table 3 (Watts and Strogatz 1998). We can see that the clustering coefficients (C) for the largest component and the university 1 are statistically large enough. The characteristic path length (L) is not too large compared to that of the random graph. However, it was not statistically significant. The reason why L did not become statistically short enough seemed to be due to the spatial and social distance between individuals.

From the above, for the largest component, it is suggested that spatial and social distance between individuals causes stratification and that as Kleinfeld's (2002a, 2002b) idea the largest component becomes small-world network that consists of a local small-world network. Or, more precisely speaking, it has a fractal structure of "almost" small-world network as Tomochi (2010) showed in his model. Moreover, when we look at the whole network that contains 247 active users, we reach the conclusion that a friendship network is "not necessarily connected" as Kleinfeld (2002a) said.

3. Modeling and Simulations

Based on what we saw in the data analysis in the previous section, a model is built under the scenario that an agent decides to befriend or not to befriend with other agents depending on the spatial and/or social distance between them. Note that, essentially, like other SNS, the number of nodes and edges in "Tomocom" varies dynamically. However, this situation is represented by introducing a probabilistic model in this paper.

From now on, the probability that a node i and a node j become friends each other is denoted as P_{ij} and it is a function of the spatial and/or social distance, that is,

$P_{ij}(SS)$: the probability that i and j become friends each other when i and j go to the same university and are in the same year,

$P_{ij}(SD)$: the probability that i and j become friends each other when i and j go to the same university but are in the different years, and

$P_{ij}(D)$: the probability that i and j become friends each other when i and j go to the different universities.

Here, Figure 2 describes the distribution of the degree in the friendship network in "Tomocom". Since the number of data is not large enough, it is not very clear, however, we can observe power law in the degree distribution (Barabkgi & Albert 1999). This is suggesting that the motivation to make friends for the active users varies over the individuals. Based on what we observed in Figure 2, it can be said that the willingness of making friends for the active uses varies. For simplicity, we assume that there are two types of agents, that is, those who are motivated to make friends and those who are not motivated to do so. The parameter a in the model denotes a ratio of those who are motivated and hence $1\text{-}a$ does a ratio of those who are not motivated.

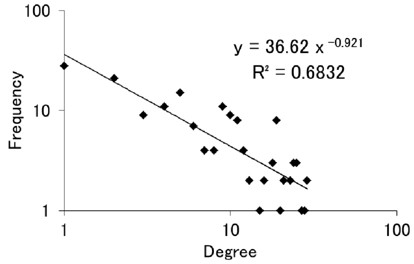

Figure 2

Based on the inequalities (1) and (2) and what has been discussed above, if a node i is motivated then it holds

$$P_{ij}(SS) > P_{ij}(SD) > P_{ij}(D) > 0. \qquad \ldots\ldots \quad (3)$$

On the other hand, if a node i is not motivated then, for simplicity, it is assumed that it should hold

$$P_{ij}(SS) = P_{ij}(SD) = P_{ij}(D) = 0. \qquad \ldots\ldots \quad (4)$$

For simplicity, let us assume that there are 10 universities and in each university there is a class that consists of 10 seniors, 10 juniors and 10 sophomores who register themselves with "Tomocom". Now we have 300 (=10*10*3) undergraduate students participating in the model. In order to represent what was happening in "Tomocom", we have set the parameters as follows and conducted simulation.

a = 0.5 (half of the students are motivated to make friends and the rest are not)

For a motivated student, it holds

$P_{ij}(SS) = 5/9$ (making 5 friends out of 9 students who are in the same university and the same year),

$P_{ij}(SD) = 1/20$ (making 1 friend out of 20 students who are in the same university but the different years), and

$P_{ij}(D) = 1/3/270$ (making 1/3 friends out of 270 students who are in the different universities and years).

For a not motivated student, it holds

$P_{ij}(SS) = P_{ij}(SD) = P_{ij}(D) = 0$ (no making friends voluntarily)

Figure 3 shows an example of a network generated by the simulation. Out of 300 registered nodes, 262 became active users who have one or more connection. The number of edges between those who are in the same university and the same year is 367, the number of edges between those who are in the same university but different years is 75, the number of edges between those who are in the different universities is 24, and hence the total number of the edges counts up 466. The colors and shape of the edges and nodes are arranged in the same manner of Figure 1. We can see that stratification by the difference of universities and years of the nodes is replicated in the simulation.

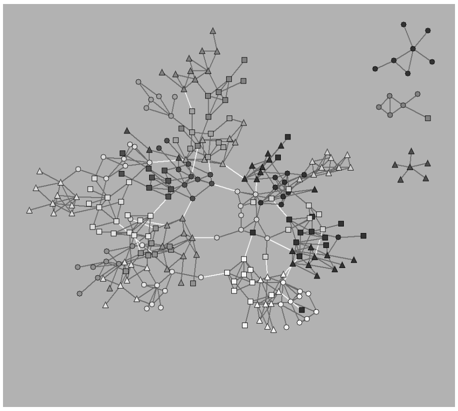

Figure 3

The largest component in Figure 3 consists of 243 nodes, 349 edges between those who are in the same university and the same year, 74 edges between those who are in the same university but different years, and 24 edges between those who are in the different universities. The clustering coefficient and average path length of the largest component are 0.2589 and 7.1989, respectively. When we compare these values with those of random graph, we have C / C_rand = 17.0301 and L / L_rand =1.7072, that is, C is large enough (statistically significant) and L is relatively small (not statistically significant, though). This means that the largest component is "almost" small world like the one we saw in the previous section.

When we take a close look at the yellow nodes (say, university 1) in the largest component, the number of nodes is 30, the number of edges is 56, C is 0.2677, and L is 2.7795. We have C / C_rand = 2.0795 and L / L_rand = 1.0764, that is, C is statistically large enough and L is statistically small enough. This means that the partial network in the largest component has the properties of a small-world network. Considering the circumstances mentioned above, it is suggested that a fractal structure of the ("almost") small-world is reproduced in the simulation.

4. Discussion

The main purpose of the research was to analyze the effects of spatial and social distance on a friendship network. We used the data obtained from

"Tomocom" which was designed for assisting and promoting interchange between undergraduate students. The data suggested that spatial and social distance between individuals caused stratification in the friendship network and brought about fractal structure of ("almost") small world. Based on what we found in the data analysis, we built a model and successfully replicated the fractal structure of small world.

To analyze if we can find the fractal structure (self-similarity) of small world in other networks is one of our future works as well as analyzing the network dynamically.

Acknowledgement

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research, 20200042, 2009.

References

Barabási, A.L., and Albert, R. 1999 "Emergence of scaling in random networks" Science 286 509-512.

Kamada, Tomihisa, and Kawai, Satoru. 1989. "An algorithm for drawing general undirected graphs" Information Processing Letters 31 (1) 7-15.

Kleinfeld, Judith S. 2002a. "Six Degrees of Separation: An Urban Myth?" Psychology Today 35 (2) : 74.

Kleinfeld, Judith S. 2002b. "The small world problem." Society 39 (2) : 61-66.

Korte, Charles, and Milgram, Stanley. 1970. "Acquaintance links between White and Negro populations: Application of the small world method." Journal of Personality and Social Psychology 15 (2) : 101-108.

Milgram, Stanley. 1967. "The Small World Problem." Psychology Today 1: 61-67.

Tomochi, Masaki. 2010. "A model of a nested small-world network." 理論と方法 25 (1) 19-29.

Travers, Jeffrey and Milgram, Stanley. 1969. "An experimental study of the small world problem." Sociometry 32: 425-443.

Watts, Duncan J. and Strogatz, Steven H. 1998. "Collective dynamics of 'small-world' networks." Nature 393: 440-442.