Evolution of Friendship Networks and Transition of Their Properties

Hiroshi Matsushima and Nobuhiro Inuzuka

Nagoya Institute of Technology,
Gokiso-cho Showa, Nagoya 466-8555, Japan

inuzuka@nitech.ac.jp, matsuhima@nous.nitech.ac.jp

Abstract. We analyze friendship networks using attendance records of students to classes. Acquire and analysis of networks are time-consuming. Hence analysis for the network evolution according time passed is highly cost. We have proposed an automated method to acquire friendship networks for students using a system which is becoming popular recently in universities. The method gives a score which measures a degree of friendship between two students based on the probability they are friends. By using these networks, we analyze how friend pairs evolve and how common properties of social networks, scale-free, cluster and small-world, change.

1 Introduction

Since the 1970s, the social network analysis (SNA) has been actively researched in sociology, and recently it gathers attention from computer scientists as well. SNA pays attention not to attributes of entities but relations among entities. For personal relationship, inter-business relationship, inter-nation relationship and WWW, researchers have been studied how network forms and evolves, who has a primary role, what factions or communities exist and what structure exists behind.

In order to begin these investigations we first acquire network data, that is, we need qualitative information among entities. Usually for this purpose we use a questionnaire and it is always time-consuming. Accordingly study on formation and evolution of networks are difficult indeed. An automated method to collect network data is desired and is useful for these studies.

Inuzuka et al.[1] introduced an automated method to collect data for friendship networks using a system which collects attendance records to classes in school. The system or similar systems are becoming popular these days. It consists of student cards with IC-chips that all students keep as their IDs and card readers equipped in all lecture rooms. For each class students check their attendance with the cards at a card reader in the room. Inuzuka et al.[1] showed that the attendance records can be used to predict the friendship relation among students.

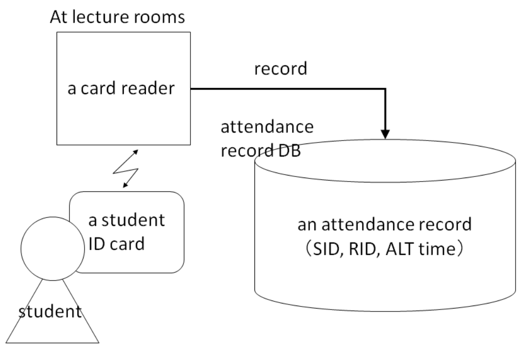


Fig. 1. The outline of CARMS system.

The purpose of this paper is to show that the method of friendship is useful to detect and analyze friendship network and their transitions. In this paper, first we introduce the system we used which collects date automatically. Section 3 reviews the method of friendship score using attendance records in university lecture class and generation of friendship network among students. Then we give analysis of temporal transition of network formation in Section 4. In Section 5 we see the common properties of social networks, that is, properties on scale-free, high cluster coefficient and small-world.

2 Class Attendance Record Management System and Friendship Relation

Nagoya Institute of Technology (NIT) installed a system to collect and manage class attendance records of students, which we call CARMS (Class Attendance Record Management System), in 2007. It aims to reduce tasks of instructors by collecting records automatically. The CARMS system consists of student ID-cards, card readers and a database management system (DBMS). A student ID card has a function of a wireless tag and keeps the information of the student ID of a card holder. A card reader has its own ID (reader-ID) and reads the information of an ID card when a holder places his/her card in the front of the reader. Each lecture room equips two or three readers near the entrances. Readers send the information as a tuple (SID, RID, ALT) to DBMS, where SID is a student ID, RID is a reader ID, and ALT is an attendance/leaving time, that is the time when the student puts close his/her card to the reader at his/her attendance to or leaving from a class. The DBMS collects and keeps all the information. An outline of the system is illustrated in Fig.1. CARMS gives lists

Table 1. Summary of records collected by CARMS system.

The period of records	2007.10.1 – 2008.3.31
# students recorded	4,403
of which first year students	942
of which second year students	936
of which third year students	929
of which fourth year students	287
# readers equipped	129
# records	864,882
of which first year students	295,700
of which second year students	242,001
of which third year students	165,588
of which fourth year students	18,463

of students who attended a class on a specified date. Although CARMS has also many other functions, we omit to describe the detail.

Table 1 shows the basic data of the system and records collected and used for our experiments. Student IDs for fourth year students appeared in records and their attendance records are fewer than other years. This is because the most of fourth year students participate only graduate research projects or take a small number of lectures. Graduate course students are also in the similar situations. Accordingly we used only the data of 1st, 2nd and 3rd year undergraduate students.

3 Friendship scores based on attendance records

Inuzuka et al.[1] utilized the character of frequency distributions of time differences of ALT (ALTD) between the each friend pair and non-friend pair(Fig. 2). The distribution of ALTD between friend pairs and one for non-friend pairs have a remarkable difference. Fig. 2 shows the distributions. The ALTD distributions have peeks at a small time length. Ones for friend pairs the peeks becomes acute.

Using the character of distributions a prediction method can be given. We take two students A and B. Let f is the event that A and B are friends each other, and $T = \{t_1, t_2, ..., t_n\}$ is a set of ALTD data between A and B. Here we assume that all t_i $(i = 1, \dots, n)$ are independent. Then, conditional probability $p(f \mid T)$ can be written as follows,

$$p(f \mid T) = \frac{p(f) \cdot p(T \mid f)}{p(T)} = p(f) \prod_{t \in T} \frac{p(t \mid f)}{p(t)}. \quad (1)$$

We move to think of the ratio, r_t, of ALTD records which are of time length t and are given by friend pairs against all ALTD records of t, that is, represented

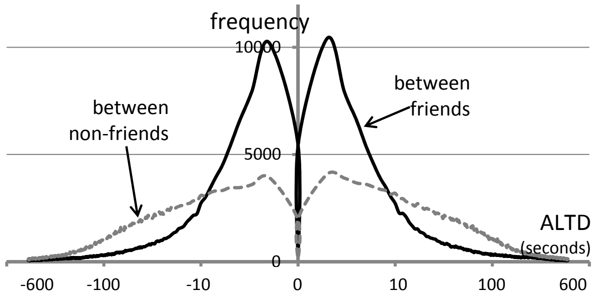


Fig. 2. Frequency distribution of ALTD

as follows,

$$r_t = \frac{\#(\text{all ALTD records of } t \text{ between friend pairs})}{\#(\text{all ALTD records of } t \text{ between all students pairs})} = \frac{X_f \cdot m_f \cdot p(t \mid f)}{X \cdot m \cdot p(t)}, \quad (2)$$

where X is the number of all of students pairs, X_f is the number of all friend pairs, m is the expected number of ALTD records produced among randomly chosen two students, m_f is the expected number of ALTD records among friend pairs, $p(t)$ is the probability that randomly chosen two students have ALTD t, and $p(t|f)$ is the probability that randomly chosen friend pair have ALTD t. Then, we can reformulate $p(t \mid f)$ as follows.

$$p(t|f) = \frac{X \cdot m \cdot p(t) \cdot r_t}{X_f \cdot m_f} = \frac{m \cdot p(t) \cdot r_t}{p(f) \cdot m_f} \quad (3)$$

When we substitute this to Equation (1) we have the following equation.

$$p(f|T) = p(f) \prod_{t \in T} \frac{m \cdot r_t}{m_f \cdot p(f)} = p(f)^{(1-n)} \left(\frac{m}{m_f}\right)^n \prod_{t \in T} r_{tf}, \quad (4)$$

where $n =\mid T \mid$. When $\bar{f}$ denotes the event that two student are not friends. Then, $p(\bar{f} \mid T)$ is following.

$$p(\bar{f} \mid T) = p(\bar{f})^{(1-n)} \left(\frac{m}{m_0}\right)^n \prod_{t \in T} (1 - r_t), \quad (5)$$

where m_0 is the expected number of ALTD records produced non-friend pair.

Then we give the friendship score by the logit of $p(f \mid T)$ as follows.

$$\operatorname{logit} p(f|T) = \log\left(\frac{p(f|T)}{1 - p(f|T)}\right) = \log(p(f|T)) - \log(p(\bar{f}|T)) \quad (6)$$

Table 2. Recall and Precision of each classes

	class K	class J	class P
Recall	75.3	86.9	65.7
Precision	60.0	88.9	78.9

Table 3. Basic information of friendship network.

term	number
Number of students	948
Number of student pairs	579552
Number of friend pair	5972

In [1] the recall and precision of the friend estimation using friendship score are reported as around 70% compared with questionnaire. When considering the difficulties to take questionnaire the prediction is useful.

4 Friendship networks

Using the friendship score method friendship networks can be generated and then we may analyze some properties among students and their friendship networks. The properties include basic information such as the number of friend pairs, and network formation and also the common properties which characterize such networks.

Basic information of friendship networks Friendship scores for freshman students in fiscal 2007 are calculated and we made their friendship networks. In order to evaluate clustering-coefficient, weights of links, i.e. friendship scores, are normalized into the range of 0 to 1. We gave weights by applying the sigmoid function to friendship scores. Fig. 3 illustrates a part of a friendship network generated. In addition, table3 summarizes the number of all students, the number of all student pairs between which there exist at least an ALTD datum less than ten minutes, which means they are potential candidates for friendship, and the number of friend pairs presumed by the friendship score.

Analysis of network evolution In order to see possibility of our method for study of evolution of friendship networks, we try to observe transition of network properties. First, we observes the transition of the number of links, the number of friendship relations in the networks.

For our purpose we generate networks from a fixed shorter period of the year and put off the period by a day. We may understand a network generated from data of a period as an averaged network along the period. By putting off

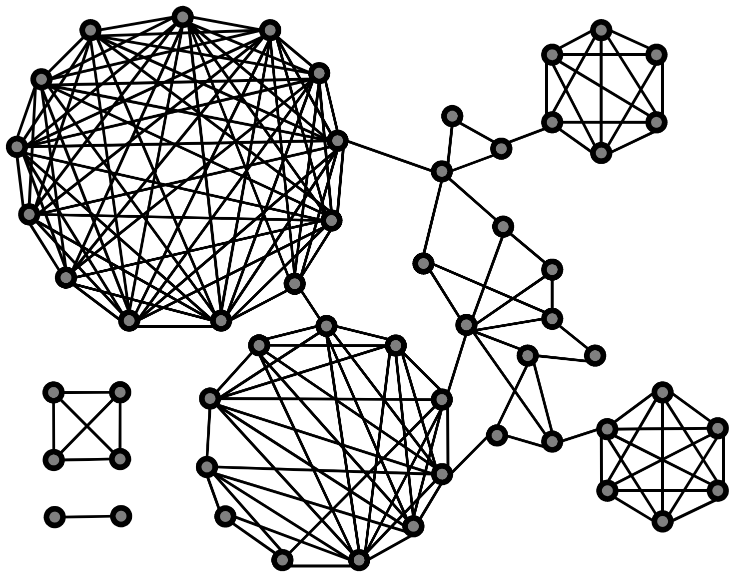
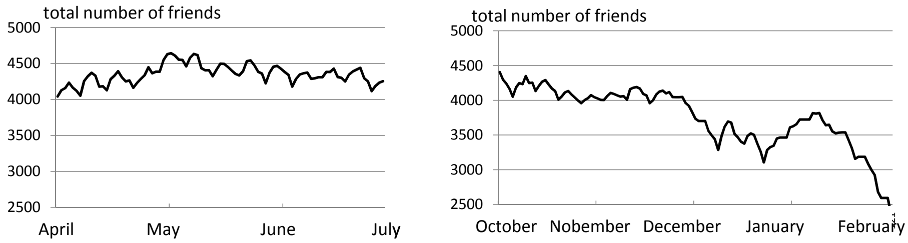


Fig. 3. A part of friendship network generated from the method.

the period used we can generated series of networks which may represent the evolution of the friendship.

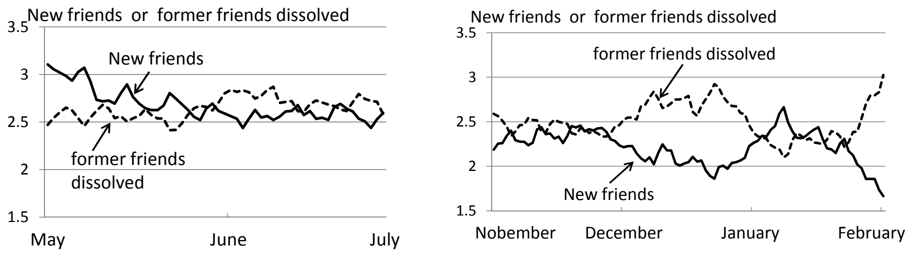
In preliminary experiments we know that a sample of three weeks is enough to generate a networks. Accordingly we generated a series of networks from data of a month (30 days) and shift the period by a day. Fig. 4 shows the transition of the number of links in the series of networks. In NIT the first semester starts 1st April and ends 31st July. Accordingly the first period is from 1st April to 30th April and the last period is from 2nd July to 31st July. The x-axis shows the start date of the periods. In the first semester, there is no substantial change but friend pairs increase by around 600. Afterward, it is stable in both first and second semester until December. December and the following month include winter holidays and collection ALDT data are disturbed. The changes in the second half of the second semester may be caused from these reasons, although we need more investigation.

Using the same network series, we analyze increase and decrease of links according to the period. That is, we observed the number of newly made friend pairs and friendship disappeared compared to a networks of a fixed period, say a month, before. In Fig. 5 (a) shows the result for the first semester and (b) shows one for the second semester. In the both graph, the solid line shows the number of new friends, that is, friendships which exist in the network after one month but do not exist before. The broken solid line shows the number of friends



(a) First semester (b) Second semester

Fig. 4. A transition of number of friend pairs.



(a) First semester (b) Second semester

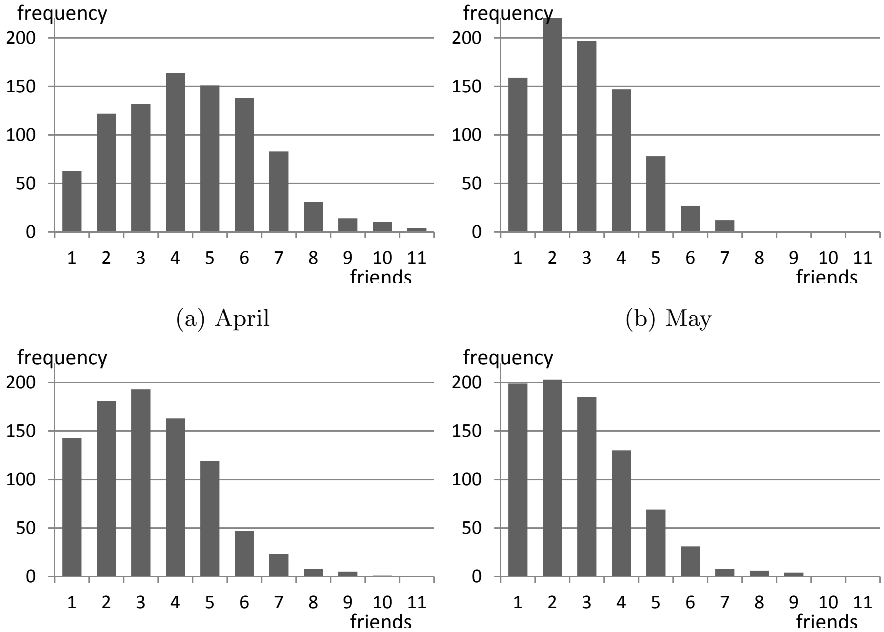
Fig. 5. Transitions of newly generated and dissolved friendships.

dissolved, that is, friendships which do not exist in the network after one month but exist before.

In early stage of first semester, newly generated friendships are larger than dissolved ones, so the number of friend pair increased in early stage of first semester. Afterward, although friendship keep changed, both of them were stable from 1.5 by 1.7. Thus, we can consider that a strength of friendship is dynamic. Of course we need carefully verify an effect of accuracy of friendship score.

5 Transitions of common properties of social network

By the previous literature, it has been clear that social networks have common properties, scale-free[2], strong cluster and small-world[3]. Scale-free shows that very few people have large number of friends but others have a few number of friends, and degree k's distribution $p(k)$ obeys $p(k) \propto k^{\gamma} (\gamma < 0)$, i.e. the power law. Strong cluster means that friends of a person are likely friend mutually. In order to measure the strength of cluster, we see the clustering-coefficient. Small-world means that everyone can reach an destination in relatively small steps, and means that average shortest path distances among nodes are short. Here we analyze transition of these properties.



(a) April (b) May

(c) June (d) July

Fig. 6. Degree distributions of friendship networks in the first semester.

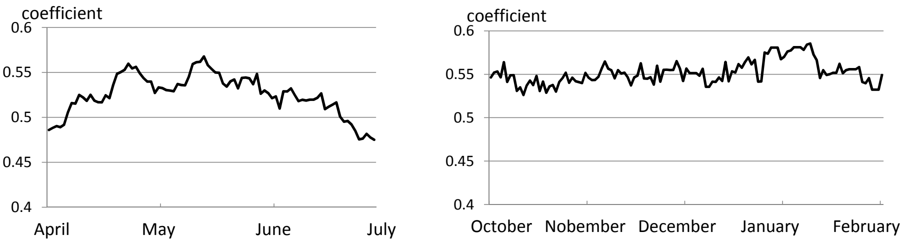
Transitions of scale-free property Fig. 6 illustrates the degree distribution of the friendship networks. The degree of a node is the number of neighboring nodes of the node, that is, the number of friends in friendship networks.

In April, the number of student who have four friends is the highest and the distribution may not obey power law. However, as time advances, the number of students who have less than three friends increases, while the number of students who have more than four friends decreases. The degree distribution in July may acquire the power law. Afterward, degree distributions become stable in second semester. Thus, friendship networks get scale-free property in around four months.

Transitions of cluster In order to analyze cluster, we use cluster-coefficient for weighted and directed graphs[6] Let $G = (V, E, c)$is a weighted and directed graph, where V is a set of vertexes, E is a set of edges and c is a weight function. The cluster-coefficient C of G can be calculated as follows[6].

$$C = avg_{v_i \in V}(C(v_i))$$

$$C(v_i) = \frac{\sum_{v_j \neq v_k \in E(v_i)} \{c(v_i, v_j) \cdot c(v_j, v_k) + c(v_i, v_k) \cdot c(v_k, v_j)\} / \{c(v_i, v_j) + c(v_i, v_k)\}}{{}_{k|E(v_i)|}\mathrm{P}_2}$$



(a) First semester (b) Second semester

Fig. 7. Transitions of cluster-coefficient

Fig. 7 shows transitions of C calculated by the networks. In first semester, cluster-coefficient increases from April to early June but decreases afterward. In second semester, it is stable around 0.55 except span affected by vacation. In a whole year, about 0.06 increases. A popular SNS networks that is generally known that cluster is high, has its cluster coefficient about 0.36. Comparing it the friendship networks have remarkably high cluster coefficient. However, it may be affected by accuracy of friendship estimation by friendship score and we need to verify it.

Transitions of small-world In order to analyze small-world, we calculate average shortest path distances of friendship networks. At first, we divide the network of each strongly connected components. A strongly connected component is a subnetwork in which every node has a directed path to all other nodes. When a network has isolated parts the average of shortest path distance is not defined. However when a network include a big component which include almost all nodes, the average of shortest path distances should be defined for the components.

In general friendship networks consist of some small components that contain only a node, i.e. students, or a few nodes and a unique large component that contain almost all nodes. We observes the shortest path distances for the unique large components.

Table 4 shows the number N, of students contained in the unique largest components of each month, and the average shortest path distances, ASPD, of networks.

Each average shortest path distances does not change much more than properties and is around six. This is the almost same result of Milgram's experiment in the 1960's[4]. Also, in December and January, average shortest path distances are longer and number of vetexes are lower than others. This is because number of ALT date in December and January are lower than others due to long vacations and this reflects accuracy of friendship estimation of friendship score. The meaning of these value need to be verified.

Table 4. Shortest average path lengths of the networks.

	April	May	June	July	Oct.	Nov.	Dec.	Jan.
N	895	909	898	885	892	866	822	802
ASPD	6.271	5.982	5.831	5.992	6.179	6.477	7.120	7.395

6 Discussions

In this paper, we made friendship networks using friendship score and analyzed transitions of network formation, transitions of three common properties of social networks and students' roles comparing questionnaires with students whose centralities are high.

A number of friendship links among students increased by end of May and network stabilized in early stage. Because strength of friendship consistently varies, friendship activities are dynamic. By analysis of transitions of three common properties, we find tendencies that degree distribution comes to acquire the power law in first semester, cluster coefficient increases in early stage and decreases before it stabilizes and shortest average path lengths do not change but are around six.

By analysis of students' role using centralities, using centralities on degrees, such as Degree and Pagerank, is predictably-effective to analyze roles on authorities.

Therefore, friendship network analysis using friendship score reflects well-known common properties of social networks and is useful to analyze transitions of friendship network and students' role.

Further interesting research topics include more detailed analysis of results, relation between friendship estimation and accuracy of friendship score and detection of new transitive properties. Also, we need to verify whether we can get the same transitional results from other networks. Moreover, it concludes suggestion of new graph models hold these properties.

References

1. N.Inuzuka, T.Nakano, K.Shimomura. Friendship Analysis Using Attendance Records to University Lecture Classes. IASK International Conference Teaching and Learning, 478-486, 2008.
2. N.Inuzuka, T.Kondo, S.Yamamoto. Analysis of Asymmetric Friendship among Students from Class Attendance Records. New Advances in Intelligent Decision Technologies (Studies in Computational Intelligence, Vol. 199), Springer, 393-405, 2009.
3. R.Albert, A.-L.Barabasí. Statical mechanics of complex networks. Review of Modern Physics, 74, 47-97, 2002.
4. S.Milgram. The small world problem. Phycology Today, 1:67, 1967.
5. D.J.Watts. Small Worlds. Princeton University Press, 1999.
6. T Suzuki. Analysis for Directed and Weighted Complex Networks. JWEIN07, 2006.