

Grouping Methods for Generating Friendship Based on Networks Properties

Ryumaru Kato
Dept of Computer
Science and Engineering
Nagoya Institute of Technology
Nagoya, Japan 466-8555
Email: kato@nous.nitech.ac.jp

Atsuko Mutoh
Dept of Computer
Science and Engineering
Nagoya Institute of Technology
Nagoya, Japan 466-8555
Email: atsuko@nitech.ac.jp

Nobuhiro Inuzuka
Dept of Computer
Science and Engineering
Nagoya Institute of Technology
Nagoya, Japan 466-8555
Email: inuzuka@nitech.ac.jp

Abstract—This paper investigates the effect of group work with the assumption of three motivators to make friends. Obeying the assumption we proposed twelve variation of methods for grouping students. The effects are evaluated by some measures from social network analysis and by the changes of real friendship networks, which are observed by a friendship prediction method. The proposed methods brought new friendship among students to classes and made rearrange of community structure.

I. INTRODUCTION

Class management is an important task of teachers. It aims to let students direct ahead for learning and concentrate subjects in class rooms. Class management is whole activities by teachers for these purposes. It includes advising to students in their study, currier direction and everyday life, construction of good relationships among students, preparing efficient learning environment in the lecture class, such as equipment and neat rooms, forming an atmosphere of classes' interest to study theme, and many things. To keep good friendship in classes is essential for these purposes. Good friendship causes efficient study in group work and discussion in classes and make good atmosphere toward study. Conversely good relationship is a result of good class work, as well. Class activity and friendship are the both sides of coin. Accordingly teachers can help students to forms good friendship through class works.

It is difficult to observe the friendship, however. To know the friendship teacher need careful observation of students' behavior, which need large effort of teaches. Fortunately the efforts can be reduced by a system. Class rooms are assisted by many electrical devices, which include automatic counting devices of students' attendance to classes. The system works with student cards tagged by IC chips. It records the time when each student attended the class. Assuming the system and data recorded for attendance, the work by Shimomura[?] gave a method to predict friendship. This paper investigates an effect of group work to form friendships and gives effective methods for this purpose using some useful concepts from social network analysis and the friendship prediction method.

The following section reviews some concepts of SNA and the friendship prediction method. We consider factors to make friendship among students from some aspects of network features in Section III. Then we design twelve methods for grouping students based on the consideration in Section IV. The methods are evaluated by measures of SNA concepts and

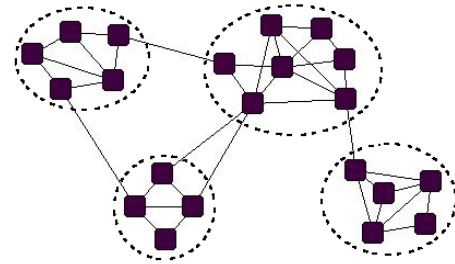


Fig. 1. An image of community structure of a network. There are dense edges in community.

from changes of friendship networks through a group work done by students with groups formed by the grouping method in Section V. Section VI gives conclusions.

II. SOCIAL NETWORK AND FRIENDSHIP NETWORKS

A. Concepts in social network analysis

For preparation this section gives some concepts on SNA. A network is represented by a graph $G = (V, E)$ which consists of a set V of nodes and a set E of edges that is a pair of the nodes connected in the network. A social network is a social structure made up of a set of actors. SNA is the methodical analysis of social networks. SNA views social relationships in terms of network theory.

We give some important concepts on network analysis, such as, centrality, dense, community structure, and cluster property, which are used for grouping. Node centrality is a concept to measure influence or importance of nodes in a network. Although there are many centrality concepts, the most simple and popular one is the degree centrality. For a network $G = (V, E)$ the degree of a node $i \in V$ in a network is the number of edges incident to i and it is denoted by k_i . The degree centrality is used as a centrality of the node. There are many other centrality concepts based on popularity, distance in the network, network flow and combination of them.

Nodes tend to create tightly knit communities characterized by a relatively high density of ties; this likelihood tends to be greater than the average probability of a tie randomly established between two nodes. A network is said to have community structure if the node of the network can be easily

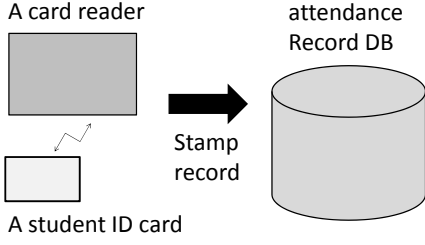


Fig. 2. The class attendance management system that consists of student ID's with IC tag, card readers and a database system.

grouped into sets of nodes that have dense connection. The number of communities in a network is typically unknown and the communities are often of unequal size and density. Despite these difficulties, however, several methods for community finding have been developed and employed with varying levels of success.

Modularity is one measure of the structure of networks[3]. It was designed to measure the strength of division of a network to communities. Let us imagine a network $G = (V, E)$ and its community partition $P = \{C_1, C_2, \dots, C_L\}$, where $C_1 \cup C_2 \cup \dots \cup C_L = V$ and $C_i \cap C_j = \emptyset$ for any $i \neq j$. Then the modularity Q of this community partition is given by the following equation,

$$Q(G, P) = \frac{1}{2|E|} \sum_{C_i \in P} (e_{ii} - a_i^2),$$

where e_{ii} is the number of edges with both ends in the same community C_i and a_i is the number of edges with at least one node in community C_i .

The cluster coefficient is also an important concept, which is high in typical social networks. It is calculated by the following C .

$$C(G) = \frac{1}{|V|} \sum_{i \in V} C_i, \quad C_i = \frac{\text{\#triangles including } i}{k_i C_2}$$

The C_i is the fraction of triangles tied including a node i in a network to the possible number of triangles, that is, the combination $k_i C_2$ of two nodes from all nodes indident to i .

B. Prediction of friendship networks

Recently colleges have installed a system to collect and manage class attendance records of students, in order to reduce cost and to make easy for student management. The system consists of student ID-cards, card readers and a database management system (see Fig. 2). 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 of a student ID, a reader ID, and the time, which is called an attendance/leaving time (ALT) to DB.

In Shimomura et al.[?] proposed a friendship score, which is a friendship strength based on ALT data. The idea is that

friend students tend to act together and then their ALT time are expected to be close. This expectation was true. They observed that the distribution of difference of ALT time, ALT D (Attendance/Leaving Time Difference), of two students strongly depends on whether they are friends or not. From this observation Shimomura et al. gave a procedure to predict friendship relation among students.

Let f be the event that a pair of students are friend pair and $T = t_1, t_2, \dots, t_n$ be a set of ALT D among them. Each ALT D datum t_i is assumed to be independent each other in [?]. Then, it gives the probability $p(f | T)$ of being friend under the condition of observation of T is as follows,

$$P(f | T) = \frac{P(f) \cdot P(T | f)}{P(T)} = P(f) \prod_{t \in T} \frac{P(t | f)}{P(t)}$$

In order to calculate the probability from observed data, [?] derives the following equation,

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

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 ALT D records produced among randomly chosen two students, and m_f is the expected number of ALT D records among friend pairs and r_t is the ratio of ALT D records of time period t from friend pairs against all ALT D records including non-friend pairs, which is given by,

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

The following logit gives a degree of friendship and it is called friendship score, which ranges in $[-\infty, +\infty]$.

$$\begin{aligned} \text{logit} P(f | T) &= \log \frac{P(f | T)}{1 - P(f | T)} \\ &= \log(p(f | T)) - \log(P(\bar{f} | T)) \end{aligned}$$

A positive friendship score between two students means that the probability of friendship between them is higher than the probability of non-friendship. Hence we may understand that a positive score corresponds to a friendship. ?? reports about 70% of the accuracy of the prediction.

In the rest of this paper we use the friendship score in order to observe friendship networks. Fig. 3 is an example observed by the procedure.

III. MOTIVATORS TO ACQUIRE FRIENDS

Here we discuss about motivator to make friends. We counts three different motivators as follows:

- The chance of two parsons who are not friend yet to communicate each other.
- The chance of two parsons who have common friends to communicate each other.
- The chance of a parson who may have a small number of friends to communicate with a parson who has many friends.

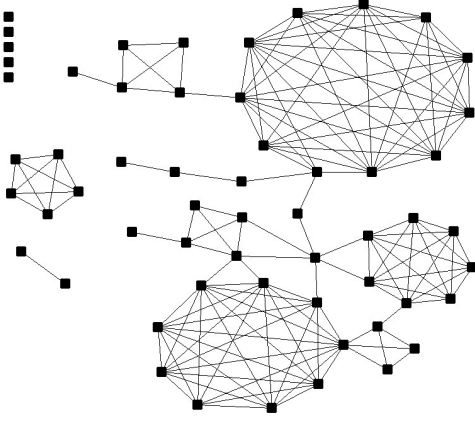


Fig. 3. An example of friendship network among students of a class. The friendships are predicted by the procedure[?] using class attendance records.

TABLE I. RESULTS OF QUESTIONNAIRE TO ASKING STUDENTS ABOUT FRIENDS IN GROUPS BEFORE AND AFTER GROUP WORK.

	the rate of parsons who had friends in a group when assigned	the rate of parsons who did or did not make friends during a grope work	
		did	did not
2008	5%	74%	21%
2009	10%	62%	28%
2010	6%	65%	29%

First we take consider the first motivator. Table I is a part of result of the questionnaire asking students who participate a group work class if they acquire friends during the class. The table gives three columns for three years of the class. The first column is the percentages of students who have friends in their assigned group initially. The second and third columns are the percentages of ‘yes’ and ‘no’ for the question. According to this questionnaire result, the lower the rate of percentages of students who has friends in the assigned group the higher the rate of the answer ‘yes’. Some of student gave an opinion that it is difficult to acquire new friends when they have friends before the class. Then the small number of friends in assigned groups can be a motivator to acquire new friends.

When we consider that a good community tends to have triangle, having a chance to make triad, that is three parsons who know each other via a parson but two of them are not friend directly, can be another motivator. That is, two parsons who have common friends but are not friends yet will become friends during a work. Then, this situation can be the second motivator.

It is guessed that a person with many friends is sociable and excels in making friends. The degree centrality may represent this view. A person with high degree has good influence for the group to which the parson belongs. He/she motives to other people to be social and make friends. When a group is small, a person with leadership is very important. Conversely a person with few friends had few opportunities of friend making. Making an asocial parson and a social person into a group causes good effect for both persons. Then this situation can be the third motivator.

Fig. 4 shows two grouping examples, where nodes are students and edges are friendship among them, and they

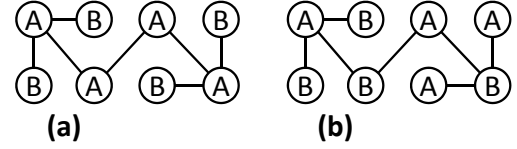


Fig. 4. Two grouping examples. Nodes are students and edges are friendship. The letters A and B mean two groups assigned.

are assigned to two groups A and B. By the examples we consider the three motivators. In grouping (a) group B has the first motivator because the grope has no friendship among members, while there are three links among members in group A. We cannot say that the whole network in (a) satisfies the first motivator. For the second motivator, two pairs of student in group A have common friends. For group B two pairs of students have common friends as well. Then the grouping (a) is weak for this motivator. For the third motivator the high degree students are only in group A of grouping (a) and then the motivator works less effectively. On the other hand, in the grouping (b), three motivators work more efficiently. There is neither friendship link in group A nor B. Both groups have four pairs of students who have common friends. Both groups include students with high degree and also ones with low degree, as well. Therefore, the grouping (b) fully satisfies the three motivators.

IV. GROUPING METHODS

In this section we give grouping methods according to top-down and bottom-up ways. A top-down method takes whole students as a group and divides the group into two groups. Each of two groups will be divided repeatedly. A bottom-up method treats each node as a group first and it unifies them until some condition satisfies. For our requirement groups given by grouping should have the equal or near size. The number of groups is given as a requirement by the work as well. Unfortunately the top down ways is difficult to control these requirements. Therefore, we take a bottom-up ways for consideration of grouping methods.

We can imagine two different schemes of grouping in bottom-up way, sequential and parallel methods, described as follows.

Sequential scheme

- 1) Choose an initial node (or, a pair of students) and it is considered as a group.
- 2) For the group, choose a required number of appropriate members evaluated by a fitness measure and make them belong to the group.
- 3) Repeat the steps 1 and 2 above until all of students are grouped.

Parallel scheme

- 1) Choose a required number of nodes (or, pairs of nodes) and each of them is considered as a group.
- 2) For every group, choose a member evaluated by a fitness measure and make it belong to the group.
- 3) Repeat the step 2 above until all of students are grouped.

TABLE II. THE COMBINATION LIST OF GROUP DIVISIONS

schemes	Fitness	Initial state
sequential	Modularity	Modularity
parallel	Cluster	High-degree node
		High and low degree nodes

Each of the sequential and parallel methods need to decide a criteria to choose initial nodes and a measure to evaluate students who belongs to groups. In the rest of this section we give three ways of criteria to choose initial nodes and two evaluation measures. Hence we have twelve different methods for grouping (see Table II).

A. Three criteria to choose initial students

We give three different criteria to choose initial student. The first criterion is to choose high degree sociable students, who may be expected be leaders.

The second and third criteria choose two core students for a group. The second one choses two students that makes low modularity, that is, we choose pairs of students satisfying that when each of pairs makes a group it gives the minimum modularity. This criterion is motivated by the first motivator.

The third criterion is to choose high degree student and low student together as core students of a group. This criterion is motivated by the third motivator.

B. Fitness

Fitness measures to determine node which should belong to a group are given in this paragraph. The measures are defined between a node and a group, which represents how suitable it is as a member of that group. We give two measures based on modularity and cluster property.

For the first motivator it is necessary to perform a grouping in which friendships do not exist as much as possible in the inside of a group. For this purpose we use, the modularity, which measures the dense of friendship in groups and then low modularity is preferable. That is, we choose the lowest combination of s and C_i in the following equation.

$$M(G, P, C_i, s) = Q(G, P - \{C_i\} \cup \{C_i \cup \{i\}\})$$

For the second motivator we give priority to a pair students with many common friends but they themselves are not friend. In order to measure this requirement, we use the following value $C(G, s)$ for is a group G , i.e. a set of students and a student s .

$$C(G, s) = \sum_{t \in G} \left(\frac{N(s) \cap N(t)}{N(s) \cup N(t)} - A(s, t) \right),$$

where $N(s)$ is set of students who are friends of s , $A(s, t)$ is 1 if s and t are friends and 0 otherwise. In the equation the first part $(N(s) \cap N(t))/(N(s) \cup N(t))$ is called the Jaccard coefficient, ranging 0 to 1. It represents the similarity between two nodes in a network. We use the term $A(s, t)$ to prevent a student who has friends in the grope to give a high value.

V. APPLYING GROUPING METHODS TO FRIENDSHIP NETWORKS

We applied the twelve grouping methods to the friendship networks of classes. The networks are predicted by the procedure described in Section II-B.

A. Evaluating grouping by the three motivators

First, we give some measures, modularity, the numbers of common friends and the variance of degrees, to evaluate whether each requirement for a grouping can be satisfied.

1) *Modularity*: The modularity measures the dense of friendship in group. The first motivator requires low dense of friendship. Hence modularity may be a measure to evaluate the strength of the first motivator, that is, the lower the modularity of groups in a network the larger the first motivator moves students make friendship.

2) *Number of common friends*: In order to evaluate the strength of the second motivator we evaluate the number of common friend of two students in a group. Grouping C_1, C_2, \dots, C_L for a class is evaluated by the following equation,

the evaluation =

$$\text{AVR}_{i \in \{1, \dots, L\}} \text{AVR}_{j, k \in C_i} \# \text{common friends of } j \text{ and } k \text{ in } C_i,$$

where AVR means the average.

3) *Variance of degree*: In order to evaluate the strength of the third motivator we observe the degrees of nodes in their friendship network and evaluate the strength of the motivator by the variance of the degree of nodes in each group. The variance is high means that there is a diversity of social ability in groups. The idea of third motivator was in keeping the diversity. Grouping C_1, C_2, \dots, C_L for a class with a friendship network G is evaluated by the following equation,

$$\text{the evaluation} = \text{avr}_{i \in \{1, \dots, L\}} \text{var}_{k \in C_i} \text{the degree of } k \text{ in } G,$$

where var means the variance.

Table III is the results of three evaluation values of grouping done by the twelve grouping methods to a freshman class of NIT. The class consists of fifty seven students and the grouping making ten groups each of which consists of five or six students. In Table III the three evaluation values are normalized, which means that the distribution of the values are adjusted to average 0 and standard deviation 1. In order to see the balance of the three evaluation values the total of three evaluations are shown, as well. In the total the modularity values are added by multiplied by -1 because they have the negative effect.

In Table III we see the first method, seq.+mod.+mod., is the best in the point of modularity. From the point of common friends the forth one, seq.+high deg.+cluster, was the best. The variance of degrees evaluates the last one, paral.+high/low+cluster, as the best. The total points, however, evaluate the second method, seq.+mod.+cluster, and the last, paral.+high/low+cluster, had good best balance.

TABLE III. COMPARISON OF EVALUATION OF THE GROUP DIVISION

	methods (scheme + initialize + fitness)	modularity	common friends	var. of deg.	total score
1	seq.+mod.+mod.	-1.650	0.047	-2.546	-0.850
2	seq.+mod.+cluster	-0.355	1.548	-0.090	1.103
3	seq.+high deg.+mod.	-1.199	0.047	-0.529	0.716
4	seq.+high deg.+cluster	1.007	1.736	-0.127	0.602
5	seq.+high deg.+mod.	-0.046	-1.079	0.726	-0.307
6	seq.+high/low+cluster	2.209	0.235	0.714	-1.261
7	paral.+mod.+mod.	-0.597	-1.455	0.924	0.067
8	paral.+mod.+cluster	-0.296	0.235	0.423	0.954
9	paral.+high deg.+mod.	0.555	-0.892	-0.523	-1.970
10	paral.+high-deg.+cluster	-0.046	0.235	-0.653	-0.373
11	paral.+high/low+mod.	-0.046	-1.079	0.726	-0.307
12	paral.+high/low+cluster	-0.246	0.422	0.955	1.624

TABLE IV. THE RESULTED NEW FRIENDS AFTER THE GROUP WORK WITH GROUPS MADE BY THREE OF PROPOSING METHODS.

classes	grouping methods	newly acquired friends	
		in total	in groups
class A	#10. paral.+high-deg.+cluster	63	7
class B	#5. seq.+ high deg.+mod.	65	5
class C	#1. seq.+mod.+mod.	46	9
average of other classes	random grouping	45.2	4.2

B. Evaluating grouping by the friend making effect during group work

With cooperation of freshman students in a year of NIT we conveyed experiments to evaluate the grouping methods. We observed effect of increase and decrease of friendship in classes before and after a group work class during three months. The class let students take group work and many discussions with five or six peers in their groups. We had three classes, A, B, and C, each of which consists of fifty six or fifty seven students. Students in a class are divided into ten groups by the proposed grouping methods, in which friendship networks are prepared using the friendship prediction method with attendance records of students for the other classes that students took before the group work class.

We have twelve different methods, while we had only three chance to apply methods to classes. We chose the following three methods for evaluation. The three methods emphasize the three motivators respectively. For the first motivator Class C was applied the 1st (sequentially + modularity + modularity). For the second motivator Class A was applied the 10th method (parallel + high degree + cluster nature). For the third motivator Class B was applied the 5th (sequentially + high and low degree + modularity).

Table IV show the result of this experiment. It shows the total numbers of newly acquired friends during the three month class in the whole class and in groups. The number means the number of friendships that does not exist before class but does exist after it. This experiment was done for class of a year. But we have data of the change of friendship before and after the same class of other three years. In these years grouping was done by the random grouping. The table shows the average of newly acquired friendship of these years as reference.

Results in Table IV shows that the newly acquired friendships in the case using the proposed grouping methods were larger than the cases of random grouping. We may say that the proposing grouping methods gave chances to make friendship. Although in Class C the number of new friends in the whole

TABLE V. CHANGES OF BASIC VALUES OF FRIENDSHIP NETWORKS BEFORE AND AFTER THE GROUP WORK.

	change of values (before → after)		
	class A	class B	class C
#friends in total	182 → 150	161 → 155	125 → 136
#friends in groups	10 → 11	4 → 8	3 → 7
modularity	-0.050 → -0.036	-0.079 → -0.051	-0.098 → -0.055
distance average	3.929 → 3.543	3.605 → 3.963	4.237 → 2.874
cluster coefficient	0.605 → 0.435	0.595 → 0.571	0.608 → 0.473

class is the level of an average year, the number of new friends in groups was high, which shows that friendship generation in a groups was active.

Table V is a summary of basic values of the friendship networks of three classes before and after the group work. Although Table IV shows generation of new friendships, the total friendships decreases in two classes, while friendship in groups increases in every class. We may understand that it is difficult to control friendships although the group work has an effect of friendship. For the changes of modularity values we can observe that networks become to fit to the groups provided. This does not contradict the decrease of friendship. Increase of the modularity may be occurred by both factors of the lost of friendships across groups and the gain of new friendships inside.

In order to understand the changes of the distance averages and the cluster coefficient, we need observation of the network in Fig. 5 (a) to (f). In Fig. 5 networks (a), (b) and (c) are the networks predicted from the attendance records of students in their classes before the group work. We can observe rigid community structure in the networks. In the network of class A we see the largest almost complete network and two or three weaker communities. The network of class B has three very strong communities in different size and other smaller groups. In the network of class C strong communities are not formed compared to the other two, while we can observe some communities.

After the group work the networks were radically changed. In Fig. 5 networks (d), (e) and (f) are the networks after the group work. The strong community structures were weakened and many ties are generated between communities. The changes of distance averages and the cluster coefficients show the changes in values. Before the group work students are separated by the community structure and then two students in different communities has the long distance, while they become close by acquiring the chance that they take communication through the bridge ties after the group work. The change of cluster coefficient evidences the same effect. When there are large complete networks the cluster coefficient is high. By breaking the strong community structure the cluster coefficients become lower.

The two bar charts in Fig. 6 compare the distributions of the numbers of new friends and the friendship scores increased before and after group work between the year applied the proposing methods and other years. Each year has three classes and each class has ten groups. The sum of the values of bar chart for a year is thirty. In Fig. 6 (a) the four left most bars count groups that does not gain friendship, the second four bars count groups that gain a friendship, and so on. We can observe that the year when we applied the proposing method

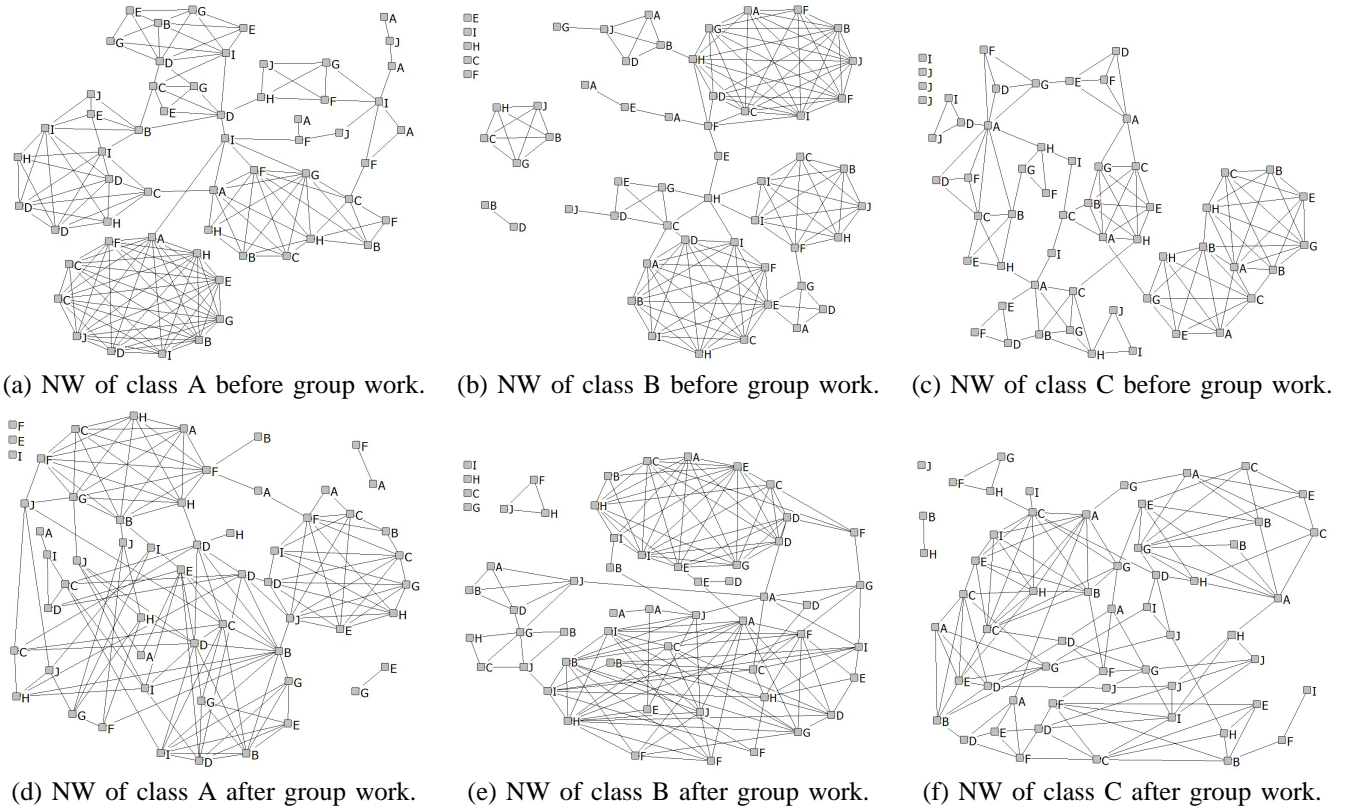


Fig. 5. Friendship networks of the classes A, B and C, before and after the group work. A letters along nodes means a group to which the node, a student, belongs at the group work.

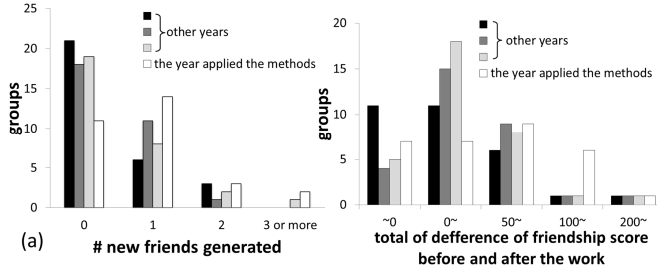


Fig. 6. Analysis of changes of friendships and friendship scores among groups applied and not applied the methods.

had slightly many groups that gain more friendship. The bar chart (b) in Fig. 6 counts groups that had some increase of friendship score. Groups are counted by ranges of below zero, 0 to 50, 50 to 100, 100 to 200, and more than 200. When we observe the chart we can see the effect of the proposing methods clearer.

VI. CONCLUSION

This paper investigated the effect of group work with the assumption of three motivators to make friends. Obeying the assumption we proposed twelve variations of methods for grouping students. Then the effects are evaluated by some measures from SNA and the changes of real friendship networks, which are observed by friendship prediction method.

We certain effect of the proposing methods, which brought new friendship among students to classes and made rearrange of community structure. For the effect of increase of friendship, the method has limited advantage. It generates new friendship while deactivates original friendship.

The dynamics of friendship is another interesting point. Our future work may include dynamics simulation for friendship according to the motivator and observation in the experiments.

REFERENCES

- [1] Baeza-Yates, R. A. and Ribeiro-Neto, B. A., "Modern Information Retrieval", ACM Press, Addison-Wesley, 1999.
- [2] Girvan, M. and Newman, M., "Community structure in social and biological networks", PNAS, 99(12), pp.7821-7826, 2002.
- [3] Newman, M. E. J. and Girvan, Y., "Finding and evaluating community structure in networks", Phys. Rev. E, 69, 026113, 2004.
- [4] Watts, D. J. and Strogatz, Y., "Collective dynamics of 'small-world' networks", Nature, vol.393, no.6684, pp.409-410, 1998.
- [5] Saito, Y., and Hiromitsu, H., eds, "Network Dynamics: Social Networks and Rational Choice", 2005.
- [6] Inuzuka, N., Takeuchi, S., and Matsushima, S., "Pattern Mining on Ego-Centric Networks of Friendship Networks", KES (4) 2011: 89-97.
- [7] Inuzuka, N., Kondo, T., and Yamamoto, S., "Analysis of Asymmetric Friendship among Students from Class Attendance Records", KES-IDT 2009.
- [8] Matsushima, H., and Inuzuka, N., "Evolution of Friendship Networks and Transition of Their Properties", KES-IDT 2012.

- [9] Inuzuka, N., Nakano, Y., and Shimomura, T., “Friendship Analysis Using Attendance Records to University Lecture Classes”, IASK International Conference on Teaching and Learning (TL 2008), pp. 478-486, 2008.