Research Article

Grouping Optimization Based on Social Relationships

Rong-Chang Chen

Department of Distribution Management, National Taichung University of Science and Technology, Taichung 404, Taiwan

Correspondence should be addressed to Rong-Chang Chen, rcchens@nutc.edu.tw

Grouping based on social relationships is a complex problem since the social relationships within a group usually form a complicated network. To solve the problem, a novel approach which uses a combined sociometry and genetic algorithm (CSGA) is presented. A new nonlinear relation model derived from the sociometry is established to measure the social relationships, which are then used as the basis in genetic algorithm (GA) program to optimize the grouping. To evaluate the effectiveness of the proposed approach, three real datasets collected from a famous college in Taiwan were utilized. Experimental results show that CSGA optimizes the grouping effectively and efficiently and students are very satisfied with the grouping results, feel the proposed approach interesting, and show a high repeat intention of using it. In addition, a paired sample *t*-test shows that the overall satisfaction on the proposed CSGA approach is significantly higher than the random method.

1. Introduction

Grouping optimization based on social relationships attracts more and more attention as social networks [1–11] have been growing rapidly in recent years. A social network is a special structure made of individuals (or organizations). It includes the ways in which individuals are connected through various social familiarities [1]. Through the analysis of relational network and the measurement of social relationships, grouping optimization is obtainable and can achieve a group objective.

In dealing with the grouping optimization problem concerning social relationships, some important issues should be addressed. Firstly, there is a deep discrepancy between the official and the secret behaviour of members [12]. A good approach to the grouping optimization should be able to find out the hidden information behind a group. Relations can be conceptualized at three levels of social complexity: individual, dyad, and group [13]. The approach should be able to disclose the hidden information in these levels. In addition,
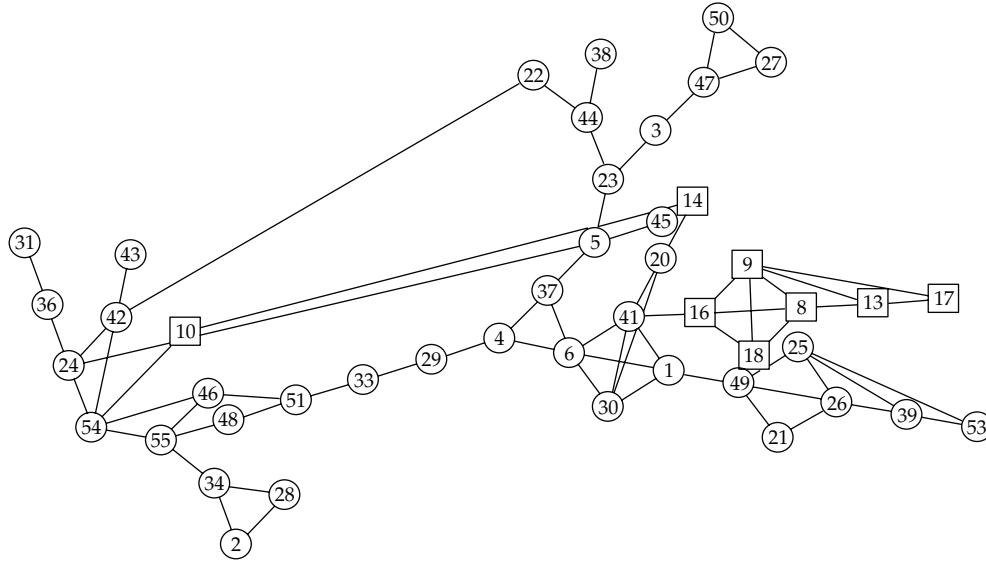
Figure 1: A social network structure.

the approach should be cost-effective. Consequently, a simple method which can explore indeed thinking of individuals is needed. Another issue is that most social networks such as Facebook employ a linear model to measure the links between group members. They use the number of connections that a node has to evaluate degree, betweenness, closeness, and network centralization [6]. However, a nonlinear relation model should be used since the real relationship tends to be nonlinear. For example, suppose that two nodes have the same number of connections. The connections for the first node are all its first choices, while for the second one the connections are its, for example, second and third choices, respectively. The disappointment in being grouped with a second choice partner over a first is unlikely to be linearly related to that of being grouped with a fourth choice over a third. A nonlinear model should be developed to suitably describe the relative importance of relationships. Thirdly, the approach should be able to help organizations improve their performances. Finally, but not the last, many grouping optimization problems subject to some constraints are known as NP-hard [14–16]. In addition, the structure of social relations usually forms a complex network, causing the problem to be very difficult to deal with. Another cause to make the grouping optimization more complicated is the constraints that require each group to include some different attributes, such as ability, specialty, gender, and position.

To tackle the above-mentioned complex problem, a new approach based on a combined sociometry [17–21] and genetic algorithm (GA) [22–28] is employed. For convenience, we call this novel approach CSGA in short. To measure the social relationships between members and to explore the hidden information behind a group, a choice-making method [29] is employed. Group members are first asked to indicate their choices (or preferences) to other members in an index system that a one stands for the first choice, two the second choice, and the like until an allowed maximum number of choices is reached. Based on the choices, a sociogram can be drawn by using a sociometry tool [17–21], as illustrated in Figure 1. Previous studies showed that sociometric choices do tend to predict such performance criteria as productivity, combat effectiveness, training ability, and leadership [30]. Moreover, some indices of social status can be calculated. In this paper, a new nonlinear model for measuring social status is presented. From the sociogram and social

status indices, the socially vulnerable or the potential bullying members can be found. Then the grouping optimization can be done according to the social status of individuals and the objective of the organization. Results from this study show that the proposed approach is effective in dealing with the grouping problem mentioned above. Furthermore, experiments founded on real data also show high satisfaction and repeat intention with this method.

The remainder of this paper is structured as follows. The grouping problem is briefly introduced in Section 2. In the subsequent section, the sociometry is briefly introduced and a new modified model for measuring relations between members is presented. Then proposed approach is presented in Section 4. Subsequently, results and discussion are presented. Finally, concluding remarks are drawn in Section 6.

2. The Grouping Problem

The grouping problem belongs to a large family of problems which partitions a set of items $\mathbf{U}$ into a collection of mutually disjoint subsets $\mathbf{U}_i$ of $\mathbf{U}$, such that $\cup \mathbf{U}_i = \mathbf{U}$ and $\mathbf{U}_i \cap \mathbf{U}_j = \emptyset$, $i \neq j$. In this study, the grouping is to optimally assign m members (individuals) to g subgroups with n members in each subgroup. Let $M = \{1, 2, \ldots, m\}$ be a set of members, $G = \{1, 2, \ldots, g\}$ a set of subgroups, $P = \{1, 2, \ldots, n-1\}$ a set of partners of a member, $A = \{1, 2, \ldots, a_l\}$ a set of attributes (or position), and $C = \{1, 2, \ldots, N_p\}$ a set of choices, where m is the number of members in a group, g is the total number of subgroups, n is the number of members in each subgroup, and a_l is the number of attribute l. The members can list a number of preferred members with whom they would like to work with or sit beside by an index system that a "1" stands for the first choice, "2" for the second choice, and the like, up to a preassigned maximum integer N_p. Note that a member cannot select himself or herself as his or her own partner. Slavin suggested that the best group size is from two to six [31]. For $i \in M$, $k \in P$, define c_{ik} as the preference given by member i to being grouped with his/her partner k and c_{ki} as the preference given by partner k to being grouped with member i. If partner k is the first choice of member i, c_{ik} is equal to one, the second choice, c_{ik} is equal to two, and the like. A lower coefficient of preference means that a member has more preference with his/her partner. If member i does not include his/her partner k in their list of preferences, then c_{ik} is assigned a relatively large penalty value b. The mathematical formulation, hence, can be expressed as follows:

$$\min \quad F = \sum_{i \in M} \sum_{j \in G} w_i R_i x_{ij}, \tag{2.1}$$

$$\text{subject to} \quad \sum_{j \in G} x_{ij} = 1 \quad \forall i \in M, \tag{2.2}$$

$$\sum_{i \in M} x_{ij} = n \quad \forall j \in G, \tag{2.3}$$

$$\sum_{i \in M} x_{ij} z_{il} = a_l \quad \forall j \in G, \ \forall l \in A, \tag{2.4}$$

$$z_{il} = \begin{cases} 1 & \text{if member } i \text{ has an attribute } l, \\ 0, & \text{otherwise,} \end{cases} \tag{2.5}$$

$$x_{ij} = \begin{cases} 1 & \text{if member } i \text{ is assigned to group } j, \\ 0, & \text{otherwise,} \end{cases} \tag{2.6}$$

$$\forall i \in M, \ k \in P \quad c_{ik} \text{ or } c_{ki} = b \quad \text{if } c_{ik} \text{ or } c_{ki} \notin C, \tag{2.7}$$

where $R_i = \sum_{k \in P} \{f(c_{ik}) + f(c_{ki})\}$, for all $i \in M$. The objective is to minimize the total scoring value, as shown in (2.1). Note that the scoring value is composed of two main parts: the priority weight w_i of member i and scoring function R_i. The priority weight can be assigned by two ways: directly assigned by the manager of the organization or calculated based on the social relationships. As for the scoring function $f(\cdot)$, a popular function used is the squared function [29]. Equation (2.2) requires that each member is assigned exactly one subgroup. Equation (2.3) ensures that one subgroup is composed of n members. Equation (2.4) requires that there are exact a_l members with attribute l in a subgroup. Note that a_l depends on attribute l. For example, if a basketball team is composed of a center (attribute 1), two forwards (attribute 2), and two guards (attribute 3), then $a_1 = 1$, $a_2 = 2$, and $a_3 = 2$, respectively.

As the number of members in a group increases, the number of possible grouping outcomes, u, grows at an unacceptably rapid rate; it grows exponentially. For example, for $n = 2$, $u = 945$ when $m = 10$, $u = 6.55E + 8$ when $m = 20$, and $u = 6.19E + 15$ when $m = 30$. The optimal solution cannot be found in polynomial time. In addition, the relationship structure becomes a complex network as m increases. To deal with the optimization problems concerning a complex network, it is impractical to use an approach like the exhaustive method since it takes a huge amount of computation time. Instead, some useful algorithms such as a genetic algorithm [22–28] for finding approximate or near-optimal solutions or strategies for improving efficiency, solution quality, exactitude, and more have been presented [32–35]. In this paper, a genetic algorithm [22–28], which is proven to be very effective in dealing with complex optimization problems, is employed to find feasible solutions.

3. The Sociometric Analysis

Sociometry was developed by Moreno in 1934. Moreno defined sociometry as "the inquiry into the evolution and organization of groups and the position of individuals within them" [17]. It is a quantified method aiming at measuring and determining social relationships in groups [18]. Sociometric explorations disclose the hidden structures that give a group its form: the alliances, the subgroups, the hidden beliefs, the ideological agreements, the stars of the show, and so on. One of Moreno's innovations in sociometry was the development of the sociogram, a systematic method for graphically representing individuals as points/nodes and the relationships between them as lines/arcs [19]. Given choices or preferences within a group, sociograms were developed and the structure and patterns of group interactions can be drawn on the basis of many different criteria [19]: social relations, channels of influence, lines of communication, and so on. Some terms in sociometry are the following:

(1) Stars: those who have many choices;

(2) Isolates: those with few or no choices;

(3) Mutual Choice (MC): individuals who choose each other;

(4) One-Way Choice: referring to individuals who choose someone but the choice is not reciprocated;

(5) Cliques: groups of three or more people within a larger group who all choose each other.

There are many indices using values instead of categories to indicate the social status of individuals. Among the most popular indices, three indices are introduced in the following paragraphs.

3.1. Status Score Index (SSI)

One of the most critical events in the history of sociometric methods is the use of both positive and negative nominations [13]. The difference between the positive and negative nominations is called status score [20]. To easily figure out the social status of a member, a relative or a normalized score called SSI was developed and defined as

$$SSI = \frac{N^{TC} - N^{TR}}{m - 1}, \tag{3.1}$$

where N^{TC} denotes the total choice number by other members, N^{TR} denotes the total rejection number by other members, and m is the total number of members in a group. Note that a member cannot choose himself or herself. Thus, $m - 1$ rather than m is used in (3.1). The value of SSI is in the range $[-1, 1]$. A higher value of SSI means that a member is more popular within a group. However, this index considers only one-way choice, and thus one cannot understand the mutual interaction between members from this index.

3.2. Index of Sociometric Status Score (ISSS)

To consider the mutual choices between members, an index called ISSS [21] is defined as

$$ISSS = \frac{1}{2} \left(\frac{N^{TC} - N^{TR}}{m - 1} + \frac{N^{MC} - N^{MR}}{N_p} \right). \tag{3.2}$$

On the right part of the equation, N^{MC} and N^{MR} represent the number of mutual choice and the number of mutual rejection, respectively. N_p is the allowed maximum number of choices that a member can make. The value of ISSS is also in the range $[-1, 1]$. A higher value of ISSS indicates that a member is more popular within a group.

3.3. Modified Index of Sociometric Status Score (MISSS)

ISSS uses a linear function to express relationship. However, it fails to consider the relative importance of different choices. For instance, satisfaction about being grouped with a first-, a second-, or a third-choice partner should be different, and the satisfaction difference between the first to the second and the second to the third is generally quite unlike. Consequently, a modified model for measuring the status score should be presented. In consideration with
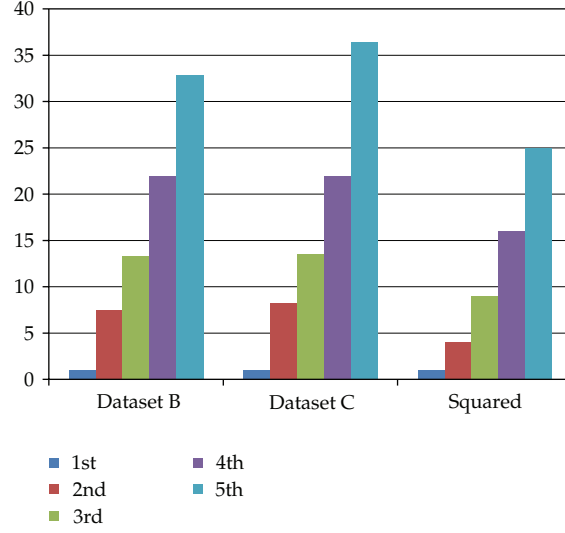
Figure 2: Characteristics of preferences among surveyed students.

this relative importance of different choices, a new index called MISSS is developed and defined as

$$
\mathrm{MISSS}_i = \frac{1}{2} \left(\frac{\sum_{p=1}^{m-1} z_{pi} w_p f\left(c_{pi}^{\mathrm{TC}}\right) - \sum_{q=1}^{m-1} z_{qi} w_q f\left(c_{qi}^{\mathrm{TR}}\right)}{m-1} \right.
$$
$$
\left. + \frac{\sum_{r=1}^{N_p} z_{ri} w_r f\left(c_{ri}^{\mathrm{MC}}\right) - \sum_{s=1}^{N_p} z_{si} w_s f\left(c_{si}^{\mathrm{MR}}\right)}{\sum_{t=1}^{N_p} w_t f(c_{ti})} \right). \tag{3.3}
$$

MISSS is composed of four parts: the contributions from the total choices, the total rejections, the mutual choices, and the mutual rejections, respectively. $z \in \{0,1\}$ and z_{pi}, z_{qi}, z_{ri}, and $z_{si} = 1$ if a member i is chosen by other members, rejected by other members, choose mutually with other members, and reject mutually with other members, respectively. w_p, w_q, w_r, and w_s are their priority weights, respectively. The value of MISSS is in the range $[-1, 1]$. A higher value of MISSS means that a member is more popular within a group. MISSS gives a more real description of relative importance of relationships. An important issue about MISSS is how to express the relations in the equation. A scoring function $f(\cdot)$, which is a function of choice (preference), is used to measure the relations. A simple method to find out the function is the use of a questionnaire asking members to provide disappointment scores of being grouped with a second-, third-, fourth-, or fifth-choice partner relative to receiving their first-choice one. To illustrate the nature of the scoring function, choices were collected from two classrooms in a college and the values were calculated. The results are shown in **Figure 2**. The scoring functions appear nonlinear rather than linear. As mentioned in **Section 1**, a linear scoring function is unreasonable. To measure the social relationships, a nonlinear model should be used.
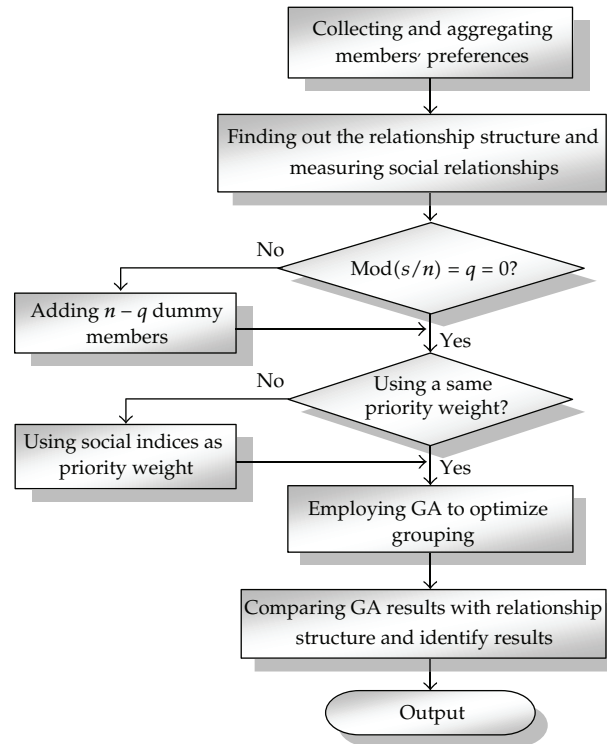
Figure 3: The proposed CSGA approach.

4. The CSGA Approach

4.1. The Procedure

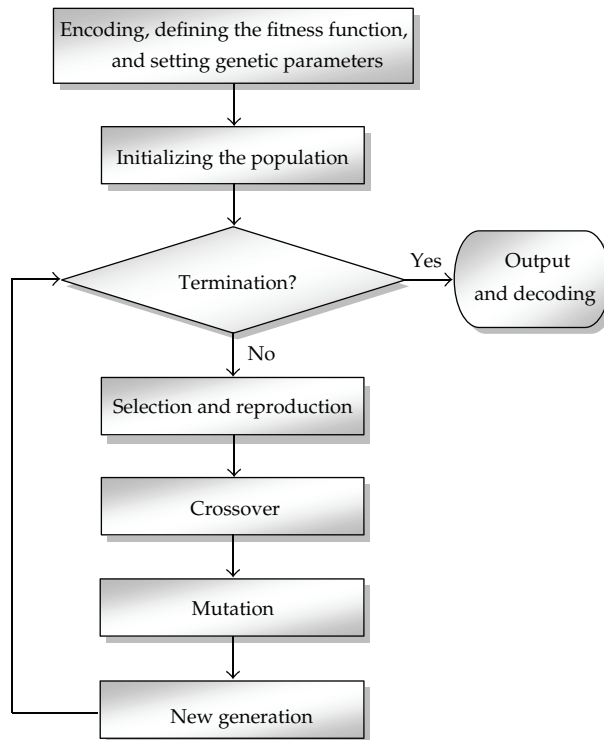The proposed CSGA approach is illustrated in Figure 3.

To begin with, the choices of members are collected and aggregated. A questionnaire is developed to collect the choices of the members and the reasons for choosing their partners. The questions are "whom in the group do you want to have a lunch or dinner with?" as a selection of lunch/dinner partner and "whom in the group do you want to sit beside and discuss with?" as a selection of learning partner. Table 1 illustrates a part of the questionnaire.

To be fair with all members, each member should fill in the same number of choices. Chen et al. [29] showed that a fewer number of choices filled in will have more chances to be assigned to their top choices. After aggregating the choices, a sociometric tool is employed to draw the sociogram and some social indices are measured.

If the remainder of $m/n = q \neq 0$, $n - q$ dummy members will be added and then grouping is performed using GA. Otherwise, GA is directly employed to group members. Adding dummy members does not influence the optimization since doing this only adds a constant value in the objective function. Note that the optimization can be done by different priority weights or the same priority weights. A higher value of the priority weight means that a member has a higher priority to be assigned his/her first choice. In the case of the same priority weight, all members have the same chance to be assigned to their top choices.

Table 1: An illustration of the questionnaire for collecting members' choices.

Priority	Index number of the member	Name	Index number of the main reason	Other reason (if available)
1	13	Don	8	
2	7	Mary	2	
3	2	Tom	17	good-looking
...	...	...	...	
N_p	1	Joe	12	

Figure 4: The flowchart of the GA program.

4.2. The GA Structure

The GA flowchart is illustrated in Figure 4. The details are depicted in the following paragraphs.

4.2.1. Encoding

The encoding of a chromosome is illustrated in Figure 5. Since there are m members, the number of genes is thus equal to m. Each gene is assigned a number which stands for a member. As illustrated in Figure 5, the values of the genes are in the order of $5, 2, 6, 10, \ldots$, and 4. If $n = 2$, that is, a subgroup has two members, then member 5 and member 2 are at the same subgroup, member 6 and member 10 are at the same subgroup, and the like.

5	2	6	10	3	7	11	23	...	4

Figure 5: Representation of a chromosome.

Table 2: A mutual choice table.

Member	Preference list	Mutual choice list
1	2, 7, 5, 4, 6	2, 6, 7
2	8, 3, 1, 7, 10	1, 7, 8, 10
3	2, 1, 9, m, 5	2, m
...	...	...
m	3, 6, 7, 8, 9	3, 6

4.2.2. Initialization of Population

The random method was employed to generate the initial solutions. Before producing the initial solutions, a mutual choice table was established. Table 2 illustrates the mutual choice. A member (say the 1st member) is first randomly selected, and the program checks with whom they choose mutually. Then from the mutual choice table, possible partners are 2, 6, and 7. Of these three possible partners, one is randomly selected to be a partner of the 1st member.

4.2.3. Evaluation of Fitness Function

To evaluate the chromosomes, the fitness value for each chromosome in the population was computed. The function used to measure the fitness is

$$\text{fitness} = \sum_{i \in S} \sum_{k \in P} w_i \left[f(c_{ik}) + f(c_{ki}) \right]. \tag{4.1}$$

A lower fitness value represents a better grouping, and the optimal fitness value is kept. After producing new chromosomes, we can evaluate the chromosomes based on the members' preferences. The fitness function contains w_i, the scoring function $f(\cdot)$, c_{ik} and c_{ki}, where w_i represents the priority weight of student i, c_{ik} represents the preference indicated by member i to partner k, and the like. If a member is grouped with a partner which is not in his/her preference list, the GA program will give c_{ik} a sufficiently large value b. As for the scoring function $f(\cdot)$, a common used function is the squared function [29]. For example, if a squared scoring function is employed, $w_i = 1$, and a subgroup is composed of three members whose partners are their 2nd and 3rd, 1st, and 3rd, and 2nd and 2nd choices, respectively, then the total score is $2^2 + 3^2 + 1^2 + 3^2 + 2^2 + 2^2 = 4 + 9 + 1 + 9 + 4 + 4 = 31$. After computing the fitness of a subgroup j, the fitness of the chromosome is calculated as
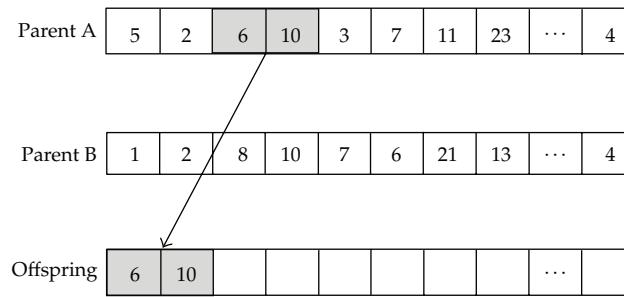
$$F = \sum_{j \in G} f_j. \tag{4.2}$$

| Parent A | 5 | 2 | 6 | 10 | 3 | 7 | 11 | 23 | ⋯ | 4 |

| Parent B | 1 | 2 | 8 | 10 | 7 | 6 | 21 | 13 | ⋯ | 4 |

| Offspring | 6 | 10 | | | | | | | ⋯ | |

Figure 6: Crossover method.

| Before | 5 | 2 | 6 | 10 | 3 | 7 | 11 | 23 | ⋯ | 4 |

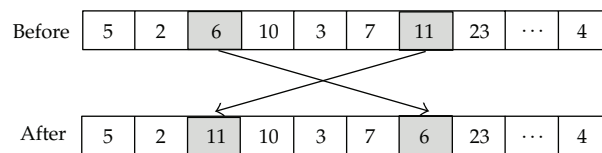| After | 5 | 2 | 11 | 10 | 3 | 7 | 6 | 23 | ⋯ | 4 |

Figure 7: Mutation operation was performed by the swap method.

4.2.4. Selection

The Roulette Wheel Method [25] was employed to select fitter individuals. The fitness values of all the chromosomes in the population were first calculated and were then sorted. A fitter chromosome with a lower value of fitness has a higher probability to be selected. For example, if there are 50 chromosomes in the population. The chromosome with the lowest fitness value has a probability of $50/(1 + 2 + 3 + \cdots + 50)$ to be selected, the second lowest has a probability of $49/(1 + 2 + 3 + \cdots + 50)$ to be selected, and the like.

4.2.5. Crossover

Two chromosomes were randomly selected first. The fitness values of these two chromosomes were compared and from the better one some genes were randomly selected and placed at the beginning positions of the offspring chromosomes, as illustrated in Figure 6. The number of the selected genes is the same as the number of members per subgroup, n. The rest of the genes in the offspring chromosome are filled in the order of fitter genes by the greedy method.

4.2.6. Mutation

In this study, the swap method was employed to mutate. The method is illustrated in Figure 7. Two genes were randomly selected and then their values were interchanged. The swap method has the advantage of avoiding value duplication. Since a member is exactly assigned one subgroup, the values in the genes should be different.

Table 3: The datasets tested.

Dataset	Students	m	N_p	Grouping for	Grouping method	
					1st	2nd
A	Senior, in their 4th year	54	9	Dining partner	CSGA	CSGA
B	Junior, in their 3rd year	55	5	Learning partner	CSGA	CSGA
C	Freshman, in their 1st year	32	5	Learning partner	Random	CSGA

4.2.7. Elitism Strategy

In order to preserve the best chromosome in every generation, a simple elitism strategy [23] was employed. The best chromosome of each generation was duplicated to the next generation to ensure that it was preserved in the present population if it was the best when compared with other chromosomes of the population. This strategy assures that the best chromosome of each generation will be, even if not better, at least equal to the best chromosome of the previous generation. In addition, the elitism strategy did not lead the GA to converge prematurely since it was applied only to one chromosome of each generation.

4.2.8. Termination Condition

The termination condition we use in this study is the generation number defined by the user. The calculation will be repeated until the number of generation reaches the preassigned value. Once the termination criterion is satisfied, the solution is displayed with a chart. We can clearly know the result from the chart.

5. Results and Discussion

To numerically investigate the influences of grouping parameters such as the number of member per subgroup, the choice number, and the number of member in a group on the grouping results, a GA program was developed by Microsoft.NET 4.0. The program was run on an ASUS K52Jc Series notebook with Intel(R) Core (TM) i5 CPU M460 @2.53 GHz and 1.86 GB RAM. The operating system is Windows 7.

To evaluate the effectiveness of the proposed approach, three real datasets including three classes of undergraduate students were tested. The numbers of students are 54, 55, and 32, respectively, as shown in Table 3. For the dataset C, the method of grouping at the first time was by random and the second is by CSGA. A comparison can thus be made between these two grouping methods. For the dataset A, a maximum allowed number of choices, N_p, was set to be 9 so that the complexity of relationship network can be observed. Students were asked to show their preferences to the dining partners in dataset A and learning partners in datasets B and C.

For easy description, the following variables are defined. The population size is represented as N_{size}, the generation number as N_g, the crossover rate as R_c, and the mutation rate as R_m. In addition, we use F_{min} as the minimal fitness value in each generation, F_{best} as the best solution at each trial. In each case, the GA program performs 30 trials. We designate the best result of 30 trials as OPT, their average value as AVG, and the coefficient of variation as C_v.

Table 4: Comparison of results between Branch and Bound and GA.

m	n	Branch and Bound		GA			
		Fitness value	Runtime (min)	OPT	AVG	C_v	Average runtime, (min)
6	2	39	<0.01	39	39	0	1.29
8	2	67	0.09	67	67	0	1.31
10	2	144	8.25	144	144	0	1.52
12	2	176	>1099.71	176	176	0	1.64
14	2	NA	*	351	351	0	1.81
20	2	NA	*	323	358.73	0.15	2.43
40	2	NA	*	1360	1474.43	0.05	4.39
60	2	NA	*	2466	2718.1	0.04	6.29
6	3	85	<0.01	85	85	0	1.41
9	3	318	1.23	318	318	0	1.86
12	3	NA	*	664	664	0	2.23
15	3	NA	*	1145	1145	0	2.69
8	4	581	0.17	581	581	0	2.06
12	4	NA	*	1234	1234	0	2.69
15	5	NA	*	3016	3016	0	4.05

*a huge amount of computation time is required if compared with GA.

5.1. Validation of GA

To ensure the validation of the GA program, some tests were made. The results from GA were compared with those from the Branch and Bound (BB) method that can obtain optimal solutions. The results are shown in Table 4. When the number of members m is small, the GA program can obtain optimal solutions that are the same as those from BB, indicating that the GA program is valid. As m increases, the BB method requires a huge number of computation time. For example, for $m = 12$ and $n = 2$, the required time is about 1100 minutes or 18.3 hours. However, the average runtime for GA is about 1.64 minutes, much smaller than that for BB shown in Table 4.

5.2. The Influences of Genetic Parameters

To evaluate the influences of genetic parameters on the results, a two-stage approach was used. The generation number and population size were decided at the first stage, while at the second stage crossover rate and mutation rate were decided. The variation of minimal fitness value at each generation with different population sizes, N_{size}, is shown in Figure 8. As we can see from the figure $N_{size} = 20$, $N_g = 30000$ can obtain good results. At the second stage, the mutation rate as well as the crossover rate were changed and tested. Experimental results show that the best parameters are $R_c = 0.9$ and $R_m = 0.05$. In the following experiments, therefore, $R_c = 0.9$ and $R_m = 0.05$ were used.

5.3. The Influences of Choice Number

To easily observe the solutions, the sociograms (relationship structure) are drawn and shown in Figure 9. Note that only mutual-choice relationships are shown in the figure. A female
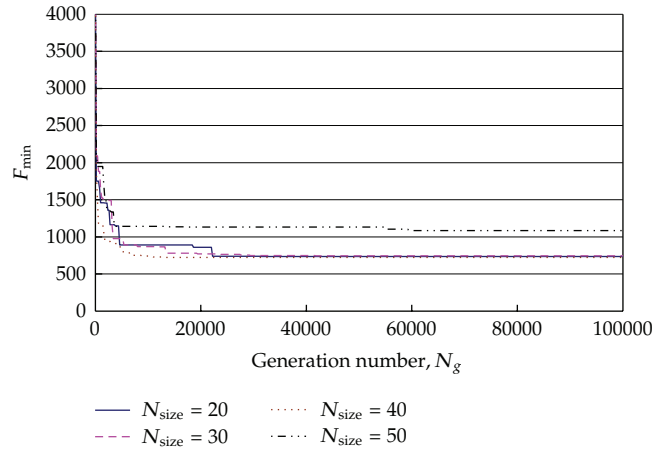
Figure 8: The variation of minimal fitness value in each generation with different population size, N_{size}.

Table 5: The influences of the number of members in a subgroup ($N_p = 5$, $N_g = 10,000$).

n	OPT	AVG	C_v	Average runtime (min)
2	801	952.57	0.10	2.09
3	2683	3170.63	0.07	2.98
4	5712	6131.40	0.04	3.97
5	9246	9548.60	0.03	5.05

member is represented with a circle, while a male member is represented with a square. From this figure we can see that the relationship network is quite complicated, especially for those with a large value of N_p. An optimal grouping is hard to be achieved by intuition or exact search methods.

5.4. The Influences of the Number of Members in a Subgroup

As the number of members in a subgroup n increases, the computation time will also increase. To observe the influences of n on the results, n was changed from two to five. Note that in Table 5 the number of choice N_p was fixed at 5, whereas in Table 6 the number of choice N_p was set to be equal to n.

5.5. The Influences of the Number of Members m in an Organization

As the number of members in a group m increases, the complexity of grouping also increases. An apparent influence is the increase of the computation time, as illustrated in Table 7. Note that the computation time increases linearly with m, showing the time efficiency of the GA program. Originally, the computation time grows exponentially. Consequently, GA is an efficient approach to solve grouping optimization problems considering social relationships.
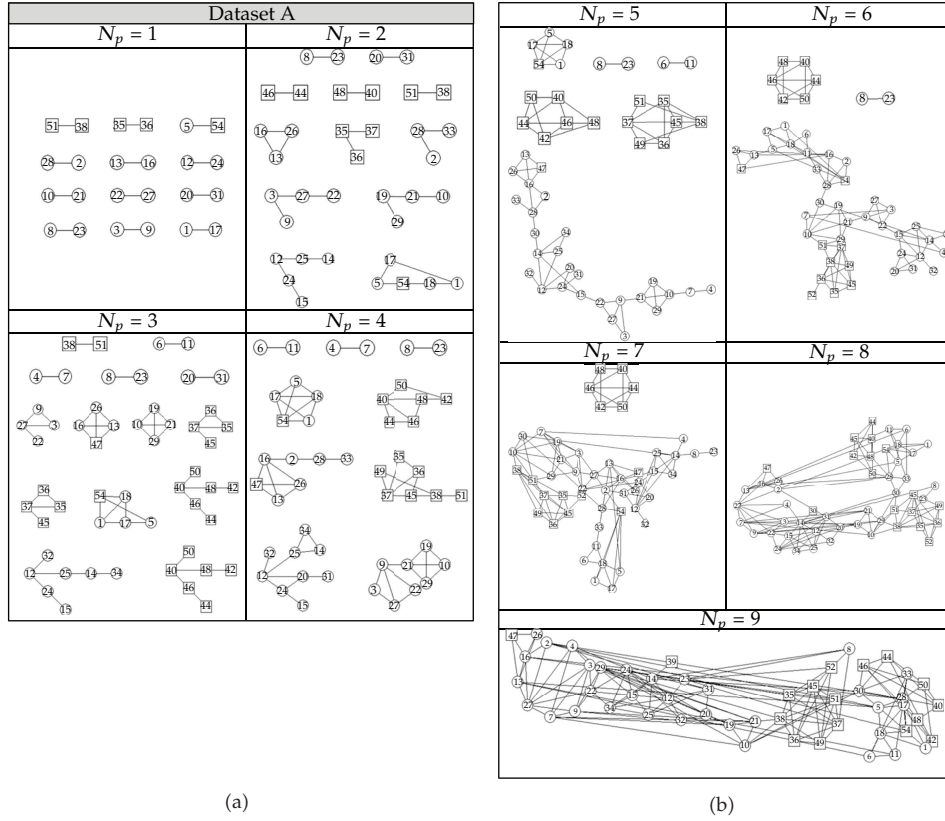
(a)

(b)

Figure 9: The sociogram of dataset A.

Table 6: The influences of the number of members in a subgroup ($N_p = n$, $N_g = 10,000$).

n	OPT	AVG	C_v	Average runtime (min)
2	1461	1532.90	0.05	2.14
3	3931	4223.93	0.04	3.01
4	6523	6583.10	0.05	4.05
5	9246	9548.60	0.03	5.05

Table 7: The influences of the number of members ($n = 2$).

m	C_v	Average runtime (min)
10	0	0.51
20	0.13	0.88
30	0.04	1.25
40	0.03	1.63
60	0.04	2.42

5.6. Homogeneous versus Heterogeneous Grouping

A good grouping approach should be able to meet the requirements of a group, such as a homogeneous or heterogeneous grouping. To investigate the effectiveness of the proposed

Table 8: The Influences of scoring function.

m	Scoring function	b	1	2	3	4	5	out	Average runtime (s)
54	$\{1^2, 2^2, 3^2, 4^2, 5^2, 100\}$	0	27.1	22.2	17.4	9.9	8.6	22.8	9.6
		50	28.4	22.2	16.5	10.0	8.2	22.7	22.4
		100	28.3	22.4	16.5	9.6	8.6	22.5	22.3
		200	28.1	21.5	16.9	10.1	8.7	22.7	23.7

Table 9: Overall satisfaction and repeat intention of the CSGA approach.

	Dataset B	Dataset C
Overall satisfaction	4.39	4.28
Repeat intention	4.52	4.09

approach, experiments were performed with different priority weights, which were based on MISSS. The result is illustrated in Figure 10. Members grouped into the same subgroup are circled. A solid line between two members means that members choose mutually. A dashed line with an arrow, on the other hand, represents a one-way choice, where the arrow-headed direction is the choice direction. The grouping can be done with different priority weights based on MISSS. For homogeneous grouping, MISSS can be used as a priority weight. For heterogeneous grouping, MISSS can be sorted and then divided into several parts, with the first part composed of the highest MISSS values, the second part composed of the second highest MISSS values, and the like. Using the proposed CSGA approach, homogeneous or heterogeneous grouping can be easily performed in fulfilling the requirements of a group.

5.7. The Influences of Scoring Function and Penalty Value

The penalty value b was changed to investigate its influences on the grouping results. The results are shown in Table 8. From the table we see that the effect of penalty value b is not very apparent. As for the effect of the scoring function, a squared function converges sooner than a linear function, as illustrated in Figure 11.

5.8. Overall Satisfaction and Repeat Intention

A survey was conducted to understand the overall satisfaction and repeat intention of using the proposed approach after the learning partner groups were decided. A typical five-level Likert item was employed. The question is "In overall, are you satisfied with the method?" and the format of the item is "very satisfied, satisfied, neither satisfied nor dissatisfied, dissatisfied and very dissatisfied." Most of the students are satisfied or very satisfied with the new method. The average points are 4.39 and 4.28, respectively, in dataset B and dataset C, indicating that the students have high satisfaction with the proposed method (see Table 9). As for repeat intention, the question is "Do you agree to use the new method to select learning partners next time?" and anchored by "strongly agree," "agree," "neither agree nor disagree," "disagree," and "strongly disagree." The average points of this question are 4.52 and 4.09, respectively. The high values indicate that the repeat intention of using the proposed method is also very high. The high overall satisfaction and the high repeat intention, in part, reflect the effectiveness of the present approach.
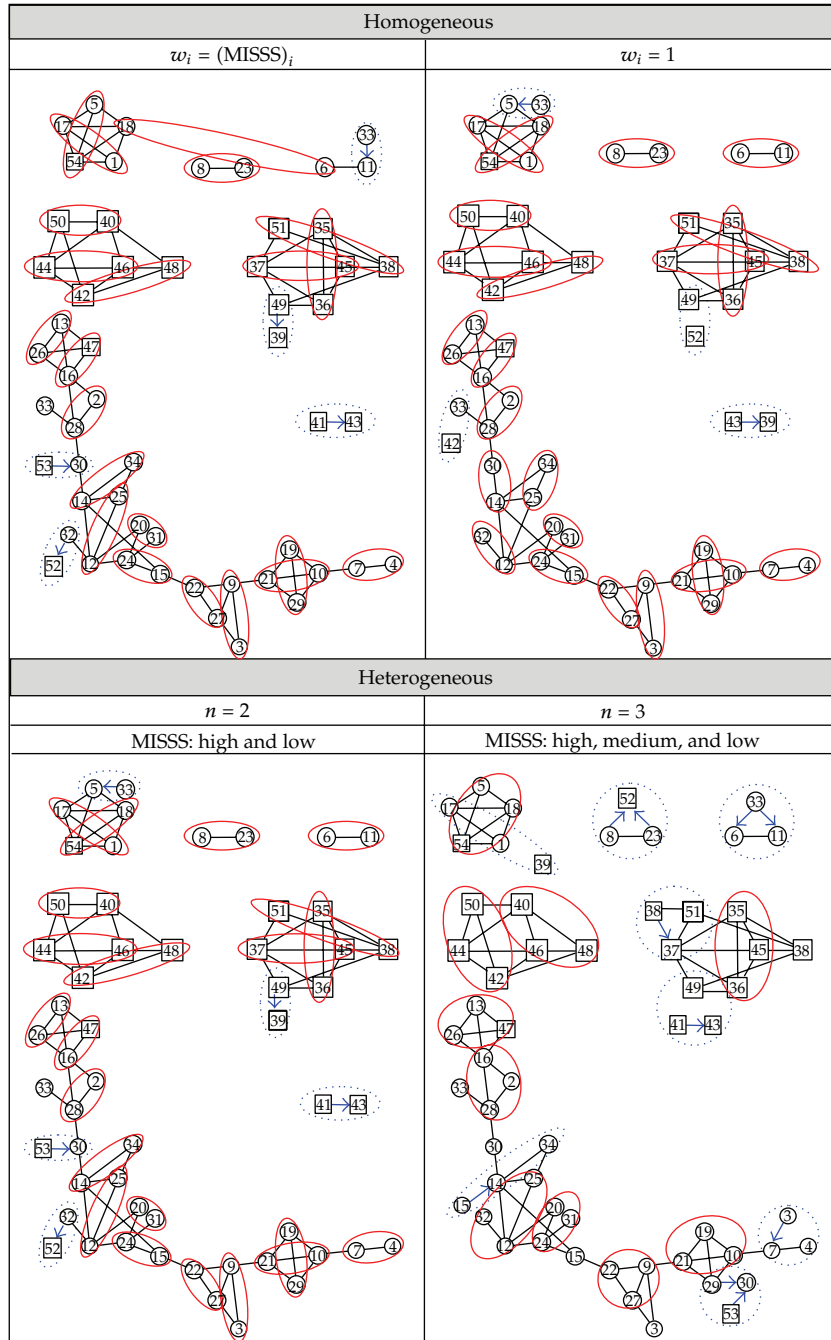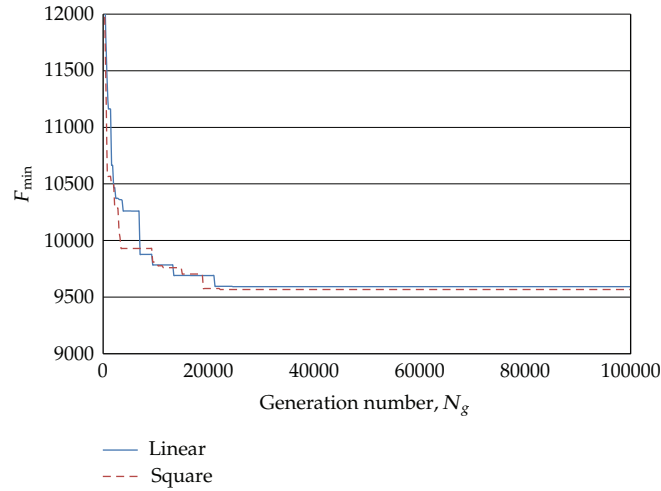
Figure 10: The grouping results are illustrated in a sociogram.

Table 10: Paired samples statistics.

Method	Mean	m	Std. deviation	Std. error mean
Random	3.50	32	.71842	.12700
CSGA	4.28	32	.68318	.12077

Figure 11: The variation of $F_{\min}$ with generation at different scoring functions.

Table 11: Paired sample t-test.

		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference	
					Lower	Upper
Pair 1	Random—CSGA	−.78125	.87009	.15381	−1.09495	−.46755

Table 12: Significance of paired sample t-test.

		t	df	Sig. (2-tailed)
Pair 1	Random—CSGA	−5.079	31	.000

A paired sample t-test was performed on dataset C to see if there is no difference between these two methods.

Hypothesis 1. There is no difference in satisfaction between using the CSGA approach and using the random method.

The result shows that the overall satisfaction of using the proposed CSGA approach is significantly higher than using the random method, as shown in Tables 10–12.

6. Conclusions

In this paper, we have employed a novel approach to group members based on their social relationships. Members are first asked to show their preferences (choices) to partners with whom they would like to be together in an index system, where a one stands for the first choice, two for the second choice, and so on up to a preassigned number. Subsequently, the choices of members are aggregated and a combined sociometry and genetic algorithm (CSGA) scheme is employed to optimize the grouping. To investigate the effectiveness of the

proposed approach, three real datasets collected from a university were used. In addition, a nonlinear model is developed to calculate the social scores of individuals. Experimental results show that CSGA can optimize the grouping effectively and efficiently and students are very satisfied with the grouping results and show high repeat intention of using it. Moreover, a paired sample *t*-test shows that the overall satisfaction on using the proposed CSGA approach is significantly higher than using the random method.

For further studies, it is recommended to apply the combined sociometry and genetic algorithm approach to solve other kinds of grouping optimization problems. Methods of rating or paired comparisons as well as peer nominations can be used to measure social relationships.

Acknowledgments

References

[1] E. Turban, D. King, J. Mckey, J. Lee, and D. Viehland, *Electronic Commerce: A Managerial Perspective*, Prentice Hall, New York, NY, USA, 2010.

[2] P. J. Lamberson, "Social learning in social networks," *The B.E. Journal of Theoretical Economics*, vol. 10, no. 1, article 36, 2010.

[3] D. Centola, "The spread of behavior in an online social network experiment," *Science*, vol. 329, no. 5996, pp. 1194–1197, 2010.

[4] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.

[5] R. S. Burt, *Structural Holes: The Social Structure of Competition*, Harvard University Press, Cambridge, Mass, USA, 1992.

[6] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Harvard University Press, Cambridge, Mass, USA, 1994.

[7] B. Wellman, *Networks in the Global Village: Life in Contemporary Communities*, Westview Press, Boulder, Colo, USA, 1999.

[8] L. C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science*, Empirical Press, Vancouver, Canada, 2004.

[9] N. C. Mullins, *Theories and Theory Groups in Contemporary American Sociology*, Harper and Row, New York, NY, USA, 1973.

[10] S. D. Berkowitz, *An Introduction to Structural Analysis: The Network Approach to Social Research*, Butterworth, Toronto, Canada, 1982.

[11] P. J. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*, Cambridge University Press, New York, NY, USA, 2005.

[12] J. L. Moreno, *The Sociometry Reader*, The Free Press, Glencoe, Ill, USA, 1960.

[13] A. H. N. Cillessen, "Sociometric methods," in *Handbook of Peer Interactions, Relationships, and Groups*, p. 83, 2011.

[14] T. Lappas, K. Liu, and E. Terzi, "Finding a team of experts in social networks," in *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining (KDD '09)*, pp. 467–476, July 2009.

[15] E. Falkenauer, *Genetic Algorithms and Grouping Problems*, John Wiley & Sons, New York, NY, USA, 1998.

[16] M. Gallego, M. Laguna, R. Martí, and A. Duarte, "Adaptive memory programming for the maximally diverse grouping problem," in *Proceedings of the 9th Metaheuristic International Conference*, 2011.

[17] J. L. Moreno, *Who shall survive? A New Approach to Problem of Human Interrelations*, Nervous and Mental Disease Publishing, Washington, DC, USA, 1934.

[18] A. H. N. Cillessen, "Chapter 5-Sociometric Methods," in *Handbook of Peer Interactions, Relationships, and Groups*, K. H. Rubin, W. M. Bukowski, and B. Laursen, Eds., 2011.

[19] H. H. Jennings, *Sociometry in Group Relations*, Greenwood, Westport, Conn, USA, 2nd edition, 1987.

[20] J. D. Coie, K. A. Dodge, and H. Coppotelli, "Dimensions and types of social status: a cross-age perspective," *Developmental Psychology*, vol. 18, no. 4, pp. 557–570, 1982.

[21] C. H. Ho, *A study of the effects on emotional experience, peer relationship and teacher-student relationship of the therapeutic approach to art activities in class in an elementary school setting*, M.S. thesis, National Pingtung University of Education, Taiwan, 2002.

[22] J. H. Holland, *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, UK, 1975.

[23] D. E. Goldberg, *Genetic Algorithm in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass, USA, 1989.

[24] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, UK, 1996.

[25] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Design*, John Wiley & Sons, New York, NY, USA, 1997.

[26] M. Gen and R. Cheng, *Genetic Algorithms and Engineering Optimization*, John Wiley & Sons, New York, NY, USA, 2000.

[27] D. A. Coley, *An Introduction to Genetic Algorithms for Scientists and Engineers*, World Scientific, Singapore, 1999.

[28] M. Gen, R. Cheng, and L. Lin, *Network Models and Optimization: Multiobjective Genetic Algorithm Approach*, Springer, New York, NY, USA, 2008.

[29] R.-C. Chen, M.-R. Huang, R.-G. Chung, and C.-J. Hsu, "Allocation of short-term jobs to unemployed citizens amid the global economic downturn using genetic algorithm," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7537–7543, 2011.

[30] V. K. Maheshwari, "Introduction to sociometry," http://www.vkmaheshwari.com/WP/?p=50, 2011.

[31] R. E. Slavin, *Cooperative Learning: Theory, Research, and Practice*, Allyn & Bacon, Boston, Mass, USA, 2th edition, 1995.

[32] S. Chen, W. Huang, and C. Cattani, "Traffic dynamics on complex networks: a survey," *Mathematical Problems in Engineering*, vol. 2012, Article ID 732698, 23 pages, 2012.

[33] S. Chen, H. Tong, and C. Cattani, "Markov models for image labeling," *Mathematical Problems in Engineering*, vol. 2012, Article ID 814356, 18 pages, 2012.

[34] S. Chen, H. Tong, Z. Wang, S. Liu, M. Li, and B. Zhang, "Improved generalized belief propagation for vision processing," *Mathematical Problems in Engineering*, vol. 2011, Article ID 416963, 12 pages, 2011.

[35] S. Chen, Y. Wang, and C. Cattani, "Key issues in modeling of complex 3D structures from video sequences," *Mathematical Problems in Engineering*, vol. 2011, Article ID 856523, 12 pages, 2011.