Expanding Students' Social Networks via Optimized Team Assignments

Alessandro Hill[1*] and Steffen Peuker[2]

[1*]Department of Industrial and Manufacturing Engineering, Organization, 1 Grand Av, San Luis Obispo, 93407, CA, USA.
[2]Department of Mechanical Engineering, Organization, 1 Grand Av, San Luis Obispo, 93407, CA, USA.

*Corresponding author(s). E-mail(s): ahill29@calpoly.edu;
Contributing authors: speuker@calpoly.edu;

Abstract

The class social network is a momentous factor when it comes to educational, personal and professional student success as well as achieving course learning outcomes. Students and teachers benefit from expanded network connectivity via augmented engagement, more inclusivity, and efficient diffusion of information. We present a novel method for positively influencing the class social network. We develop an in-class grouping strategy based on optimization and sociocentric network analysis that pragmatically expands the students' social networks. In contrast to existing routines, our technique focuses on maximizing individual student opportunities to establish new ties. Based on the knowledge of existing connections, our procedure systematically optimizes the overall number of new ties that can be established during a team project. Our data-driven approach is designed for practical use in class. We show that the underlying combinatorial problem of maximizing unrelated intra-team students can be modeled as a bin packing variant. Using an integer programming formulation, we demonstrate the efficient spreadsheet implementation. We discuss model extensions to account for high-density networks, team balancing, and teammate forcing and forbidding, allowing for hybridization using existing grouping techniques. In an empirical study, we provide evidence for the efficacy of our approach using data from 10 industrial engineering classes with 253 students and 77 project teams - in both face-to-face and virtual modes. We demonstrate the impact of our grouping method compared to random-assignment,

self-selection, and maximizing existing intra-team ties. We report an impressive 62% increase of ties compared to only 17% when self-assigning.

Keywords: Education, class social networks, student collaboration, dynamic social network analysis, mathematical optimization

1 Introduction

Student success is one of the principal goals in higher education. Beyond degree completion and achieving learning outcomes this also includes the holistic development of students regarding their social and personal development. Establishing and maintaining a healthy and diverse peer network is an important factor to accomplish the latter. Furthermore, effective learning, social, and general well-being can be seen as additional positive implications from having a healthy and diverse peer network.

Working in teams is a crucial pillar when it comes to the creation of strong networks. Moreover, it is well understood that teamwork is a vital component in active learning and effectively equips students with collaboration skills which are so important on the job [1]. Teamwork has notable potential to support the creation of an inclusive environment and to increase the corresponding student awareness of diversity at universities [2]. However, only little work addresses the desire to pragmatically extend the number of different collaborators for each student in order to amplify diverse and inclusive collaboration. This applies to both course modes—virtual and face-to-face. The Accreditation Board for Engineering and Technology (ABET) requires under Criterion 3 that student outcomes include "an ability to function effectively on a team..." and therefore all accredited engineering programs have to include teamwork into their curriculum.

Teams are typically built by either letting students self-select their team members or through the instructor [3]. Teacher-assigned student teams are either randomly grouped, or a selection mechanism including well-studied criteria such as the students' educational performance, demographic characteristics, gender etc. is applied [4].

The interdisciplinary work presented in this paper integrates knowledge on class social network structures into the process of assigning students to project teams. The developed approach, by contrast, directly fosters the generation of new connections between students. To this end, students are preferably combined with students that do not know each other well. As a result, individuals benefit from being on a diverse team, improving their ad-hoc collaboration skills, and expanding their networks. We suggest a multi-step method that captures existing ties in a class first. This initial social network is modelled assuming two nodes–representing two students–are connected if the students know each other. In a second step, a mathematical optimization model is used to group students with the goal of maximizing the potential for new ties. The

proposed intervention aims at maximizing ties for one-mode sociocentric social networks in group settings. Finally, we carefully compare a snapshot of the post-project social network to the pre-project social network.

The approach can be combined with classical grouping preferences or requirements, and, to a certain extent, it includes a randomized component. Furthermore, it can be used in non-educational environments when forming, for example, sports teams, committees, task forces, or room mates. A comprehensive team assignment procedure spanning the collection of social network data, the solution of the mathematical problem, and the visual presentation of all steps in order to engage the students is developed and presented. We describe how to mathematically formulate the underlying combinatorial problem, and employ integer programming to obtain optimal solutions. Furthermore, for the sake of instructor friendliness, we show how the techniques can be carried out solely using spreadsheets and freely available mathematical programming solver plugins. Our procedure can partially be automated, while still being very flexible with respect to instructor, student, and project characteristics.

We report concepts and findings from a two-year research project including the practical implementation and analysis of the new method. To evaluate our new method we performed a case study using real-life data from 10 college-level courses. We explore both, the potential theoretical network growth and the actual network evolution. In contrast to existing team formation studies, we carefully analyze the impact of multiple grouping strategies in dynamic real-world social networks. Besides clique and independent subgroups, we quantify network density, components and diameter. Additionally, we simulate alternative assignment methods (self-assigned, instructor-assigned) using the collected data. Corresponding student surveys prove that the method is well-received in class.

The research hypothesis is that creating project teams while avoiding pre-existing intra-team connections between students significantly increases the density of the in-class social network. Our main contributions can be summarized as follows.

- We present a novel model for assigning students to teams based on social network analysis, and describe extensions and hybridization techniques.
- We elaborate an effective mathematical optimization framework that can be applied using spreadsheets.
- We conducted an empirical study and present the impact of our approach, using data from more than 250 students in 10 industrial engineering classes over the course of one year, in both face-to-face and virtual settings.
- We analyze and compare related assignment methods such as random, self, and clique-oriented to the optimized assignment method.

An overview of related work in the scientific literature is given in Section 2 The used concepts from social network analysis and our data collection method are described in Section 3. We explain the optimization approach including the mathematical formulation and possible model extensions in Section 4.

The empirical study and its results are described in Section 5, followed by a summary of our findings in Section 6.

2 Related Literature

In this section, we provide an overview of relevant work that overlaps with our interdisciplinary research topic. We reference literature on optimization-based approaches, single-team formation problems, educational grouping aspects, classroom social networks, and classroom management.

The rich applications of operations research in education have already been outlined as early as in the 60s by [5]. Techniques have been suggested to assign students to teams. [6] suggests a simple assignment scheme that takes into account student preferences and aims at optimizing the worst case preference fulfillment. The problem of creating groups of maximal diversity has been tackled by several authors. Approximate grouping methods, also called heuristics, have been applied to efficiently form teams (e.g., [7], [8], [8], [9]). [10] suggests a heuristic grouping method based on Tabu Search with a special focus on allowing additional side requirements by the instructor. The approach was tested in 4 graduate-level design project classes. An actual software tool to heuristically assign student to teams is developed by [11]. [12] focus on optimized student groups with the objective to have high heterogeneity. The authors study a model with an arbitrary set of generic student characteristics and apply two nature-inspired heuristics to solve the problem. Similarly, [13] study the problem of forming online learning communities with high student compatibility. Diverse, recurring MBA student grouping using local search techniques is suggested by [14]. Note that in contrast to our exact approach, no guarantee is provided for the quality of the team formation with respect to the predefined goals. Exact methods for a variant of the maximal diversity assignment problem are considered by [15] and [16]. [17] optimize student-to-team assignments under temporal student availability constraints using Constraint Programming. In an iterative optimization-based procedure, students are sequentially assigned to multiple class projects in [18]. The authors simultaneously consider two objectives: Combining students with similar performance history, and exposing students as many class mates as possible. In contrast to our approach, the model does not take into account the underlying social network by allowing collaboration between students that already know each other.

The problem of forming one single strong team online in a social network has been studied with various objectives. Team member compatibility and skill composition are relevant factors [19]. Two potential team members are typically presumed to be compatible, when they are capable of communicating efficiently based on their existing relationship. The identification of a team with a suitable leader has been considered by [20]. Multiple optimized teams with respect to sociometric factors and skill requirements are sought by [21] using exact and heuristic methods. In our work, we provide a comprehensive analysis based on real world data and analyze the potential of multiple

grouping methods. A related approach is described by [22] who heuristically partition college classes into teams based on student preferences to work with class mates. They report the positive feedback obtained by students.

Group work is generally seen as a powerful representative collaborative learning technique ([23]). [24] performed a meta-analysis on the effects of small-group learning on student achievement, persistence, and attitudes in classes in undergraduate science, mathematics, engineering, and technology (STEM). Their conclusion regarding student achievement shows that students who learned in small groups demonstrated greater achievement than students in traditional instruction: "Students who learn in small groups generally demonstrate greater academic achievement, express more favorable attitudes toward learning, and persist through [STEM] courses or programs to a greater extent than their more traditionally taught counterparts. The reported effects are relatively large in research on educational innovation and have a great deal of practical significance.". They also report equally positive increases in academic performance for women and men, STEM majors and nonmajors, first-year and other students when participating in collaborative learning. In addition "the positive effects of small-group learning were significantly greater for members of underrepresented groups (African Americans and Latinas/os)." [25] conducted a systematic review of 2,506 published and unpublished citations of cooperative learning in secondary (the last years of high school) and early postsecondary science classrooms (community colleges and the first two years of college and university instruction) published between 1995 and 2007. Thirty of these studies met his criteria for inclusion in the meta-analysis he presented. He concludes "The results of this review indicate that cooperative learning improves student achievement in science. The overall mean effect size was .308, a medium effect...If the intervention was structured in a particular fashion, the effect on student achievement was greater than that for an unstructured intervention". [26] study the positive learning impact of team work in over 500 engineering and computer science courses. They conclude students working in teams—compared to students working individually—are significantly more likely to agree that the course had achieved its stated learning objectives. Project work is seen as an important method in inductive teaching [27] and exposing students to team projects improves the attitude towards working in teams ([28]). Teams can work effectively in virtual settings as well [29]. Self-and peer evaluation of teams is discussed in detail in [30]. We note that study program accreditation guidelines such as ABET require students to demonstrate "an ability to function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives" (Criterion 3).

As described above, there is general agreement that cooperative learning, collaborative learning, and team-based learning are effective pedagogical approaches to increase student learning. These approaches have in common

that they use student teams ([31]). Grouping of students is the first step in initiating any type of cooperative, collaborative or team work. There are three main methods instructors use to assign students to teams: self-selection, instructor formed and randomly assigned. Self-selection allows students to chose with whom they want to be in a team. Self-selection can lead to a better team experience by students, better group dynamics and perceived better learning ([32], [33], [34], [35]). The opposite selection method to self-selection is team formation by the instructor. There are multiple factors instructors might use to, for example, create diverse teams. Research shows that instructor formed teams can also perform well ([36], [37], [38]). Randomly assigning students to teams is one method which does not require student or instructor input and is often used for expediency for short-term team assignments. Student teams that are randomly assigned tend to have worse team experiences compared to self-selected teams ([32]). The group formation problem has also been studied in Mobile Computer Supported Collaborative Learning (MCSCL) environments. An overview of related work can be found in [39]. None of the aforementioned grouping methods optimize grouping of students based on maximizing their social network connections, which is the topic of this study.

The dynamics of social networks in undergraduate cohorts is studied by [40]. The goal of their study was to measure the multidimensional and dynamic aspects of social networks between students and to add data to a growing and diverse group of educational network studies. [41] present an overview of the role that social networks play in students' learning experiences, the construction of students' social networks, the evolution of these networks, and their effects on the students' learning experience in a university environment based on a thorough literature review. Their conclusion shows more research is required to better understand the role of social learning presence on student satisfaction, student motivation and other attitudinal factors.

The considered problem is related to classroom management such as the assignment of classroom seats. In [42], the authors report a student-chosen seating order which doubles the overall disruptive behavior in a class compared to when the instructor assigns seats. Similarly, it is known that the seating arrangement has a positive impact on the number of questions asked by students [43].

3 The Class Social Network

Our novel grouping approach considers a social network which mirrors connections between students in a class. In this section, we review the relevant concepts from social network analysis, clearly define the type of tie which we are interested in, and describe how we efficiently collect the corresponding student data.

Social networks describe social interactions and, more generally, personal relationships between individuals. Examples for relationships are numerous, including friendship, shared interests, geographical proximity, and relatives. In

social network analysis, abstract network structures are implemented to model and analyze the complex real-world system which consists of these connections. Properties of interest are typically related to connectivity within the network, and subgroups of individuals with common characteristics. In such a network model, individuals are represented by nodes. A relationship between two individuals is associated with an edge, or tie, that connects the corresponding nodes. The resulting network is also called a graph. It is considered symmetric when the relationship is symmetric—the two connected persons are mutually experiencing the considered relationship. Real-world networks that represent friendships between individuals are typically asymmetric [44, 45]. However, we model the class network using an undirected network, which can be motivated as follows. Class network data is naturally collected in a directed form since ties are surveyed egocentrically. There are two ways to interpret asymmetric connections. On can drop one-sided connections, assuming that the tie might not be sufficiently strong. However, we decided to complement the network by adding the reverse connection, resulting in a symmetric relationship. On the one hand-side, we presume that one student indicating a tie is a good enough sign for meaningful prior interaction. On the other hand-side, we believe that we batter account for connections that were accidentally overseen by respondents, and incomplete survey data due to poor participation.

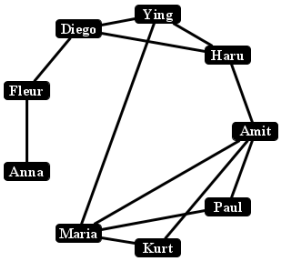
In this work, we consider dynamic networks in which student ties continuously change over the course period. We focus on two temporal points for network analysis: The pre-class state and the post class state. Edges may be created, crash, and recover, whereas nodes are considered permanent. This snapshot-based analysis of the network evolution is carried out by studying two static networks [46]. In contrast to studies that aim at capturing the continuous network evolution ([47, 48]) or long-term trends ([49, 50]), we are specifically interested in discrete change over time. In fact, we conduct an exploratory comparison of the initial pre-project class social network and the post-project network.

For an introduction to social networks and their analysis, we refer to [51]. The following example illustrates a class social network.

Example

Consider the social network depicted in Figure 1 (left). The 9 nodes correspond to individuals Anna, Amit, Fleur, Diego, Haru, Kurt, Maria, Paul, and Ying. The nodes are connected via 12 (symmetric) edges. In this example, the node degrees are: 1 (Anna), 4 (Amit), 1 (Fleur), 3 (Diego), 3 (Haru), 2 (Kurt), 4 (Maria), 2 (Paul), and 3 (Ying).

As described above, every student in the class is represented by a node in the network. The definition of what constitutes a relationship between students is essential for the data acquisition and interpretation of results. We model the class social network using a relationship between two individuals which is based on either personal nature, or prior in-depth school-related interaction, on or off campus. Ties emerging from notable direct communication are of



#	Name	1 Anna	2 Amit	3 Fleur	4 Diego	5 Haru	6 Maria	7 Paul	8 Kurt	9 Ying
1	Anna									
2	Amit									
3	Fleur									
4	Diego									
5	Haru									
6	Maria		X					X	X	X
7	Paul									
8	Kurt									
9	Ying									

#	Name	1 Anna	2 Amit	3 Fleur	4 Diego	5 Haru	6 Maria	7 Paul	8 Kurt	9 Ying
1	Anna			X						
2	Amit					X	X	X	X	
3	Fleur	X			X					
4	Diego			X		X				X
5	Haru		X		X					X
6	Maria		X					X	X	X
7	Paul		X				X			
8	Kurt		X				X			
9	Ying				X	X	X			

Fig. 1 An example for a class social network with 9 nodes and 12 edges (left), the corresponding survey form filled out by student Maria (center), and the complete survey data resulting in the adjacency matrix (right).

interest for our study since our intention is to identify connections which are beyond loose contact. In order to capture the class social network and its evolution, a survey is given to students in each course twice per term: Once at the beginning, and once at the end of the term. Students are asked to provide binary information about their ties to class mates. The survey explicitly states the following instructions: "Please insert an X for the people that you *know*. Knowing someone means to have worked with the person or have repeatedly interacted with this person beyond surface-level."

To illustrate the data collection process let us consider the hypothetical class shown in Figure 1. Each student fills out the survey at the beginning of the term. Figure 1 (center) shows how Maria filled out the form. Maria indicated she knows four students in the class: Amit, Paul, Kurt and Ying. All individual survey responses are consolidated into the so-called adjacency matrix depicted in Figure 1 (right), which encodes the entire network. A graphical representation of the connections between students is shown in Figure 1 (left).

In our empirical study, we shared the anonymized data after the first survey in form of a network diagram with the class to motivate student participation. An example for a class social network graphical representation from one of our investigated classes can be found in the appendix.

4 Optimizing Team Assignments

In this section, we present the optimization approach we used to optimally assign students to teams. We apply this technique to the social network which is obtained through the data collection described in Section 3. We first develop the core mathematical model and show how it can be applied to the classroom setting in Section 4.1. This model is used in our empirical study in Section 5. A discussion of a model variant that delivers more accurate results in the case of very dense class social networks is presented in Section 4.2. Further model extensions and variants of practical relevance are elaborated in Section 4.3.

4.1 Mathematical Model

We introduce the optimal student assignment problem (OSAP), which asks for an assignment of each student to precisely one team, while respecting the desired team sizes, and minimizing the existing intra-team connections. Note that this problem can be interpreted as the problem of identifying viable teams that potentially result in filling a maximal number of missing ties in the existing social network. The problem is closely related to the classical assignment problem (AP), with uniform assignment weights, which can be solved efficiently by the Hungarian method as presented by [52]. In the AP, a number of so-called agents (our students) need to be uniquely assigned to given tasks. The tasks can be interpreted as the various places over all teams that a student can be assigned to. The objective in the AP is to minimize pre-defined assignment costs, whereas the objective in the OSAP considers the overall number of existing ties within the teams. The considered variant is significantly harder than the simple assignment problem as a result of the additional ties that depend on the actual team composition. For an overview of assignment problems we refer to [53]. Moreover, the OSAP can be seen as a variant of the bin-packing problems under conflict constraints [54]. For the latter, the social network can be seen as the conflict graph—the individual ties are the conflicts—and the teams represent the bins that items can be packed into. Research on this class of problems goes back to [55] who considered the special case of a single bin where not all items need to be assigned. Conflicts are also used in the classical graph coloring problem [56]. Translated to the OSAP, a minimal number of teams is sought such that a student assignment without inter-team ties exists. These bin-packing and coloring problems are extremely difficult to solve since they fall into the category of NP-hard optimization problems. The OSAP is NP-hard, since it could be used to determine whether a graph can be colored with a given number of colors. Furthermore, we mention scheduling with time constraints [57], which yields con-natural models. In these generalizations of bin-packing problems, given jobs need to be assigned to machines avoiding conflicts such that the number of machines is minimized.

Solving an instance of the considered OSAP by hand is only possible for very small student classes. When dealing with 10 or more students it is typically impossible to guarantee that the best-possible assignment is used. Therefore, we model the student assignments and the social network mathematically, and use integer programming (IP) from mathematical optimization in order to efficiently compute optimal assignments. IP is a methodology which is widely used to solve complex design and planning problems across many industries. It is based on mathematical reasoning and returns an assignment of values to so-called decision variables which optimizes a user-specified objective function. Additional constraints can be provided to model application-related side requirements. All expressions, which have to be linear in the variables, are integrated in a corresponding so-called integer program, which describes (or formulates) the problem. In practice, highly sophisticated solver software

is used to compute solutions. If the computation time permits, a best-possible, or optimal, solution is returned. For a thorough discussion on integer programming, we refer to [58].

In the following, we present an IP formulation for the OSAP. For a class social network with node set N and edge set E, let $M = \{1, 2, ..., m\}$ be the set of desired teams. The minimal team size of team $k \in M$ is denoted by u_k and its maximal size is denoted by o_k, respectively ($0 \leq u_k \leq o_k \leq \mathsf{N}$). Note that in the case of fixed team sizes, we have $u_k = o_k \ \forall k \in M$. If we require regular team sizes, then $u_k = u_{k'} = o_k = o_{k'} \ \forall k, k' \in M$ holds. To model the potential assignment of a node $i \in N$ to a team $k \in M$, we use a binary assignment variable $x_{i,k}$. Its possible values are interpreted as follows.

$$x_{i,k} = \begin{cases} 0 & \text{if node } i \text{ is not assigned to team } k \\ 1 & \text{if node } i \text{ is assigned to team } k \end{cases}$$

Additionally, we introduce a binary conflict variable $y_{i,j}$ for every existing tie $\{i, j\} \in E$, such that:

$$y_{i,j} = \begin{cases} 0 & \text{if nodes } i \text{ and } j \text{ are not assigned to the same team} \\ 1 & \text{if nodes } i \text{ and } j \text{ are assigned to the same team} \end{cases}$$

Then the problem of assigning all nodes to teams of appropriate sizes while minimizing the number of existing intra-team ties can be formulated as the following integer program.

$$(F) \quad \textit{Minimize} \quad \sum_{\{i,j\} \in E} y_{i,j} \tag{1}$$

$$\textit{Subject to} \quad \sum_{k \in M} x_{i,k} = 1 \qquad \forall i \in N, \tag{2}$$

$$u_k \leq \sum_{i \in N} x_{i,k} \leq o_k \qquad \forall k \in M, \tag{3}$$

$$x_{i,k} + x_{j,k} \leq y_{i,j} + 1 \qquad \forall \{i, j\} \in E, k \in M, \tag{4}$$

$$x_{i,k} \in \{0, 1\} \qquad \forall i \in N, k \in M, \tag{5}$$

$$y_{i,j} \in \{0, 1\} \qquad \forall i, j \in N. \tag{6}$$

Objective function (1) minimizes the number of conflicts that are detected, which is measured by summing over the conflict variables. Unique-assignment constraints (2) assure that every student is assigned to exactly one team. Inequalities (3) force the team sizes to be within the corresponding lower and upper bounds. Variable-linking constraints (4) require conflict variable $y_{i,j}$ to

take a value of 1 if both students, i and j, are assigned to the same team. Note that in an optimized assignment, variables $y_{i,j}$ will be set to zero if possible, since we are minimizing the sum of these variables. The sets of binary variables are defined in (5) and (6).

For our example class from Section 3, we assume that $M = \{1, 2, 3\}$, and that all teams have precisely three members; i.e., $u_k = o_k = 3\ \forall k \in M$. The corresponding complete IP formulation (F) is provided in the appendix. An optimal student assignment is depicted in Figure 2 (left). The minimal number

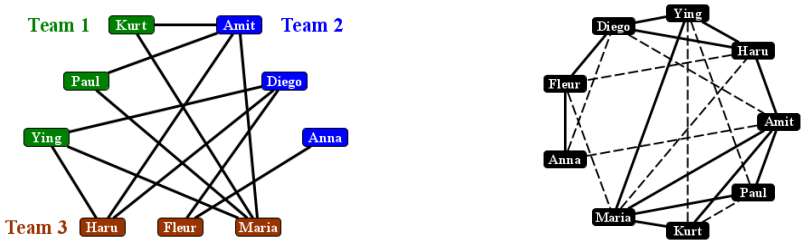


Fig. 2 An optimal team assignment for our 9-student example class (left) and the expected social network expansion with 69% new ties (right).

of intra-team ties is achieved, and the objective function value equals zero. For a given fixed group size σ and γ groups, the maximum number of intra-team ties can be computed by $\gamma\sigma(\sigma - 1)/2$. Assuming team members effectively collaborate within the teams, the potential post-project social network is given in Figure 2 (right). In this example, the number of ties increases from 13 to at least 22 (169%), and the potential for new ties is 3*3*(3-1)/2=9.

4.2 The Dense Case

For social networks that are dense or of a particular complicated structure, a certain amount of violations is unavoidable. In other words, some students will know each other within the groups, even after optimizing the assignments. Recall the approach described in Section 4.1 which focuses on the minimization of the overall number of intra-team connections. This class-oriented goal can be seen as a global strategy, and may result in an irregular distribution of an inevitable number of existing ties among different teams. Hence, students in some teams may not benefit from our idea, since they know most or even all students within their team. Others, however, may get to know a maximum number of peers. To demonstrate this, consider the optimal assignment for another example class social network with 9 students and 21 ties as depicted in Figure 3 (left). The optimal assignment found using formulation (F) contains 3 intra-team conflicts, all occurring in team 2. Neither Astrid, Ottilie nor Mario will be exposed to students that they do not know yet—they form a so-called clique. On the other hand, the alternative optimal assignment illustrated in Figure 3 (right) seems favorable because the three inevitable conflict violations are spread over different teams. Consequently, every student gets to know at

least one other student. However, there is no incentive provided in formulation (F) to return such a conflict-balanced assignment.

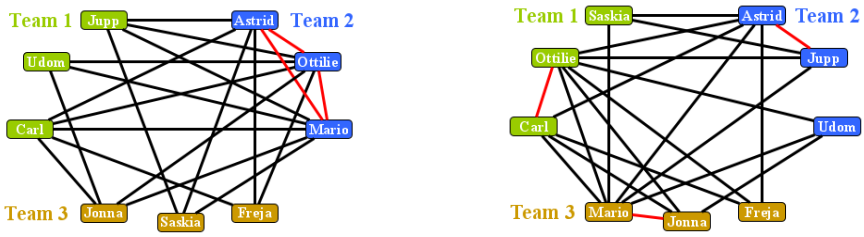


Fig. 3 Optimal team assignments with minimal overall intra-team conflicts (left) and minimized maximal intra-team conflicts over all teams (right).

To overcome this weakness of formulation (F), we describe a model variant which aims at minimizing the maximum number of conflict violations over all teams. Formulating this model variant requires a more detailed encoding of the team assignments to account for the resulting min-max type objective. To this end, we first introduce the following alternative set of binary conflict variables that are team-dependent, for nodes $i, j \in N$ and team $k \in M$.

$$y^k_{i,j} = \begin{cases} 0 & \text{if nodes } i \text{ and } j \text{ are not both assigned to team } k \\ 1 & \text{if nodes } i \text{ and } j \text{ are both assigned to team } k \end{cases}$$

In addition, we define an integer variable $z \geq 0$ that is used to represent the maximum over the sums of conflicts in the individual teams. The min-max version of the OSAP can then be formulated as the following integer program.

$$(F') \quad \textit{Minimize} \quad z \tag{7}$$

$$\textit{Subject to} \quad (2),\ (3),\ (5)$$

$$x_{i,k} + x_{j,k} \leq y^k_{i,j} + 1 \qquad \forall \{i,j\} \in E, k \in M, \tag{8}$$

$$\sum_{\{i,j\} \in E} y^k_{i,j} \leq z \qquad \forall k \in M, \tag{9}$$

$$y^k_{i,j} \in \{0,1\} \qquad \forall i,j \in N, k \in M, \tag{10}$$

$$z \in \{0, 1, ..., \mathsf{E}\}. \tag{11}$$

The objective function (7) minimizes the maximal intra-team violation represented by variable z. Constraints (2), (3), and variables (5) are inherited from formulation (F). The inequalities (8), which link assignment and intra-team conflict variables, are an adaptation of the inequalities (4) in formulation (F).

Variable z is forced to equal at least the maximum number of violation in the teams in linking it with the inequality (9). The new variables are defined in (10) and (11). Note that an optimal assignment found by formulation (F') may have an increased number of overall intra-team conflicts, compared to optimizing using F. Not all these intra-team ties are necessary, and their elimination can be achieved by exchanging objective (7) by the following function.

$$Minimize \quad z + \frac{1}{\mathsf{M}+1} \sum_{\{i,j\} \in E, k \in M} y_{i,j}^{k} \tag{12}$$

Objective (12) implements a ranked optimization of the objectives from formulation (F) and (F'). It prioritizes the maximum intra-team violations and considers the minimization of the overall conflicts as secondary objective. Formulation (F') is larger in terms of number of variables. In our empirical study in Section 5, we use formulation (F) since optimally assigned teams without violations can be found for all classes.

4.3 Model Extensions and Variants

The model presented above may serve as a base model that can be extended to suffice additional practical requirements. Finally, it is the instructor's freedom and responsibility to adjust the assignment model in order to accommodate student needs and other factors. In this section, we present a broad set of features allowing for flexible hybridization of the core social-network-oriented approach. Using these modifications, instructors can, for example, force the assignment of selected students to specific teams, or let students partially assign themselves to teams. Moreover, the integration of the different modifications into the mathematical formulation (F) is described.

4.3.1 Pre-Assigning Students

Probably the most intuitive side constraint is that a student $i \in N$ must be assigned to a team $k \in M$. This can be motivated by the case that project topics are predefined and the participation of student i in the corresponding team k is crucial, or the student strongly prefers to work on this particular project topic. We can accommodate this requirement by fixing the corresponding assignment via the addition of the constraint $x_{i,k} = 1$ to the integer program.

4.3.2 Balancing Teams

Team project requirements may necessitate that a certain number of team mates hold specific attributes. These could stem from a desired team composition with respect to the students (e.g., study major, required skills, personal background, gender, etc.). Say, team $k \in M$ requires at least l students fulfill a certain property Π, and let $\Pi(N) \subseteq N$ denote all eligible students. Then the

following constraint guarantees that at least l eligible students are assigned accordingly, and allows the hybridization of our approach.

$$\sum_{i \in \Pi(N)} x_{i,k} \geq l \tag{13}$$

Note that inequality (13) can also be used by to setting the number of allowed students to u on the right hand side of the equation and changing the relationship to less than or equal to. Applying such a hard constraint to each team supports globally balanced teams.

4.3.3 Forcing Team Mates

In some cases, two or more students are required to be on the same team. This can be due to the nature of the project (e.g., industry collaboration, senior design project) or special student-related circumstances (e.g., students' location). In addition, some students may be allowed to self-select their team mates, leading to a hybrid model. Say, students i and j have to be team mates, then the addition of the following constraints ensures this.

$$x_{i,k} + \sum_{k' \in M : k' \neq k} x_{j,k'} \leq 1 \qquad \forall k \in M \tag{14}$$

Inequalities (14) imply that if student i is assigned to a team k, then student j must not be assigned to any other team. Hence, they have to be on the same team. Note that when multiple students need to be on the same team, this can be enforced by adding all corresponding pairwise constraints.

4.3.4 Averting Team Mates

Instructors may want to offer students to indicate if they don't want to be with a particular student or students on a team before teams are assigned. Similarly, instructors may want to avoid pairing certain students for various reasons. This leads to the opposite of forcing team mates, as suggested in Section 4.3.3. Consider students i and j who should not be on the same team. This can be achieved by adding the following constraints to formulation (F).

$$x_{i,k} + x_{j,k} \leq 1 \qquad \forall k \in M \tag{15}$$

4.3.5 Maximizing Intra-Team Ties

For projects which are time-critical, the initial setup cost of getting to know each other might hinder the students from accomplishing the project goals in time. In this case, an optimization approach which is contrary to our main idea can be implemented allowing students to start working effectively immediately. We can minimize the time spent for getting to know each other by maximizing

the existing ties within the teams—in other words, teams are formed with as many students that already know each other as possible.

The objective contrary to objective (1) in formulation (F) is the maximization of ties in the formed teams. For specific projects which require the team members to know each other, this approach may be well-suited. This can be done by replacing inequalities (4) in formulation (F) by the following inequalities.

$$x_{i,k} + x_{j,k} \leq y_{i,j} + 1 \qquad \forall \{i,j\} \notin E, k \in M \tag{16}$$

Note that in inequalities (16), variables $y_{i,j}$ are interpreted in a reverse way. In other words, a conflict is observed if two team members do not know each other. We will evaluate this alternative objective in our empirical study in Section 5.

4.4 Spreadsheet Implementation

Our intention is to emphasize the practical applicability of our method. In contrast to customized implementations for related student assignment problems (see Section 2), we fully carried out the team planning in the 10 classes using a spreadsheet-based model[1]. The IP solver included in Excel cannot be used to solve the formulations above, since it is limited in terms of number of formulation variables (200) and constraints (100). Formulation (F) for a class with 31 students and 8 teams contains 7936 variables and 7735 constraints. Therefore, we used the freely available OpenSolver plug-in[2]. OpenSolver interfaces several commercial and non-commercial IP solvers which do not impose limits regarding the formulation size. In our experiments, we found out our problems could not be solved to optimality using the default open-source solver COIN-OR CBC within a reasonable time. To overcome this, we used the commercial Gurobi[3] IP solver, which is available for research under an academic license and can be used with OpenSolver. The formulations introduced above can also be used directly in combination with other IP solvers.

5 Results and Analysis

To demonstrate the effectiveness of the introduced OSAP model, a case study is presented using real-life data from 10 college-level courses. In the following, we first describe the used data set (Section 5.1), and discuss participation-related properties of the survey outcomes (Section 5.2). Our main results on grouping methods potentials and practical impact are analyzed in Sections 5.3-5.5. We close this section with a summary of the student perception in Section 5.6.

[1] The template spreadsheet is available from the authors upon request.
[2] https://opensolver.org
[3] https://www.gurobi.com

5.1 Class Data

We use real data from 10 sections (A-J) of four different industrial engineering courses over a time period of four terms. Our study includes 253 surveyed students and 77 formed student teams. The detailed class data is given in Table 1. Four classes (A-D) were held as face-to-face classes, and six classes (E-J) were taught virtually in a synchronous mode. In six courses, we implemented our new optimization grouping method (OPT), and students in the other four courses were allowed to self-select (SELF) into teams yielding our control set which we use for the evaluation of the overall impact. The courses included in our case study are Operations Research II (IME 305), Manufacturing Automation (IME 356), Simulation (IME 420), and Advanced Operations Research (IME 541). The number of class students is given in column n. The response rates at the beginning and at the end of the term can be found in column RR_0 and column RR_1, respectively. The group size ranges from 2 (column s_{min}) to 5 (column s_{max}) students, and is 3.3 students on average (column s_{avg}).

Table 1 Classes used in our case study including properties, social network survey response rates, and grouping method.

#	Term	Course	Class Mode	n	RR_0	RR_1	#Groups	s_{min}	s_{avg}	s_{max}	Method
A	1	IME 420	Face-To-Face	31	100.0	87.1	8	3	3.9	4	OPT
B	2	IME 420	Face-To-Face	26	100.0	100.0	7	3	3.7	4	OPT
C	2	IME 305	Face-To-Face	28	85.7	71.4	10	2	2.8	3	SELF
D	2	IME 305	Face-To-Face	30	76.7	86.7	10	3	3.0	3	OPT
E	3	IME 420	Virtual	28	100.0	85.7	7	4	4.0	4	SELF
F	3	IME 420	Virtual	25	100.0	84.0	7	3	3.6	4	OPT
G	3	IME 356	Virtual	20	70.0	55.0	10	2	2.0	2	OPT
H	3	IME 541	Virtual	9	100.0	66.7	4	2	2.3	3	SELF
I	4	IME 420	Virtual	30	90.0	93.3	7	4	4.3	5	SELF
J	4	IME 305	Virtual	26	92.3	92.3	7	3	3.9	4	OPT
All				253	91.5	82.2	77	2	3.3	5	

5.2 Participation, Incomplete Data, and Asymmetry

Students voluntarily provided their network data in our Institutional Review Board (IRB) approved survey (see Section 3). We observed that the response rates tend to be lower at the end of the term compared to the beginning of the term. On average, the beginning of the term response rate was 91.5%, and 82.2% at the end of the term. Even though the average response rate is relatively high (87%), some students may have refused, forgotten or didn't have the time to fill out the surveys. As a result the network data is incomplete: On one hand, some existing ties may not be captured, on the other hand, only one student of a connected pair may have indicated the tie. Furthermore, a student might not indicate a tie with a class mate intentionally even though the other class mate does specify a tie. Both cases result in asymmetric network

information. Relate to link-reconstruction as described by [44]. To resolve this, we assume that a connection exists between two individuals if at least one person indicates knowing the other person. The number of ties that emanate from this symmetrization is illustrated in Figure 4 (left). On average, about half of the symmetric ties are non-mutual. In Figure 4 (right), we can see a

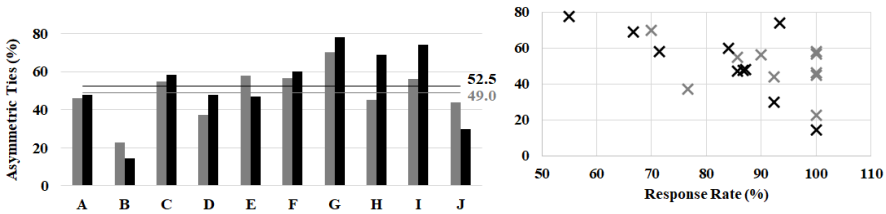


Fig. 4 The relative number of asymmetric ties in pre-project (grey) and post-project (black) class surveys for the ten different classes (left); response rates versus the asymmetric ties (right).

reduced survey participation tends to increase the observed asymmetric ties in the model.

5.3 Grouping Method Potentials

We are interested in comparing our optimized assignment method to two different grouping approaches—random and self-select. The question is: How many ties can be created as a result of project collaboration compared to the theoretical maximum upper bound (UB) for the desired number of teams and team sizes? For a single team of size k, at most $k(k-1)/2$ new ties can be established. The overall potential for a class is then obtained by summing up this value for all teams. Note: We do not use the relative increase in overall class ties with respect to the existing ties here. In addition, for this hypothetical analysis of the potentials, only the initial class social network survey is required. As a consequence, the resulting potentials are independent from the teaching mode. The actual in-class social network growth rates are discussed in Section 5.4.

We conducted a Monte-Carlo simulation to understand the potential impact if random assignment would be used. For every class, we drew 100 samples of randomly assigned students to teams and evaluated the intra-team potentials. The obtained distributions of the overall potentials from the Monte-Carlo simulation are depicted in Figure 5.

A trivial lower bound (LB) on the number of new ties is zero, which could occur in the case of an extremely dense social network. To obtain a practical lower bound on the potential which takes into account the existing network, we conduct an additional optimization-based analysis. Contrary to optimizing for a minimum number of intra-team ties (see Section 4), we maximize the overall intra-team ties. This can be achieved by the following model

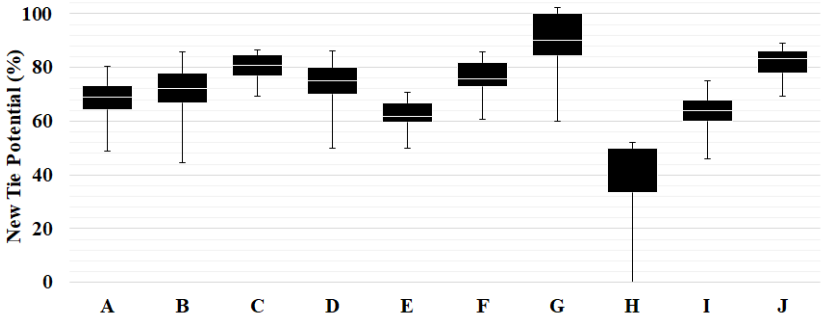


Fig. 5 Grouping results from the Monte-Carlo simulation: The distributions of the relative number of potential new ties from 100 randomly generated team formations for each class.

transformation. Instead of using all existing student ties, we use the complementary conflict set. In other words, add a conflict between student i and student j if and only if i and j are not connected. Then we use formulation (F) to compute an assignment with maximum intra-team ties, and denote this method as OPT(INVERSE) (see also Section 4.3.5). This approach could also be described as clique-oriented. Furthermore, we include the optimized assignment method (OPT), self-assignment (SELF), and random grouping (RANDOM) in our analysis.

We illustrate the relative potentials for the different methods in Figure 6. The average lower bound on the potential obtained through OPT(INVERESE) is 21.1%. For the four classes in which we let students self-select (C,E,H,I), we observe an average potential of 49.2%. Randomly assigning students to teams yields a significantly higher potential of 71.5% on average. However, our optimized team formation method increases the potential to an outstanding 98.3% on average. In most cases, we can even find teams without intra-team ties at all.

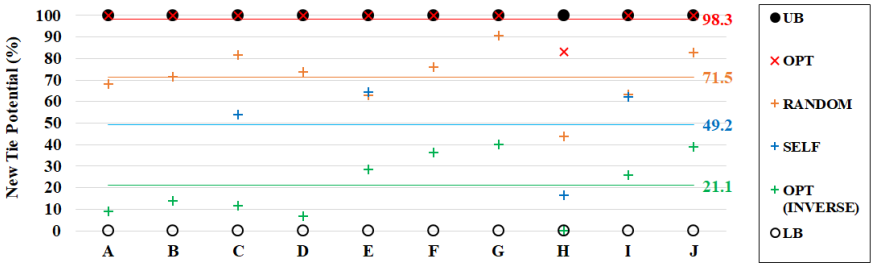


Fig. 6 The relative potential (with respect to the theoretical maximum) for new ties using the different grouping methods for the classes and the corresponding overall average values.

5.4 Grouping Impact Analysis

In the following, we present the impact of our optimization-based grouping on the class social network by comparing the results to the cases where students self-assign to teams (see Table 1 for used methods, OPT or SELF). We first focus on the overall increase of ties, degrees and the corresponding network density. A subgroup and connectivity analysis is presented in Section 5.5. The summary of the grouping methods impacts is given in Table 2. A notable increase of student ties is obtained when using our new method compared to student self-assignment—more than 3.5 times the ties are created when optimizing the teams (61.5% vs. 16.9%). There is a notable difference when allowing students to choose their team mates when comparing the two different teaching modes—Face-to-Face versus Virtual. New ties are dramatically rare in the virtual setting (12.2% vs. 31%). This difference is reduced from 18.8 to 5.2 percentage points when assigning optimized teams (58.9% vs. 64.1%), demonstrating significant increases in student ties regardless of teaching mode—33.1% increase for the investigated face-to-face courses and 46.7% increase for virtual courses. The actual impact might be even higher as a result of the survey response rates (see Section 5.2).

Table 2 The average relative increase in ties for different class modes (Face-to-Face and Virtual) and team member assignment methods (Student Self-Assignment and Optimized Assignment); the factor of augmentation through optimization in the different mode settings.

Increase of Student Ties (%)			**Factor**
Mode	**SELF**	**OPT**	
Face-to-Face	31.0	64.1	x2.1
Virtual	12.2	58.9	x4.8
All	16.9	61.5	x3.6

Student Degrees

The degree of a node, or student, measures the number of direct connections to class mates. The higher the degree, the better the student is embedded into the class socially. We measured average student degrees before and after the term and illustrate this data in Figure 7. It can be seen that the average degree increased by 23% (6.5→8.0) in classes where students self-assigned. Our optimized grouping helped to obtain a 62% increase in average degree (6.1→9.9).

Network Density

The density of a network relates the number of edges to the number of nodes. It is computed by dividing the number of network edges by the maximum

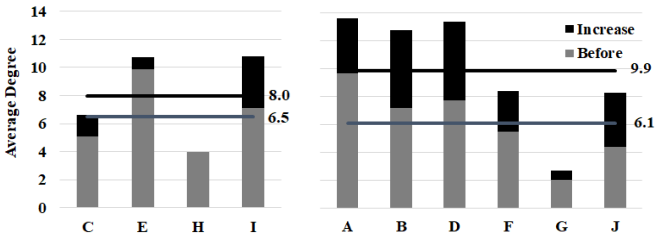


Fig. 7 The initially observed average degree (grey) and the increase (black) after the team projects for the different classes and assignment methods (SELF: left; OPT: right).

possible number of edges: $2\mathsf{E}/(\mathsf{N}(\mathsf{N}-1))$. Note that, for each new tie, two students establish a new connection. The example network in Figure 1 (left) has a density of $12/36 \approx 0.33$. As illustrated in Figure 8 (right), the average density doubled (0.2→0.4) in classes with optimized teams, whereas it only increased by 33% when students self-assigned (0.3→0.4).

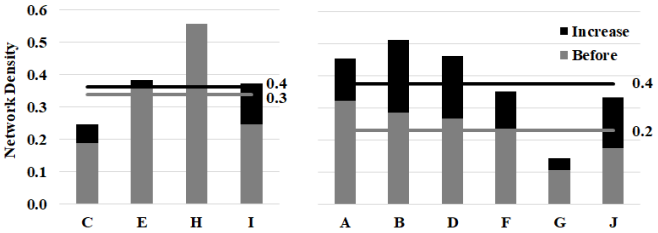


Fig. 8 The initially observed average density (grey) and the increase (black) after the team projects for the different classes and assignment methods (SELF: left; OPT: right).

5.5 Network Subgroups, Components and Diameter

An important aspect of social network analysis is the identification of cohesive subgroups. To better understand the effect of our grouping methodology, we consider the following two well-known measures[4] and compare the priori and posteriori networks according to them (see also [51]).

Cliques

A clique is a sub-set of nodes were all nodes—and therefore the corresponding individuals the nodes represent—are connected to each other. The clique number measures the size of a largest sub-group that is a clique. A high clique number can be seen as an indicator for a strong community. Figure 9 (left) illustrates the changes of clique numbers after finishing the team projects. For the self-assignment courses the clique number increased (C, I), stayed the

[4]Computations are performed using the MAPLE Graph Theory package (https://www.maplesoft.com).

same (H), or decreased (E). For the optimized assignments, the clique number stayed the same (F) or increased (A, B, D, G, J). On average, an increase in the clique number of 19.6% is observed for optimized assignments, and an increase of 13.9% when letting the student self-assign (17.3% total).

Independent Set

An independent set is a set of nodes for which there does not exist any connection between the corresponding individuals. The independence number counts the number of nodes in a largest independent set in the network. Figure 9 (right) illustrates the change of independence number after finishing the team projects. On average, a decrease of 34.3% in the Independence Number is observed for optimized assignments, and an average decrease of only 11.8% is observed when letting the student self-assign.

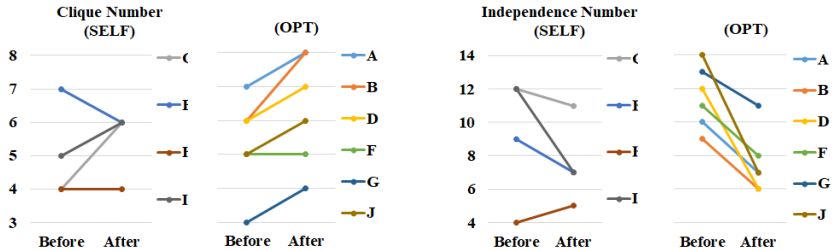


Fig. 9 Change of maximum student clique size when letting students self-assign (SELF) and optimizing (OPT) teams (left), and change of maximum independent student sub-set when letting students self-assign and optimizing teams (right).

Furthermore, we are interested in how the studied grouping methods affect network connectivity. The creation of new ties obviously decreases the number of individuals a message would have to go through in order to reach another student. To quantify the corresponding impact, we use the following measures.

Connected Components

Two nodes in the network are connected if there is a set of ties that (indirectly) connect them. Note that for some networks, there are nodes which are not connected, directly or indirectly. A connected component is a sub-set of nodes for which all nodes are connected. One self-assigned class and three optimized classes—out of the 10 classes we studied—had an initially disconnected network; i.e., more than one connected component, and hence pairs of students which were not connected. After the team project, all networks in all classes studied were connected.

Diameter

The diameter of a network is a measure related to how fast information spreads. Let us assume we know the length of the shortest (indirect) connection in the

network for each pair of individuals. The diameter of the network is the largest of these connection distances. For disconnected networks the diameter is based on the largest connected component. This invariant can be seen as a global measure to quantify the worst case of efficient (indirect) communication. In our study, we observed initial class diameters ranging from 2 to 5, and 3.6 on average. In three cases, the diameter was shortened by an average of 30%, but only when optimized team assignments have been implemented.

5.6 Student Perception

The written student feedback regarding the students' team experience was largely positive. Below are some selected comments from the mandatory team project evaluations in classes where we optimized the team assignments. Students were asked to "take a few minutes to carefully evaluate each team member", including themselves. Additionally, an open-ended field, titled "Other comments (optional)" was provided, not explicitely soliciting grouping-related feedback.

- "One of the best teams of my IME [Industrial and Manufacturing Engineering] experience." [Class A]
- "This group is one of the most efficient and hardworking groups I have ever been a part of in Cal Poly [California Polytechnic State University]. Everyone was attentive and supportive to everyone's ideas and everyone contributed an equal amount. No one was left with an unequal workload, which is a common thing that happens in group projects. Overall, I am so pleased to end this quarter with two new friends and peers in IE [Industrial Engineering]." [Class B]
- "It was a nice experience to work with others in my major that I have never gotten the chance to work with before, even as a 4th Year Industrial Engineering major (I thought I knew everyone in my year!). I applaud for having that experience."[Class B]
- "Really quite liked this group; for a bunch of people that did not know each other, and in the case of XXX [name removed], came from rather different places, we all got along well and had a surprisingly enjoyable time considering how messy our simulation model was." [Class A]
- "We overall did not work well as a group." [Class A]
- "I truly feel that everyone did their fair share of the work. We had many zoom meetings where we all collaborated and worked together. While I was initially sad that the groups were assigned, it ended up working out really well for our team. We worked very well together and I was glad to meet some new classmates." [Class E]
- "I really liked getting to work with new people!" [Class E]
- "I made three great friends throughout this project, thank you for a great quarter professor!" [Class J]

As the comments indicate, students had a good team experience overall—except for one comment from class A—and students noticed their social network increased (e.g., "new friends", "meet some new classmates"). We

could not find any indication from student comments that using the optimized approach to assign students to teams had a negative impact on students. The student comment indicating their team "did not work well" is not necessarily related to the grouping method used and does not mean the students didn't get to know each other.

6 Conclusion

We presented a new approach for assigning students to project teams. Our overall goal is to maximize the students' opportunities to get to know other students through collaboration during class projects. To this end, we collected information on the class social network before the team projects commenced and grouped the students—using our new approach—such that the number of preexisting ties between students in each team is minimal. We showed the underlying optimization problem is challenging, but it can be solved efficiently by an integer linear programming method as presented. Several practical model extensions were suggested allowing the integration of existing grouping paradigms, as well as additional requirements emanating from instructor and student preferences, or other limitations.

An empirical study was conducted including data from more than 250 students from 10 classes over one full calendar year—including both, virtual and face-to-face teaching modes. We investigated the potentials for new ties between students by comparing our optimized team compositions to other grouping methods—random and self-select—and reported the achieved increase in student ties at the end of the term.

Through a Monte-Carlo simulation study we found 30% more ties on average can be obtained by our approach compared to randomly assigning students into teams. The comparison to the classes where students were allowed to self-assign to teams demonstrates that our new method effectively doubles the number of newly obtained ties—students got to know twice as many students. In the post-project analysis, we were able to quantify the impact of the team structures on the class social network. Through our method, we demonstrated an increase of the number of ties by an impressive 62% compared to only 17% when allowing students to self-assign to groups.

Students qualitative comments indicate students had a good team experience and students noticed that their social network increased by making "new friends". We could not find any indication from student comments that using the optimized approach to assign students to teams had a negative impact on students.

Our approach is not limited to educational settings but can also be applied to various other settings including corporate operations, workshops, sports, professional development courses, committees, and social events. In the considered college environment, we suggest the incorporation of social network aspects in seating arrangements, seating assignments, and class section assignments as future research directions.

Acknowledgments

We are indebted to Nian Cheng and Jose Macedo for supporting the collection of class social network data and letting us generate project teams in their classes. Furthermore, we thank John Pan for his valuable comments.

References

[1] Burke, A.: Group work: How to use groups effectively. Journal of Effective Teaching **11**(2), 87–95 (2011)

[2] Forslund Frykedal, K., Hammar Chiriac, E.: Student collaboration in group work: Inclusion as participation. International Journal of Disability, Development and Education **65**(2), 183–198 (2018)

[3] Johnson, R.T., Johnson, D.W.: Active learning: Cooperation in the classroom. The annual report of educational psychology in Japan **47**, 29–30 (2008)

[4] Webb, N.M.: Assessing students in small collaborative groups. Theory into practice **36**(4), 205–213 (1997)

[5] Shepherd, W.G.: Operations research in education. Management Science **11**(4), 13–19 (1965)

[6] Proll, L.: A simple method of assigning projects to students. Journal of the Operational Research Society **23**(2), 195–201 (1972)

[7] Weitz, R., Lakshminarayanan, S.: An empirical comparison of heuristic methods for creating maximally diverse groups. Journal of the operational Research Society **49**(6), 635–646 (1998)

[8] Gallego, M., Laguna, M., Martí, R., Duarte, A.: Tabu search with strategic oscillation for the maximally diverse grouping problem. Journal of the Operational Research Society **64**(5), 724–734 (2013)

[9] Palubeckis, G., Ostreika, A., Rubliauskas, D.: Maximally diverse grouping: an iterated tabu search approach. The Journal of the Operational Research Society **66**(4), 579–592 (2015)

[10] Hübscher, R.: Assigning students to groups using general and context-specific criteria. IEEE Transactions on Learning Technologies **3**(03), 178–189 (2010)

[11] Layton, R.A., Loughry, M.L., Ohland, M.W., Ricco, G.D.: Design and validation of a web-based system for assigning members to teams using

instructor-specified criteria. Advances in Engineering Education **2**(1), 1–28 (2010)

[12] Zervoudakis, K., Mastrothanasis, K., Tsafarakis, S.: Forming automatic groups of learners using particle swarm optimization for applications of differentiated instruction. Computer Applications in Engineering Education **28**, 282–292 (2019)

[13] Thanh, T.N., Morgan, M., Butler, M., Marriott, K.: Perfect match: Facilitating study partner matching. In: SIGCSE '19: Proceedings of the 50th ACM Technical Symposium on Computer Science Education, pp. 1102–1108 (2019)

[14] Banchero, M., Robledo, F., Romero, P., Sartor, P., Servetti, C.: Max-diversity orthogonal regrouping of mba students using a grasp/vnd heuristic. In: Mladenovic, N., Sleptchenko, A., Sifaleras, A., Omar, M. (eds.) Variable Neighborhood Search, pp. 58–70. Springer, ??? (2021)

[15] Holmberg, K.: Formation of student groups with the help of optimisation. Journal of the Operational Research Society **70**(9), 1538–1553 (2019)

[16] Schulz, A.: The balanced maximally diverse grouping problem with block constraints. European Journal of Operational Research, In press (2021)

[17] Roeder, T.M., Saltzman, R.M.: Schedule-based group assignment using constraint programming. INFORMS Transactions on Education **14**(2), 63–72 (2014)

[18] Celik, B.G., Brunnhoeffer, G.C.: A stepwise multi-assignment spreadsheet model to select student teams. International Journal of Engineering & Technology **2**(3), 247–258 (2013)

[19] Anagnostopoulos, A., Becchetti, L., Castillo, C., Gionis, A., Leonardi, S.: Online team formation in social networks. In: Proceedings of the 21st International Conference on World Wide Web. WWW '12, pp. 839–848. Association for Computing Machinery, New York, NY, USA (2012)

[20] Kargar, M., An, A.: Discovering top-k teams of experts with/without a leader in social networks. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. CIKM '11, pp. 985–994. Association for Computing Machinery, New York, NY, USA (2011)

[21] Gutiérrez, J.H., Astudillo, C.A., Ballesteros-Pérez, P., Mora-Melià, D., Candia-Véjar, A.: The multiple team formation problem using sociometry. Computers & Operations Research **75**, 150–162 (2016)

[22] Chen, R.-C.: Grouping optimization based on social relationships. Mathematical Problems in Engineering **2012** (2012)

[23] Felder, R.M., Brent, R.: Cooperative learning in technical courses: Procedures, pitfalls, and payoffs. Technical report, North Carolina State University (1994)

[24] Springer, L., Stanne, M.E., Donovan, S.S.: Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. Review of Educational Research **69**(1), 21–51 (1999)

[25] Romero, C.C.: Cooperative learning instruction and science achievement for secondary and early post-secondary students: A systematic review. PhD thesis (2009)

[26] Oakley, B.A., Hanna, D.M., Kuzmyn, Z., Felder, R.M.: Best practices involving teamwork in the classroom: Results from a survey of 6435 engineering student respondents. IEEE Transactions on Education **50**(3), 266–272 (2007)

[27] Prince, M.J., Felder, R.M.: Inductive teaching and learning methods: Definitions, comparisons, and research bases. Journal of Engineering Education **95**(2), 123–138 (2006)

[28] Levi, D., Rinzel, L., Cadiz, D., Cacapit, M.: Effects of education and team projects on student attitudes toward team work. Technical report, California Polytechnic State University, San Luis Obispo (1998)

[29] Johnson, S.D., Suriya, C., Won Yoon, S., Berrett, J.V., La Fleur, J.: Team development and group processes of virtual learning teams. Computers & Education **39**(4), 379–393 (2002)

[30] Ohland, M.W., Loughry, M.L., Woehr, D.J., Bullard, L.G., Felder, R.M., Finelli, C.J., Layton, R.A., Pomeranz, H.R., Schmucker, D.G.: The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self-and peer evaluation. Academy of Management Learning & Education **11**(4), 609–630 (2012)

[31] Felder, R., Woods, D., Stice, J., Rugarcia, A.: The future of engineering education: Part 2. Teaching methods that work. Chemical Engineering Education **34**(1), 26–39 (2000)

[32] Bacon, D.R., Stewart, K.A., Silver, W.S.: Lessons from the best and worst student team experiences: How a teacher can make the difference. Journal of Management Education **23**(5), 467–488 (1999)

[33] Chapman, K.J., Meuter, M., Toy, D., Wright, L.: Can't we pick our own

groups? The influence of group selection method on group dynamics and outcomes. Journal of Management Education **30**(4), 557–569 (2006)

[34] Myers, S.A.: Students' perceptions of classroom group work as a function of group member selection. Communication Teacher **26**(1), 50–64 (2012)

[35] Rusticus, S.A., Justus, B.J.: Comparing student- and teacher-formed teams on group dynamics, satisfaction, and performance. Small Group Research **50**(4), 443–457 (2019)

[36] Rienties, B., Alcott, P., Jindal-Snape, D.: To let students self-select or not: That is the question for teachers of culturally diverse groups. Journal of Studies in International Education **18**(1), 64–83 (2014)

[37] Feichtner, S.B., Davis, E.A.: Why some groups fail: a survey of students' experiences with learning groups. Organizational Behavior Teaching Review **9**(4), 58–73 (1984)

[38] Oakley, B.A., Felder, R.M., Brent, R., Elhajj, I.: Turning student groups into effective teams. Journal of Student Centered Learning **2**(1), 9–34 (2004)

[39] Amara, S., Macedo, J., Bendella, F., Santos, A.: Group formation in mobile computer supported collaborative learning contexts: A systematic literature review. Journal of Educational Technology & Society **19**(2), 258–273 (2016)

[40] Vörös, A., Boda, Z., Elmer, T., Hoffman, M., Mepham, K., Raabe, I.J., Stadtfeld, C.: The swiss studentlife study: Investigating the emergence of an undergraduate community through dynamic, multidimensional social network data. Social Networks **65**, 71–84 (2021)

[41] Liccardi, I., Ounnas, A., Pau, R., Massey, E., Kinnunen, P., Lewthwaite, S., Midy, M.-A., Sarkar, C.: The role of social networks in students' learning experiences. ACM Sigcse Bulletin **39**(4), 224–237 (2007)

[42] Bicard, D.F., Ervin, A., Bicard, S.C., Baylot-Casey, L.: Differential effects of seating arrangements on disruptive behavior of fifth grade students during independent seatwork. Journal of Applied Behavior Analysis **45**(2), 407–411 (2012)

[43] Marx, A., Fuhrer, U., Hartig, T.: Effects of classroom seating arrangements on children's question-asking. Learning Environments Research **2**(3), 249–263 (1999)

[44] Krause, R.W., Huisman, M., Steglich, C., Snijders, T.: Missing data in cross-sectional networks – an extensive comparison of missing data

treatment methods. Social Networks **62**, 99–112 (2020)

[45] Kitts, J.A., Leal, D.F.: What is(n't) a friend? dimensions of the friendship concept among adolescents. Social Networks **66**, 161–170 (2021)

[46] Borgnat, P., Fleury, E., Guillaume, J.-L., Magnien, C., Robardet, C., Scherrer, A.: Evolving networks. In: Mining Massive Data Sets for Security vol. 19, pp. 198–203. IOS Press, ??? (2008)

[47] Palla, G., Barabási, A.-L., Vicsek, T.: Quantifying social group evolution. Nature **446**(7136), 664–667 (2007)

[48] Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: Densification and shrinking diameters. ACM transactions on Knowledge Discovery from Data (TKDD) **1**(1), 2–42 (2007)

[49] Albert, F., Koltai, J., David, B.: The increasing importance of friends: Changes in core discussion network composition in post-communist hungary between 1997–2015. Social Networks **66**, 139–145 (2021)

[50] Volker, B., De Cuyper, R., Mollenhorst, G., Dirkzwager, A., van der Laan, P., Nieuwbeerta, P.: Changes in the social networks of prisoners: A comparison of their networks before and after imprisonment. Social Networks **47**, 47–58 (2016)

[51] Borgatti, S.P., Everett, M.G., Johnson, J.C.: Analyzing Social Networks. Sage, ??? (2018)

[52] Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Research Logistics Quarterly **2**(1-2), 83–97 (1955)

[53] Burkard, R., Dell'Amico, M., Martello, S.: Assignment Problems. Society for Industrial and Applied Mathematics, ??? (2012)

[54] Sadykov, R., Vanderbeck, F.: Bin packing with conflicts: a generic branch-and-price algorithm. INFORMS Journal on Computing **25**(2), 244–255 (2013)

[55] Yamada, T., Kataoka, S., Watanabe, K.: Heuristic and exact algorithms for the disjunctively constrained knapsack problem. Information Processing Society of Japan Journal **43**(9), 2864–2870 (2002)

[56] Jensen, T.R., Toft, B.: Graph Coloring Problems. Wiley Series in Discrete Mathematics and Optimization, vol. 39. Wiley, ??? (2011)

[57] Jansen, K., Öhring, S.: Approximation algorithms for time constrained scheduling. Information and Computation **132**(2), 85–108 (1997)

[58] Conforti, M., Cornuéjols, G., Zambelli, G., *et al.*: Integer Programming. Graduate Texts in Mathematics, vol. 271. Springer, ??? (2014)

Appendix A Example for Mathematical Formulation

The IP formulation (F) for computing the 3 optimized teams in the example class from Section 4.1 is given in an LP file format below. The formulation contains 39 binary variables and 48 constraints. Note that Inequality (3) is split into two inequalities to comply with standard LP notation for solvers.

```
/*Objective (1)*/

MINIMIZE y_{1,3} + y_{2,5} + y_{2,6} + y_{2,7} + y_{2,8} + y_{3,4} + y_{4,5} + y_{4,9} + y_{5,9} + y_{6,7} + y_{6,8} + y_{6,9}


SUBJECT TO


/*Equation (2)*/

x_{1,1} + x_{1,2} + x_{1,3} = 1                                          x_{4,1} + x_{4,2} + x_{4,3} = 1
x_{2,1} + x_{2,2} + x_{2,3} = 1                                          x_{5,1} + x_{5,2} + x_{5,3} = 1
x_{3,1} + x_{3,2} + x_{3,3} = 1                                          x_{6,1} + x_{6,2} + x_{6,3} = 1

x_{7,1} + x_{7,2} + x_{7,3} = 1
x_{8,1} + x_{8,2} + x_{8,3} = 1
x_{9,1} + x_{9,2} + x_{9,3} = 1


/*Inequality (3)*/

x_{1,1} + x_{2,1} + x_{3,1} + x_{4,1} + x_{5,1} + x_{6,1} + x_{7,1} + x_{8,1} + x_{9,1} >= 3
x_{1,2} + x_{2,2} + x_{3,2} + x_{4,2} + x_{5,2} + x_{6,2} + x_{7,1} + x_{8,1} + x_{9,1} >= 3
x_{1,3} + x_{2,3} + x_{3,3} + x_{4,3} + x_{5,3} + x_{6,3} + x_{7,3} + x_{8,3} + x_{9,3} >= 3


/*Inequality (3)*/

x_{1,1} + x_{2,1} + x_{3,1} + x_{4,1} + x_{5,1} + x_{6,1} + x_{7,1} + x_{8,1} + x_{9,1} <= 3
x_{1,2} + x_{2,2} + x_{3,2} + x_{4,2} + x_{5,2} + x_{6,2} + x_{7,1} + x_{8,1} + x_{9,1} <= 3
x_{1,3} + x_{2,3} + x_{3,3} + x_{4,3} + x_{5,3} + x_{6,3} + x_{7,3} + x_{8,3} + x_{9,3} <= 3


/*Inequality (4)*/

x_{1,1} + x_{3,1} <= y_{1,3} + 1                                         x_{1,2} + x_{3,2} <= y_{1,3} + 1
x_{2,1} + x_{5,1} <= y_{2,5} + 1                                         x_{2,2} + x_{5,2} <= y_{2,5} + 1
x_{2,1} + x_{6,1} <= y_{2,6} + 1                                         x_{2,2} + x_{6,2} <= y_{2,6} + 1
x_{2,1} + x_{7,1} <= y_{2,7} + 1                                         x_{2,2} + x_{7,2} <= y_{2,7} + 1
x_{2,1} + x_{8,1} <= y_{2,8} + 1                                         x_{2,2} + x_{8,2} <= y_{2,8} + 1
x_{3,1} + x_{4,1} <= y_{3,4} + 1                                         x_{3,2} + x_{4,2} <= y_{3,4} + 1
x_{4,1} + x_{5,1} <= y_{4,5} + 1                                         x_{4,2} + x_{5,2} <= y_{4,5} + 1
x_{4,1} + x_{9,1} <= y_{4,9} + 1                                         x_{4,2} + x_{9,2} <= y_{4,9} + 1
x_{5,1} + x_{9,1} <= y_{5,9} + 1                                         x_{5,2} + x_{9,2} <= y_{5,9} + 1
x_{6,1} + x_{7,1} <= y_{6,7} + 1                                         x_{6,2} + x_{7,2} <= y_{6,7} + 1
x_{6,1} + x_{8,1} <= y_{6,8} + 1                                         x_{6,2} + x_{8,2} <= y_{6,8} + 1

x_{1,3} + x_{3,3} <= y_{1,3} + 1
x_{2,3} + x_{5,3} <= y_{2,5} + 1
x_{2,3} + x_{6,3} <= y_{2,6} + 1
x_{2,3} + x_{7,3} <= y_{2,7} + 1
x_{2,3} + x_{8,3} <= y_{2,8} + 1
x_{3,3} + x_{4,3} <= y_{3,4} + 1
x_{4,3} + x_{5,3} <= y_{4,5} + 1
x_{4,3} + x_{9,3} <= y_{4,9} + 1
x_{5,3} + x_{9,3} <= y_{5,9} + 1
x_{6,3} + x_{7,3} <= y_{6,7} + 1
x_{6,3} + x_{8,3} <= y_{6,8} + 1


/*Variables (5)*/
```

```
bin x_{1,1}
bin x_{1,2}
bin x_{1,3}
bin x_{2,1}
bin x_{2,2}
bin x_{2,3}
bin x_{3,1}
bin x_{3,2}
bin x_{3,3}
bin x_{4,1}
bin x_{4,2}
bin x_{4,3}
bin x_{5,1}
bin x_{5,2}
bin x_{5,3}
bin x_{6,1}
bin x_{6,2}
bin x_{6,3}
bin x_{7,1}
bin x_{7,2}
bin x_{7,3}
bin x_{8,1}
bin x_{8,2}
bin x_{8,3}
bin x_{9,1}
bin x_{9,2}
bin x_{9,3}

/*Variables (6)*/

bin y_{1,3}
bin y_{2,5}
bin y_{2,6}
bin y_{2,7}
bin y_{2,8}
bin y_{3,4}
bin y_{4,5}
bin y_{4,9}
bin y_{5,9}
bin y_{6,7}
bin y_{6,8}
bin y_{6,9}
```

Appendix B Example for Social Network and Grouping Effect

An example class social network snapshots from our case study—before and after the team project—with optimized team assignments (Class J, IME 305, Operations Research II) is presented below. The number of ties among the 30 students increased from 57 to 108, as illustrated in Figure B1. In Figure B2, the 32 newly created intra-group ties (left) which presumably stem from project interaction, and the 30 new ties which were established outside of the projects (right) are depicted.

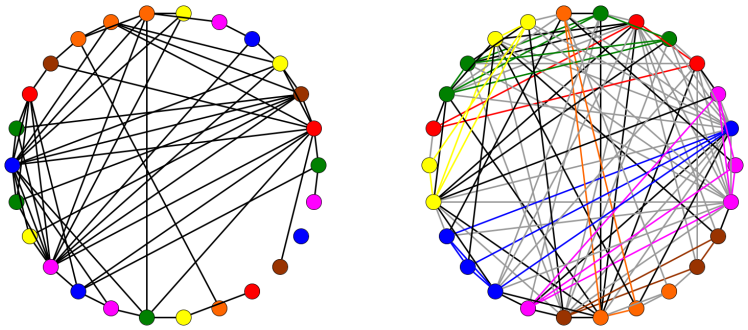


Fig. B1 The anonymized social network and the optimized grouping of 30 students in class J before (left) and after the term (right), indicated by 7 different group colors.

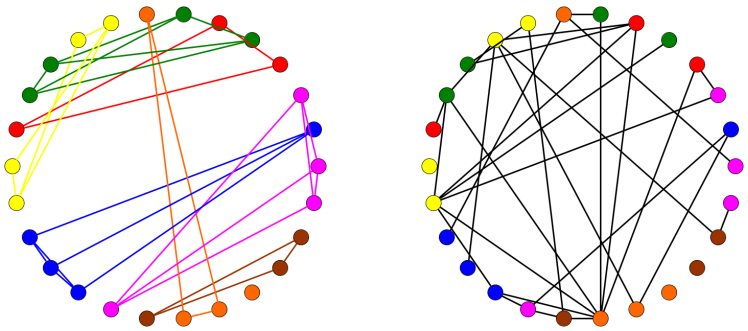


Fig. B2 The 32 new intra-group ties (left) and the 30 generic new inter-group ties (right) after the term, indicated by 7 different group colors.