

Sponsor Group Assignment at Pomona College as a Constraint Satisfaction Problem

Sarah Jundt and Shannon Lubetich
Pomona College

Abstract

Pairing college roommates is a laborious process. We confront this problem, as well as the problem of creating hall groups known as "sponsor groups," by formulating them as Constraint Satisfaction Problems. Students are prohibited from being paired together based on preferences submitted on housing forms. Our results return successful roommate and sponsor group pairings, with sponsor groups having more intragroup similarity than intergroup similarity.

Introduction

The first year of college can be a life-changing experience, and it is often greatly affected by the people students live with. At Pomona College, sponsor groups consisting of 10 to 20 freshmen living together are an essential aspect of first-year culture. Roommates and sponsor groups are paired manually at Pomona by a small number of Head Sponsors who have to sift through hundreds of applications. If roommate pairs and sponsor groups could be generated by a computer, less manual work would be required and less time would be spent.

We propose a quicker and more efficient approach to this pairing and grouping process. By formulating these groupings as Constraint Satisfaction Problems, we are able to produce successful assignments of students into roommate pairs and sponsor groups where students are more similar to others in their sponsor group than to those outside their group.

Related Works

SARAH TAKE IT AWAY!

Background

Sponsor groups are created at Pomona College by first considering individual students and matching each student with a roommate. Next, each roommate pair is matched with four to nine other pairs in order to form one of thirty 10-20 person sponsor groups. Criteria for groupings are flexible, but sponsor groups tend to consist of people who could be classified as "similar." Before coming to Pomona, first-years fill out

a housing form on which they provide qualitative information about their habits and quantitative information about the extent to which several different traits matters in their roommate or hall. Every single individual has a list of these quantitative preferences, which range from 1, indicating not important, to 10, indicating very important. These statements are as follows:

- It's important for me to have a roommate who...
 1. Is serious about studying and will make studying a priority of our room.
 2. Allows me to have visitors over as often as I'd like and doesn't mind having people over in our room.
 3. Will be my friend who confides in me and likes to do a lot of things together.
 4. Doesn't let school take over our entire lives, and knows how to have a good time.
 5. Shares responsibility for keeping our room neat.
 6. Respects my need for privacy and will allow me some time to myself.
 7. Respects my property and doesn't borrow my things without asking.
 8. Has similar sleep habits (i.e. windows opened/closed, absolute quiet, no light, etc.)
- It's important for me to live in an area where...
 9. There are people who share backgrounds and cultures similar to my own.
 10. People are aware of, sensitive to, and willing to discuss multicultural issues.

Statements 1 through 8 ask the student explicitly about qualities of their future roommate, whereas statements 9 and 10 refer primarily to the student's future sponsor group.

The Class of 2017 comprises of 414 students: 196 males and 218 females. Each student submitted a housing form and the forms from this class comprises our data set.

The problem of assigning students to sponsor groups is easily formulated as a constraint satisfaction problem. Constraint Satisfaction Problems can be solved using the Backtracking algorithm. Backtracking assigns variables to values in a depth-first manner. Backtracking considers one variable X at a time, assigns it to a value D in its domain, and then

adds the variable assignment $X = D$ to an overall assignment of all variables. Then, all constraints are checked to guarantee that this new assignment $X = D$ does not violate any constraints on the variables that have previously been assigned. If a constraint is violated, the assignment of $X = D$ is removed from the overall assignment and X is assigned to some other value D' in its domain. If no constraints are violated and a full assignment of all variables is reached, then a complete and successful assignment is returned.

Backtracking can be improved with the addition of intelligent ordering of domains, heuristics to choose the next variable to assign, and an inference to remove inconsistent values from the domains of variables. To create even groupings of variables among domain values, domains can be ordered such that those with fewer variables assigned to them are chosen first. The Minimum Remaining Values (MRV) heuristic assigns variables with fewer possible assignment values first. This ensures that less backtracking is necessary, speeding the algorithm. Inference can be used after an assignment to eliminate all possible values from other unassigned variables' domains, working in conjunction with MRV to speed the algorithm by minimizing the occurrence of assignments that violate constraints involving previously assigned variables.

System Description

The problem of placing students into roommate pairs and sponsor groups is configured as two separate constraint satisfaction problems. Students are first placed into roommate pairs and are then placed into sponsor groups in an attempt to emulate the human process for such sponsor group pairings. Each of these placements is performed using backtracking and each returns an assignment satisfying constraints designed to help create sponsor groups with students who are similar in essential ways.

All constraints focusing on the "preferences" of students refer to the students' numeric ranking of their preferences about the traits of their roommate or surroundings as specified on their housing forms.

These preferences are stored as a feature vector, along with the student's self-specified gender. Each student has a vector as follows:

$$[M/F, P_1, P_2, \dots, P_{10}] \text{ where } 1 \leq P_i \leq 10$$

The first entry in the vector refers to the gender of the student and is important because Pomona currently only pairs same-gender roommates. That is, someone who specifies male gender will be guaranteed a roommate who also specifies male gender. Because of this distinction, we run our roommate pairing algorithm twice, once on all students with female gender and once on all students with male gender. Gender affects sponsor group assignments as well, as Pomona attempts to balance the number of males and females in each sponsor group. The remaining values in the feature vector are numeric preferences from 1 to 10 ranking the importance of statements as presented and discussed in the previous section.

Creating Roommate Pairs

The constraints for this problem are based on maximizing the difference between the preferences of students within a roommate pair.

In formulating the problem of pairing roommates as a Constraint Satisfaction Problem, the variables to be assigned are defined to be unique ID numbers of every student. Each variable has the domain of the ID numbers of every other student, thus representing all potential roommates for a given student.

To assign students to roommate pairs, we initially modify the domains of all students such that the domain of student X contains only ID numbers of students who are compatible, according to our constraints, with student X . We then run backtracking on these pairs with the heuristic MRV. As students are assigned a roommate, both students in the newly-assigned roommate pair are removed from the domains of all other students.

Constraints are created based on the students' preferences. To determine whether students are compatible, we iterate through their feature vectors and compare the numerical values corresponding to their preferences. If any of these preferences differ by more than a constant amount C , these students cannot be roommates. C is defined to be the largest constant such that roommate pairs can be assigned by Backtracking. C is defined in this manner so that no specific preference is weighted more heavily and the similarity between roommates is as high as possible. For the Pomona College class of 2017, C is 4 for females and 5 for males.

Backtracking search is then applied to generate roommate pairs, with the ordering of the students to be assigned chosen intelligently using the Minimum Remaining Values heuristic in order to reduce the running-time of backtracking. To select a roommate assignment for each student, values in the domain are randomly ordered so that ID number does not affect the chances of any particular pairing. This allows us to return multiple different, yet still constraint satisfying, pairings over multiple runs. We then use forward-checking to remove the two roommates involved in this pairing from all other students' domains, because since they have been paired, they cannot be matched as roommates with anyone else. If this removal process leaves an unassigned student with an empty domain, we must backtrack and attempt to assign our current student to a different roommate that will result in pairing everyone successfully. Once a complete assignment is found of all students into roommate pairs, this assignment is returned as a success.

Creating Sponsor Groups

The constraints for this next step are based on maximization of similarity between the preferences of roommate pairs within a sponsor group. For this problem, we combine roommate pairings into a single identifier (the lower of the two students' ID numbers) with a single associated feature vector (where the entries are an M or F to represent the gender of the two students followed by the average of the numerical values in their feature vectors).

Pomona College creates 30 sponsor groups for first-year students. Thus, for this Constraint Satisfaction Problem, ev-

ery variable is a unique ID number that represents a roommate pair, and the domain for every variable are the numbers 0 through 29 representing the labels of sponsor groups. Roommate pairs are then assigned to sponsor groups.

Here, we incorporate constraints by building a constraint satisfaction graph. The preferences specified by students on their housing forms, as described in the previous section, do not necessarily apply as strongly to a student's hallmates as to their roommates. Some preferences are clearly more relevant to creating sponsor groups than others, so we rank some as more important than others.

The most important preferences are those which pertain directly to sponsor groups:

- There are people who share backgrounds and cultures similar to my own.
- People are aware of, sensitive to, and willing to discuss multicultural issues.

The second most highly-ranked preferences are those which discuss roommates but which strongly indicate some aspect of a sponsor group dynamic:

- Is serious about studying and will make studying a priority of our room.
- Will be my friend who confides in me and likes to do a lot of things together.
- Doesn't let school take over our entire lives, and knows how to have a good time.

The medium-priority preferences are those which discuss roommates but which may have some bearing on a sponsor group dynamic:

- Allows me to have visitors over as often as I'd like and doesn't mind having people over in our room.
- Respects my need for privacy and will allow me some time to myself.

The unlisted preferences are those of low priority, which we judge to have little or no bearing on a sponsor group dynamic.

The difference between higher-ranked preferences is required to be within a smaller constant difference than the requirement for lower priority preferences. In our constraint satisfaction graph, two roommate pairs are involved in a constraint of not being able to occupy the same sponsor group if any of their averaged preferences are different by the specified constant for that preference. For the Class of 2017, these constants are: 4 for the most important preferences, 4 for the highly-ranked preferences, 6 for the medium-priority preferences, and 7 for the low-priority preferences.

Backtracking search is then applied to generate sponsor groups, with the next roommate pair to be assigned chosen using the MRV heuristic. In order to select a sponsor group for each student pairing, we reorder the domain from smallest to largest sponsor group to attempt to assign a student to a sponsor group that is more empty first before assigning to a more populated group.

There are several constraints we incorporate along the way to yield valid sponsor groups. We enforce a minimum size of 10 students (5 roommate pairs) and a maximum size

of 20 students (10 roommate pairs), because these are the actual range of sizes of sponsor groups at Pomona College. Additionally, Pomona attempts to have a balanced number of males and females in sponsor groups, so we try to form groups with as close to a 1:1 ratio as possible. This ratio is accomplished by only assigning a roommate pair of gender M to a sponsor group if that sponsor group already does not have more pairs of M than F gender, and vice versa for assigning a roommate pair of gender F.

We also use forward-checking to determine if our newly assigned roommate pair X is involved in a constraint with any other roommate pair Y by checking if there is an edge between these two pairs, X and Y , in our constraint satisfaction graph. If this is the case, then the sponsor group that X was just assigned to is removed from the domain of Y , because we have a constraint that specifies that they cannot be in the same sponsor group.

If any constraints are not successfully met, either on sponsor group size or gender ratio or in the constraint satisfaction graph, the assignment is a failure. The assignment is removed as we backtrack and attempt to assign our roommate pair to a different sponsor group that will satisfy all constraints. Once a complete assignment is found of all roommate pairs into sponsor groups, this assignment is returned as a success.

Results

We were able to successfully produce roommate pairings and sponsor groups satisfying our specified constraints. Because we do not have data on how these students were actually paired and grouped, there is no way to test how "correct" our approach is as compared to how students were paired by the Pomona College housing committee. For roommate pairs, we can examine if we found a successful matching and how strict or loose our constraints have to be in order to achieve success. For sponsor groups, our main goal is to create groupings that have preferences more similar to each other than to the overall average of all students outside of the group.

Roommate Pairs

We are able to assign students to roommate pairs based on gender and a constant maximum difference in preferences. Interestingly, we found that the smallest constant difference C for which students can be successfully paired into roommates differs depending on genders for the Class of 2017. For females, this occurs when $C = 4$; for males, this occurs when $C = 5$. This means that females can be subject to more strict constraints so that none of the preferences between roommates differ by more than 4 points and still find successful pairings, whereas males require that none of the preferences differ by more than 4 points.

Sponsor Groups

For sponsor groups, our main goal is to create groupings in which the roommate pairs have similar preferences to one another. We evaluate our CSP results based on the similarity of students to the average student within their sponsor group

as compared to the average student in the class. This involves calculating the difference of a roommate pair's preferences and their sponsor group's average preferences and comparing this to the difference of their preferences and the preferences of all the students. This method of evaluation encapsulates the idea that, to be successful, sponsor groups should contain roommate pairs who all have something in common. Students within a sponsor groups should thus be more similar to other students within that sponsor group than they are to the student body population as a whole.

For each student, we determine the difference between the student's preferences and the average preferences of the student's sponsor group (the intragroup difference). We then determine the difference between the student's preferences and the average preferences of all students (the intergroup difference). We find these differences for each of the ten preferences of the student.

We then compute the average of these differences over the entire student population.

From the Class of 2017, we successfully generated 67 roommate and sponsor group assignments and found the average intragroup difference and intergroup differences across all of these assignments, as seen in Table 1. The differences for preferences were averaged based on their importance: that is, all of the most important features' inter and intra-group differences were averaged together as the differences for the "most important" preferences.

Preference Importance	Intragroup Difference	Intergroup Difference
Most	0.84	1.41
High	1.01	1.31
Mid	1.12	1.25
Low	1.22	1.36

Table 1: Average Preference Differences

All intra-group differences are lower than the inter-group differences, suggesting that our sponsor group assignments successfully place students in sponsor groups with similar students. The most important features have the most significantly decreased intragroup difference as compared to the intergroup differences, suggesting that our CSP correctly emphasizes making sponsor groups with higher similarity for the most important preferences.

Conclusion

We are able to successfully place Pomona College first-years into roommate pairs and sponsor groups by formulating two different constraint satisfaction problems and solving each using Backtracking. Roommates are assigned such that none of the students' preferences, as indicated on their housing forms with a number from 1-10, differ by more than a constant C . Constraints for sponsor group assignments are designed based on a hierarchical system of the importance of different preferences to the success of a sponsor group. Higher constraints are placed on more important

preferences.

Over 67 sponsor group assignments, the average intra-group difference was lower than the average intergroup difference for all preferences. Higher-priority preferences, which have more significant constraints assigned to them, have a larger decrease in the intragroup difference. Although it is difficult to evaluate assignments of students to sponsor groups, these results suggest that the constraint satisfaction formulation and solution we use produces sponsor groups with people who are most similar in the most crucial ways.

References

S. Gravel, V. Elser. "Divide and concur: A general approach to constraint satisfaction." *Physical Review E* 78, 036706 (2008). American Physical Society.

A.K. Mackworth, E.C. Freuder. "The complexity of some polynomial network consistency algorithms for constraint satisfaction problems." *Artificial Intelligence*, 25 (1985), pp. 65-74