A Greedy Algorithm for Solving Meeting Mixing Problems

Adrian Corduneanu
Cyrus C. Hsia
Ryan O'Donnell
hsia@math.toronto.edu
{ a.corduneanu, ryan.odonnell } @utoronto.ca
University of Toronto
Toronto, Ontario M5S 3G3
Canada

Advisor: Nicholas A. Derzko

Introduction

We consider the problem of how to best arrange a large number of people into small discussion groups so that the groups are well mixed. It is important to have well-mixed groups because any meeting runs the risk of being controlled or directed by a dominant personality. Thus, we wish to ensure schedules give different mixes of people for each group.

This problem relates directly to the case of An Tostal Corporation. The company wants to place its board members in small groups within each session so that the board members are well mixed throughout the day. The company's schedule must also satisfy other constraints:

- At the morning session, there is a senior officer assigned to each group, and no board member is to be placed with any senior officer more than once.
- A percentage of the board members are in-house members, and the company wishes that no group should have a disproportionate number of in-house members.

To solve the problem, we first develop a scoring system for schedules. There must be a schedule or schedules that achieve the best possible score; but the total number of schedules is astronomical, so we cannot check all of them. Consequently, the problem reduces to finding as good a schedule as possible in

The UMAP Journal 18 (3) (1997) 331–342.

a short period of time. To find good schedules, we wrote a computer program in C that uses a greedy algorithm. The algorithm places the board members into the schedule one by one, at each step making the schedule placement that gives the best possible score. Our algorithm also uses a switching procedure to improve on the greedy placement at each step.

Assumptions

- The day is split into a number of sections, and each section has a fixed number of sessions and a fixed number of groups per session. We assume that the senior officer constraint applies in a fixed subset of the sections. For these sections, we assume that the number of senior officers is the same as the number of discussion groups, and that there are at least as many discussion groups in an officer-led section as there are meeting sessions (not assuming this makes the problem unsolvable).
- The secretary in charge of scheduling can change the parameters of our computer program to reflect the discussion day at hand.
- It is not vitally important to get the perfect mix of group members; close-to-optimal solutions are acceptable.

Choosing an Incidence Scoring System

We need to develop some scoring system to provide a total ordering on the set of all possible configurations; this way, there will be some best configurations, and our goal is a configuration with a score that is as good as possible. Although there are a number of side constraints imposed by the An Tostal Corporation (relating to the in-house members and the senior officers), we assume that our computer algorithm will take care of these and that the main criterion—mixing well—is what our score should measure.

The obvious first choice for an incidence score is just the *incidence sum*. However, this scoring system turns out to have a major flaw: The minimal incidence sum is easy to achieve, but in many cases the incidence elements vary widely. An incidence matrix composed mostly of 1s and 2s is better than one with lots of 0s but also lots of 5s, 6s, and 7s; we don't want to minimize just overall incidence, we want to try to minimize everyone's incidence, which presumably involves keeping everyone's incidence about the same. At the other extreme, the variance or standard deviation of the incidence elements in the lower incidence matrix could be used as the scoring system. However, using just variance does nothing to keep the overall incidence low; it merely keeps all the incidence elements close. As a compromise, the scoring system

Table 1.

Definitions and notations (in parentheses, values for the An Tostal problem instance).

Day	The time block that encompasses all the meetings for a given problem.
Session	A block of time within a day, to which meetings are devoted. Each person goes to only one meeting during a session.
Groups	Actual meetings attended by sets of board members.
Section	A set of sessions with the same number of groups per session.
Configuration	An assignment of people into groups so that each person is in one and exactly one group per session (sometimes referred to as a *schedule*).
S	The number of sections. (2)
$N_1, \ldots, N_S$	The number of sessions per section. (3, 4)
$G_1, \ldots, G_S$	The number of groups per session. (6, 4)
B	The number of people to be scheduled (board members). (29)
I	The number of board members who are in-house members. We number the board members $1, \ldots, B$, and $1, \ldots, I$ are the in-house members. (9)
O	The number O such that the senior officer constraint applies to sections $1, \ldots, O$. For simplicity, we number the groups in any officer-led section $1, \ldots, G$ and regard each officer as leading the same group number for the entire section. Thus, a configuration satisfies the officer constraint if and only if every board member is in a different group number (within each officer-led section). (1)
Incidence	The incidence of person X with person Y is the number of times that X is in the same group as Y within a given configuration.
Incidence matrix	The incidence matrix for a given configuration is the $B \times B$ matrix $IM = (a_{ij})$, where a_{ij} is the incidence of person i and person j. Elements of this matrix are *incidence elements*.
Lower incidence triangle	Since any incidence matrix has zeros down its main diagonal and is symmetric, we often consider just the lower triangle minus the diagonal, i.e., a_{ij} for $j < i$.
Incidence score	A rating given to the configuration that reflects how well mixed it is, a function of the configuration's lower incidence triangle. Lower scores are better.
Incidence sum	The sum of the incidence elements in the configuration's lower incidence matrix, which gives the total number of common memberships over all groups.
Optimal configuration	A configuration with the lowest possible incidence score.

that we use is a sum of squares:

$$\text{the incidence score of } IM = \sum_{1 \le i \le j \le B} a_{ij}^2.$$

This scoring system requires the incidence elements to be close to one other. At the same time, if we can bound this score, we can also automatically bound the incidence sum using the Cauchy-Schwartz inequality (see **Theorem 3**).

Theoretical Lower Bounds for Scores and Sums

We wish to find a theoretical lower bound for all possible incidence sums for a given day. We first find a closed form for this total.

Theorem 1. *For any configuration, the incidence sum is*

$$\sum_{\textit{all groups } G} \binom{n_G}{2},$$

where n_G is the number of people in the group G.

Proof: The sum of the incidence elements in the lower incidence matrix of a configuration is the same as the sum over all possible pairs in all the groups. For each group with n_G people, there are $\binom{n_G}{2}$ pairs. The total number of pairs for all groups is the indicated sum. □

This closed form allows us to calculate a lower bound in terms of the problem parameters.

Theorem 2. *For a given day, the incidence sum of configurations is bounded below by*

$$\frac{1}{2} B^2 \sum_{i=1}^{S} \frac{N_i}{G_i} - \frac{1}{2} B \sum_{i=1}^{S} N_i.$$

(For An Tostal, this value is 529.25.)

Proof: From the previous theorem, we have that a configuration's incidence sum is

$$\sum_{\text{all groups } G} \binom{n_G}{2} = \sum_{\text{all groups } G} \frac{n_G^2 - n_G}{2} = \frac{1}{2} \sum_{\text{all groups } G} n_G^2 - \frac{1}{2} B \sum_{i=1}^{S} N_i.$$

To get a lower bound for this value, we apply the Cauchy-Schwartz inequality [Plank and Williams 1992, 46] to each section, since the number of groups may differ across sections. We have

$$\left(\sum 1^2\right)\left(\sum n_G^2\right) \ge \left(\sum n_G\right)^2 = B^2 N_i^2,$$

where the sums are each over all groups G of section i. This yields

$$(G_i N_i)\left(\sum n_G^2\right) \geq B^2 N_i^2,$$
$$\sum n_G^2 \geq \frac{N_i}{G_i} B^2,$$

with equality holding iff all the n_G are equal (this may not be achievable in our discrete case, as these numbers must be integers). So taking the sum of all these inequalities over all the possible sections, we have

$$\sum_{\text{all groups } G} n_G^2 \geq B^2 \sum_{i=1}^{S} \frac{N_i}{G_i}.$$

Thus, we reach the conclusion claimed, with equality holding when in each section the groups have the same number of people. □

This minimal value cannot be achieved when the number of people cannot be divided evenly among the groups in the sessions. However, by distributing the people as evenly as possible among the groups, the sum will be as small as possible.

Fact. *The minimal sum for An Tostal Corporation is 532.*

Proof: Distributing the 29 people as evenly as possible among the 6 groups in each morning session we get groups of 4, 5, 5, 5, 5, and 5 (in some order); for the 4 groups in the afternoon sessions, we get groups of 8, 7, 7, and 7 (in some order). The resulting incidence sum is then

$$3\left[\binom{4}{2} + 5\binom{5}{2}\right] + 4\left[\binom{8}{2} + 3\binom{7}{2}\right] = 532. \qquad \square$$

Theorem 3. *For a given day, a configuration's incidence score is bounded below by*

$$\frac{B}{2(B-1)}\left(B\sum_{i=1}^{S}\frac{N_i}{G_i} - \sum_{i=1}^{S} N_i\right)^2.$$

For An Tostal, this value is about 689.9.

Proof: By using the Cauchy-Schwartz inequality and **Theorem 2**, we have that

$$\left(\sum_{1\leq i\leq j\leq B} a_{ij}^2\right)(1+\cdots+1) \geq \left(\sum_{1\leq i\leq j\leq B} a_{ij}\right)^2$$
$$\geq \left(\frac{1}{2}B^2\sum_{i=1}^{S}\frac{N_i}{G_i} - \frac{1}{2}B\sum_{i=1}^{S} N_i\right)^2,$$

so

$$\left(\sum_{1\le i\le j\le B} a_{ij}^2\right)\binom{B}{2} \ge \frac{1}{4}B^2\left(B\sum_{i=1}^{S}\frac{N_i}{G_i}-\sum_{i=1}^{S}N_i\right)^2$$

and

$$\left(\sum_{1\le i\le j\le B} a_{ij}^2\right) \ge \frac{B}{2(B-1)}\left(B\sum_{i=1}^{S}\frac{N_i}{G_i}-\sum_{i=1}^{S}N_i\right)^2. \qquad \square$$

The first line demonstrates that the incidence score can be used to bound the incidens sum from above. Hence, incidence scores are also bounded below and incidence sums are squeezed in between. So, if we achieve a small incidence score, we also achieve a small incidence sum.

In our case, the minimum sum of squares cannot be achieved, for similar reasons that the minimum sum could not be achieved. We can still calculate the minimum possible sum of squares by distributing the B people as evenly as possible among the groups of a session. By doing so, we can find a closed form that may be achieved.

Theorem 4. *The minimum incidence score possible given a fixed minimum incidence sum MS and a fixed number of people B is*

$$(2d+1)MS - d(d+1)\binom{B}{2},$$

where

$$d = \left\lfloor \frac{MS}{\binom{B}{2}} \right\rfloor.$$

For An Tostal, this minimum incidence score is 784.

Proof: We wish to make the incidence elements as close as possible, which means we must have a number of them, say a, which have value d, and the rest, say b, which have value $(d+1)$. The total $(a+b)$ must be the total number of pairs of people, $\binom{B}{2}$. The minimum sum of the incidence elements, MS, is then $da+(d+1)b$. Solving for a and b in the two equations, we get $b = MS-\binom{B}{2}$ and $a = (d+1)\binom{B}{2} - MS$. This gives as our incidence score the value

$$ad^2 + b(d+1)^2 = (2d+1)MS - d(d+1)\binom{B}{2},$$

which upon substitution for a and b, and some manipulation, yields the desired result. $\square$

Fact. *Any incidence matrix for the An Tostal case must contain at least some values greater than 1 in the optimal distribution.*

Proof: By the previous **Fact**, the optimal distribution has sum of incidence elements 532. The greatest number of distinct incidences is $\binom{29}{2} = 406$. Thus, at least one person will have an incidence of at least 2. $\square$

This fact means that in our results for the An Tostal problem, we expect to see at least 2s and 1s in any incidence matrix.

Theoretical Bounds on Computer Run Time

We show that the total number of possibilities is exponential in B.

Theorem 5. *The total number of possible configurations is*

$$\left(\prod_{i=1}^{S} G_i^{N_i}\right)^B .$$

For An Tostal, this is $(6^3 4^4)^2 9 > 3 \times 10^{137}$.

Proof: For each of the B people, this person may be in any of the G_i groups for any of the N_i sessions for all possible sections i ranging from 1 to S. $\square$

We can do better than this by considering only cases in which every group in a session has at least one member. However, even if we have an even distribution (e.g., 4, 5, 5, 5, 5, 5 / 8, 7, 7, 7 for An Tostal), we have a corresponding bound.

Theorem 6. *The total number of possible configurations with even distribution is bounded below by*

$$\prod_{i=1}^{S} \left[\frac{B!}{\left(\left\lceil \frac{B}{G_i} \right\rceil !\right)^{G_i}}\right]^{N_i} .$$

For An Tostal, this is

$$\left[\frac{29!}{(5!)^6}\right]^3 \left[\frac{29!}{(8!)^4}\right]^4 > 3 \times 10^{105}.$$

Proof: In the even distribution, each group in a session of section i has at most $\lceil B/G_i \rceil$ people. Given a group of $n_1 + \cdots + n_k$ numbers with n_1 1s, ... , and n_k

ks, the total number of ways of rearranging them is given by the multinomial coefficient

$$\binom{n_1 + \cdots + n_k}{n_1, \ldots, n_k}.$$

Therefore, in each session, the total number of possible group placements is bounded below by

$$\frac{B!}{\left(\left\lceil \dfrac{B}{G_i} \right\rceil !\right)^{G_i}}.$$

Thus, the total number of possible configurations with even distribution is bounded below by the claimed quantity. $\square$

How We Implement a Greedy Algorithm

The algorithm that we use to generate configurations has two main ingredients, which we call *Greedy Placement* and *Switching*. The principal ingredient is Greedy Placement; Switching is merely a tweak to give slightly better scores.

Greedy Placement proceeds through the B board members one by one; on the ith iteration, it finalizes the placement of person i in all the groups. The placements are "greedy," that is, we "[make] the choice that looks best at the moment" [Cormen et al. 1996, 329]. At each iteration, the algorithm looks at every possible way to place the person subject to the senior officer restriction and chooses the one that leads to the best possible incidence score. The first I board members are the in-house members, so Greedy Placement distributes the in-house members as evenly as it distributes all the board members. Thus, if the algorithm is successful overall, it also satisfies the criterion that no group should have a disproportionate number of in-house members.

Finally, we describe the Switching add-on. After every iteration of Greedy Placement, we do Switching. Switching looks at our current configuration and tries to find a case where, within a given session, it is possible to switch the placement of two board members and consequently get a better score. If there is such a case, we make the switch and reevaluate the configuration to see if there is another useful switch. We continue making switches until we are in a state from which any switch would be detrimental; then we move on to the next iteration of Greedy Placement. In this way, we get a score that is a local minimum after every iteration. Although many switches could take place at each iteration, causing an unpredictable increase in running time, we found that the average number of switches per iteration was about 1.

Switching introduces two complexities.

- We must make sure not to make any switches that would cause the senior officer restriction to be violated.

- If we allow all other switches, we run the risk of switching in-house members with non-in-house members, which destroys our argument that the number of in-house members per group will not be disproportionately high. To counter this, we restrict Switching to switch in-house members only with in-house members and non-in-house members with non-in-house members.

Justification for Our Algorithm

- Trying every possible configuration is impossible, as **Theorems 4–5** show.
- Our algorithm is fast. When the An Tostal day was run on an SGI Challenge, our algorithm took less than 45 sec. Even when it was run on a Pentium 90 MHz with 32 MB RAM, it took only about 7 min. Thus, if any board members don't show up, or extra board members do, the secretary can surely calculate a new assignment in under an hour.
- Our algorithm is flexible. All parameters to the problem (i.e., every variable defined near the beginning of this paper) can be altered simply in a text file.
- Our algorithm always gives a configuration that satisfies the senior officer criterion. Since the configuration satisfies the main criterion of mixing well (see below), it also does not have a disproportionate number of in-house members in any group.
- Our algorithm produces configurations that satisfy the main criterion (mixing well). To demonstrate this, we created some different days (parameter setups). We ran our algorithm for these cases and compared the scores produced by the algorithm with calculated lower bounds from **Theorems 2** and **3** (see the **Results** section below).

Results

To determine the effectiveness of our algorithm, we ran a number of test days. In all cases, the incidence sum found by the algorithm is very close to the minimal theoretical sum; this means that the total number of common memberships is essentially minimized. Also, in each case the incidence score is quite small. In cases similar to the An Tostal problem, the scores found never exceeded the theoretical bound by more than 14%. Even in a huge test case, the score exceeded the theoretical bound by only 29%. This shows that another important constraint is achieved by our algorithm: Each board member meets others a similar number of times.

We present in **Table 2** the results of our greedy algorithm as run on the An Tostal day. **Table 3** shows the distribution of incidence elements and gives other data.

Table 2.

Recommendation to An Tostal.

The in-house members are numbers 1 through 9.

	Morning Section		
	Group 1 Officer 1	Group 2 Officer 2	Group 3 Officer 3
Session 1	1 4 14 22 25	2 9 12 21 28	3 10 15 23 27
Session 2	2 7 18 19 27	1 8 13 23 25	4 9 17 20 29
Session 3	3 9 17 23 26	4 10 19 20 24	1 11 12 18
	Group 4 Officer 4	Group 5 Officer 5	Group 6 Officer 6
Session 1	5 6 18 20 26	7 11 17 24 29	8 13 16 19
Session 2	3 11 16 22 28	5 10 14 21 26	6 12 15 24
Session 3	2 13 14 15 29	6 8 16 22 27	5 7 21 25 28

	Afternoon Section			
	Group 1	Group 2	Group 3	Group 4
Session 4	1 5 9 16 23 24 27	2 6 10 13 17 22 26 28	3 3 7 12 14 19 20 25	4 8 11 15 18 21 29
Session 5	1 3 6 17 19 21 27 29	2 5 11 15 20 22	8 9 10 14 18 24 25 28	4 7 12 13 16 23 26
Session 6	1 2 10 15 16 17 25	7 8 9 20 21 26 27	3 5 12 13 18 22 24 29	4 6 11 14 19 23 28
Session 7	1 7 15 19 22 26 28 29	2 3 4 16 18 21 24	6 9 10 11 13 25 27	5 8 12 14 17 20 23

Table 3.

Distribution of incidence elements.

Incidence	0	1	2	3	4+
Number of incidence elements	33	226	134	13	0

Mean: 1.39	Standard Deviation: 0.67
Incidence Score: 879	Incidence Sum: 533
Incidence Score Lower Bound: 784	Incidence Sum Lower Bound: 532

Note that the officer constraint is indeed satisfied, the in-house members are in as even a distribution as possible, and that the resultant score was very close to its theoretical lower bound.

Limitations of Our Model

- Our algorithm does not guarantee an optimal incidence score. We have found cases where an optimal solution is known but is not found by our algorithm.
- Our greedy algorithm with switching may take too much time for very large parameter sizes.
- Our algorithm also does not guarantee a minimal incidence sum. On the other hand, the incidence sums we got in our test were very close and should really be good enough.
- We were able to provide only theoretical lower bounds on incidence scores; these lower bounds may not be achievable.
- Our algorithm is not easy to change. A secretary could not easily change it to use a different incidence scoring system or to consider additional constraints.

Conclusion

We generalized the problem to allow for any number of board members and in-house members; any values for sections, sessions, and groups; and any number of officer-led sessions. We defined a scoring system to evaluate possible configurations that successfully encapsulated the well-mixing criterion. Following this, we determined theoretical lower bounds on scores in terms of the problem parameters. We created a computer program that uses a modified greedy algorithm to come up with good schedules according to our scoring system; the program also took care of the other An Tostal constraints. We used this algorithm to provide the An Tostal Corporation with a well-mixed schedule for their original problem. Finally, we ran tests to verify that our algorithm came up with schedule scores close to the theoretical lower bounds. We believe that the results more than adequately achieved the criteria specified by the original problem, and that our algorithm is a valuable tool for use in scheduling similar planning days.

References

Cormen, Thomas H., et al. 1996. *Introduction to Algorithms*. Cambridge: MIT Press.

Plank, A.W., and N.H. Williams. 1992. *Mathematical Toolchest*. Canberra: Australian International Centre for Mathematics Enrichment.