Dickinson College
Dickinson Scholar

Honors Theses By Year | Honors Theses

5-22-2011

Mathematical Techniques for Assigning First-Year Seminars

Thanh Thien To
Dickinson College

Follow this and additional works at: http://scholar.dickinson.edu/student_honors

 Part of the Mathematics Commons

Recommended Citation
To, Thanh Thien, "Mathematical Techniques for Assigning First-Year Seminars" (2011). *Dickinson College Honors Theses.* Paper 117.

This Honors Thesis is brought to you for free and open access by Dickinson Scholar. It has been accepted for inclusion by an authorized administrator. For more information, please contact scholar@dickinson.edu.

Mathematical Techniques for Assigning First-Year Seminars

By

Thanh T. To

Submitted in partial fulfillment of Honors Requirements

for the Department of Mathematics and Computer Science

Dickinson College, 2010-2011

Professor Dick Forrester, Supervisor

Professor Jeffrey Forrester, Reader

Professor Barry Tesman, Reader

17 May 2011

The Department of Mathematics and Computer Science at Dickinson College hereby accepts this senior honors thesis by Thanh T. To, and awards departmental honors in Mathematics.

Richard Forrester (Advisor)	Date
Jeffrey Forrester (Committee Member)	Date
Barry Tesman (Committee Member)	Date
Tim Wahls (Department Chair)	Date

Department of Mathematics and Computer Science
Dickinson College

May 2011

Abstract

Mathematical Techniques for Assigning First-Year Seminars

by
Thanh T. To

Every first-year student at Dickinson College is required to take a first-year seminar. The summer before the students arrive they each select six seminars among all the seminar choices. Each student is then assigned to a seminar from their list, if possible. Currently, this process is performed manually, which is tedious and time consuming. This research is concerned with utilizing mathematical techniques to assign first-year students to seminars. Specifically, we develop a technique that not only assigns students to seminars, but also seeks to balance the gender ratio and international student ratio of the students in the classes. In addition, we use simulation to study how the number of seminars each student chooses affects our ability to make an assignment.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof. Dick Forrester for his expertise in the field of operations research and his extraordinary guidance and incessant support throughout this research, without which I could not have accomplished it. Second, I want to thank Prof. Barry Tesman and Prof. Jeff Forrester for serving on my committee and for providing their feedback and comments on the progress of the research. Next, I would like to thank Prof. Bob Winston and Prof. Shalom Staub for their insights into how seminar assignment is currently performed at Dickinson College. In addition, I also want to thank Mr. Terry Beard in LIS for providing us with the First-Year Seminar data. Last, I would like to thank Prof. Dave Richeson for encouraging me to be a math major as well as to pursue honors research in mathematics.

Table Of Contents

Chapter 1

INTRODUCTION & BACKGROUND

1.1. Introduction

The First-Year Seminar (FYS) program at Dickinson College contributes significantly to the experience of first-year students. While each seminar investigates a specific or general topic chosen by a faculty member, the seminars all focus on five core goals: "critically analyze information and ideas; examine issues from multiple perspectives; discuss, debate and defend ideas, including their own views, with clarity and reason; develop discernment, facility and ethical responsibility in using information; and create clear academic writing." ("First-year seminar") Each first-year student is required to take a first-year seminar. The summer before they arrive, incoming students have to provide a list of six seminars from the approximately 42 available, which is in no order, that they are interested in taking. Each student is then assigned to a seminar on their list if possible.

Currently, the assignment of first-year seminars at Dickinson is, for the most part, a manual task, which is tedious and time-consuming. More importantly, it does not always guarantee student or faculty satisfaction with all criteria such as gender balancing. This research is concerned with utilizing mathematical techniques to assign first-year students to seminars. Specifically, we develop a technique that not only achieves an assignment of students to seminars, but also seeks to balance the gender and number of international students in the classes. In addition, we use simulation to study how the number of seminars each student chooses affects our ability to make an assignment.

1.2. Operations Research and Mathematical Modeling

Operations Research (OR), also known as decision science, is concerned with utilizing scientific techniques to make decisions. Some applications of OR include goods distribution management (minimizing the cost of distributing goods), airlines scheduling, and truck routing. OR was developed during World War II as British military leaders realized the need for better management of radar deployment, as well as convoy and antisubmarine operations (Winston, 1994). That knowledge was then applied outside the military and became increasingly popular in industry, inspiring many other scientists to conduct research relevant to the field. As a result, OR is now widely implemented in business, industry, and government agencies.

The typical OR modeling process consists of the following steps: First, we need to define the problem of interest. Then, we collect all relevant data to formulate a mathematical model to represent the problem. The next step is to utilize an algorithm to solve the model. Then, the model is tested and refined until we are satisfied with the result. The results of the research are then presented to the "clients." Finally, we implement the model and evaluate any feedback resulting from this implementation.

Example 1.2.1

To illustrate the mathematical modeling step of an OR problem, let us consider the following example.

> The BW Bakery produces and sells two types of chocolate: white and dark chocolate. Each piece of dark chocolate requires 1 unit of sugar and 5 units of cocoa whereas each piece of white chocolate requires 5 units of sugar and 4 units of cocoa. The bakery gains a profit of $3 and $4 for selling one piece of dark and white chocolate, respectively. They have 22 units of sugar and 35 units of cocoa available

for production. Determine the combination of products they should make to maximize their profit.

Our objective here is to maximize the profit from selling dark and white chocolate. First of all, we need to define the *decision variables*, which are symbols that represent the quantities of interest. In this case, they are the number of pieces of dark and white chocolates. Thus, let x_1 and x_2 be the quantities of dark and white chocolates we should produce, respectively. It follows that our *objective function*, which is the function we need to maximize/minimize to achieve the goal stated in the problem, would be:

$$\text{Maximize: } z = 3x_1 + 4x_2 \tag{1.1}$$

Note that 3 and 4 are the corresponding profit we get from selling one unit of x_1 and one unit of x_2. The values of x_1 and x_2 are limited by the availability of sugar and cocoa. These *constraints*, which are the inequalities or equalities that represent the restrictions on the variables, can be characterized by the following system of inequalities:

$$1x_1 + 5x_2 \leq 22 \tag{1.2}$$

$$5x_1 + 4x_2 \leq 35 \tag{1.3}$$

$$x_1 \geq 0 \tag{1.4}$$

$$x_2 \geq 0 \tag{1.5}$$

It is clear that the objective function and all of the constraint inequalities are linear, and so we say that the objective function (1.1) along with its constraints (1.2) – (1.5) forms a *linear programming* (LP) *problem*. Our goal is to find values of x_1 and x_2 that maximize the objective function (1.1) subject to the constraints (1.2) – (1.5). There are various ways to solve this problem, but the most common approach is the simplex method.

The simplex method, which was developed by George Dantzig in 1947, is listed as one of the top ten algorithms of the twentieth century by the journal *Siam News* (Cipra,

2000). The idea behind this method is that if an optimal solution of a linear program exists, it will occur at a vertex, or corner point, of the *feasible region* (set of feasible points). The algorithm travels along the edge of the feasible region, from vertex to vertex, until it reaches the optimal solution. The simplex algorithm can solve linear programs with thousands of constraints and variables (Winston, 1994).

For this particular problem, we utilized the Excel Solver to find the optimal solution, which is $x_1 = 4$, $x_2 = 3$. In other words, to maximize profit, we should produce 4 pieces of dark chocolate and 3 pieces of white chocolate.

1.3. Network Flow Models

In this section we introduce a specific type of model called a *network flow program*. A network is a set of points and line segments in which certain pairs of points are connected by the line segments. The points are called *nodes* (or *vertices*) and the line segments are called *arcs* (or *edges*). The arcs typically represent possible movement from one node to another. Thus, they may have a "flow" of some type through them, e.g., the volume of water transmitted through a certain pipe. If the flow through an arc is restricted to only one direction, the arc is defined as a *directed arc*. Graphically, a directed arc will have an arrowhead indicating the direction of the flow. Similarly, an *undirected arc* can have flow go in both directions and we can present it by putting arrowheads at both of its ends or we can simply put no arrowheads at all.

A network flow model is a *directed graph*, which means it only consists of directed arcs, whose arcs receive flow restricted by their *capacity*. The capacity of an arc determines the maximum flow through it. A flow network has three types of nodes: source, sink and

transshipment nodes. A *source node* (or supply node) has the property that the flow out of the node exceeds the flow into the node. In contrast, a *sink node* (or demand node) receives more flow into it than flow out of it. A *transshipment node* receives an equal number of incoming and outgoing flow. The maximum flow problem, which is often referred to as a max-flow problem, is concerned with determining the maximum amount of flow that can be sent from the source node to the sink node.

We illustrate these concepts by considering an example from the textbook "Introduction to Mathematical Programming," by Winston and Venkataramanan (Winston, & Venkataramanan, 2003).

Example 1.3.1.

Fly-by-Night Airlines must determine how many connecting flights daily can be arranged between Juneau, Alaska, and Dallas, Texas. Connecting flights must stop in Seattle and then stop in Los Angeles or Denver. Because of limited landing space, Fly-by-Night is limited to making the number of daily flights between pairs of cities as shown in Table 1.1. We wish to set up a max-flow problem whose solution will tell the airline how to maximize the number of connecting flights daily from Juneau to Dallas.

Table 1.1: Arc Capacities for Fly-by-Night Airlines

Cities	Maximum Number of Daily Flights
Juneau – Seattle (J, S)	3
Seattle – L.A. (S, L)	2
Seattle – Denver (S, De)	3
L.A. – Dallas (L, D)	1
Denver – Dallas (De, D)	2

The above problem can be illustrated as the following network:

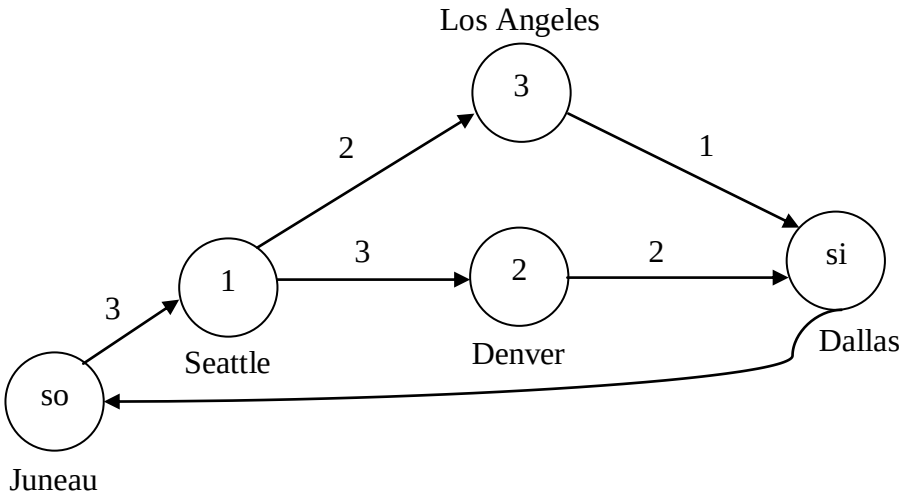


Figure 1.1: Network for Fly-by-Night Airlines

In Figure 1.1, Juneau is the source and Dallas is the sink since all flights "flow out of" Juneau and no flight "flows into" Juneau. Similarly, all flights "flow into" Dallas but no flight "flows out of" Dallas. There is no direct flight from Juneau to Dallas. Hence, the flow between these two nodes is zero. Any flight from Juneau to Dallas must transit at the other 3 airports (Seattle, Los Angeles, and Denver) on the graph. Therefore, Seattle (node 1), Denver (node 2), and Los Angeles (node 3) are transshipment nodes since flights only stop in those places temporarily. The number on each arc indicates the arc capacity in the direction of the arrowhead. For instance, the maximum number of daily flights from Denver to Dallas is 2.

We can use Figure 1.1 to maximize the number of connecting flights daily from Juneau to Dallas by finding the maximum amount of flow that is possible from the source node (Juneau) to the sink node (Dallas). We also mention that according to the well-known *Integral Flow Theorem* (also known as flow integrality theorem, or integrality theorem), if each edge in a flow network has an integer capacity, then there exists a maximum flow such that all the flows are integers. This theorem ensures, for example, that we will not have a

situation where the optimal solution would be to have 1.5 flights from Seattle to Los Angeles.

One of the most famous techniques for solving the max-flow problem is the *augmenting path algorithm*, which was developed by Ford and Fulkerson in 1956 (Ford and Fulkerson, 1956). A detailed explanation of this algorithm can be found in Appendix A.

Interestingly, we can formulate the max-flow problem as a linear programming problem. First we need to define x_{ij} = number of daily flights from node *i* to node *j*. Given the information from Table 1.1, we obtain the following values:

$x_{so1} = 3$; $x_{12} = 3$; $x_{13} = 2$; $x_{2si} = 2$; and $x_{3si} = 1$.

Note that a flow is feasible only if it is nonnegative and its value through each arc is less than the arc capacity; moreover, for any transshipment node *i*, the flow out of *i* must be equal to the flow into *i*. Let *v* be the number of daily flights from Juneau to Dallas. Our goal is to maximize *v*, and subject to the following constraints:

$$\left.\begin{array}{l} 0 \le x_{so1} \le 3 \\ 0 \le x_{12} \le 3 \\ 0 \le x_{13} \le 2 \\ 0 \le x_{2si} \le 2 \\ 0 \le x_{3si} \le 1 \end{array}\right\} \text{Arc Capacity Constraints}$$

$x_{so1} = x_{13} + x_{12}$ (Node *1* flow constraint)

$x_{13} = x_{3si}$ (Node *3* flow constraint)

$x_{12} = x_{2si}$ (Node *2* flow constraint)

$v = x_{so1}$ (Node *so* flow constraint)

$v = x_{3si} + x_{2si}$ (Node *si* flow constraint)

This linear programming representation of the max-flow problem can be solved using the simplex method. It can be shown that the optimal solution to this linear program is $v = 3$, $x_{so1} = 3$, $x_{13} = 1$, $x_{12} = 2$, $x_{3si} = 1$, $x_{2si} = 2$. That is, in order for Fly-by-Night to maximize their number of daily flights from Juneau to Dallas, they should send one flight that connects via Juneau-Seattle-L.A.-Dallas, and two flights that connect via Juneau-Seattle-Denver-Dallas.

Chapter 2

THE BASIC ASSIGNMENT MODEL

2.1. Max-flow Representation

In this section we present a basic model of the first-year seminar assignment problem. In OR, assignment problems deal with assigning *n* assignees to *n* tasks where each assignee is given exactly one task. There is a cost occurred when assignee *i* is assigned task *j* and the goal is to minimize the total cost. A basic assignment problem can be presented with a network flow model: every assignee and every task is represented as a node and there would be arcs connecting assignee nodes to task nodes. In addition, we need to add two dummy nodes: one is the so-called *source* and the other is the so-called *sink*.

Our FYS assignment problem is a type of assignment problem, but instead of trying to minimize the total "cost", we want to maximize the number of students assigned to a seminar of their choice. Thus, it can be considered as a maximum flow problem where the flow represents the actual assignments. Here is the graphical representation of the FYS assignment problem:

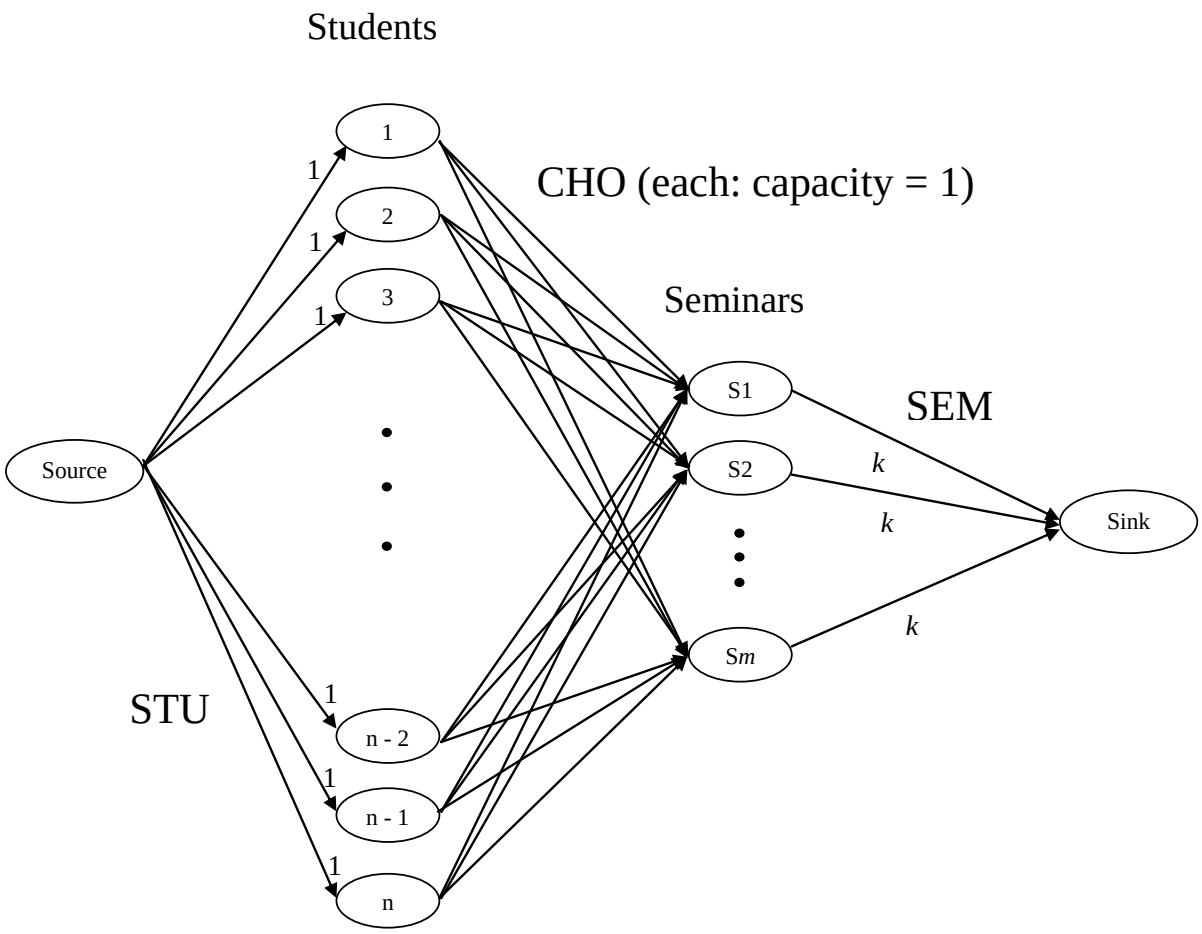


Figure 2.1: The graphical representation of our basic FYS assignment model

In Figure 2.1, the ovals represent the nodes. "Source" and "Sink" nodes obviously represent the source and the sink. All the other nodes, Student (1, 2, ..., *n*) and Seminar (S1, S2, …, S_m), are transshipment nodes. These nodes are connected by arcs whose directions are specified by the arrowhead at the end of each arc. There are *n* arcs coming from the Source where *n* is the total number of first-year students. There are six arcs coming from each Student node to six Seminar nodes indicating the six choices made by every first-year student. For example, if Student 10 selected the six seminars 1, 4, 10, 14, 23, and 40, there would be an arc from Student node 10 to the six corresponding seminar nodes. Each

Seminar node is connected to the Sink by an arc whose capacity is 16, which is the maximum number of students allowed in each seminar. All the other arcs have a capacity of 1. Our objective is to send as much flow as possible from the source node to the sink node.

The flow integrality theorem ensures that our model will yield a meaningful solution, if that solution exists, since all of our arcs have integral capacities. To see this, note that there is at most one unit of flow that can travel from the Source node to each Student node. Let's take the flow from the Source node to node 1 (Student 1) as an example. This one unit of flow must travel along one and only one arc to a Seminar node (because the flow must be an integer, i.e., 0 or 1), say S_j, which implies Student 1 is assigned to Seminar S_j. The capacity of 16 on the arcs connecting each Seminar node to the sink ensures that no more than sixteen students can be assigned to a seminar. The FYS assignment problem has a solution if and only if all students are assigned to one of their six requested seminars. Thus, the max-flow value, *v*, will equal *n* if there is a feasible assignment. If $v < n$, then it is not possible to assign students to seminars such that every student is enrolled in one of their six choices.

2.2. Linear Programming Representation

The network flow model discussed in the previous section can be formulated as the following linear programming problem:

Maximize v

Subject to

$$STU_i - \sum_j CHO_{ij} = 0 \text{ for each Student } i \qquad (2.1)$$

$$\sum_i CHO_{ij} - SEM_j = 0 \text{ for each Seminar } j \qquad (2.2)$$

$$\sum_j SEM_j = v \qquad (2.3)$$

$$\sum_i STU_i = v \tag{2.4}$$

$$STU_i \leq 1 \text{ for all Student } i \tag{2.5}$$

$$0 \leq CHO_{ij} \leq 1 \text{ for all Student } i \text{ and Seminar } j \tag{2.6}$$

$$SEM_j \leq 16 \text{ for all Seminar } j \tag{2.7}$$

In this model,

STU_i represents the amount of flow coming out from node Student *i*,

CHO_{ij} = 1 if Student *i* chose Seminar *j*, and

SEM_j equals the number of students enrolled in Seminar *j*.

Recall that the STUDENTS nodes and SEMINARS nodes are transshipment nodes. Hence, constraint equations (2.1) and (2.2) are needed to satisfy the requirement of a transshipment node, i.e., the amount of flow into a node must be equal to the amount of flow out of that node. Similarly, constraint equations (2.3) and (2.4) are needed to satisfy the condition of the source and sink nodes. The amount of flow that comes out of the source ($\sum_i STU_i$) must be equal to the amount of flow that goes into the sink ($\sum_j SEM_j$); we denote this amount as *v*. Obviously, we want to maximize *v*, and hope that an optimal solution will have $v = n$, which would indicate a feasible assignment. The set of inequalities (2.5), (2.6), and (2.7) are capacity constraints: (2.5) ensures that each student is assigned to at most 1 seminar and (2.6) makes sure all the flow (choice) is nonnegative and one student can pick at most one seminar. We need (2.7) because each seminar can have at most 16 students.

Note that we do not need to explicitly enforce that the values of STU_i, CHO_{ij}, and SEM_j are integers because of the integrality theorem, and therefore the formulation could be solved using the simplex method.

2.3. Basic Implementation

We implemented the linear programming representation of the first-year seminar assignment problem of Section 2.2 in the Mosel Modeling language. Mosel is an algebraic modeling language for mathematical programming that allows the user to express common mathematical programming structures, which can then be solved using an optimizer.

The actual first-year selection data from the students enrolling in 2010 (the class of 2014) was then inputted into Mosel and solved using the Xpress optimizer. Xpress is the mathematical programming solver that is bundled with Mosel, and is designed to solve linear, integer, and convex quadratic programming problems.

The solver was able to find a solution to the model within a few seconds. That is, the solver was able to assign the students of the class of 2014 to seminars so that they were assigned one of their chosen six seminars. Below is the output from our Mosel code.

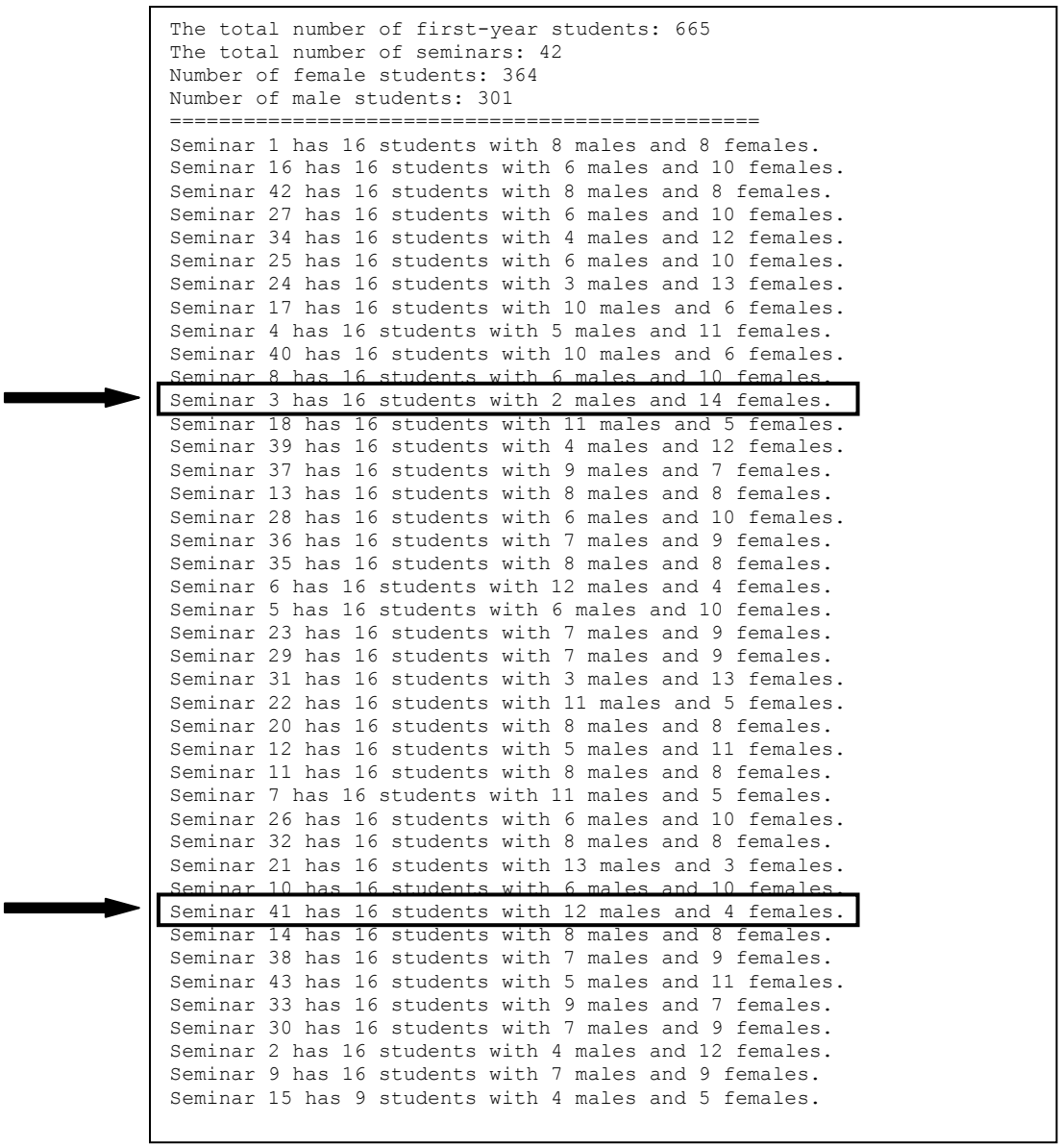


Figure 2.2: Output from the Mosel code obtained by using the basic assignment model to solve 2010 data

While this now allows us to determine an assignment of students to seminars in a scientific manner, it produced a result that was undesirable from a gender-balance standpoint. For instance, in our solution, the ratio of males to females is 3:1 in Seminar 41**.** In contrast, the ratio of males to females is 1:7 in Seminar 3. Therefore, while the assignment model of Section 2.2 quickly provided a solution to the basic assignment problem, the solution it

determined was certainly not optimal from the standpoint of gender composition. In Chapter 3 we will improve our assignment model so that it will take gender into consideration when assigning students to seminars.

2.4. Simulation

As discussed in the Introduction, we are interested in determining how our ability to make an assignment of students to seminars is affected by how many seminars students are required to choose. Recall that currently, students are required to select six seminars that they would be interested in enrolling in. Some students might find it attractive to provide fewer than six seminars (as they would be more likely to get one they are very interested in). However, the concern would be whether or not it would be likely to find an assignment of students to seminars if they were required to select fewer than six seminars. Note that we will not consider gender in our analysis here. This will be addressed in Chapter 3.

In order to test this, we used a *simulation* to determine our ability to find an assignment using the basic assignment model of Section 2.2. Specifically, we repeatedly generated random data that mimic how students actually chose their six seminars, and then determined whether or not an assignment of students to seminars exists for the randomly generated student selection data. This would then allow us to estimate how likely it is for us to find an assignment if the students were required to select fewer than six seminars.

Before we can run a simulation we need to find a model that mimics how students actually select their seminars. We were given access to the seminar selections of first-year students for the years 2008, 2009, and 2010. However, the data for 2008 had a number of significant issues, and therefore we decided not to utilize it in our study. For example,

because of a technical glitch, the selections of almost 100 students were not recorded by the system.

2.4.1. Model for the Class Entering In 2010

Figure 2.3 below is a chart showing the popularity of the 42 possible seminars for the entering class in 2010. Each seminar is represented by a column showing the percentage of total number of choices all students made. For example, of the 675 first-year students in the complete 2010 data set (which includes selection data for students that was added after an initial assignment of students to seminars was made by the college), 274 students picked seminar A1 as one of their six choices. Assuming all of these in-coming students selected six seminars, the total number of seminar selections among all students was $675 \times 6 = 4{,}050$. Thus, the percentage of students selecting seminar A1 was $274/4{,}050 \approx 6.8\%$.

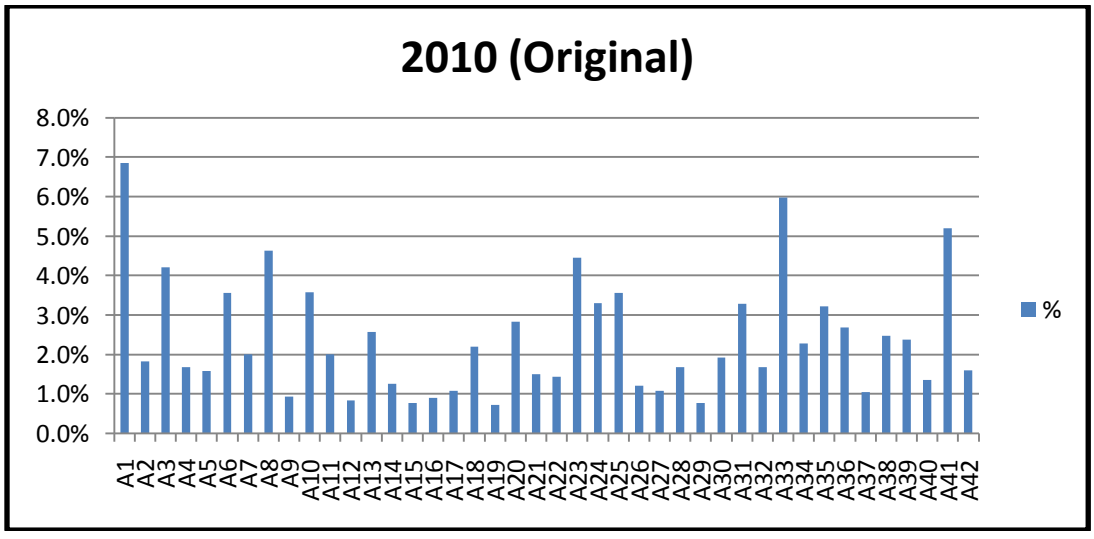


Figure 2.3: The distribution of probability a seminar is chosen in 2010

This chart clearly indicates that some seminars are more popular than others, and therefore we need to develop a probabilistic model that captures this. We began by

renumbering the seminars, where under the new numbering system, the most popular seminar is numbered one, the second most popular seminar is numbered two, etc. This leads to the following right-skewed distribution.

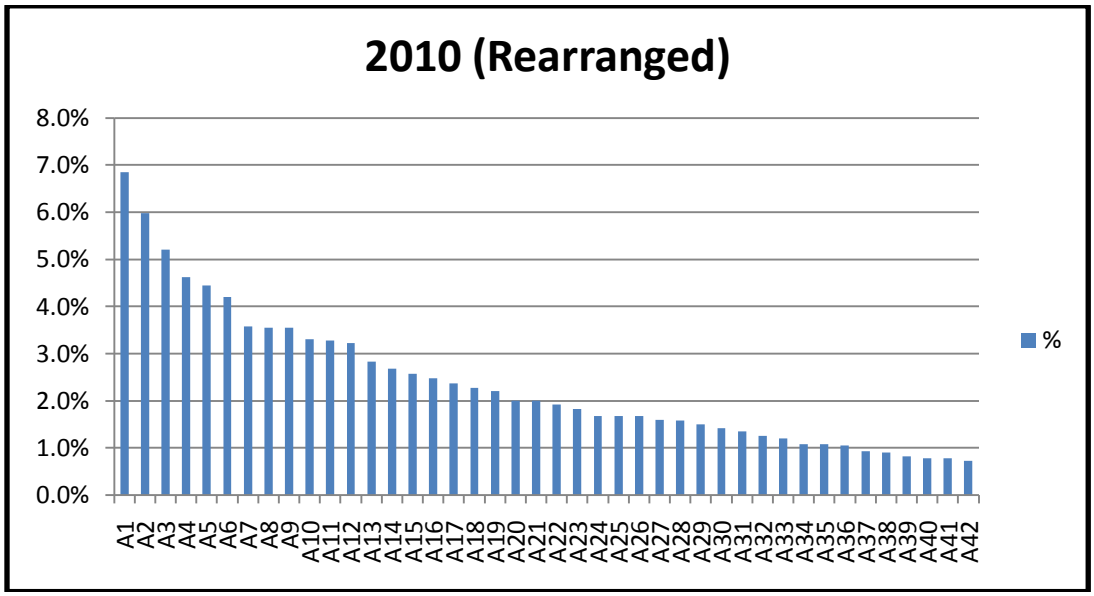


Figure 2.4: Renumbering of seminars in Figure 2.3 from most popular to least popular

Since the numbering of the seminars is arbitrary, it was this distribution that we attempted to model. Specifically, we developed a discrete probability distribution that tries to capture the main features of this data. The resulting model is presented in the table and figure below.

Table 2.1: Probability model for 2010 FYS popularity

2010 SEMINAR	Probability of each seminar in the range getting picked
1 - 2	6.4%
3 - 6	4.6%
7 - 12	3.4%
13 - 19	2.5%
20 - 30	1.7%
31 - 42	1.0%

To come up with the above model, we grouped the right-skewed distribution of Figure 2.4 into groups according to their probability of being selected. An illustration of our model is presented in the figure below.

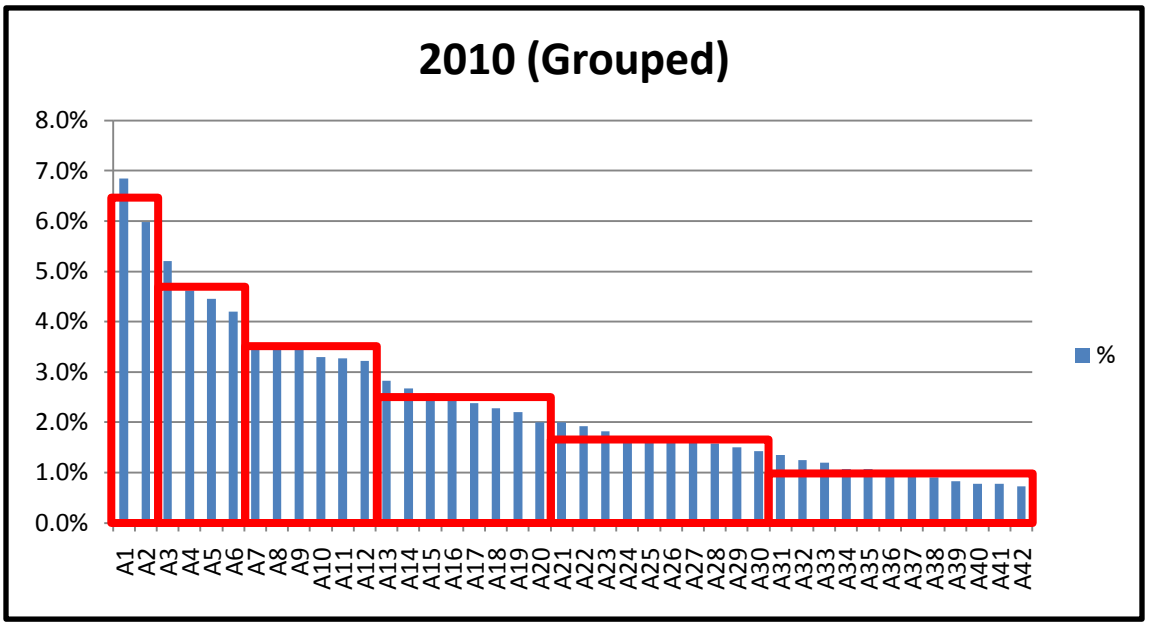


Figure 2.5: Grouped version of Figure 2.4

In order to compute the probability that a desired assignment is possible if each student picks x seminars, say 5 seminars, we utilize both the probability distribution above and the Mosel implementation of the assignment problem of Section 2.2. Specifically, we simulated the seminar selections of 675 students (the number of first-years included in the initial 2010 data set) by using a random number generator to choose seminars under the above model. Each randomly generated data set was then solved using our Mosel model, and we determined whether or not a feasible assignment exists. This process was repeated 200 times, and the number of successes, all students are assigned to a seminar of their choice, is recorded to compute the probability in question.

We performed this simulation for when students pick 6, 5, 4, and 3 seminars, and our results are given in the table below.

Table 2.2: Success rate for 2010 data

Number of Seminars Picked	Success Rate
6	100%
5	100%
4	100%
3	57.5%

The success rate is defined as the number of times all students are assigned one seminar of their choice over 200 trials.

Interestingly, this simulation shows that it is still extremely likely that an assignment of students to seminars will exist even when students only select four seminars, but then abruptly becomes unlikely when selecting three seminars. Given that most students would prefer to select fewer seminars, this seems to suggest that the college may want to consider reducing the number of selections to five, or possibly four. However, these results assume that student seminar selection follows the 2010 data. In addition, this simulation does not take gender into account. In other words, while an assignment might exist when students select as few as four seminars, it might be impossible to achieve a balance of gender in each of the seminars.

2.4.2. Model for the Class Entering In 2009

For the 2009 data, we repeated the same process used for the model based on 2010 data. In 2009, there were 596 first-year students in the complete data set (which includes selection data for students that was added after an initial assignment of students to seminars was made by the college), with 41 possible seminars. During this particular year, the

capacity of the seminars was set to 15, so we adjusted our model to reflect this. Figure 2.5 shows the popularity of the 41 seminars for that year.

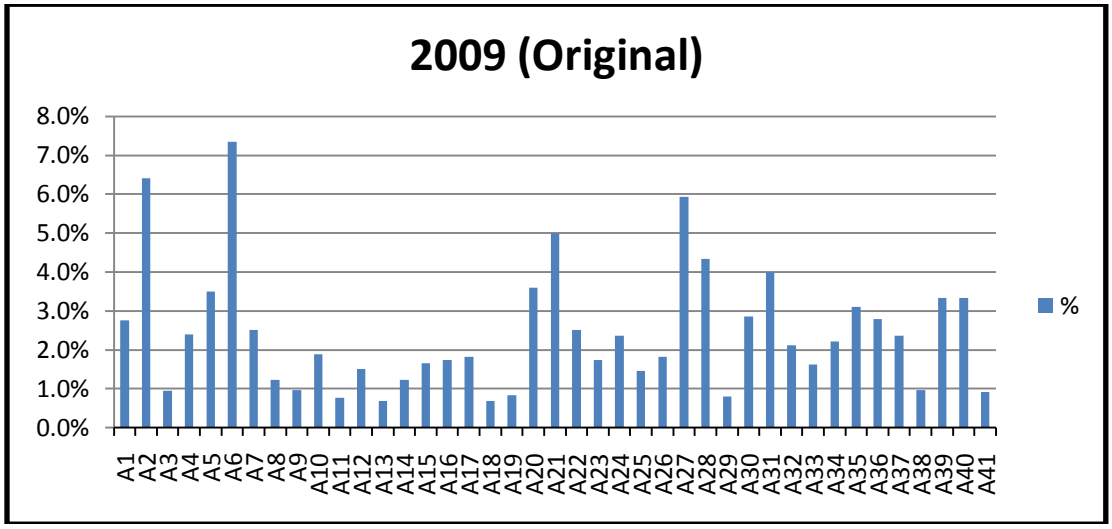


Figure 2.6: The distribution of probability a seminar is chosen in 2009

This chart clearly indicates that some seminars are more popular than others, and therefore we need to develop a probabilistic model that captures this. Again, we began by renumbering the seminars, where under the new numbering system, the most popular seminar is numbered one, the second most popular seminar is numbered two, etc. This leads to the following right-skewed distribution.

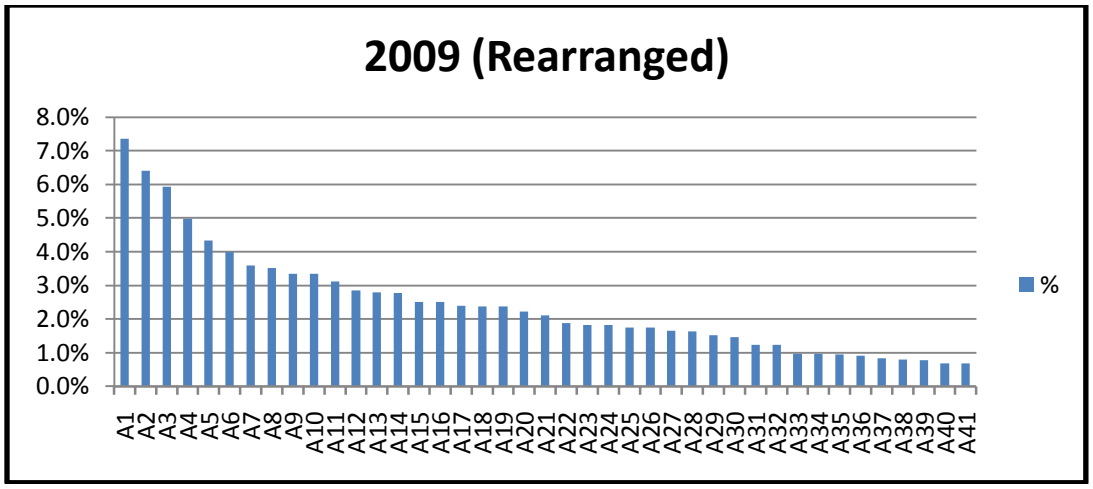


Figure 2.7: Renumbering of seminars in Figure 2.6 from most popular to least popular seminars

We can see that the distribution of 2009 seminar popularity is slightly different from that of 2010. We then grouped the seminars based on their popularity. This resulted in the following chart (Figure 2.8):

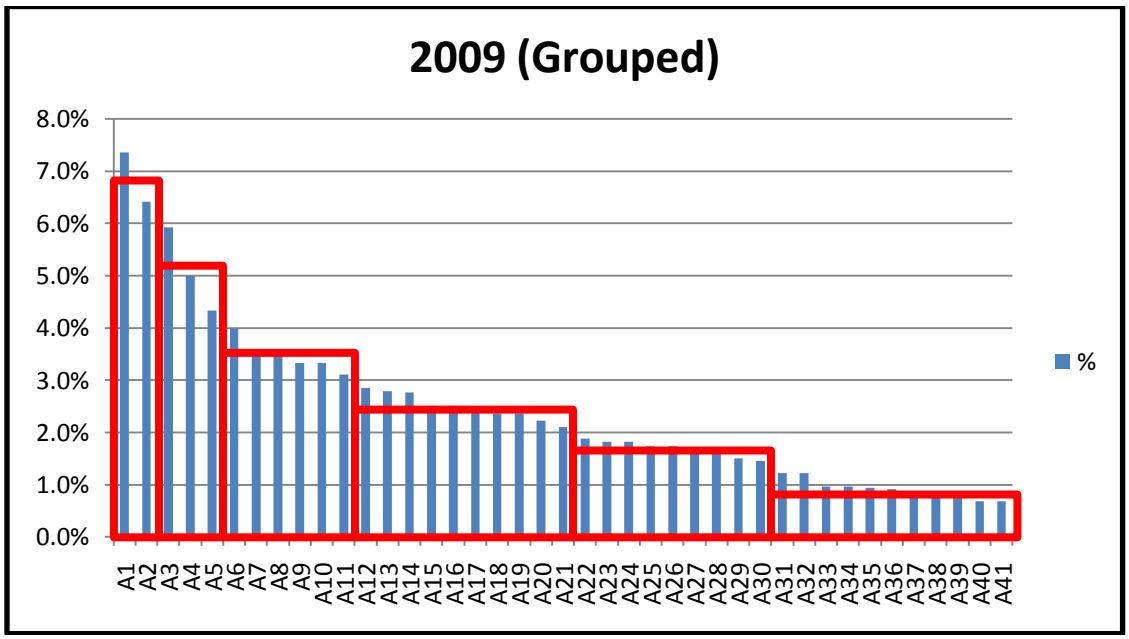


Figure 2.8: Grouped version of Figure 2.7

Figure 2.8 visually helped us develop the following discrete probability distribution that captures the characteristics of 2009 FYS popularity data:

Table 2.3: Probability model for 2009 FYS popularity

2009 SEMINAR	Probability of each seminar in the range getting picked
1-2	6.9%
3-5	5.1%
6-11	3.5%
12-21	2.5%
22-30	1.7%
31-41	0.9%

Table 2.4 below represents the "success rate" for the model of 2009 data.

Table 2.4: Success rate for 2009 data

Number of Seminars Picked	Success Rate
6	100%
5	100%
4	100%
3	50.5%

Recall that the success rate is defined as the number of times all students are assigned one seminar of their choice over 200 trials.

Note that our results for the model based on 2009 data support the conclusion we made at the end of Section 2.4.1. That is, it would likely be feasible for us to reduce the number of selections to five, or possible four, and still be able to find an assignment.

2.4.3. Conclusions from Simulations

Despite the difference in seminar popularity distributions for the 2009 and 2010 data, the assignment success rate of the models for each year suggests that the college could reduce the number of seminars students need to select, and have it still be likely that an

assignment of students to seminars would exist. However, as mentioned above, our assignment model did not take gender into consideration. Therefore, it is not clear yet what effect reducing the number of seminar choices will have on our ability to find an assignment with a good gender balance. We will revisit this conclusion in the next chapter where we take gender into account.

We also point out that while our models of the 2009 and 2010 seminar selection data appear to capture the characteristics of the actual data, they are limited in that they do not take into account possible *correlations* between seminar selections. For example, it could be the case that a student who selected a seminar related to sustainability would be more likely to select other seminars related to the topic. Our model does not take this into consideration, and such correlations could affect the assignment success rates. It would be interesting to attempt to create a model that takes into consideration these possible correlations; however, it would be difficult to model given the limited amount of data available.

Chapter 3

THE ASSIGNMENT MODEL WITH GENDER

3.1. Max-flow Representation with Gender

Given that male and female students may have various perspectives on the same issue, a good gender balance could possibly enrich the diversity of opinions contributing to a seminar. Moreover, it could make it more comfortable for, say a male student, if he does not happen to be assigned to a female dominant seminar, and vice versa. Therefore, it is vital that we achieve a good gender balance among all seminars. More importantly, the *college* wants gender balance in the seminars.

As we have mentioned above, the basic assignment model of Section 2.2 does not take gender into consideration when assigning students to seminars. Therefore, we need to generate a new model that allows us to keep track of the gender of students. We now discuss our improved max-flow model that takes gender into account. Specifically, the model below will allow us to keep track of the gender of the students enrolled in the seminars.

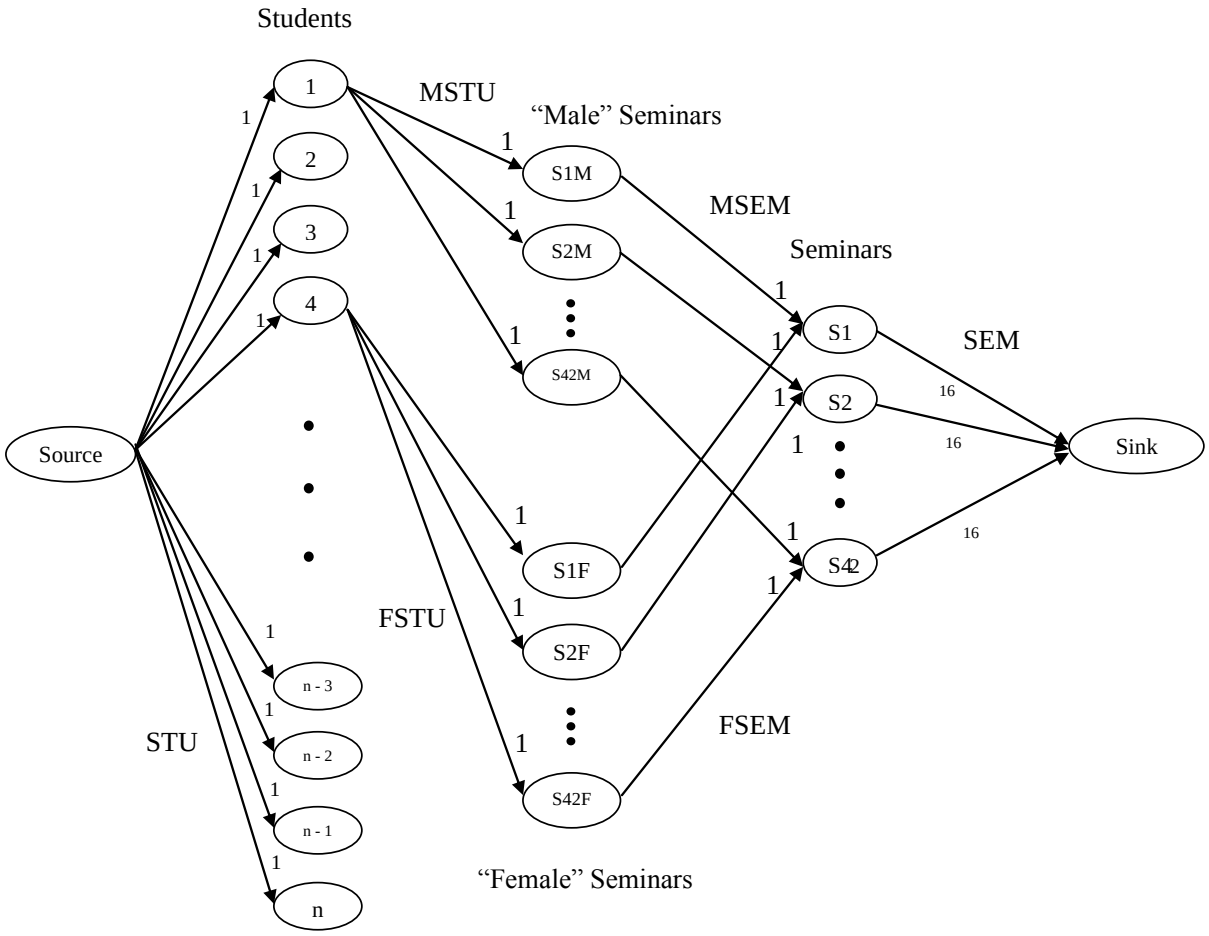


Figure 3.1: The graphical presentation of our assignment model with gender

It should be noted that this graph is a modified version of the graph in Figure 2.1. There are two new sets of transshipment nodes added to this graph: "Male" seminars nodes and "Female" seminar nodes. The former set of nodes is denoted as S_iM where i is the index of the seminar. Similarly, the later set of nodes is denoted as S_iF where i is the index of the seminar. If a student x is a male student, there would be 6 arcs, each with a capacity of 1, coming from node Student(x) to the 6 corresponding S_iM nodes. Similarly, if a student y is a female student, there would be 6 arcs with a capacity of 1 coming from node Student(y) to the 6 corresponding S_iF nodes. Moreover, there is a flow from each of these "Male" and

"Female" seminars nodes to their corresponding Seminar node. For instance, there is a flow $MSEM_{S1M,S1}$ from S_1M to S_1 and a flow $FSEM_{S1F,S1}$ from S_1F to S_1, etc. The flow from node S_iM to node S_i tells us how many male students there are in seminar S_i. Similarly, the flow from node S_iF to node S_i tells us how many female students there are in seminar S_i. Thus, these two new sets of nodes allow us to keep track of the gender ratio in each seminar.

3.2. Linear Programming Representation with Gender

The max-flow model of the previous section can be represented as the linear programming problem below.

Maximize v

Subject to

$$STU_i - \sum_j MSTU_{i,j} = 0 \text{ for all Male students } i \quad (3.1)$$

$$STU_i - \sum_j FSTU_{i,j} = 0 \text{ for all Female students } i \quad (3.2)$$

$$\sum_i MSTU_{i,j} - MSEM_j = 0 \text{ for all Seminar } j \quad (3.3)$$

$$\sum_i FSTU_{i,j} - FSEM_j = 0 \text{ for all Seminar } j \quad (3.4)$$

$$MSEM_j + FSEM_j - SEM_j = 0 \text{ for all Seminar } j \quad (3.5)$$

$$0 \le MSTU_{i,j} \le 1 \text{ for all Male Student } i \text{ and Seminar } j \quad (3.6)$$

$$0 \le FSTU_{i,j} \le 1 \text{ for all Female Student } i \text{ and Seminar } j \quad (3.7)$$

$$0 \le STU_i \le 1 \text{ for all Student } i \quad (3.8)$$

$$MSEM_j \le 16 \text{ for all Seminar } j \quad (3.9)$$

$$FSEM_j \leq 16 \text{ for all Seminar } j \tag{3.10}$$

$$SEM_j \leq 16 \text{ for all Seminar } j \tag{3.11}$$

$$\sum_j SEM_j = v \tag{3.12}$$

$$\sum_i STU_i = v \tag{3.13}$$

In this model,

STU_i represents the amount of flow coming out from node Student *i*,

$FSTU_{ij}$ = 1 if Female Student *i* chose Seminar *j* (and 0 otherwise),

$MSTU_{ij}$ = 1 if Male Student *i* chose Seminar *j* (and 0 otherwise),

$MSEM_j$ equals the number of male students enrolled in Seminar *j*,

$FSEM_j$ equals the number of female students enrolled in Seminar *j*, and

SEM_j equals the number of students enrolled in Seminar *j*.

Note that the model above does not actually make assignment decisions based on gender, but does allow us to keep track of the gender of students assigned to seminars. Since the nodes representing STUDENTS, "MALE" SEMINARS, "FEMALE" SEMINARS, and SEMINARS are transshipment nodes, constraint equations (3.1) – (3.5) are needed to satisfy the requirement of a transshipment node, i.e., the amount of flow into a node must be equal to the amount of flow out of that node. Similarly, constraint equations (3.12) and (3.13) are needed to satisfy the condition of a source and sink node. The amount of flow coming out of a source ($\sum_i STU_i$) must be equal to the amount of flow coming into a sink ($\sum_j SEM_j$); we denote this amount as *v*. Obviously, we want to maximize *v*. The set of inequalities (3.6) – (3.11) are capacity constraints: (3.8) ensures that each student is assigned to at most 1 seminar and (3.6) and (3.7) make sure all the flow (choice) is nonnegative and one student

can pick a seminar once. We need (3.9) – (3.11) because each seminar can have at most 16 students.

3.3. Initial Implementation

As a starting point to taking gender into account, we set the capacity of each MSEM arc and each FSEM arc as 9:

$$MSEM._j \leq 9 \text{ for all Seminar } j$$

$$FSEM._j \leq 9 \text{ for all Seminar } j$$

This enforces that no more than 9 of any one gender will be allowed into a course. As mentioned above, since each arc coming from a Seminar node to the Sink has a capacity of 16, the number of students assigned to a certain seminar, say S_i, can never exceed 16 even if we allow two units of flow with capacity of 9 units, one from S_iM and the other from S_iF, into S_i.

Note that this method will not guarantee that all students would get a seminar of their choice because in its current form our model will never allow more than 9 of any one gender into a course. Given that there are considerably more females than males in recent first-year classes, an assignment might only exist if more than 9 females are allowed into a course . In fact, while we know an assignment exists for the 2009 data set, the model was not able to find an assignment when the above two constraints are included. This was related to a few issues, including that only 9 male students chose Seminar 28 during that year. Moreover, if we distribute the 243 male students in 2009 evenly, each seminar will on average have $243/41 \approx 6$ males. In the next section we take a different approach to achieving a gender balance.

3.4. Improved Implementation

In the previous section we used constraints to try to achieve a gender balance in the assignment, which was not successful. In this section we take a different approach where we modify the objective function to take gender into account. Note that the quantity $|MSEM_j - FSEM_j|$ represents the gender imbalance in Seminar *j*, and therefore we could attempt to minimize the sum of these quantities. That is, we could change the objective function of the model in Section 3.2 to

$$\text{minimize } \sum_j |MSEM_j - FSEM_j|$$

while adding the constraint

$$v = n,$$

where n is the total number of first-year students. Note that the objective will attempt to find an assignment that makes the total gender imbalance as small as possible, while the $v = n$ constraint dictates that the model will only have a feasible solution if an assignment of all students exists. It turns out that while there are algorithms that can handle objective functions with absolute values, they do not perform very well. Therefore, we will instead utilize the objective function below, which attempts to minimize the sum of the *squared* gender-differences.

$$\text{minimize } \sum_j (MSEM_j - FSEM_j)^2 \qquad (3.14)$$

This quadratic objective function essentially accomplishes the same goal as the one with the absolute values, and is preferable because optimizers can handle quadratic objective functions easier than ones involving absolute value. However, a major requirement for

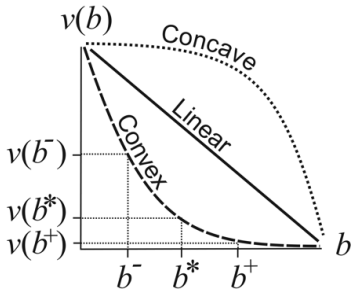
optimizers to handle quadratic programs is that the optimization problem is *convex*. This concept is discussed in the section below.

3.4.1. Convex Programs

Convex programming covers a wide range of programming problems. It examines the case when the constraint functions form a *convex set* and the objective function is either *convex* or *concave*, depending on whether it is a minimization or maximization problem, respectively. We have two new terminologies here: convex/concave function and convex set. Here is a definition from Hillier & Lieberman (2005):

> "$f(x_1, x_2, x_3, \ldots, x_n)$ is a *convex function* if, for each pair of points on the graph of $f(x_1, x_2, x_3, \ldots, x_n)$, the line segment joining these two points lies entirely above or on the graph of $f(x_1, x_2, x_3, \ldots, x_n)$. It is a *strictly convex function* if this line segment actually lies entirely above this graph except at the endpoints of the line segment. *Concave functions* and *strictly concave functions* are defined in exactly the same way, except that *above* is replaced by *below*."

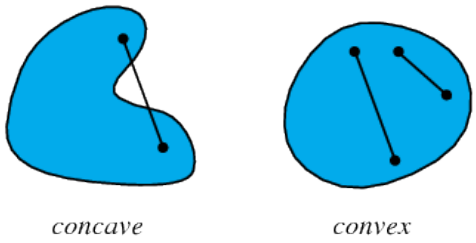
An illustration of the shape of convex and concave functions can be found below.



http://www.pnas.org/content/103/24/9113/F3.expansion.html

Figure 3.2: Illustration of convex, concave, and linear functions

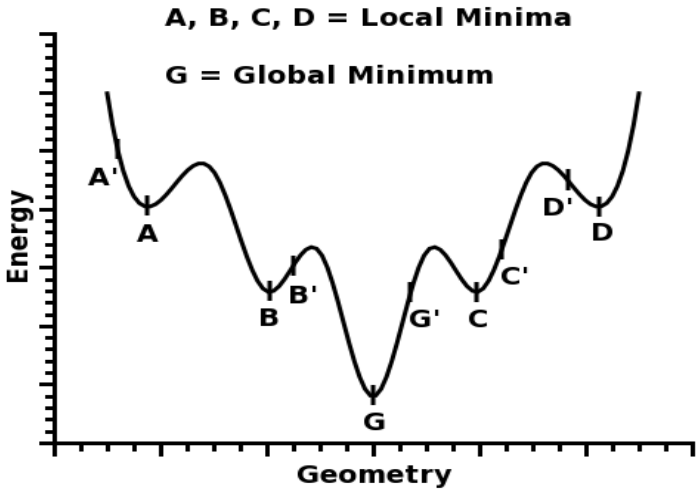
A *convex set* is a set of points in which for each pair of points belonging to the set, the line segment connecting those two points is completely contained within the set. Below is a graphical presentation of a convex set:



http://mathworld.wolfram.com/Concave.html

Figure 3.3: Illustration of some convex and concave sets

A very important characteristic of a convex function is that its local minimum is also its global minimum. This is not necessarily true for non-convex functions. In Figure 3.4 we can see that the function represented there is not a convex function since the line segment connecting B and D intersects the graph. It is obvious that the function has multiple local minima and only one unique global minimum. When solving for global minimum of a non-convex function, there is some chance that the optimization algorithm will be "deceived" when it reaches a local minimum. If we let the solver run for a long time, it may still only reach point B or C. Hence, we will never know if the solution yielded is a global minimum.



http://www.sparkle.pro.br/tutorial/geometry

Figure 3.4: Illustration of a non-convex function

It is a well-known fact that sums of convex functions are also convex. Furthermore, it can be shown that for any x and y, $f = (x - y)^2$ is a convex function by checking if the second order partial derivatives $\frac{\partial^2 f(x,y)}{\partial x^2}$, $\frac{\partial^2 f(x,y)}{\partial y^2}$, and $\frac{\partial^2 f(x,y)}{\partial x^2}\frac{\partial^2 f(x,y)}{\partial y^2} - \frac{\partial^2 f(x,y)}{\partial x \partial y}^2$ are greater than or equal to zero. For $f = (x - y)^2$, the values of these second order partial derivatives are 0, 2, and 2, respectively, and hence, f is convex (for further details on this see Appendix 2 of Hillier, & Lieberman (2005)). Since each term of our objective function $w = \sum_j \left(MSEM_j - FSEM_j\right)^2$ is of this form, it follows that the entire objective function is convex.

It is very important that our objective function is convex because it ensures that the solution we find when minimizing w is the optimal solution, i.e. the global minimum, and not just a local minimum.

3.4.2. Quadratic Programs

A quadratic program (QP) is a model where all the constraints are linear, but the objective function is quadratic. Various algorithms have been developed to deal with convex quadratic programs. One such algorithm is a modified version of the simplex method. This algorithm involves the construction of the Karush-Kuhn-Tucker conditions (or KKT conditions) for the quadratic objective function and transforming these conditions into a form such that linear programming can be applied to solve the program (Hillier, & Lieberman, 2005). We refer interested readers to Section 12.7 of Hillier & Lieberman (2005).

3.4.3. Integer Programs

Since our objective function is no longer linear, the integrality theorem does not hold. Thus, it is not guaranteed that the algorithm will yield integer solutions. We could easily end up with an optimal solution with a fractional flow, say 3/4, from Student *i* to Seminar *j*, which would indicate that 3/4 of Student *i* is assigned to Seminar *j*. For this application, it obviously does not make sense to have a non-integer solution. Therefore, we need to add additional restrictions that the variables must be integers. Specifically, in the model we need to set these extra conditions: the flow through each arc must be an integer. Note that this implies that the flow on STU_i, $MSTU_{i,j}$, and $FSTU_{i,j}$ must be binary because they each have a capacity of 1. This leads us to another concept: integer programming.

A linear integer programming problem (IP) is a linear program where "some or all of the variables are required to be nonnegative integers" (Winston, 1994). An IP where all variables have to be integers is called a pure integer programming problem. If some of the variables of an IP are allowed to be non-integer then we call that IP a mixed integer programming problem (MIP).

It turns out that integer programming problems are significantly more difficult to solve than linear programs. The typical solution method, which is called the *branch-and-bound method*, involves systematically enumerating all the possible integer solutions by estimating bounds of certain solution candidates by solving linear programs. In other words, in order to solve a single integer program, the algorithm must solve many linear programs (possibly hundreds of thousands of LP's). We therefore want to emphasize that by adding the integer restrictions on our variables we have significantly increased the difficulty of

finding an optimal solution to our assignment model. A detailed explanation of the branch-and-bound method and an example is included in Appendix B.

3.5. Solutions to Previous Years Data

We implemented the assignment model of Section 3.2, using the quadratic objective function (3.14), in Mosel. The model was solved for the 2009 and 2010 selection data on a Dell Optiplex 740, running Windows 7, which is equipped with a 2.7 GHz AMD Athlon processor and 2.00 GB of memory.

As mentioned in the previous section, integer programs are significantly more difficult to solve. In fact, for the 2010 data the Xpress optimizer was not able to find a solution after letting the computer run for 12 hours. However, we have the ability to stop the optimizer during its execution, and then have it provide the best solution it was able to find up to that point. While the solution provided after stopping Xpress prematurely is not technically optimal, it is likely to be a good-quality solution. In other words, while the assignment it determined might not be the best, it is likely to be a solution that is more than satisfactory from a practical point of view.

Note that the quality of our approximate solution should improve the longer we let the optimizer run. We began by inputting the data from 2010, using a capacity of 16 students per seminar (which was the capacity for that year), and stopped the optimizer after one minute. The Mosel output is provided in the figure below. Note that "Gender Penalty" is the value of the quadratic objective, which we are trying to minimize. Therefore, smaller values of "Gender Penalty" indicate a better quality solution.

```
The total number of first-year students: 665
The total number of seminars: 42
Number of female students: 364
Number of male students: 301
================================================
All students were assigned
================================================
Gender Penalty: 173
================================================
The largest gender gap is: 3
The second largest gender gap is: 3
The third largest gender gap is: 3
================================================
Seminar 1 has 16 students with 7 males and 9 females.
Seminar 16 has 16 students with 7 males and 9 females.
Seminar 42 has 16 students with 7 males and 9 females.
Seminar 27 has 16 students with 7 males and 9 females.
Seminar 34 has 16 students with 7 males and 9 females.
Seminar 25 has 16 students with 7 males and 9 females.
Seminar 24 has 16 students with 7 males and 9 females.
Seminar 17 has 15 students with 9 males and 6 females.
Seminar 4 has 16 students with 7 males and 9 females.
Seminar 40 has 16 students with 7 males and 9 females.
Seminar 8 has 16 students with 7 males and 9 females.
Seminar 3 has 16 students with 7 males and 9 females.
Seminar 18 has 16 students with 7 males and 9 females.
Seminar 39 has 16 students with 7 males and 9 females.
Seminar 37 has 16 students with 7 males and 9 females.
Seminar 13 has 16 students with 7 males and 9 females.
Seminar 28 has 16 students with 8 males and 8 females.
Seminar 36 has 16 students with 7 males and 9 females.
Seminar 35 has 16 students with 7 males and 9 females.
Seminar 6 has 16 students with 7 males and 9 females.
Seminar 5 has 16 students with 7 males and 9 females.
Seminar 23 has 16 students with 7 males and 9 females.
Seminar 29 has 16 students with 7 males and 9 females.
Seminar 31 has 16 students with 7 males and 9 females.
Seminar 22 has 13 students with 8 males and 5 females.
Seminar 20 has 16 students with 7 males and 9 females.
Seminar 12 has 16 students with 7 males and 9 females.
Seminar 11 has 16 students with 7 males and 9 females.
Seminar 7 has 16 students with 7 males and 9 females.
Seminar 26 has 16 students with 7 males and 9 females.
Seminar 32 has 16 students with 7 males and 9 females.
Seminar 21 has 16 students with 7 males and 9 females.
Seminar 10 has 16 students with 7 males and 9 females.
Seminar 41 has 15 students with 7 males and 8 females.
Seminar 14 has 16 students with 7 males and 9 females.
Seminar 38 has 16 students with 7 males and 9 females.
Seminar 43 has 16 students with 7 males and 9 females.
Seminar 33 has 15 students with 8 males and 7 females.
Seminar 30 has 16 students with 7 males and 9 females.
Seminar 2 has 16 students with 7 males and 9 females.
Seminar 9 has 16 students with 7 males and 9 females.
Seminar 15 has 15 students with 9 males and 6 females.
```

Figure 3.5: Output from the Mosel code obtained by using the gender model on 2010 data (after 1 minute)

First, note that the optimizer was able to find an assignment of students to seminars so that all students were assigned one of their choices. Second, you will notice a vast improvement in terms of the number of gender imbalances when compared to the assignment

generated by Xpress in Section 2.3 (see Figure 2.2 in Section 2.3). However, there are still a number of classes that have somewhat unsatisfactory gender gaps (the absolute difference between the number of male and female students in a seminar). For example, seminars 15, 17, and 22 have gender gaps of 3. The overall Gender Penalty was 173, which is a measure of the quality of our solution.

It should be the case that if we let the optimizer run longer we will likely find a better solution, i.e., a smaller Gender Penalty and less unsatisfactory gender ratios. To see how the solution improves, we re-ran our tests and stopped the optimizer after 2, 3, 4, 5, and 30 minutes to compare the solutions. Our results are provided in the table below.

Table 3.1: 2010 data

Number of minutes the model was run	Gender Penalty	Largest Gender Gap	2nd Largest Gender Gap	3rd Largest Gender Gap
1	159	4	4	4
2	119	2	2	2
3	119	2	2	2
4	119	2	2	2
5	119	2	2	2
30	119	2	2	2

Interestingly, while the solver was able to find an improved solution after two minutes, the solution did not improve thereafter. This solution yielded a Gender Penalty of 119 and the largest gender gap was 2. The actual Mosel output for this new solution is presented in the figure below.

```
The total number of first-year students: 665
The total number of seminars: 42
Number of female students: 364
Number of male students: 301
================================================
All students were assigned
================================================
Gender Penalty is: 119
================================================
The largest gender gap is: 2
The second largest gender gap is: 2
The third largest gender gap is: 2
================================================
Seminar 1 has 16 students with 8 males and 8 females.
Seminar 16 has 15 students with 7 males and 8 females.
Seminar 42 has 16 students with 7 males and 9 females.
Seminar 27 has 16 students with 7 males and 9 females.
Seminar 34 has 16 students with 8 males and 8 females.
Seminar 25 has 16 students with 7 males and 9 females.
Seminar 24 has 16 students with 7 males and 9 females.
Seminar 17 has 16 students with 7 males and 9 females.
Seminar 4 has 16 students with 7 males and 9 females.
Seminar 40 has 16 students with 7 males and 9 females.
Seminar 8 has 16 students with 7 males and 9 females.
Seminar 3 has 16 students with 7 males and 9 females.
Seminar 18 has 16 students with 8 males and 8 females.
Seminar 39 has 16 students with 8 males and 8 females.
Seminar 37 has 16 students with 7 males and 9 females.
Seminar 13 has 15 students with 7 males and 8 females.
Seminar 28 has 16 students with 7 males and 9 females.
Seminar 36 has 16 students with 7 males and 9 females.
Seminar 35 has 16 students with 7 males and 9 females.
Seminar 6 has 16 students with 8 males and 8 females.
Seminar 5 has 16 students with 7 males and 9 females.
Seminar 23 has 15 students with 7 males and 8 females.
Seminar 29 has 15 students with 7 males and 8 females.
Seminar 31 has 16 students with 7 males and 9 females.
Seminar 22 has 16 students with 7 males and 9 females.
Seminar 20 has 16 students with 8 males and 8 females.
Seminar 12 has 15 students with 7 males and 8 females.
Seminar 11 has 16 students with 7 males and 9 females.
Seminar 7 has 16 students with 8 males and 8 females.
Seminar 26 has 16 students with 7 males and 9 females.
Seminar 32 has 16 students with 7 males and 9 females.
Seminar 21 has 16 students with 7 males and 9 females.
Seminar 10 has 16 students with 7 males and 9 females.
Seminar 41 has 16 students with 7 males and 9 females.
Seminar 14 has 15 students with 7 males and 8 females.
Seminar 38 has 16 students with 7 males and 9 females.
Seminar 43 has 16 students with 7 males and 9 females.
Seminar 33 has 16 students with 7 males and 9 females.
Seminar 30 has 16 students with 7 males and 9 females.
Seminar 2 has 16 students with 7 males and 9 females.
Seminar 9 has 16 students with 7 males and 9 females.
Seminar 15 has 15 students with 7 males and 8 females.
```

Figure 3.6: Output from the Mosel code obtained by using the gender model to solve 2010 data (after 30 minutes)

While the solver was unable to prove that the solution above is optimal, we are actually able to prove that a Gender Penalty of 119 is in fact the smallest possible objective

value for this data set. In other words, the solution in Figure 3.6 must be an optimal solution. To see this, consider the following analysis.

Given that we had more females than males in 2010, it is impossible to achieve a perfect gender balance of 8 males and 8 females in all seminars if every seminar has 16 students. In other words, it is infeasible to achieve a Gender Penalty of 0. Therefore, the best we can do is to spread out the number of males and females evenly among the seminars.

Let us start with distributing females first (it does not matter whether we distribute males or females first). Recall that in 2010 we had 364 females. Note that $364/42 = 8 \times 42 + 28$, which means that ideally $42 - 28 = 14$ seminars will have 8 females and 28 seminars will have $8 + 1 = 9$ females. Similarly, there were 301 males in 2010, and given that $301/42 = 7 \times 42 + 7$, this implies that ideally $42 - 7 = 35$ seminars will have 7 males and 7 seminars will have 8 males.

Since the class size cannot exceed 16, we can only "pair" the seven 8-male seminars with seven of the 8-female seminars. As a result, the best possible assignment we can do regardless of how students actually chose their seminars in 2010 is to have 7 seminars with 8 males and 8 females, 28 seminars with 7 males and 9 females, and $42 - 7 - 28 = 7$ seminars with 7 males and 8 females. Thus, the most "ideal" distribution of males and females in 2010 is presented in the table below.

Table 3.2: The most "ideal" distribution of males and females in 2010

Seminar	Number of Females	Number of Males	Total Number of Students	Number of Males – Number of Females (α)	α^2
1	9	7	16	2	4
2	9	7	16	2	4
3	9	7	16	2	4
4	9	7	16	2	4
5	9	7	16	2	4
6	9	7	16	2	4
7	9	7	16	2	4
8	9	7	16	2	4
9	9	7	16	2	4
10	9	7	16	2	4
11	9	7	16	2	4
12	9	7	16	2	4
13	9	7	16	2	4
14	9	7	16	2	4
15	9	7	16	2	4
16	9	7	16	2	4
17	9	7	16	2	4
18	9	7	16	2	4
19	9	7	16	2	4
20	9	7	16	2	4
21	9	7	16	2	4
22	9	7	16	2	4
23	9	7	16	2	4
24	9	7	16	2	4
25	9	7	16	2	4
26	9	7	16	2	4
27	9	7	16	2	4
28	9	7	16	2	4
29	8	8	16	0	0
30	8	8	16	0	0
31	8	8	16	0	0
32	8	8	16	0	0
33	8	8	16	0	0
34	8	8	16	0	0
35	8	8	16	0	0
36	8	7	15	1	1
37	8	7	15	1	1
38	8	7	15	1	1
39	8	7	15	1	1
40	8	7	15	1	1
41	8	7	15	1	1
42	8	7	15	1	1
Sum	364	301	665	63	119

Notice that the best possible gender penalty we can obtain for the 2010 data is 119, which is the sum of the last column in Table 3.2. Thus, 119 is a lower bound on the objective value, which we achieved in the solution presented in Figure 3.6. Hence, our solution must be optimal.

We next applied our technique to the data from the first-year students entering in 2009. During this year there were only 582 students, so the capacity of the 41 seminars was set to 15 in constraints (3.9) – (3.11) (as opposed to the capacity of 16 in 2010). We began by terminating the optimizer after two minutes, and the Mosel output is given in the figure below.

```
The total number of first-year students: 582
The total number of seminars: 41
Number of female students: 339
Number of male students: 243
================================================
All students were assigned
================================================
Gender Penalty is: 248
================================================
The largest gender gap is: 3
The second largest gender gap is: 3
The third largest gender gap is: 3
================================================
Seminar 6 has 15 students with 7 males and 8 females.
Seminar 15 has 15 students with 6 males and 9 females.
Seminar 17 has 14 students with 6 males and 8 females.
Seminar 19 has 14 students with 6 males and 8 females.
Seminar 22 has 15 students with 6 males and 9 females.
Seminar 26 has 15 students with 6 males and 9 females.
Seminar 36 has 15 students with 6 males and 9 females.
Seminar 35 has 15 students with 6 males and 9 females.
Seminar 10 has 15 students with 6 males and 9 females.
Seminar 5 has 15 students with 7 males and 8 females.
Seminar 2 has 15 students with 6 males and 9 females.
Seminar 37 has 15 students with 6 males and 9 females.
Seminar 28 has 15 students with 6 males and 9 females.
Seminar 11 has 10 students with 4 males and 6 females.
Seminar 4 has 14 students with 6 males and 8 females.
Seminar 27 has 15 students with 7 males and 8 females.
Seminar 23 has 15 students with 6 males and 9 females.
Seminar 21 has 15 students with 6 males and 9 females.
Seminar 31 has 15 students with 6 males and 9 females.
Seminar 39 has 15 students with 6 males and 9 females.
Seminar 34 has 15 students with 7 males and 8 females.
Seminar 30 has 15 students with 6 males and 9 females.
Seminar 1 has 15 students with 6 males and 9 females.
Seminar 24 has 15 students with 6 males and 9 females.
Seminar 32 has 15 students with 6 males and 9 females.
Seminar 40 has 15 students with 6 males and 9 females.
Seminar 20 has 15 students with 6 males and 9 females.
Seminar 7 has 15 students with 7 males and 8 females.
Seminar 16 has 15 students with 7 males and 8 females.
Seminar 14 has 14 students with 6 males and 8 females.
Seminar 12 has 14 students with 6 males and 8 females.
Seminar 33 has 15 students with 6 males and 9 females.
Seminar 9 has 14 students with 6 males and 8 females.
Seminar 25 has 15 students with 7 males and 8 females.
Seminar 3 has 4 students with 1 males and 3 females.
Seminar 8 has 15 students with 6 males and 9 females.
Seminar 41 has 14 students with 6 males and 8 females.
Seminar 38 has 14 students with 6 males and 8 females.
Seminar 18 has 10 students with 4 males and 6 females.
Seminar 29 has 14 students with 6 males and 8 females.
Seminar 13 has 12 students with 5 males and 7 females.
```

Figure 3.7: Output from the Mosel code obtained by using the gender model to solve 2009 data (after 2 minutes)

This solution is actually quite undesirable for a number of reasons, including the fact that a large number of classes have three more females than males. However, another point of concern is that Seminar 3 has only has 4 students and Seminars 11 and 18 have only 10

students each. This was possible because even when the capacity of the seminars is set at 15 students, there are many extra seats because of the smaller size of the first-year class (if the capacity of all seminars is 15, we have $41 \times 15 - 582 = 33$ extra seats).

Note that in 2009 there was only 1 male student that chooses Seminar 3. Thus, since the optimizer is attempting to minimize the gender gap, and there can never be more than 1 male in Seminar 3, the next best thing is to have a smaller class size.

It seems clear that we want to ensure that the class sizes are a little more even, so we need to determine a way to handle this in our model. We were unable to come up with a way of handling this within our objective function, i.e., create a penalty when the class sizes are uneven that the optimizer would try to minimize. Therefore, we decided to add constraints to the problem to set a lower limit on class sizes. Note that the largest lower class limit is 14 students, i.e., it is not possible to assign 15 students to all seminars. This follows because we have 582 students / 41 seminars ≈ 14.2 students/seminar**.** We decided to enforce a lower class limit of 13 students, as opposed to 14, because while it may be possible to find an assignment where all classes have at least 14 students, it would not allow as much flexibility in terms of gender balancing. Specifically, we added the following constraints to our model:

$$SEM_j \geq 13 \text{ for all Seminar } j \qquad (3.14)$$

After adding constraints (3.14), we re-ran our tests and stopped the optimizer after 1, 2, 3, 4, 5, 30, and 120 minutes to compare the solutions. Our results are provided in the table below.

Table 3.3: 2009 data

Number of Minutes the Model Was Run	Gender Penalty	Largest Gender Gap	2nd Largest Gender Gap	3rd Largest Gender Gap
1	342	11	4	3
2	332	11	5	3
3	324	11	3	3
4	320	11	3	3
5	320	11	3	3
30	318	11	3	3
120	314	11	3	3

The first observation from our table is that the largest gender gap was 11, even after allowing the optimizer to run for 2 hours. However, this is to be expected because only 1 male signed up for Seminar 3, and we enforced that all seminars have at least 13 students (12 females – 1 male = gender gap of 11). Next, you will note that even in our best solution, the second and third largest gender gaps are both 3. Clearly these results are not as satisfactory as compared to those in 2010. This results from the fact that there were a few seminars that were very unpopular among males in 2009. For example, in addition to the 1 male student who selected Seminar 3, there were only 5 males that selected Seminar 11, and 7 males that selected Seminar 18.

Once again, the results shown above might not be optimal, and it might be possible to achieve a better solution if we let the solver run longer. However, from a practical standpoint, the result we get after letting the model run for 30 minutes is good enough. Below is the Mosel output for the solution after running for 30 minutes.

```
The total number of first-year students: 582
The total number of seminars: 41
Number of female students: 339
Number of male students: 243
================================================
All students were assigned
================================================
Gender Penalty is: 314
================================================
The largest gender gap is: 11
The second largest gender gap is: 3
The third largest gender gap is: 3
================================================
Seminar 6 has 15 students with 7 males and 8 females.
Seminar 15 has 14 students with 6 males and 8 females.
Seminar 17 has 14 students with 6 males and 8 females.
Seminar 19 has 14 students with 6 males and 8 females.
Seminar 22 has 14 students with 6 males and 8 females.
Seminar 26 has 15 students with 6 males and 9 females.
Seminar 36 has 15 students with 6 males and 9 females.
Seminar 35 has 15 students with 6 males and 9 females.
Seminar 10 has 14 students with 6 males and 8 females.
Seminar 5 has 14 students with 6 males and 8 females.
Seminar 2 has 15 students with 6 males and 9 females.
Seminar 37 has 15 students with 6 males and 9 females.
Seminar 28 has 14 students with 6 males and 8 females.
Seminar 11 has 13 students with 5 males and 8 females.
Seminar 4 has 14 students with 6 males and 8 females.
Seminar 27 has 15 students with 7 males and 8 females.
Seminar 23 has 14 students with 6 males and 8 females.
Seminar 21 has 14 students with 6 males and 8 females.
Seminar 31 has 15 students with 7 males and 8 females.
Seminar 39 has 14 students with 6 males and 8 females.
Seminar 34 has 14 students with 6 males and 8 females.
Seminar 30 has 15 students with 6 males and 9 females.
Seminar 1 has 14 students with 6 males and 8 females.
Seminar 24 has 14 students with 6 males and 8 females.
Seminar 32 has 14 students with 6 males and 8 females.
Seminar 40 has 14 students with 6 males and 8 females.
Seminar 20 has 14 students with 6 males and 8 females.
Seminar 7 has 15 students with 6 males and 9 females.
Seminar 16 has 14 students with 6 males and 8 females.
Seminar 14 has 14 students with 6 males and 8 females.
Seminar 12 has 13 students with 5 males and 8 females.
Seminar 33 has 14 students with 6 males and 8 females.
Seminar 9 has 14 students with 6 males and 8 females.
Seminar 25 has 15 students with 7 males and 8 females.
Seminar 3 has 13 students with 1 males and 12 females.
Seminar 8 has 14 students with 6 males and 8 females.
Seminar 41 has 14 students with 6 males and 8 females.
Seminar 38 has 14 students with 6 males and 8 females.
Seminar 18 has 14 students with 6 males and 8 females.
Seminar 29 has 14 students with 6 males and 8 females.
Seminar 13 has 14 students with 6 males and 8 females.
```

Figure 3.8: Output from the Mosel code obtained by using the gender model to solve 2009 data (after 30 minutes)

3.6. Simulation

In Section 2.4 we performed a simulation to study how the number of seminars that students choose affects our ability to make an assignment. We found that if the college only required students to select as few as four courses, it was still extremely likely that we could find an assignment of students to seminars. However, this simulation did not take gender into consideration. While finding an assignment might be possible if we reduce the number of required seminar selections, it might be much more difficult to balance the genders of the classes.

In this section we perform a new simulation that takes gender into consideration. The idea is male and female students will tend to have different interests in choosing seminars. Hence, their probability distribution for seminar popularity will differ.

We created a probability model for male seminars first, and then created a probability model for female seminars by "aligning" the probability of each group of female seminars based on their corresponding male seminar probability.

3.6.1. Model for the Class Entering In 2010

To begin with, we repeated the exact same process that we did with the simulation of the basic assignment model in Section 2.4. However, instead of modeling the probability distribution of seminar popularity *among all students*, we first model the probability distribution of seminar popularity *among male students*. The model for male students is provided in Figure 3.9 and Table 3.4 below.

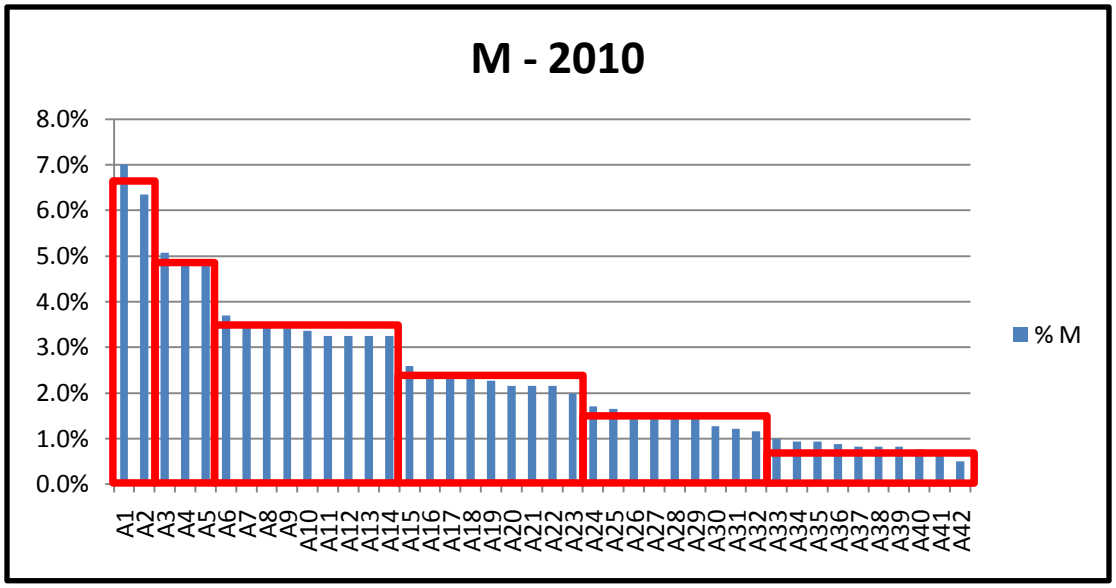


Figure 3.9: Probability model for 2010 male seminar popularity

Table 3.4: Probability model for 2010 male seminar popularity

2010 MALE SEMINAR	Probability of each male seminar getting picked
1 - 2	6.7%
3 – 5	4.9%
6 – 14	3.4%
15 – 23	2.3%
24 – 32	1.4%
33 – 42	0.8%

After creating the model for male seminar popularity, we analyzed the distribution of seminar popularity *among female students*, which is shown in the figure below. Note that the seminar names in Figure 3.10 correspond to those of Figure 3.9, which means this distribution is arranged according to how popular the seminars are among male students, where A1 is the most popular.

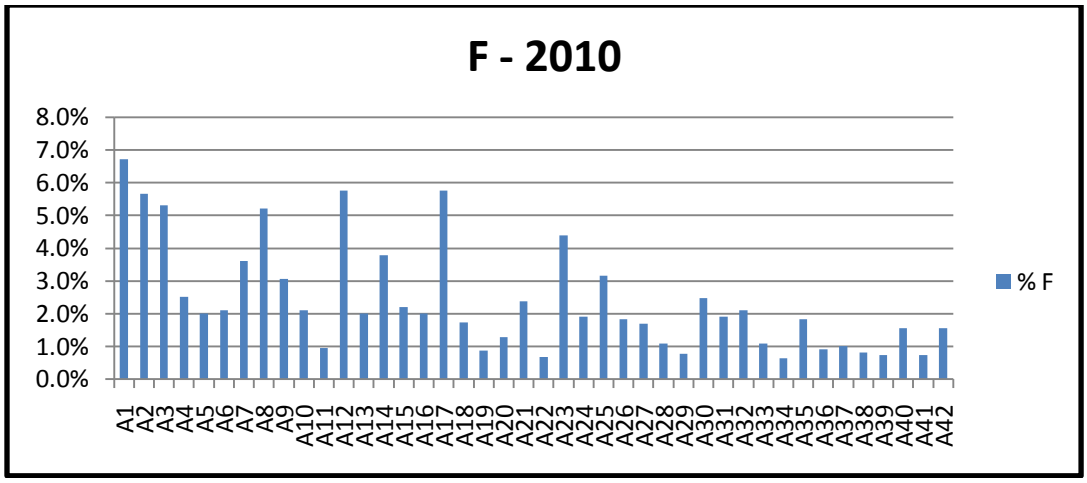


Figure 3.10: Distribution of 2010 female seminar popularity

By comparing Figures 3.9 and 3.10, we can see that there are distinct differences between males and females in terms of selecting seminars. While Seminar A1 was the most popular for both males and females, Seminars A12 and A17 are much more popular among females, for example.

The model we created for female seminar popularity is presented in Figure 3.11 and Table 3.5. Note that we rearranged the seminars so that in the graph the most popular seminar among females is listed first, the second most popular is listed next, and so on.

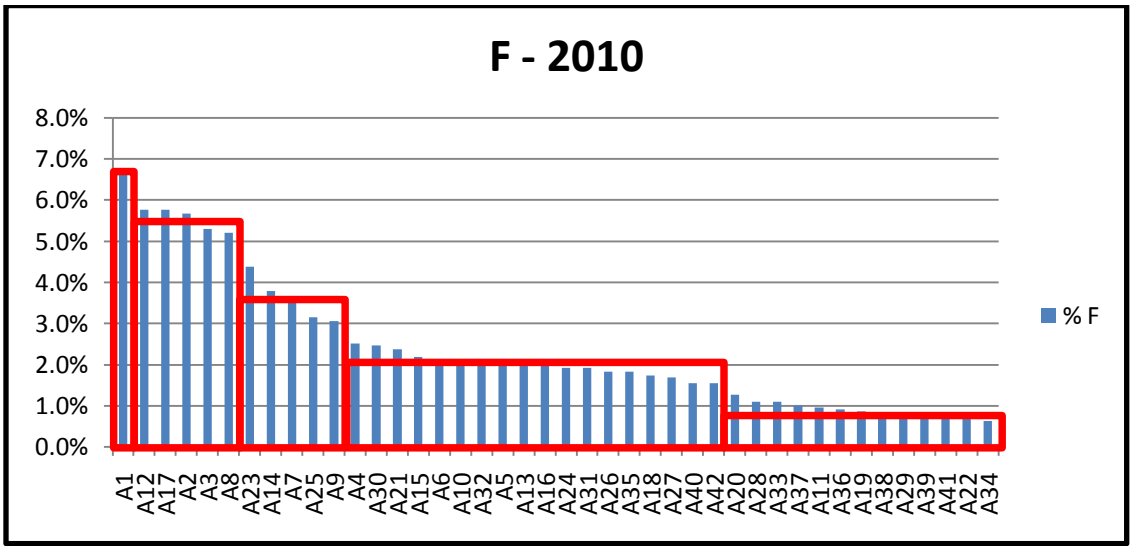


Figure 3.11: Rearranged and grouped version of Figure 3.10

Table 3.5: Probability model for 2010 female seminar popularity

2010 FEMALE SEMINAR	Probability of each female seminar getting picked
1	6.7%
12, 17, 2, 3, 8	5.5%
23, 14, 7, 25, 9	3.6%
4, 30, 21, 15, 6, 10, 32, 5, 13, 16, 24, 31, 26, 35, 18, 27, 40, 42	2.0%
20, 28, 33, 37, 11, 36, 19, 38, 29, 39, 41, 22, 34	0.9%

3.6.2. Model for the Class Entering In 2009

The model for male students starting in 2009 is provided in Figure 3.12 and Table 3.6 below.

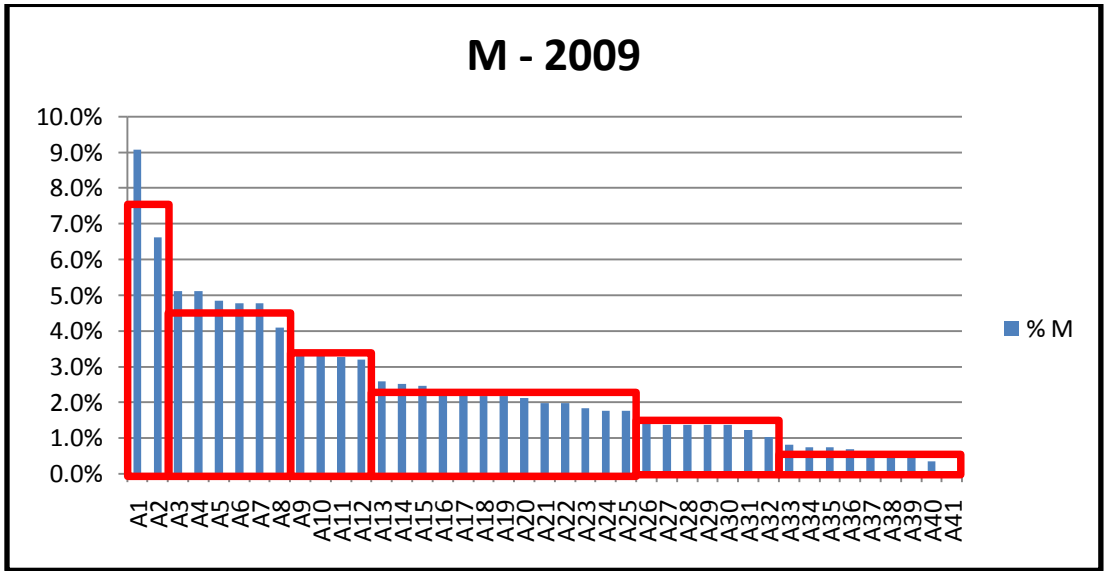


Figure 3.12: Probability model for 2009 male seminar popularity

Table 3.6: Probability model for 2009 male seminar popularity

2009 MALE SEMINAR	Probability of each male seminar getting picked
1 - 2	7.8%
3 – 8	4.8%
9 – 12	3.3%
13 – 25	2.2%
26 – 32	1.3%
33 – 41	0.6%

The distribution of female seminar popularity is provided in Figure 3.13, with the actual probability model provided in Figure 3.14 and Table 3.7.

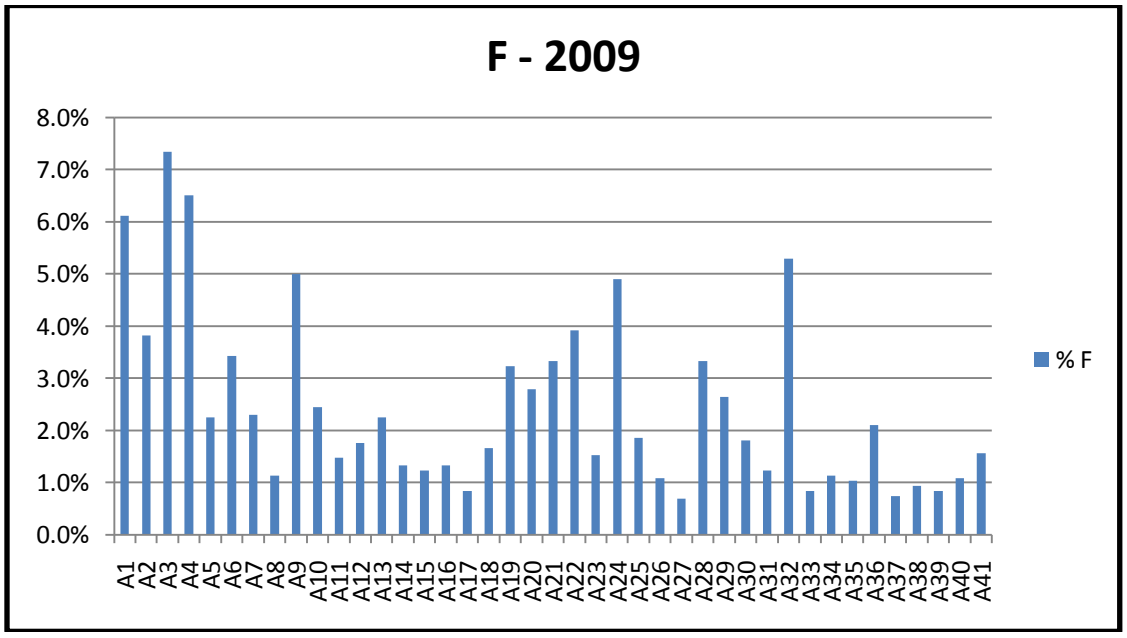


Figure 3.13: Distribution of 2009 female seminar popularity

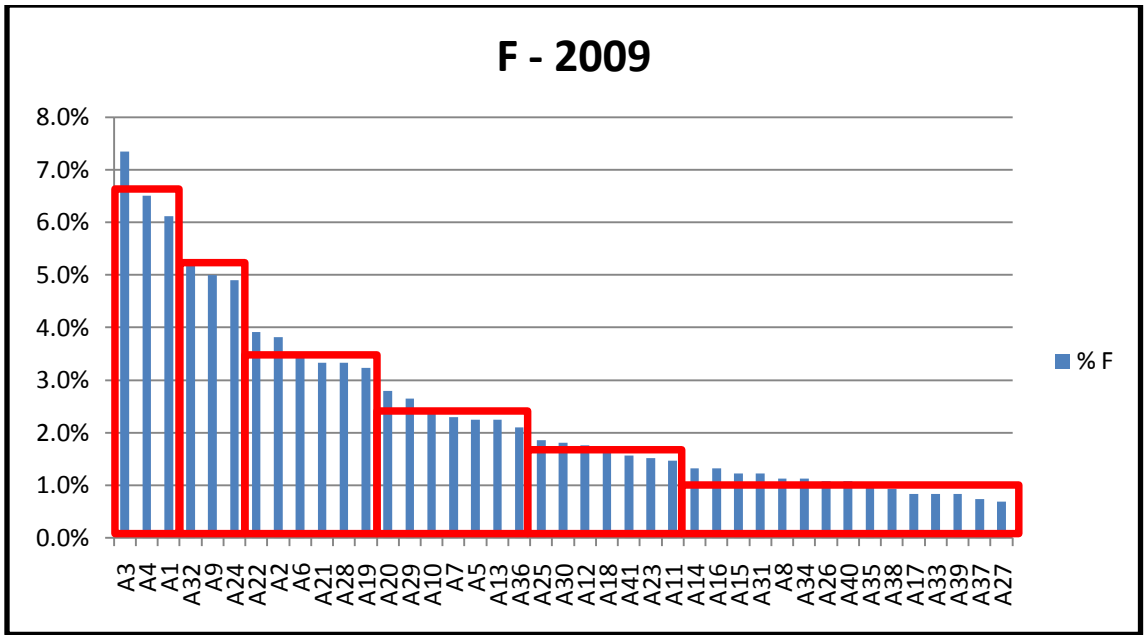


Figure 3.14: Rearranged and grouper version of Figure 3.13

Table 3.7: Probability model for 2009 female seminar popularity

2009 FEMALE SEMINAR	Probability of each female seminar getting picked
3, 4, 1	6.7%
32, 9, 24	5.1%
22, 2, 6, 21, 28, 19	3.5%
20, 29, 10, 7, 5, 13, 36	2.4%
25, 30, 12, 18, 41, 23, 11	1.7%
14, 16, 15, 31, 8, 34, 26, 40, 35, 38, 17, 33, 39, 37, 27	1.0%

3.6.3. Results and Observations

We performed a simulation using the 2010 and 2009 models created in the last section. For the 2010 simulation we generated a simulated seminar selection data set using the following steps 665 times (one for each student in 2010):

1. Randomly assign a gender to each student using the actual gender percentages in 2010, i.e., $P(\text{male}) = 301/665 \approx 0.45$ and $P(\text{female}) = 364/665 \approx 0.55$.
2. If the student is a male, randomly select x seminars using the male model in Table 3.3. Otherwise, randomly select x seminars using the female model in Table 3.4.
3. Set seminar capacities to 16 (which is the actual standard class size in 2010).

For the 2009 simulation we generated a simulated seminar selection data set using the following steps 582 times (one for each student in 2009):

1. Randomly assign a gender to each student using the actual gender percentages in 2009, i.e., $P(\text{male}) = 243/582 \approx 0.42$ and $P(\text{female}) = 339/582 \approx 0.58$.
2. If the student is a male, randomly select x seminars using the male model in Table 3.5. Otherwise, randomly select x seminars using the female model in Table 3.6.
3. Set seminar capacities to 15 (which is the actual standard class size in 2009).

For our simulation, we are not interested in determining if an assignment exists for a particular number of seminar selections, which was performed in Section 2.4, rather we are interested to see how gender balancing comes into play when we reduce the number of selections.

For both 2009 and 2010, we randomly generated ten different seminar selection data sets (we refer to each data set as a "trial"), and solved each using the gender-balancing model of Section 3.4 when students chose 6, 5, and 4 seminars. For each simulation, we terminated the optimizer after 4 minutes, and the top three largest gender gaps were recorded. Our results are presented in the table below.

The first column "Year" tells what year of data the model is based on. The second column "Number of Seminars Picked" tells how many seminars we assume each student is

required to pick. These two columns represent the framework of assumptions on which we based our simulations. The third column "Trial #" tells the order (within each of the 10 simulations) of the trial we are looking at. The three rightmost columns indicate the gender gaps of 3 seminars that have the largest gender gap among all the seminars within each trial.

For example, the second row can be read as "Based on 2010 data, if each student picks 6 seminars, in the first trial we obtained the 3 largest gender gaps as 2, 2, and 2."

Table 3.8: Results for running simulation of gender models based on 2010, 2009, and 2008 data.

Year	Number of Seminars Picked	Trial #	Largest Gender Gap	2nd Largest Gender Gap	3rd Largest Gender Gap
2010	6	1	2	2	2
		2	4	4	2
		3	2	2	2
		4	2	2	2
		5	2	2	2
		6	2	2	2
		7	2	2	2
		8	2	2	2
		9	2	2	2
		10	2	2	2
	5	1	2	2	2
		2	2	2	2
		3	2	2	2
		4	2	2	2
		5	2	2	2
		6	6	2	2
		7	2	2	2
		8	4	4	2
		9	6	2	2
		10	4	4	4
	4	1	4	2	2
		2	6	2	2
		3	6	6	4
		4	2	2	2
		5	6	2	2
		6	10	4	2
		7	Not All Students Were Assigned		
		8	10	2	2
		9	Not All Students Were Assigned		
		10	2	2	2
2009	6	1	3	3	3
		2	5	3	3
		3	2	2	2
		4	3	3	3
		5	3	3	3
		6	7	5	5
		7	5	3	3
		8	5	4	4
		9	5	3	3
		10	5	2	2
	5	1	5	3	3

		2	3	3	3
		3	7	4	3
		4	3	3	3
		5	7	3	3
		6	5	3	3
		7	5	3	3
		8	5	3	3
		9	11	7	3
		10	3	3	3
	4	1	7	5	5
		2	7	5	5
		3	Not All Students Were Assigned		
		4	9	7	5
		5	Not All Students Were Assigned		
		6	Not All Students Were Assigned		
		7	5	5	5
		8	9	5	5
		9	9	7	7
		10	11	3	3

From Table 3.8 we can see that for each year, as we reduce the number of required seminar choices, it tends to increase the size of the top three largest gender gaps. For example, in 2010, the largest gender gap over all trials is 4 when *Number of Seminars Picked* is 6, but the largest gender gap over all trials is 6 when *Number of Seminars Picked* is 5. This trend is even more pronounced in the 2009 data.

Interestingly, there were 2 trials in 2010 and 3 trials in 2009 where there were no possible assignments when the students only choose 4 seminars. This is curious because in our simulations in Section 2.4 that did not consider gender, we never had a situation where there was not a feasible assignment when students choose 4 seminars, even out of 200 trials. It is difficult to say exactly why this is the case, however, it should be clear that utilizing two different models (one for males and one for females) would result in a seminar selection data set that is different when using a single model. This suggests that reducing the number of seminar choices to 4 selections might not be a good idea in that an assignment might not in fact exist.

In conclusion, while the college could consider reducing the number of required seminar selections, it would likely make it more difficult to find a good gender balance.

Chapter 4

THE ASSIGNMENT MODEL WITH GENDER AND INTERNATIONAL STUDENT CONSIDERATIONS

Given the emphasis on global education at Dickinson College, it is critical that each first year seminar is assigned with some international students who bring along with them a different range of perspectives. In our model, we define someone as an "international student" if he/she is not a U.S. citizen, so this includes U.S. permanent residents, dual citizens, and international students. While this classification is somewhat arbitrary, we felt that this was sufficient in terms of attempting to diversify perspectives.

4.1. Graphical Representation with Gender and International Students

We begin by presenting a new max-flow network that will allow us to take both gender and citizenship status into consideration when making assignments. This graph is presented in the figure below.

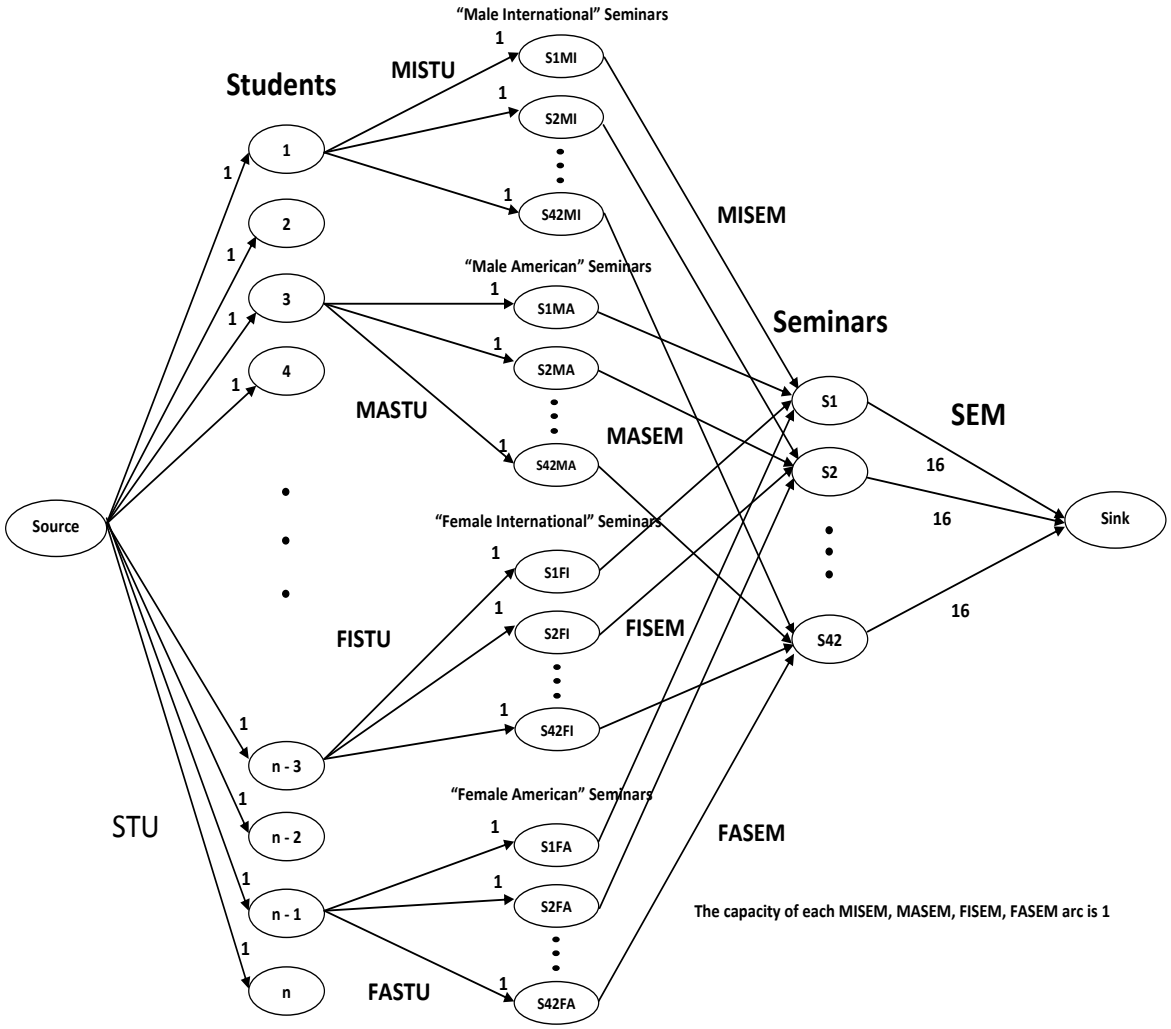


Figure 4.1: The graphical presentation of our assignment model with gender and international student consideration

This graph can be thought of as a modification of the network in Figure 2.1, where there are four new sets of nodes, instead of two like in Figure 3.1: "Male American" seminars nodes, "Male International" seminar nodes, "Female American" seminar nodes and "Female International" seminar nodes. They are denoted as S_iMA, S_iMI, S_iFA, and S_iFI, respectively, where *i* is the index of the seminar. If a student *x* is a male international student, there would be six arcs with each having a capacity of 1 coming from node Student(*x*) to six corresponding S_iMI nodes. Similarly, if a student *y* is a female American student, there would be six arcs with each having a capacity of 1 coming from node Student(*y*) to six

corresponding S_iFA nodes. Moreover, there is a flow from each of these "Male American," "Male International," "Female American," and "Female International" seminars nodes to their corresponding Seminar node. For instance, there is a flow $MASEM_{S1MA,S1}$ from S_1MA to S_1, a flow $FASEM_{S1FA,S1}$ from S_1FA to S_1, a flow $MISEM_{S1MI,S1}$ from S_1MI to S_1, a flow $FISEM_{S1FI,S1}$ from S_1FI to S_1, etc. The flows from node S_iMA and node S_iFA to node S_i tells us how many male and female American students there are in seminar S_i, respectively. Similarly, the flows from node S_iMI and node S_iFI to node S_i tells us how many male and female international students there are in seminar S_i, respectively. Thus, these four new sets of nodes allow us to keep track of the gender ratio and U.S. citizenship ratio in each seminar.

4.2. Mathematical Programming Representation with Gender and International Students

We begin by providing the linear programming constraints for the max-flow model in Figure 4.1. Specifically, the max-flow problem is subject to the following constraints.

$$STU_i - \sum_j MISTU_{i,j} = 0 \text{ for all Male International students } i \quad (4.1)$$

$$STU_i - \sum_j FISTU_{i,j} = 0 \text{ for all Female International students } i \quad (4.2)$$

$$STU_i - \sum_j MASTU_{i,j} = 0 \text{ for all Male American students } i \quad (4.3)$$

$$STU_i - \sum_j FASTU_{i,j} = 0 \text{ for all Female American students } i \quad (4.4)$$

$$\sum_i MISTU_{i,j} - MISEM_j = 0 \text{ for all Seminar } j \quad (4.5)$$

$$\sum_i FISTU_{i,j} - FISEM_j = 0 \text{ for all Seminar } j \quad (4.6)$$

$$\sum_i MASTU_{i,j} - MASEM_j = 0 \text{ for all Seminar } j \quad (4.7)$$

$$\sum_i FASTU_{i,j} - FASEM_j = 0 \text{ for all Seminar } j \quad (4.8)$$

$$MASEM_j + FASEM_j + MISEM_j + FISEM_j - SEM_j = 0 \text{ for all Seminar } j \quad (4.9)$$

$$0 \le MISTU_{i,j} \le 1 \text{ for all Male International Student } i \text{ and Seminar } j \quad (4.10)$$

$$0 \le FISTU._{i,j} \le 1 \text{ for all Female International Student } i \text{ and Seminar } j \quad (4.11)$$

$$0 \le MASTU_{i,j} \le 1 \text{ for all Male American Student } i \text{ and Seminar } j \quad (4.12)$$

$$0 \le FASTU_{i,j} \le 1 \text{ for all Female American Student } i \text{ and Seminar } j \quad (4.13)$$

$$0 \le STU_i \le 1 \text{ for all Student } i \quad (4.14)$$

$$MISEM_j \le 16 \text{ for all Seminar } j \quad (4.15)$$

$$FISEM._j \le 16 \text{ for all Seminar } j \quad (4.16)$$

$$MASEM_j \le 16 \text{ for all Seminar } j \quad (4.17)$$

$$FASEM_j \le 16 \text{ for all Seminar } j \quad (4.18)$$

$$SEM_j \le 16 \text{ for all Seminar } j \quad (4.19)$$

$$\sum_j SEM_j = v \quad (4.20)$$

$$\sum_i STU_i = v \quad (4.21)$$

In this model,

STU_i represents the amount of flow coming out from node Student i,

$FASTU_{ij}$ = 1 if Female American Student i chose Seminar j,

$FISTU_{ij}$ = 1 if Female International Student i chose Seminar j,

$MASTU_{ij}$ = 1 if Male American Student *i* chose Seminar *j*,

$MISTU_{ij}$ = 1 if Male International Student *i* chose Seminar *j*,

$MISEM_j$ equals the number of male international students enrolled in Seminar *j*,

$MASEM_j$ equals the number of male American students enrolled in Seminar *j*,

$FISEM_j$ equals the number of female international students enrolled in Seminar *j*,

$FASEM_j$ equals the number of female American students enrolled in Seminar *j*, and

SEM_j equals the number of students enrolled in Seminar *j*.

It follows that,

$MISEM_j + FISEM_j$ equals the number of international students enrolled in Seminar *j*,

$MASEM_j + FASEM_j$ equals the number of Americans students enrolled in Seminar *j*,

$MISEM_j$ + M$ASEM_j$ equals the number of male students enrolled in Seminar *j*, and

F$ISEM_j + FASEM_j$ equals the number of female students enrolled in Seminar *j*.

Note that the model above does not actually make assignment decisions based on gender and U.S. citizenship, but does allow us to keep track of the gender and U.S. citizenship of students assigned to seminars. Since the nodes representing STUDENTS, "MALE INTERNATIONAL" SEMINARS, "FEMALE INTERNATIONAL" SEMINARS, "MALE AMERICAN" SEMINARS, "FEMALE AMERICAN" SEMINARS, and SEMINARS are transshipment nodes, constraint equations (4.1) – (4.9) are needed to satisfy the requirement of a transshipment node, i.e., the amount of flow into a node must be equal to the amount of flow out of that node. Similarly, constraint equations (4.20) and (4.21) are needed to satisfy condition of the source and sink node. The amount of flow coming out of the source ($\sum_i STU_i$) must be equal to the amount of flow coming into the sink ($\sum_j SEM_j$); we denote this amount as *v*. Obviously, we want to maximize *v*. The set of inequalities (4.10) – (4.19) are capacity constraints; (4.14) ensures that each student is assigned to at most 1

seminar, and the set of inequalities (4.10) – (4.13) make sure all the flow (choice) is nonnegative and one student can pick a seminar once. We need (4.15) – (4.19) because each seminar can have at most 16 students.

We now need to develop an appropriate objective function that will help us to achieve our goals of balancing both gender and citizenship status. Recall that in Section 3.4 we defined the Gender Penalty (GP) as the quantity $\sum_j (MSEM_j - FSEM_j)^2$, where $MSEM_j$ and $FSEM_j$ represented the number of males and females in Seminar j, respectively. The analogous gender penalty function for the model in Figure 4.1 is

$$\text{GP} = \sum_j [(MISEM_j + MASEM_j) - (FISEM_j + FASEM_j)]^2 ,$$

where $(MISEM_j + MASEM_j)$ now represents the number of males in Seminar j and $(FISEM_j + FASEM_j)$ represents the number of females in Seminar j. Note however that we are also interested in balancing the number of international students in each seminar. Toward this end, we define the Nationality Penalty (NP) as

$$\text{NP} = \sum_j [(MISEM_j + FISEM_j) - (MASEM_j + FASEM_j)^2 ,$$

where $(MISEM_j + FISEM_j)$ represents the number of international students in Seminar j and $(MASEM_j + FASEM_j)$ represents the number of American students in Seminar j.

It seems clear that we are now interested in minimizing both the Gender Penalty and the Nationality Penalty. However, it is quite possible that these two criteria, balancing gender and balancing the number of international students, are in conflict with each other. This is what is known as a *multi-criteria optimization problem*.

A typical strategy for handling problems of this form is to simply take a linear combination of the two objective functions, Gender Penalty and Nationality Penalty, using weights that correspond to the importance of each criterion. For example, we can define our objective function as

$$c_1 \sum_j [(MISEM_j + MASEM_j) - (FISEM_j + FASEM_j)]^2 \; + \\ c_2 \sum_j [(MISEM_j + FISEM_j) - (MASEM_j + FASEM_j)^2 \,,$$

where the weights c_1 and c_2 are chosen depending on the order of importance of the two criteria. For our particular problem, we decided that balancing the gender of the seminars was the top priority, while balancing the number of international students was secondary. For this reason, we decided to set c_1 = 100 and c_2 = 1, yielding the following quadratic objective function (which we will be minimizing):

$$100 \sum_j [(MISEM_j + MASEM_j) - (FISEM_j + FASEM_j)]^2 \; + \\ \sum_j [(MISEM_j + FISEM_j) - (MASEM_j + FASEM_j)^2 \qquad (4.22)$$

Note our choice of weights is somewhat arbitrary, but seemed to work well in practice. Also, we mention that this quadratic objective function is convex. Furthermore, we will need to enforce that all of the variables are integer, which will make this problem challenging to solve given the large number of variables and constraints in the model. In fact this problem has 10 variables and 21 constraints, which are 8 more constraints and 4 more variables than the gender model. As a result, this model is harder to solve compared to the gender model.

4.3. Solutions to Previous Years Data

We implemented the assignment model of the previous section, utilizing the quadratic objective function (4.22), and solved for the 2009 and 2010 data set. The results for the 2010 selection data set are presented in the table below, where the solver was terminated after 1, 2, 3, 4, 5, 30, and 120 minutes. Note that the seminar capacity was set to 16 as was done during that year.

Table 4.1: 2010 data

Number of Minutes the Model Was Run	Gender Penalty	Nationality Penalty	Number of Seminars with ___ International Students			
			0	1	2	3
1	187	6707	2	5	28	7
2	125	6683	0	6	30	6
3	125	6683	0	6	30	6
4	123	6655	0	5	32	5
5	123	6655	0	5	32	5
30	119	6649	0	4	34	4
120	119	6651	0	2	38	2

It is interesting to observe how the solution improves as the solver is allowed to work longer, indicated by the smaller values of Gender Penalty and Nationality Penalty. Note that by 30 minutes, the solver was able to achieve a Gender Penalty of 119, which is the same value reached in Section 3.5, where we did not take into consideration the number of international students.

It is also interesting to note how the solver is able to more evenly distribute the number of international students in the seminars as it is allowed to work longer. In fact, by 120 minutes the optimizer has managed to distribute the international students so that 38 seminars have 2 international students, 2 seminars have 1 international student, and 2 seminars have 3 international students. Thus, the international students are almost perfectly

evenly distributed among the 42 seminars. We also mention that the Nationality Penalty actually increased when the solution time was increased from 30 minutes to 120 minutes. We attribute this to round-off error. The actual Mosel output for the solution achieved after 120 minutes is provided in the figure below.

```
The total number of first-year students: 665
Number of female students: 364
Number of male students: 301
Number of American students: 581
Number of international students: 84
================================================
All students were assigned
================================================
Gender Penalty is: 119
Nationality Penalty is: 6651
================================================
Seminar 01 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 16 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 42 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 27 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 34 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 25 has 15 students enrolled: 7 males and 8 females; 14 Americans and 1 international
Seminar 24 has 16 students enrolled: 8 males and 8 females; 14 Americans and 2 international
Seminar 17 has 15 students enrolled: 7 males and 8 females; 14 Americans and 1 international.
Seminar 04 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 40 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 08 has 16 students enrolled: 8 males and 8 females; 13 Americans and 3 international.
Seminar 03 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 18 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 39 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 37 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 13 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 28 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 36 has 15 students enrolled: 7 males and 8 females; 13 Americans and 2 international.
Seminar 35 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 06 has 16 students enrolled: 8 males and 8 females; 14 Americans and 2 international.
Seminar 05 has 15 students enrolled: 7 males and 8 females; 13 Americans and 2 international.
Seminar 23 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 29 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 31 has 16 students enrolled: 7 males and 9 females; 13 Americans and 3 international.
Seminar 22 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 20 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 12 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 11 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 07 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 26 has 16 students enrolled: 8 males and 8 females; 14 Americans and 2 international.
Seminar 32 has 16 students enrolled: 8 males and 8 females; 14 Americans and 2 international.
Seminar 21 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 10 has 16 students enrolled: 8 males and 8 females; 14 Americans and 2 international.
Seminar 41 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 14 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 38 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 43 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 33 has 15 students enrolled: 7 males and 8 females; 13 Americans and 2 international.
Seminar 30 has 15 students enrolled: 7 males and 8 females; 13 Americans and 2 international.
Seminar 02 has 16 students enrolled: 7 males and 9 females; 14 Americans and 2 international.
Seminar 09 has 16 students enrolled: 8 males and 8 females; 14 Americans and 2 international.
Seminar 15 has 15 students enrolled: 7 males and 8 females; 13 Americans and 2 international.
```

Figure 4.2: Mosel output obtained by using the gender and international student ratio to solve 2010 data

We next applied our technique to the data from the first-year students entering in 2009, where the seminar capacity was set to 15 as was actually done during that year. In addition, we also set the lower class limit to 13 as we did in Section 3.5. Our results are presented in the table below.

Table 4.2: 2009 data

Number of Minutes the Model Was Run	Gender Penalty	Nationality Penalty	Number of Seminars with ___ International Students					
			0	1	2	3	4	5
1	364	5886	2	18	18	2	1	0
2	322	5900	4	12	24	0	1	0
3	320	5890	1	18	21	0	1	0
4	320	5890	1	16	22	0	1	0
5	318	5892	3	14	23	0	1	0
30	316	5894	2	19	17	2	0	1
120	314	5908	2	17	20	1	1	0

There are a number of interesting observations that can be made from the table of results above. First, note that by 120 minutes the solver was able to reduce the Gender Penalty to 314, which is the value achieved in Section 3.5, when we did not take nationality into consideration. Second, the Nationality Penalty does not strictly decrease as the solver is allowed to work longer. This is not altogether surprising in that we weighted the Gender Penalty so that minimizing gender gap was a top priority. Therefore, the solver was able to make the Gender Penalty smaller at the expense of increasing the Nationality Penalty. Third, the solution with regards to the spread of international students is not quite as good as that in 2010. It appears that this is directly related to what seminars international students selected.

The actual Mosel output for the solution achieved after 120 minutes is provided in the figure below.

```
The total number of first-year students: 582
Number of female students: 339
Number of male students: 243
Number of American students: 518
Number of international students: 64
================================================
All students were assigned
================================================
Gender Penalty is: 316
Nationality Penalty is: 5894
================================================
Seminar 06 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 15 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 17 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 19 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 22 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 26 has 15 students enrolled: 6 males and 9 females;  13 Americans and 2 international.
Seminar 36 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 35 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 10 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 05 has 15 students enrolled: 6 males and 9 females;  13 Americans and 2 international.
Seminar 02 has 15 students enrolled: 6 males and 9 females;  13 Americans and 2 international.
Seminar 37 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 28 has 15 students enrolled: 7 males and 8 females;  13 Americans and 2 international.
Seminar 11 has 13 students enrolled: 5 males and 8 females;  13 Americans and 0 international.
Seminar 04 has 15 students enrolled: 6 males and 9 females;  13 Americans and 2 international.
Seminar 27 has 15 students enrolled: 6 males and 9 females;  12 Americans and 3 international.
Seminar 23 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 21 has 15 students enrolled: 6 males and 9 females;  12 Americans and 3 international.
Seminar 31 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 39 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 34 has 15 students enrolled: 6 males and 9 females;  13 Americans and 2 international.
Seminar 30 has 15 students enrolled: 6 males and 9 females;  13 Americans and 2 international.
Seminar 01 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 24 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 32 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 40 has 15 students enrolled: 6 males and 9 females;  14 Americans and 1 international.
Seminar 20 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 07 has 15 students enrolled: 7 males and 8 females;  14 Americans and 1 international.
Seminar 16 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 14 has 14 students enrolled: 6 males and 8 females;  14 Americans and 0 international.
Seminar 12 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 33 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 09 has 13 students enrolled: 6 males and 7 females;  12 Americans and 1 international.
Seminar 25 has 15 students enrolled: 7 males and 8 females;  13 Americans and 2 international.
Seminar 03 has 13 students enrolled: 1 males and 12 females; 12 Americans and 1 international.
Seminar 08 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 41 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
Seminar 38 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 18 has 13 students enrolled: 6 males and 7 females;  8 Americans and 5 international.
Seminar 29 has 14 students enrolled: 6 males and 8 females;  13 Americans and 1 international.
Seminar 13 has 14 students enrolled: 6 males and 8 females;  12 Americans and 2 international.
```

Figure 4.3: Mosel output obtained by using the gender and international student ratio to solve 2009 data

Chapter 5

CONCLUSIONS AND FUTURE WORK

5.1. Conclusions

We have developed several models that help to assign first-year students to seminars, including the *basic assignment model*, the *gender balance model*, and the *gender and international student balance model*. In addition, we also developed probabilistic models to simulate the seminar selections of first-year students based on the 2009 and 2010 data.

For the *basic assignment model*, we successfully achieved a solution for both the 2009 and 2010 data sets, which means every student was assigned a seminar out of their six choices. The solutions were obtained within seconds since our basic model is a fairly small linear program. In terms of the simulation, we determined that if the college reduces the number of seminar choices to as low as four, it is still highly likely that a full assignment can be achieved.

Our *gender balance* and *gender and international student balance models* are modifications of the *basic model*, in which the objectives are convex quadratic functions. In order to solve these models, we used the mixed-integer programming solver in Xpress. Given the large number of integer variables, we were not able to find an optimal solution to the 2009 and 2010 data sets within a reasonable amount of time. However, the solutions obtained after stopping the optimizer early were satisfactory from a practical standpoint.

For the entering class of 2010, the assignment we determined has every student assigned a seminar out of their six choices, and the largest gender gap among all seminars is

two. In addition, almost every seminar is assigned with two international students, while a few have one or three. In other words, we were not only able to assign students to one of their six choices, but we were also able to achieve a very good quality assignment in terms of gender and nationality balance.

Unfortunately, the assignment we determined for the 2009 data was not quite as good of quality as that for 2010. This appears to be due to the fact that males and females did not select seminars as evenly as they did in 2010. For example, recall a seminar only had one male student register for it. This was also true with regard to international students. This made it more challenging to find an assignment with a good balance of gender and numbers of international students. However, we still feel our model worked well, and that the solution was of good quality, at least in consideration of characteristics of the data.

In terms of our simulation that included gender, we determined that while the college could reduce the number of seminar selections without compromising our ability to find an assignment, it could make it much more difficult to balance gender. Therefore, we suggest that the college should keep the number of seminar choices at six.

5.2. Future Work

There is a wide range of topics that we could further explore. For instance, we could develop more sophisticated models when doing the simulation that take into account the *correlation* between seminars selections as mentioned in Section 2.4.3. This would require a good classification of the topic and theme of the seminars and of how seminars attract different genders.

Another topic we could explore is to balance other criteria such as region of the country, political affiliation, etc. These would be more challenging from a modeling

standpoint because there could be more than two classifications. For example, what does it even mean to say we want to balance political affiliation if we have say, three classifications: Democrat, Republican, and Independent?

Finally, we are also interested in tackling the problem where the first-year students submit a list of seminars that are ranked in order of preference. While the college does not make assignments in this manner (although it used to), it is an interesting problem nonetheless.

APPENDIX A: THE MAXIMUM FLOW PROBLEM

The maximum flow problem is one of the five most important network problems, which include the shortest-path problem, the minimum spanning tree problem, the minimum cost flow problem, and the problem of obtain the most economical way to complete a project within its deadline.

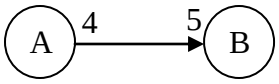
A general description of a typical maximum flow problem is as follows.

- The graph of a maximum flow problem is a directed graph where flow through an arc is in a direction indicated by its arrowhead and each arc has a capacity. (A directed path from node A to node B is a sequence of connecting arcs whose direction is toward node B, so that flow from node A to node B along this path is feasible.)
- All flow starts from one node, called the source, and end at another node, called the sink. All of the remaining nodes are transshipment nodes.
- At the source, all arcs point away from the source node. At the sink, all arcs point towards the sink node.
- The objective here is to maximize the amount of flow from the source to the sink regardless of the "distance" traveled. This objective v can be understood as either the amount of flow leaving the source or the amount of flow entering the sink.

There are a number of algorithms that can be used to solve the maximum flow problem. We will discuss the augmenting path algorithm, which was developed by Ford and Fulkerson in 1956 (Ford and Fulkerson, 1956). This algorithm is derived from two concepts:

residual network and augmenting path. Note that this algorithm assumes that the arc capacities are either integers or rational numbers.

Residual network can be understood as a "leftover" network after some flow has been assigned to the arcs. Specifically, it shows the residual capacities, which are the remaining arc capacities for assigning additional flow. To illustrate this, suppose we have two nodes A and B, and the arc going from A to B has a capacity of 9. If we assign a flow of 5 through this arc, then the residual capacity is 9 – 5 = 4 for any additional flow assignment through A → B. A graphical representation is as follows



The number on an arc next to a node tells the residual capacity for flow from that node to the other node. In the above example, the number 5 on the left next to node B means we can assign a flow of up to 5 units from node B back to node A.

To construct a residual network, the original directed network will be transformed into an undirected network, which means all the arcs now become undirected arcs. At first, the arc capacity in the original direction remains the same and the arc capacity in the opposite direction is zero. Once a flow of x units is assigned to an arc, x is then subtracted from the residual capacity in the same direction and added to the residual capacity in the opposite direction.

An augmenting path is a directed path from the source to the sink in the residual network in which the residual capacity of each arc is greater than zero. Since these residual capacities might be different, the minimum of them is the residual capacity of the augmenting path (the feasible amount of flow can be added to the entire path).

The augmenting path algorithm is a repetition of the following three steps

1. Identify an augmenting path by finding some directed path from the source to the sink such that all arcs on this path have positive residual capacity. In large networks, finding an augmenting path can be a difficult task. To deal with this, we can use a procedure called the fanning-out procedure. We first identify all nodes that are connected to the source by a single arc with a positive residual capacity. Then, for each of these nodes, we continue to identify all nodes that can be connected to them by a single arc with positive residual capacity. Note that these newly identified nodes are not reached before. Repeat this process with the new nodes as they are reached. In the end, we can identify possible paths, where residual capacities on all arcs are positive, from the source to the sink.
2. Identify the residual capacity *a* of this augmenting path. Add a flow of *a* to this path in the original network.
3. Decrease by *a* the residual capacity of each arc on this augmenting path. Increase by *a* the residual capacity of each arc in the opposite direction on this augmenting path.

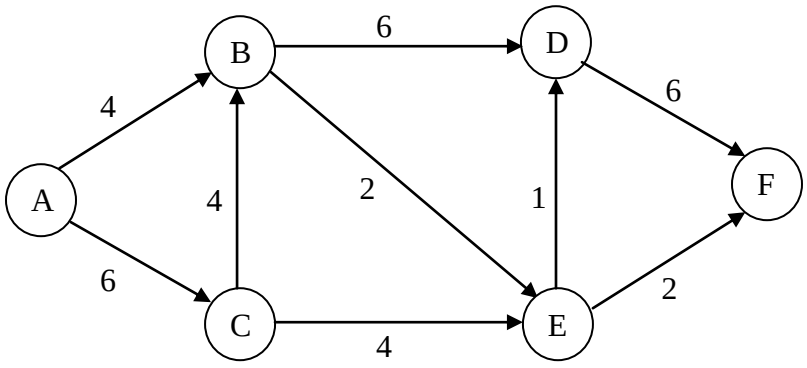
This process is repeated until there is no longer an augmenting path. This algorithm yields the optimal solution because it allows augmenting paths to cancel some previously assigned flow in the original network. Thus, we can always find a better combination of flow assignments despite some previous arbitrary selection of paths.

A famous theorem, known as the max-flow min-cut theorem, helps us recognize when optimality has been reached without extensive search for a nonexistent augmenting path. This theorem states that for all network with a single source and a single sink, the

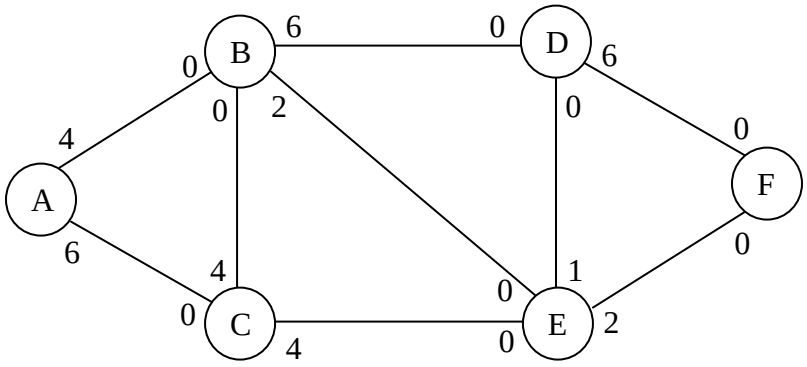
maximum feasible flow from the source to the sink equals the minimum cut value for all cuts of the network. There are two new definitions here: a cut is a set of directed arcs containing at least one arc from every directed path from the source to the sink. For each cut, the cut value is the sum of all residual capacities in the original direction of all arcs of that cut. Hence, to apply this theorem, we compare the amount of flow from the source to the sink for any feasible flow pattern with the minimum cut value we can find on the graph. If they are equal, we have found the optimal flow pattern.

Let's solve the following problem to illustrate the augmenting path algorithm and the max-flow min-cut theorem:

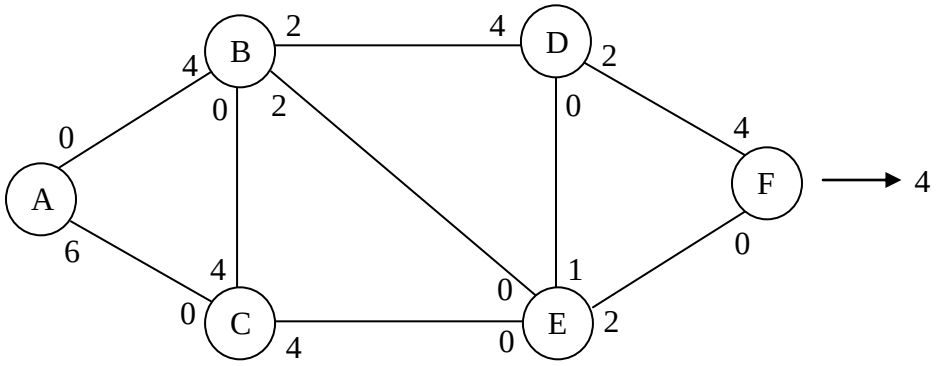
Given the following graph, maximize the flow from Node A (the source) to Node F (the sink):



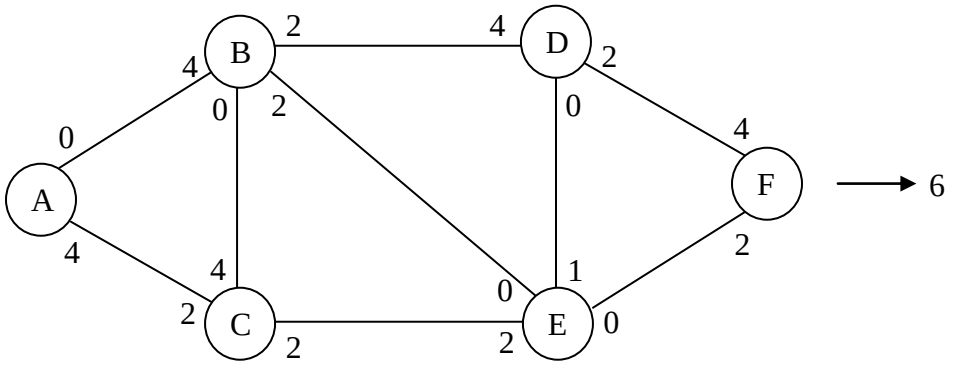
Step 1: Convert the original graph to an undirected graph



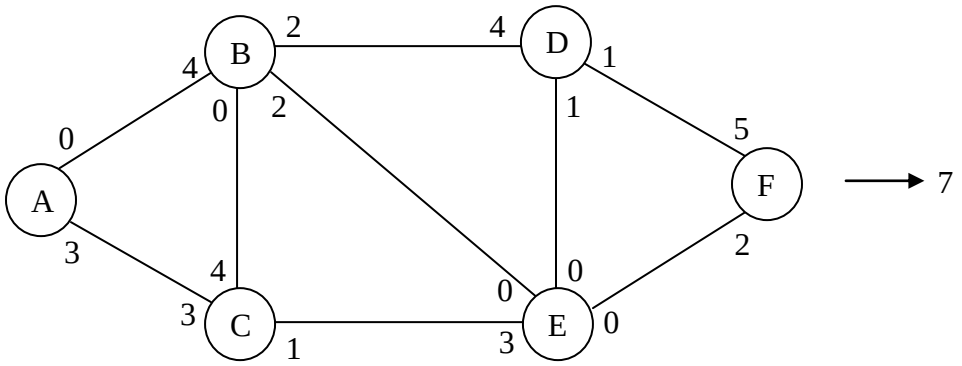
Step 2: A → B → D → F: 4



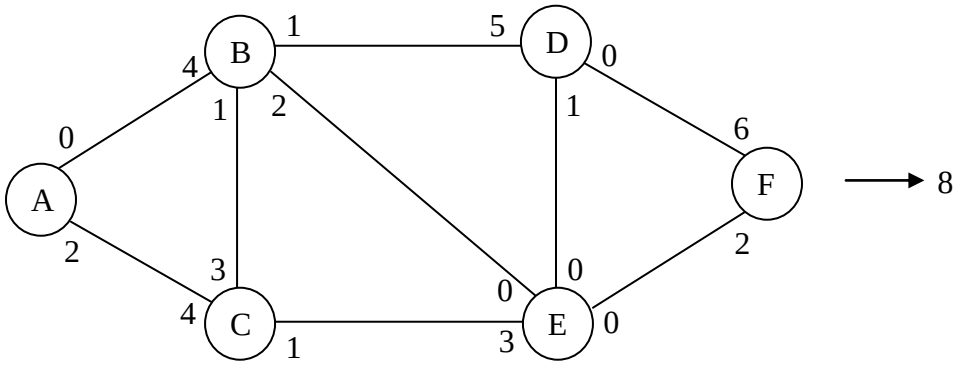
Step 3: A → C → E → F: 2



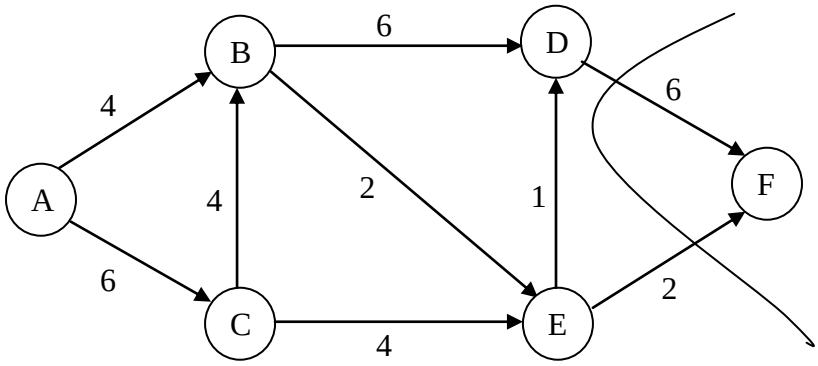
Step 4: A→ C →E → D → F: 1



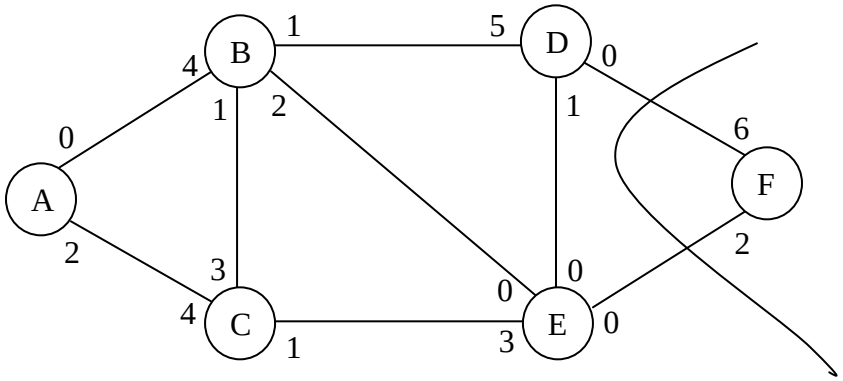
Step 5: A→ C →B → D → F: 1



By the max-flow min-cut theorem, we know that 8 is the optimal solution. Observe the following cut:



This cut has a value of 8, and it is the minimum cut value we can obtain from the given graph. Also notice that at step 5, the same cut gives us a cut value of 0, which means we have achieved the optimal result.



APPENDIX B: INTEGER PROGRAMS

For a simple IP problem where the feasible region is small and bounded, we could literally enumerate all the possible integer solutions, and then compare their objective values. If for example, the problem was a maximization problem, then simply choose the feasible solution that yields the largest value. While such a strategy might be theoretical possible, this technique is impractical for even small-sized integer programs because of the exponential number of feasible solutions with regards to the number of integer variables.

One of the ideas to solve an IP is to "relax" its integer constraints. The term "relax" here means that we would at first ignore all the integer restrictions and treat the IP as an linear program. This newly created LP is called the LP relaxation of the IP. It can be deduced that the feasible region of an LP is larger than and in fact contains that of an IP. It may be possible to solve the LP relaxation of the IP first, then round the solution to get the integer solution for the IP. This method, however, may yield an integer solution of poor quality or one that is not even contained in the feasible region.

One of the most widely used techniques for solving integer programs is called the branch-and-bound technique. The idea is to divide the original problem into smaller and smaller sub-problems until these sub-problems can be solved. Every branch-and-bound algorithm consists of three basic steps: branching, bounding, and fathoming (Hillier, & Lieberman, 2005). Branching is dividing the entire set of feasible solutions into smaller and smaller subsets. We then "bound" and "fathom" each resulting subset. Bounding can be considered as finding the upper bound for the best possible solution founded in a maximization problem. Fathoming is to dismiss a subset if its bound signals that the subset

cannot contain an optimal solution for the original problem. Thus, the algorithms *implicitly* enumerates certain solutions as opposed to explicitly enumerating them.

To illustrate the process of branching, bounding and fathoming, let us consider the following BIP problem from Hillier & Lieberman book:

Example: Branch and Bound

Maximize $Z = 9x_1 + 5x_2 + 6x_3 + 4x_4$

Subject to $6x_1 + 3x_2 + 5x_3 + 2x_4 \leq 10$

$x_3 + x_4 \leq 1$

$-x_1 + x_3 \leq 0$

$-x_2 + x_4 \leq 0$

x_j is binary for j = 1, 2, 3, 4

Branching: Since all the variables are binary, one straightforward way to branch the set of solutions is to assign fixed value to one of the variables, say x_1, to obtain two subsets, one of which contains $x_1 = 1$, and the other contains $x_1 = 0$. Hence, we have two new sub-problems as follows.

Sub-problem 1: Fix $x_1 = 0$

Maximize $Z = 5x_2 + 6x_3 + 4x_4$

Subject to $3x_2 + 5x_3 + 2x_4 \leq 10$

$x_3 + x_4 \leq 1$

$x_3 \leq 0$

$-x_2 + x_4 \leq 0$

x_j is binary, for j = 2, 3, 4

Sub-problem 2: Fix $x_1 = 1$

Maximize $Z = 9 + 5x_2 + 6x_3 + 4x_4$

Subject to $3x_2 + 5x_3 + 2x_4 \leq 4$

$x_3 + x_4 \leq 1$

$x_3 \leq 1$

$-x_2 + x_4 \leq 0$

x_j is binary, for j = 2, 3, 4

Note that some sub-problems can be fathomed immediately whereas the rest of the sub-problems require further branching by setting the next variable to be 0 or 1. In reality, it is important to determine the order of variables to be assigned fixed values. However, in this problem, for the purpose of simplicity, we will repeat this branching process in the natural order of x_1, x_2, x_3, and x_4.

Bounding: We need to layer-by-layer bound the original problem and its subsequent sub-problems, and the sub-problems of those sub-problems, and so on. In order to do this, we solve the LP relaxation of the IP. Recall that to obtain an LP relaxation of an IP, we simply

replace the constraint that x_j is binary for j = 1, 2, 3, 4 with $x_j \leq 1$ and $x_j \geq 0$. Then, using the simplex method we will obtain the following optimal solution to the LP relaxation of the original IP:

$(x_1, x_2, x_3, x_4) = (5/6, 1, 0, 1)$ with $Z = 16.5$

Since we want an integer solution and the coefficients are integers, $Z = 16$, not 16.5, is the bound for all feasible solutions to the original BIP problem. Next, we will find the bounds for the two sub-problems 1 and 2. Using the same process we obtain:

LP relaxation of sub-problem 1: $(x_1, x_2, x_3, x_4) = (0, 1, 0, 1)$ with $Z = 9$

LP relaxation of sub-problem 2: $(x_1, x_2, x_3, x_4) = (1, 4/5, 0, 4/5)$ with $Z = 16.5$

Thus, $Z \leq 9$ is the bound for sub-problem 1 and $Z \leq 16$ is the bound for sub-problem 2.

Fathoming: At the first branching and bounding round, we need to determine the *incumbent Z**, which is the best feasible solution found so far, at this step. In our example, *Z** at the first branching is 9, the bound for sub-problem 1, since the solution for the sub-problem 1 is integer. We can dismiss (fathom) a branch in one of the three following ways: if its bound is less than or equal to *Z**; if its LP relaxation has no feasible solutions; or if the optimal solution for its LP relaxation is integer. Returning to our example, we can now fathom Sub-problem 1 since its solution is integer and its bound is now the incumbent and proceed to branch Sub-problem 2 into two sub-problems 3 and 4 since its bound is greater than the incumbent. Sub-problem 3 will have x_2 fixed at 0 and sub-problem 4 will have x_2 fixed at 1. Bounding these two sub-problems we obtain the bounds for sub-problem 3 and sub-problem 4 are 13 and 16, respectively. However, since both of these bounds are greater than Z* and the other two fathoming tests fail, we need to determine which sub-problem to

be branched next. Since the bound for sub-problem 4 is greater than that of sub-problem 3, we will branch sub-problem 4 next. Repeat the three above steps until there are no remaining sub-problems and the optimal solution is the current incumbent. The optimal solution for this example is $(x_1, x_2, x_3, x_4) = (1, 1, 0, 0)$ with $Z = 14$.

Note that there are other techniques such as branch-and-bound algorithm for mixed integer programming or branch-and-cut algorithm for BIP that are not discussed here. Interested readers should refer to OR textbooks such as those of Winston or Hillier & Lieberman.

References

Books:

Hillier, F, & Lieberman, G. (2005). *Introduction to operations research*. New York, NY: McGraw-Hill.

Winston, W. (1994). *Operations research: applications and algorithms*. Belmont: Duxbury Press.

Winston, W, & Venkataramanan, M. (2003). *Introduction to mathematical programming*. Brooks/Cole.

Journals:

Ford, L, & Fulkerson, D. (1956). Maximal flow through a network. *Canadian Journal of Mathematics, 8,* 399-404.

On-line Journals:

Cipra, B. (2000). The best of the 20th century: editors name top 10 algorithms. *SIAM News, 33*(4). Retrieved 24 April, 2011 from http://www.siam.org/news/news.php?id=637

Web Site:

First-year seminar. (n.d.). Retrieved 29 September, 2010 from http://dickinson.edu/academics/first-year-programs/seminars/First-year-Seminar/