Math 353	Optimization Spring 2017	Lab 3

Facility Location

Pair Names: ____________________

1 Motivation

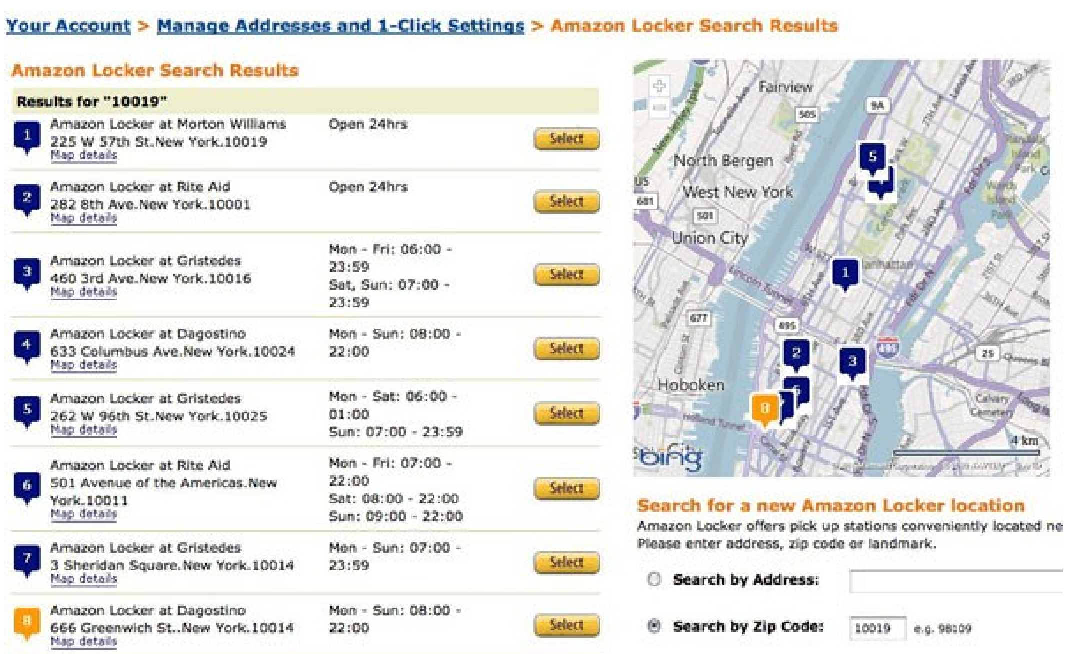
In Section 1 of the course packet, in the Sudoku and Graph Coloring problems from Homework, and in discussing the 3-SAT problem in lecture, we have seen examples where integrality was used in a powerful way: a $\{0,1\}$ -valued variable is used to describe either doing something or not in the solution.

Today we will explore the power of the integrality requirement by tackling a slightly larger problem facing Amazon.com. Navigate to the following article in WIRED Magazine (and read it):

`http://www.wired.com/business/2012/08/amazons-lockers-move-frontlines-of-retail-war-to-back-of-7-eleven/`

This article is from several years ago, and Amazon Lockers are now available in a wider range of cities. Also, Google subsequently purchased a San Francisco start-up called Buffer-Box which has a basically identical business plan.

Here is a map of Amazon Locker Locations in the 10019 area code (Manhattan, New York):



2 Lab 3: Facility Location

We will describe the discrete version of the Facility Location Problem which is sometimes called the "k-Center Problem."

Input to the Facility-Location Problem consists of:

- **Locations:** A set of N locations which we index by $i \in \{1, 2, ..., N\}$.
- **Opening Costs:** For each location i, a cost to open a facility at that location, c_i.
- **A Budget:** A maximum budget B available for opening facilities.
- **Distances:** A distance d_{ij} between each pair of locations.

The Objective is then to choose a set of locations to open facilities at (so that the openings cost at most B in total) such that the average distance from a randomly-chosen location to the nearest open facility is minimized.[1]

For now, assume that the cost of opening a facility is fixed at 1 regardless of location.

- Create your own small instance of this problem (say, one with 8 locations). You should specify the entire input.

- First, suppose that you are allowed to open 2 facilities, then suppose that you are able to open 3. How does the objective function change? Describe why this change takes place **in terms of the set of feasible solutions.**

[1] Assume this random choice has uniform probability across the N locations

- Draw a picture of a larger set of locations, using the Euclidean distance on the paper for your d_{ij}. Does it make sense that this problem is sometimes called the k-Center Problem? Similar models describe various types of data clustering problems, where given some distance metric on data points, the goal is to divide the data points into k clusters (so that within each cluster the data points are fairly similar).

2.1 Writing Facility Location as an IP

- What types of decisions do you have to make explicitly to describe a solution for the Facility Location Problem? What types of decisions are you making implicitly in the Facility Location Problem *(Hint: what will you need to know to figure out how well your solution is doing for the objective of minimizing average distance to a open facility)?*

  If you think about this for more than 10-15 minutes without making progress, discuss with your neighbors or with the Instructor.

- Using integrality, create appropriate decision variables.

- Formulate the constraints. There should be several types of constraints so that your decision variables take on their intended values (*Hint: think back to 3-SAT where decision variables related to the truth of clauses depended on the value of the decision variables for literals in that clause. You may also find it useful to think forward to the objective and come back to this part.*)

- Formulate the objective.

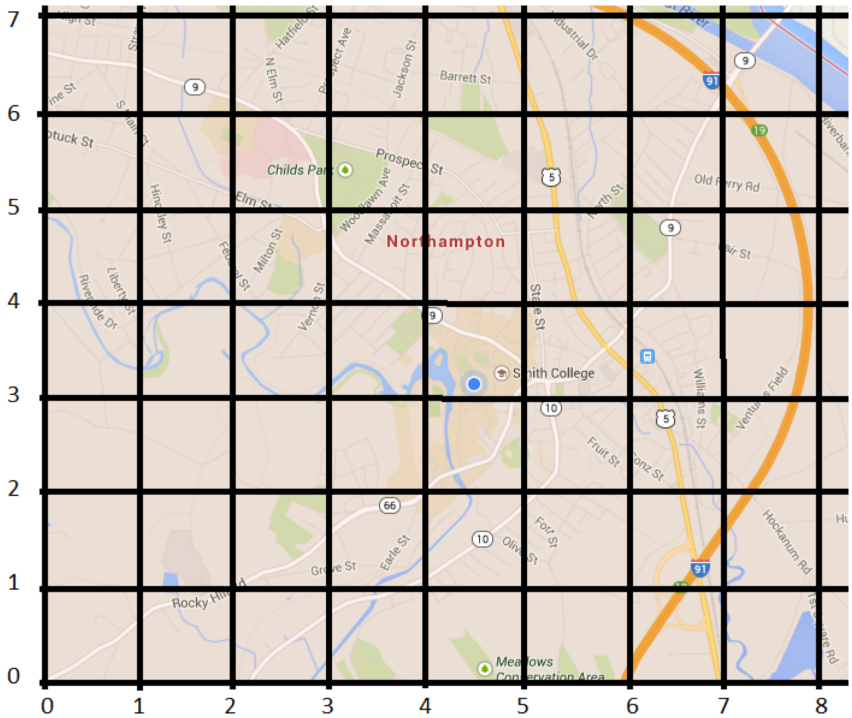
At this point, check in with the instructor before moving on.

2.2 Expanding Amazon's locker system to Northampton

Suppose that Amazon has decided to expand its locker system to Northampton (a real crime hotspot), and has hired you as a consultant.

The success of such a locker system will be determined by how convenient people find it to get to the nearest locker. Amazon has decided that it is willing to place 3 lockers around Northampton. It is your job to find the optimal placement of these lockers.

Assume that the possible locations are the 72 grid points below, and use Euclidean distance (as a crow flies) between grid points. For the objective, minimize the average distance from a uniformly-chosen location(grid point) to the nearest locker.



- In order to solve this IP we'll need data on the pairwise distances between every pair of grid points.

  Conveniently, the data fed into AMPL models need not always be specified explicitly: the values can be computed. Using this method is preferable to explicitly specifying a large matrix of distance values, particularly if we want to tinker with the objective later. For more information, read section 7.5 in the AMPL book at your leisure.

  For now, try to parse the meaning of the following relevant code:

  ```
  param distance {p1 in X, p2 in Y,v1 in X,v2 in Y} =
  sqrt((abs(p1-v1))^2+(abs(p2-v2))^2), >=0;
  ```

  Note that for computed parameters such a declaration goes in the `.mod` file.

- Using your work formulating the Facility Location Model as an IP, create a `.mod` and `.dat` file to compute the optimal placement for 3 lockers. To get started, notice that you will probably want to index each location by a pair that describes grid coordinates, rather than naming each of the 72 locations. **Be sure to specify CPLEX as your solver.**

  If you work for more than 30 minutes and are having issues getting you mod file to work, see the Instructor for prepared files.

  What is the optimal set of 3 lockers to open? What is the resulting average distance?

Draw the solution you found for 3 lockers onto your map of Northampton!

- Place 5 lockers. What happens to your objective function? Why?

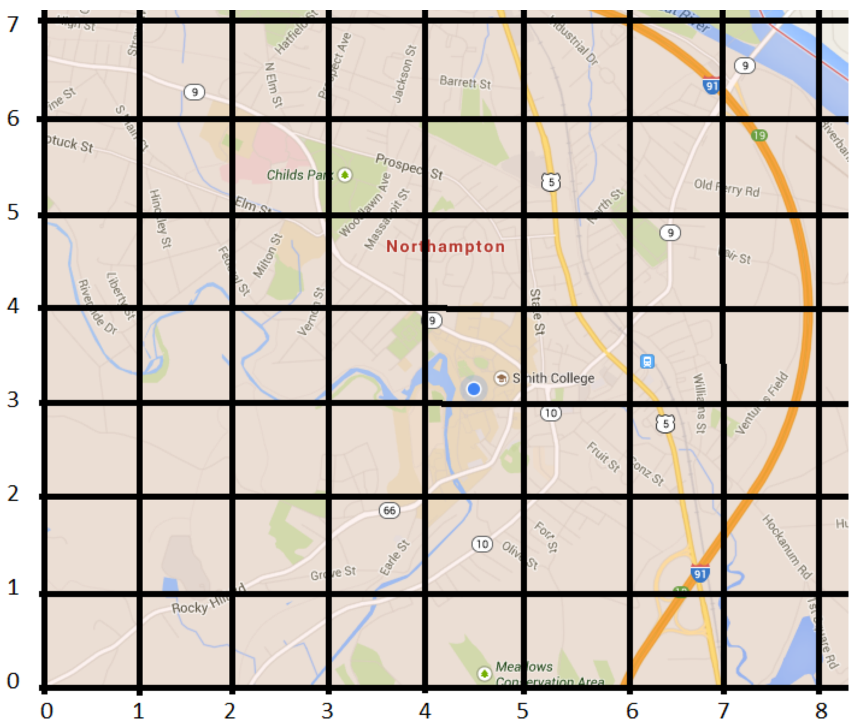
- Place 72 lockers. What is your optimal objective? Does this make sense?

- It seems unlikely that someone will be trying to reach an Amazon Locker from the middle of the Connecticut River. Suppose that you are given data that encodes the population distribution of Northampton. How might you change your existing objective to reflect "minimizing average distance to a locker"? Explain below.

- While you don't have a true population distribution for Northampton at hand, building on the previous part, introduce a parameter `popdist` to explore how much the location of the optimal solution changes as the population becomes noisily distributed. The following code may be helpful.

```
param popdist{i in X, j in Y}= max(Normal(1,0.8),0);
```

Choose a substantial standard deviation and optimize the placement of 3 lockers. Draw the optimal solution below. (If you do this repeatedly you should get different solutions, but you will not since AMPL will repeatedly use the same random seed generator. To reset the random seed type the command `option randseed 0;`). After solving for the new locker locations, display the parameter `popdist`. Can you see any connection between the population distribution and the optimal solution?



- Now "Double the resolution" of your AMPL procedure.

  That is, replace `set X := 0 1 2 3 4 5 6 7 8;` with
  `set X := 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17;`
  and analogously for Y .

  This gives you a significant increase in the number of variables. How many integer variables do you now have?

  Try running the model now. Can you notice a slight lag?

- Now that you have spent some time exploring this problem, suggest additional refinements to improve this model (including improvements you can imagine with regards to the data being used for Locker placement). What additional practical constraints can you imagine?

  Notice that this entire page is available for your suggestions...

- In this lab we have considered distances that correspond to either a physical distance (or some measure of travel time between two locations). To finish, read a short description about *Hamming Distance*:

  `http://planetmath.org/hammingmetric`

  Suppose that in a scientific experiment you recorded 1,000 binary sequences of length 200. Propose 2 questions that could you investigate using the ideas of *facility location*.