INEG 3613: Introduction to OR
Spring 2020
Course Project

1 Overview

Students will work in groups of three on this assignment. This assignment is worth 100% of the overall project grade.

2 Purpose

The purpose of this assignment is to solve a realistic problem by applying the OR process from end to end. In order to be successful on this assignment, each team must figure out what elements of the problem are important, formulate an optimization model to solve the problem, gather data to populate the optimization model, and then use optimization software in order to perform analysis of the problem. Be creative! If you have a great idea about how to solve this problem that doesn't fit within the structure I've given, please feel free to discuss with me—Honestly, I'm more interested in students *creating* and *investigating* than I am in having everyone do exactly what I've spelled out. The assignment is intentionally open ended in more than a few places, so I expect that each group's assignment will look significantly different.

3 Background

Hub-and-spoke designs are common in transportation systems wherein *entities* (e.g., freight or passengers) need to be delivered through the system from many origins to many destinations. In such a system, efficiencies can be gained by consolidating entities at locations designated as *hubs*. Hub networks are particularly common in air freight and parcel distribution.

4 Problem Description

As an employee of a major logistics company, you have been tasked with performing analysis to support the selection of hub locations for a nationwide distribution network. The company distributes product to and from 42 cities (see the "Cities" worksheet of "ProjectData.xlsx"), and a forecast has been provided (see the "O-D Demand" worksheet of "ProjectData.xlsx") of the number of cargo units to be transported annually between each pair of cities. Although not perfectly realistic, you may assume all cargo units represented in this dataset are of the same size and weight.

The efficiency of the hub network depends on (i) in which cities hubs are located and (ii) how cargo is routed. By applying optimization, you are seeking to minimize the annual costs due to (i) and (ii). To assist with your analysis, I have provided a matrix (see the "Distance" worksheet of "ProjectData.xlsx") that specifies the distance $d_{i,j}$ (in miles) from each city $i = 1, \ldots, 42$ to each city $j = 1, \ldots, 42$.

Regardless of whether you choose to locate hubs, you may always choose to route cargo directly from its origin to destination. Alternatively, if you locate hubs, the cargo may be routed through one or two hub cities on the way from its origin to destination. All of the cargo having city i as its origin (destination) must, if it is to be routed through the hubs, enter (exit) the hub system through a single hub city to which city i is assigned. Transportation costs from routing cargo are incurred for each of the following three cases:

1. If neither city i nor city j is a hub, each unit of cargo routed directly from city i to city j costs $\$0.40d_{i,j}$.
2. If either city i or city j is a hub (but not both), each unit of cargo routed directly from city i to city j costs $\$0.22d_{i,j}$.
3. If both city i and city j are hubs, each unit of cargo routed directly from city i to city j costs $\$0.05d_{i,j}$.

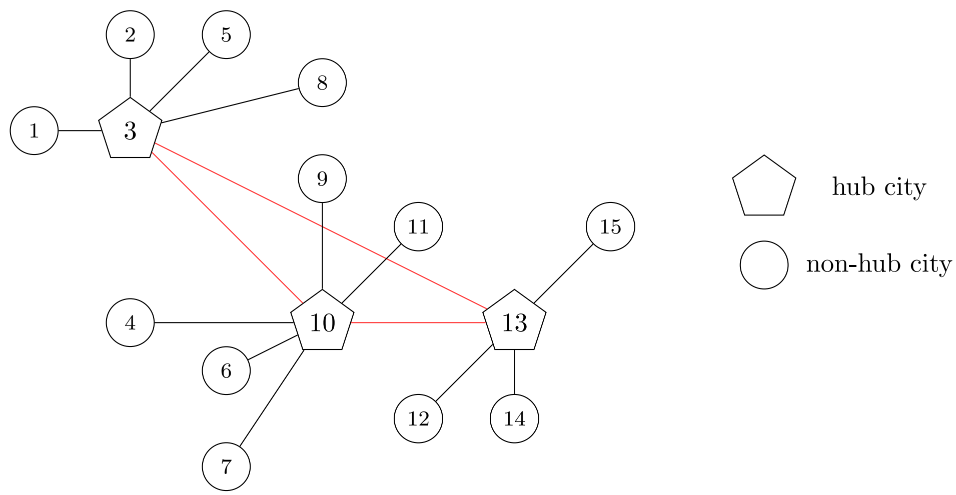

Figure 1: Illustration of the assignment of non-hub cities to hub cities.

Figure 4 provides an example to illustrate the routing cost. In Figure 4, the black lines depict each non-hub city's connection to its assigned hub city and the red lines depict connections between all pairs of hub cities. Thus, the cost to transport one unit of cargo with origin city 8 and destination city 6 is (i) $\$0.40 d_{8,6}$ if the unit is routed directly from city 8 to city 6 or (ii) $\$0.22 d_{8,3} + \$0.05 d_{3,10} + 0.22 d_{10,6}$ if the unit is routed through the hubs in cities 3 and 10. The cost to transport one unit of cargo with origin city 11 and destination city 7 is (i) $\$0.40 d_{11,7}$ if the unit is routed directly or (ii) $\$0.22 d_{11,10} + 0.22 d_{10,7}$ if the unit is routed through the hub in city 10.

The (annualized) cost of leasing and operating a hub is uncertain, but it is known to depend on the total throughput (i.e., the total number of units moved through the hub annually). The "Hub Cost" worksheet of "ProjectData.xlsx" provides 43 paired observations on hub throughput and annualized hub leasing and operation cost, obtained using a simulation model of hub operations.

5 Context and Additional Background

A Google Scholar search for "hub location optimization" might give you some ideas and/or provide evidence as to the importance of these problems.

6 Dr. Sullivan's Advice

I recommend beginning with a simple optimization model. Rather than try to model the whole problem at once, you could start by making a simple model to route cargo from each origin to each destination (i.e., leaving out some of the details in regards to the selection of hub locations). *Hint*: For cargo with origin city i and destination city j, you might use some of the following variables, imposing an artificial "capacity" on the more economically attractive variables to enable testing the functionality of the other variables.

- $v_{i,j}$ = the number of units transported directly from origin city i to destination city j without using hubs
- $w_{i,h,j}$ = the number of units transported from origin city i to destination city j by first passing through hub city h
- $x_{i,h,k,j}$ = the number of units transported from origin city i to destination city j by first passing through hub city h and then passing through hub city k

After verifying functionality of your model's routing capabilities, you could expand the model to incorporate hub location decisions and relate those decisions to the routing decisions. Similarly, although the "Hub

Cost" data appear to be nonlinear, you need not begin by trying to solve a nonlinear model: Rather than embed a full capacity-to-cost function into your optimization model, why not begin by assuming there are a few different "sizes" of hubs that you could build in each city?

7 Final Deliverables

Each group will submit one carefully written Microsoft Word document that is no more than fifteen pages in length. Each team's report should be uploaded to blackboard before the posted deadline. In addition, each team should upload any AMPL model and/or data files used in solving the problem. *The report should be cohesive, telling and illustrating the story of how you modeled the problem, solved the model, and interpreted the results.* Be sure to address each of the following topics within a separate section of the report.

Section 1: Preliminary analysis. Perform the following analysis prior to formulating an optimization model:

> (i) Using the given data, determine the annual cost that would result from not using any hubs. That is, what would be the cost if all cargo were shipped directly from its origin city to its destination city?
>
> (ii) Using the given data, determine the annual cost that would result from building a hub node in every city. (Note: Because each city would be assigned to its own hub node in this case, the hub node for a given city would only need to be large enough to handle the incoming and outgoing cargo for that city.) In order to address this question, you will need to investigate the relationship between the annualized cost of building and operating a hub and its throughput capacity using the "Hub Cost" data.

The analysis above will help to generate ideas for how to model the problem, and it will also be useful for verification and validation of your model.

Section 2: Scope the problem for optimization. Read the problem description in Section 4 of this document, and begin thinking about how you will model the problem. In this section of the report, explain (i) what decisions must be made, (ii) by what criteria the decisions will be evaluated, and (iii) what constraints limit the decisions. *There should be no math in this section.*

Section 3: Formulate an optimization model. Formulate a mathematical program to optimize hub location and routing decisions. The mathematical program should be stated in general terms (i.e., no input data specified) so that any problem instance could be solved using your formulation. Include the mathematical program in your report (in mathematical form, not AMPL form) and summarize what each of the components do (i.e., be sure to define your variables in words, explain what each constraint does, explain what the objective does, and so on). It is understood that you have to make some assumptions and/or simplifications along the way to ensure you can populate and solve the model. Be sure to state and justify any assumptions and/or simplifications you have made.

Section 4: Gather and synthesize the data. How will you populate the parameters of your model? At a minimum, you should use the provided demand, distance, and cost data. If you also use other published sources of data, be sure to cite them.

Section 5: Solve and analyze the model. Implement and solve the model(s) using AMPL/CPLEX. Interpret the results within the context of the original problem and your preliminary analysis, providing illustrations and/or figures if appropriate. Are the hub location decisions (and assignments of non-hubs to hubs) reasonable, considering the geographical relationships of the cities under consideration? To what extent were the hubs relied upon for routing? How would the results (e.g., cost, number of hubs, proportion of units routed through hubs) change in response to a change in the parameters (e.g., demands, transportation costs per-mile, costs to operate a hub of a given size). You many need to solve a number of instances of the problem to answer this question thoroughly.

Section 6: Extend the model. The hope is that you'll observe something in the previous sections that provokes an idea about how we could better understand/solve the problem. Some potential ideas are given below:

1. If you formulated the model using my advice in Section 6 of this document, it required more than 42^4 variables. Here, you might consider trying to formulate the model using significantly fewer variables. Why? Although not a guarantee, models with fewer variables may solve faster. *Hint*: Try letting x_{ij}^k, v_{ij}^k, and w_{ij}^k denote the amount of cargo originating at city $k = 1, \ldots, 42$ that flows directly from city $i = 1, \ldots, 42$ to city $j = 1, \ldots, 42$ with the following distinctions: x_{ij}^k is used if both i and j are hubs; v_{ij}^k is used if exactly one of $\{i, j\}$ is a hub; and w_{ij}^k is used if neither i nor j is a hub.

2. Could you improve upon your solution from Section 5 by considering hub locations that were not part of the initial list of solutions?

3. Assuming your original model took a long time to solve, is there a different model or code you could use to generate a high-quality solution in less time? For example, suppose you instead solved a suitably-adapted p-median or fixed-charge facility location problem to identify hub locations? What annualized cost would result for the hub network if you located facilities in this manner?

Choose **one** idea (either mine or one of your own) for further investigation. If your idea leads to a new model, implement the new model and run experiments to investigate the impact of your idea. Alternatively, this section could instead consist of numerical experimentation above and beyond what was required for the previous section. Add a section to the report that describes your idea and summarizes how you extended the model and/or what new experiments you ran. In this section, you should also summarize and discuss the results.

8 Given Data

The Excel file "ProjectData.xlsx" contains the relevant data. The contents of this file are summarized below.

Cities. This worksheet contains the following data for 42 cities:

- CityID: An integer in $\{1, \ldots, 42\}$ that is unique to each city.
- City: The name of the city.
- State: The state in which the city resides.
- Latitude (Deg): The city's latitude in degrees.
- Longitude (Deg): The city's longitude in degrees.

O-D Demand. This worksheet contains the forecasted annual demand (in units of cargo) from each origin city "From CityID" to each destination city "To CityID."

Distance. This worksheet contains the distance (in miles) between each pair of cities.

Hub Size. This worksheet contains 43 simulated observation pairs on annual throughput and annualized leasing and operation cost.

9 Intermediate Deliverables

The final report is due on **Thursday, April 30**; however, it is expected that Sections 1–4 of the report will be complete by **Thursday, April 9** and Section 5 will be complete by **Thursday, April 23**.

Each team *must* schedule one 30-minute meeting with Dr. Sullivan between Thursday, April 9 and Tuesday, April 14. During this meeting, the team will show its progress in completing Sections 1–4 and discuss a plan for completing Section 5.

In addition, each team *may* schedule additional meetings with Dr. Sullivan. It is strongly advised, but not required, that each team meet with Dr. Sullivan between Thursday, April 16 and Friday, April 24 to discuss its progress in completing Section 5 and a plan for completing Section 6.

10 Rubric

Each team's final report will be graded according to the following rubric:

	Unsatisfactory (0–4)	**Needs Improvement (5–6)**	**Adequate (7–8)**	**Excellent (9–10)**	**Grade**
Prelim. Analysis		Fails to make progress on one of (i)–(ii)	Makes significant progress on both (i)–(ii)	Adequately answers both of (i)–(ii)	
Scope the Problem		Adequately answers 1 of (i)–(iii) with no supporting arguments drawn from the specific application	Adequately answers 2 of (i)–(iii) with weak supporting arguments from the application	Answers (i)–(iii) with strong supporting arguments drawn from the specific application	
Model Formulation		Numerous errors with correctness, not stated in general terms	Few errors, stated in general terms	Correct, stated in general terms	
Data Collection		Lacking in quantity, irrelevant to problem	Relevant data but lacking in quantity, errors in citation	Thorough and relevant to problem, appropriately cited	
Solution and Analysis		One instance solved, no significant observations or insights drawn	Fewer instances solved, provides basic observations about solutions	Instances correctly solved, provides significant insights into the specific problem	
Extend the model		Reasonable idea posed but little progress	Reasonable idea posed and incorporated into model	Reasonable idea analyzed using model	
Report Content		Factual errors, inaccurate or incomplete descriptions of model and/or data, conclusions not justified by analysis	Conclusions justified by supporting analysis, but descriptions contain some inaccuracies	Strong conclusions justified by supporting analysis, all descriptions accurate	
Report Quality		Lacks detail or cohesion, frequent grammatical or spelling errors, inconsistent or unprofessional formatting, fails to use figures and tables effectively	Lacks detail in some areas, some parts disjointed, few typographical errors in spelling and grammar, fails to use figures and tables effectively	Enjoyable to read, cohesive and informative, free of grammatical errors, consistent formatting, uses figures and tables effectively	
Creativity		Directly applies well-known models and/or the modeling ideas provided in this document	Attempts to tailor ideas/models to description; goes above and beyond the ideas provided in this document	Successfully tailors ideas/models to application	

In addition, each team member will be graded on his/her contribution to the team's efforts.