

Finding Optimal Airport Hubs

Jackie Hudepohl

Tiffany Zhu

Karishma Desai

Table of Contents

Background	2
Assumptions	2
Data	3
Initial Model	5
Integer Program Formulation	6
Results with Constant Hub Costs	7
Results with Variable Hub Costs	8
Conclusion	9
References	10

Background

Many airlines use airport hubs to concentrate passenger traffic and flight operations to a particular airport. Hubs often serve as layover points for passengers going to a final destination. An airline operates flight from non-hub cities to hub cities or through hub cities if the final destination is not a hub city. This model creates an economies of scale that allows airlines to fly between a city pair (through a hub) which would not have been economical if the flight were to be nonstop. It is useful for an airline to have a hub so that they can provide more flights that cater to their passengers' needs and to maximize their revenue.

Airlines incur many operating costs. Some of these costs include salaries, repairs, and maintenance. An airline deals with extra costs by having a hub airport. They have to hire more employees and facilities which increases costs.

In this project, we wanted to find what the optimal number and location for an airline hub would be by minimizing the operating costs of the airline. The top twenty airports for flight frequency were used as candidates for optimal hub locations. This report includes two methods of solving this problem. The first method, using the hub-and-spokes model, is a simpler model and is done in Python using brute force. The second method, using integer programming, has more complexity due to the inclusion of discounted flights and varied hub operating costs. The solution to the complex program was found using the IP solver in Matlab.

Assumptions

To simplify the model, we made a few assumptions. The first assumption is that the airline only has one type of aircraft, a large narrow body aircraft such as a Boeing 737-800. One type of aircraft was used to standardize the weight of the aircraft which was useful for determining one of the operating costs.

Another assumption we made is that the airline only has two operating costs: landing fees and delay fees. These two costs vary depending on the airport. In one of the iterations of our complex model, a constant cost was used for all of the airports. One more assumption is that the cost of a flight from one city to another city is determined by the distance between the two cities

and the frequency of flights that go between those cities. This is done to apply a value to each flight which can be interpreted as a monetary cost.

We also assumed that the airline only has domestic flights and that all the available flights go between the top 20 airports. Our last assumption was that flights that have a layover at a hub are discounted by 25%, but direct flights are not discounted. The reasoning behind is that passengers who are willing to sacrifice their time by choosing an indirect flight are rewarded with a discount and passengers who want direct flights have to pay full price. To find the optimal solution, the integer program chooses the cheaper option between the discounted flight and the full price flight.

Data

The airports that could be chosen as an optimal hub were the top 20 airports with the highest flight frequency. Flight frequency is determined by the number of flights that leave a particular airport.

Rank	Code	Name
1	ATL	Atlanta, GA
2	DFW	Dallas, TX
3	ORD	Chicago - O'Hare, IL
4	CLT	Charlotte, SC
5	DEN	Denver, CO
6	LAX	Los Angeles, CA
7	PHX	Phoenix, AZ
8	MSP	Minneapolis, MN
9	SEA	Seattle, WA
10	DTW	Detroit, MI
11	SFO	San Francisco, CA
12	IAH	Houston, TX
13	LAS	Las Vegas, NV
14	SLC	Salt Lake City, UT
15	PHL	Philadelphia, PA
16	SYR	Syracuse, NY

17	EWR	Newark, NJ
18	LGA	Queens, NY
19	DCA	Arlington, VA
20	BOS	Boston, MA

Table 1. Top 20 most frequent airports in the USA.

To calculate the base cost for a plane to fly from city i to city j , we used the frequency of flights from i to j and distance from i to j . These two number values were multiplied to create a cost for a flight. We placed the values into matrices to make it easier for the solver to calculate the results.

For the operation costs of the airline, we used landing fees and delay costs. A landing fee is how much it costs an airline to have their plane land at an airport. The fee varies based on the weight of a plane; the fee is per thousand pounds of the plane's weight. For this project, we used a Boeing 737-800 which weighs 174,200lbs to figure out the total fees. Landing fees vary for each airport. For the delay costs, the average delay time for a flight at each of the airports was found. This amount was multiplied by the cost of delay per minute which was \$65.43.

Airport	Landing Fee Costs	Delay Costs	Total Cost
ATL	\$142.84	\$1,897.47	\$2,040.31
DFW	\$445.95	\$2,224.62	\$2,670.57
ORD	\$1,466.94	\$2,420.91	\$3,887.85
CLT	\$151.55	\$1,897.47	\$2,049.02
DEN	\$789.13	\$1,962.90	\$2,752.03
LAX	\$763.00	\$1,962.90	\$2,725.90
PHX	\$339.69	\$1,766.61	\$2,106.30
MSP	\$421.56	\$2,159.19	\$2,580.75
SEA	\$621.89	\$1,635.75	\$2,257.64
DTW	\$599.25	\$2,224.62	\$2,823.87
SFO	\$848.35	\$2,224.62	\$3,072.97
IAH	\$344.92	\$2,159.19	\$2,504.11
LAS	\$405.89	\$2,028.33	\$2,434.22
SLC	\$348.40	\$2,028.33	\$2,376.73

PHL	\$654.99	\$2,224.62	\$2,879.61
SYR	\$1,700.00	\$2,682.63	\$4,382.63
EWR	\$266.53	\$2,290.05	\$2,556.58
LGA	\$1,754.19	\$2,813.49	\$4,567.68
DCA	\$681.12	\$2,290.05	\$2,971.17
BOS	\$444.21	\$2,159.19	\$2,603.40

Table 2. Operating costs for the top 20 most frequent airports.

Initial Model

To start, we formulated a simpler model. In the initial model we used a Hubs and Spokes model and minimized the cost which was frequency times distance. We also gave a discount factor for flying through a hub, we used chose to discount by 25%, so we minimized the cost of the frequency times distance then times the one minus the discount if there was a hub. The frequency is the frequency of flight from one airport to another and the distance is the distance from one airport to another. Therefore we want to minimize the distances people travel and the frequencies of the airplanes (which gives a measure of what is spent on operating flights). We wrote Python code to simulate the model, we could change the number of hubs by giving the code an input. However, we used a brute force method to solve the Hubs and Spokes model so it took several minutes to run it on more than one hub. We used two other inputs, a 20x20 frequency matrix and a 20x20 distance matrix. Both are 20x20 matrices because like stated previously, we used the top 20 airports based on their frequency of flights. So each column and each row represents an airport and the diagonals are zero because going to and from the same airport gives zero frequency and has zero distance. The distance matrix is symmetric since if you're given two airports, i and j, going from i to j and j to i gives the same distances. We calculated these distances using a website that finds the number of miles it is to fly from one airport to another. The frequency matrix on the other hand is not symmetric because going from i to j and j to i does not necessarily give the same number of flights because one may be more popular of a destination than the other.

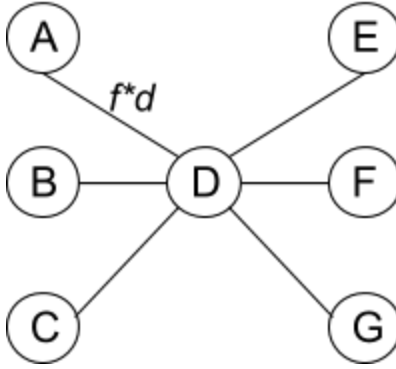


Figure 1. Example of the Hubs and Spokes Model. Node D is the hub and each edge is frequency times distance.

Now we will describe how the graph looked like in further detail and how the code worked. Each of the 20 airports was a node and if we only had one hub, then we'd first select the first node and make that a hub. Making a node a hub means that now every other node connects to it to ensure that every flight goes through the hub first before going to its final destination. We illustrate this in a simple example in Figure 1, where node D is the hub. The weight every edge is the frequency times distance. So if we still look at the graph in Figure 1, to get the data for the labeled edge the frequency, f , would be found in row A column D of the frequency matrix (if we are using A as the starting point) and similarly the distance, d , would be found in column D of the distance matrix. The code would calculate all of the costs with D being the hub and store that cost value and the node that was selected as hub. Then it would move on and select another node to be the hub and repeat. Once it found all the costs, it would find the minimum cost and output its associated hub.

We ran our code with one hub and two hubs. For hub we got the optimal solution to be ORD (Chicago). With two hubs, we got ORD (Chicago) and DEN (Denver). Since our costs just depended on the distances and frequencies, these centrally located hubs gives a reasonable answer.

Integer Program Formulation

In order to solve for the optimal number of hubs, we formulated the problem as an integer program. We used these variables to represent constants in our problem:

c_i = cost of opening airport i as a hub

d_{ij} = distance from airport i to airport j

f_{ij} = relative frequency of flights from airport i to airport j

And allowed the following to be binary decision variables:

$x_{ij} = \mathbb{1}(\text{fly directly from airport } i \text{ to airport } j)$

$y_{ijk} = \mathbb{1}(\text{fly from airport } i \text{ to airport } k \text{ through hub } j)$

$h_i = \mathbb{1}(\text{airport } i \text{ is a hub})$

Then we can formulate the objective function as:

$$\min \sum_{i=1}^{20} \sum_{j=1}^{20} x_{ij} d_{ij} f_{ij} + \sum_{i=1}^{20} \sum_{j=1}^{20} \sum_{k=1}^{20} 0.75 y_{ijk} (d_{ij} + d_{jk}) f_{ik} + \sum_{i=1}^{20} c_i h_i$$

The first term represents the cost of all direct flights, the second term is the cost of all flights which go through a hub, and the third term quantifies the fixed cost of operating an airport as a hub. We are treating distance as a proxy for cost, so all quantities here are in terms of distance. We also make the assumption that when flying through a hub, the total distance flown both into and out of that hub should be discounted by 25 percent.

We also introduced the following constraints in order to make the solution fit the real world:

$$x_{ijk} \leq h_j \text{ (layovers must be at a hub)}$$

$$x_{ik} + \sum_{j=1}^{20} y_{ijk} = 1 \forall i, k \text{ (fly between every pair of cities)}$$

$$x_{ii} = 0 \forall i \text{ (do not fly to same airport)}$$

$$y_{iji} = 0, y_{ijj} = 0, y_{iij} = 0 \forall i, j \text{ (be careful with layover variables)}$$

Results with Constant Hub Costs

Initially, we did not let the fixed cost of operating an airport as a hub vary with each airport, i.e.

$$c_i = c \forall i$$

As we were unsure how to convert hub cost in dollars to flight cost in terms of distance, we let this constant c vary. Note the magnitude of c is inversely related to the optimal number of hubs. The following table summarizes our results for different choices of c :

c	Number of Hubs	Hubs
50	1	DEN
20	2	DEN, DCA
15	3	DEN, DCA, ORD
10	4	DEN, DCA, ORD, LAS
8	5	DEN, DCA, LAS, DFW, DTW
5	6	DEN, DCA, ORD, LAS, DFW, DTW

Table 3. Summary of integer program results using a constant cost for setting up a hub but different numbers of hubs and values of c .

Results with Variable Hub Costs

Next, we allowed the fixed hub cost to vary across airports by introducing a different scaling factor for each airport. We based the total hub cost on data we found for landing fees and delay costs at each airport. Atlanta was the cheapest airport to operate as a hub, so we used this as the baseline cost. As an example, LaGuardia was the most expensive, with total hub costs 2.24 times those for Atlanta. So we represented the total hub cost for LaGuardia as $2.24c$, where c denotes the baseline cost of operating Atlanta as a hub and 2.24 is the scaling factor for LaGuardia. As in the previous case, we let the baseline cost c vary to deal with our uncertainty in converting between dollar costs and distance costs. The introduction of a scaling factor for each airport led to somewhat different results, summarized in the table below:

c	Number of Hubs	Hubs
50	1	DEN
20	2	DEN, DTW
11.5	3	DEN, DTW, LAS
10	4	DEN, LAS, DTW, CLT
5	5	DEN, LAS, DTW, DCA, DFW
4	6	DEN, LAS, DTW, DCA, DFW, ATL

Table 4. Summary of integer program results using a different costs for setting up a hub but different numbers of hubs and values of c .

Conclusion

Using frequency, distance, and a discount factor, we computed the costs in our initial model and found the most optimal hubs, if we were to have two, to be Chicago and Denver. Since the way we calculated cost relied a lot on distance, our results looked reasonable. However, our cost function from this initial model seemed too simple and could be improved since distance and frequency are not the only factors that should be taken into account.

We then formulated the problem as an integer program, first with a fixed hub cost that was constant across airports. With two hubs, our model found Denver and Washington Reagan to be optimal. The different results from the brute force approach make sense because our integer program formulation does not discount direct flights where the origin or destination is a hub (as we did in our brute force approach); the discount only applies when routing two-leg flights through a hub. Finally, we allowed the fixed hub cost to vary with airport based on data about landing fees and delay costs at each airport. This modification led to the choice of Denver and Detroit as the two optimal hubs.

One future improvement to our model would be allowing the hub discount rate to vary with each airport. Currently, we use a 25 percent hub discount rate for all airports. However, a more realistic way to model this rate would be having the discount rate be an increasing function of available capacity, i.e. the more unused capacity an airport has, the more of a discount the airport should provide in order to incentivize carriers to use the airport as a hub.

References

- Airlines For America. <http://airlines.org/data/per-minute-cost-of-delays-to-u-s-airlines/>. Accessed Nov. 2016.
- Bruffey, Wade. Most Frequently Delayed Airports, 2 June 2016.
<http://airlines.wanderbat.com/stories/14024/most-frequently-delayed-airports>. Accessed Nov. 2016.
- Bureau of Transportation Statistics. www.transtats.bts.gov/. Accessed Nov. 2016.
- World Airport Codes. Fubra Limited, 2016, www.world-airport-codes.com/distance/. Accessed Nov. 2016.
- http://aci-na.org/sites/default/files/dtw_pt2_0.pdf
- <http://www.cltairport.com/AirCargo/Documents/2014QuickCaller.pdf>
- <https://dwuconsulting.com/images/Ratebook/SLC%20Jul15%20Rate%20Book.pdf>
- <https://dwuconsulting.com/images/Ratebook/LAX%20Jul15%20Rate%20Book.pdf>
- https://en.wikipedia.org/wiki/Minneapolis%E2%80%93Saint_Paul_International_Airport
- <http://www.flychicago.com/OHARE/EN/ABOUTUS/FACTS/Pages/Financial-Data.aspx>
- http://www.flydenver.com/sites/default/files/rules/120_fees_charges.pdf
- https://fly2houston-production.s3.amazonaws.com/media/filer_public/d2/58/d2582408-5824-4a6d-b876-051a3cb13920/iah_and_hou_fy17_rate_comparisons.pdf
- <https://www.massport.com/hanscom-field/overview/aircraft-operating-parking-fees/>
- <https://www.mccarran.com/DoingBusiness/Finance/Annual-Report-2015/index.html#33/z>
- http://www.mwaa.com/sites/default/files/2015_cafr.pdf
- http://one.atl.com/docs/BusinessInformation/Reports/FY15_Comprehensive_Annual_Financial_Report.pdf
- <https://www.panynj.gov/airports/pdf/scheduleofcharges-swf.pdf>
- <https://www.panynj.gov/airports/pdf/scheduleofcharges-lga.pdf>
- <https://www.phl.org/Documents/Business/PNE/PNEAppH.pdf>
- https://sfoconnect.com/sites/default/files/summary_airport_charges_fy15-16.pdf
- <https://skyharbor.com/docs/default-source/default-document-library/phx-annual-report-2014.pdf?sfvrsn=0>