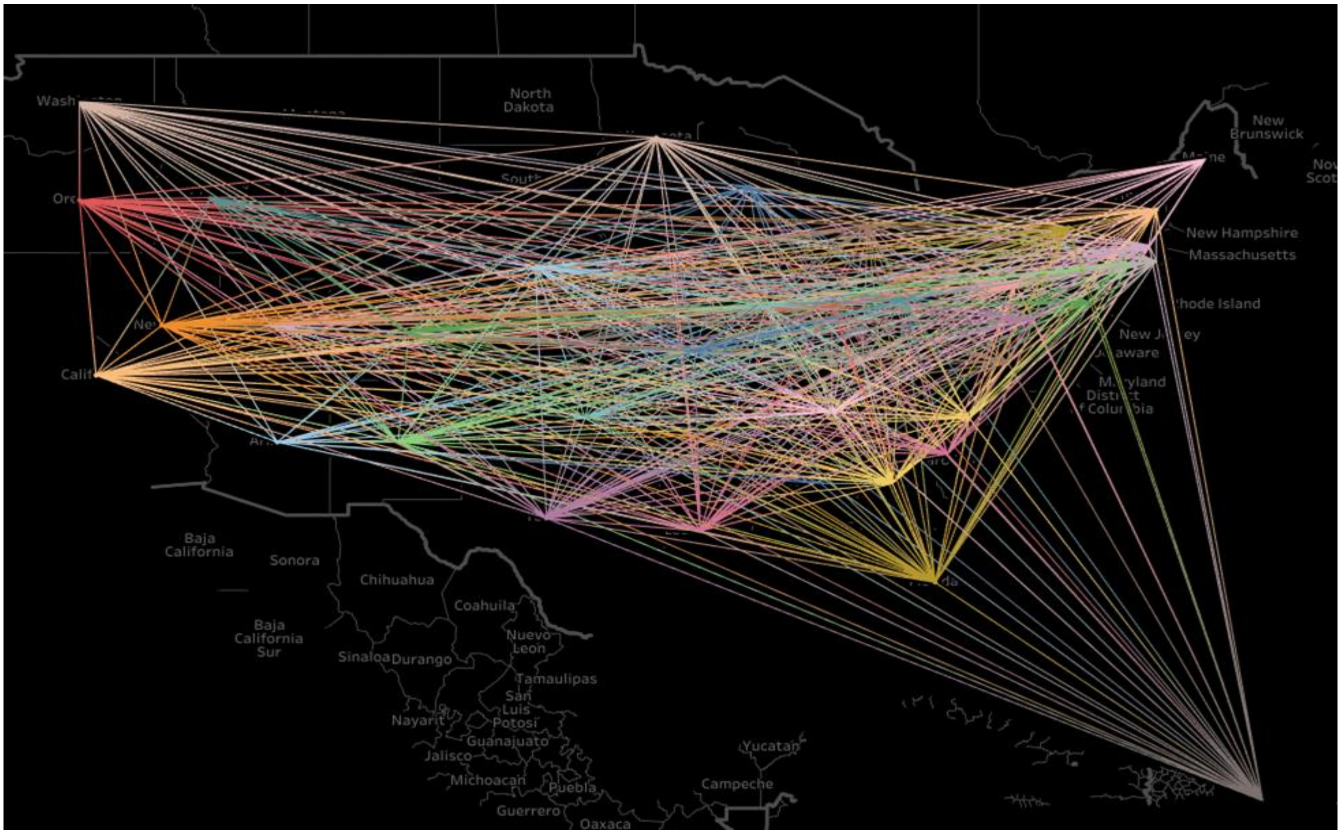


AIRLINE NETWORK ANALYSIS

SEMESTER PROJECT FOR IDS-564

In this project, we apply the principles of Graph Theory and Social Network to investigate how the airline network behaves. We then do an analysis of network structure to see how an airline network grows and how this study can help improve productivity and efficiency of the overall airline network.

Authors: Abhishek Mishra, Rashmi Choudhary, Anuj Tiwary, Shubham Sirothia

ABSTRACT

In United States, the airline network not only supports the travel of people from one place to another, but it also supports transportation of cargoes and parcels. In situations of a network failure, where airports can be shut down due to any of the reasons such as - a snow storm, a terrorist attack or a technical glitch the result could be significant in terms of economic loss. In this project, we analyze how when a critical airport gets shut down the average travel time of a network will increase. Some of these airports include Chicago's O-Hare airport and New York's John F Kennedy airport.

We also analyze which cities would be disconnected from the other cities because of this airport shutdown. We use different centrality measures to assess the importance of an airport and then do an analysis around the most central or important ones.

The project also tries to analyze the network growth and then suggest a route that can help reduce the average commute time. We use the triadic closure property and link prediction phenomena to come up with this analysis.

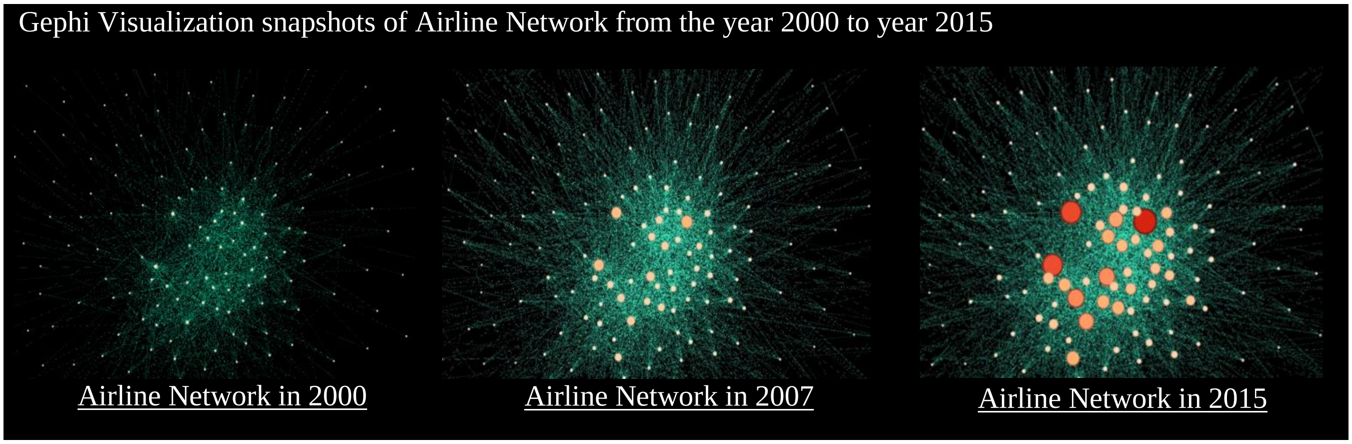
We apply all our analysis on a network that comprises 303 airports and 4115 flights within the United States. The data has been obtained from the website of BTS - The Bureau of Transportation and Statistics, that maintains airline traffic data and from FAA – Federal Aviation Administration. We have used the airline data from the year 2000 to the year 2015 and the passenger data from the year 2005 to 2015 for our analysis.

We have used graphical visualization tools – TABLEAU and GEPHI for visualizing the network. We have used statistical software R and its 'igraph' library to come with up all our analysis.

Keywords: Modularity, Betweenness, Clustering, PageRank, Hub, Authority

1. INTRODUCTION

Gephi Visualization snapshots of Airline Network from the year 2000 to year 2015



Airline Network in 2000

Airline Network in 2007

Airline Network in 2015

The US airline transportation network grew rapidly over the past 10 years. In 2006 there were 28 million scheduled flights, whereas the number has jumped to 35 million in the year 2015 – a 25% jump. That's is why a study to check the capacity of how the airports can address the increasing load ten to fifteen years down the line becomes vital.

To study the airline network, we identify the centrality measures which signify the critical nodes that help improve the overall connectivity and efficiency of the network. The key centrality measures we will work with are Degree centrality, Betweenness centrality and PageRank. Researchers have also come up with new centrality measures such as Delay centrality, Travel Time centrality and Commuter Flow centrality that we look forward to include in our future research to evaluate even more precise impact of airport failure.

Often, in Policy Making & Benchmarking, decision makers find it challenging to evaluate how to spend the $75 billion FAA budget for 2016 to 2021 for airport infrastructure development? Their main objectives are to overcome the congestion and delay problem faced by commuters. When establishing funding priorities, authorities need to acknowledge the importance of connectivity. Should they build a new airport or should they expand an existing airport is always a tough call to make. Using our work, we can evaluate whether inclusion of a new airport enhances the efficiency of our network more or expansion of an existing airport will be more beneficial.

We summarize our project work in the steps below:

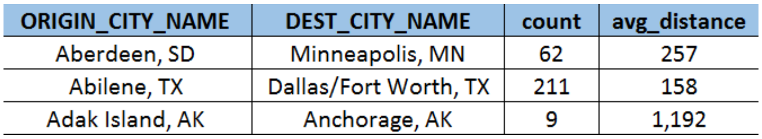
- Calculation of Centrality measures and Identifying Modularity
- Evaluating the Impact of airport failures
- Network characteristics
- Network growth model
- Triadic Closures
- Link Prediction

2. RELATED WORK

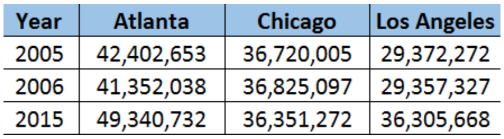
Node centrality or finding the most important node has always been an important task in Transportation Network analysis. Tore Opsahl [7] has used centrality measures to list down airports in the descending order of their importance. With rising fuel costs, increased commuter and cargo flow the dependency on airlines has increased significantly. Y.Y. Cheng et al. [8] in their research developed centrality measures that incorporate *commuter flow* and *travel time delay* to determine critical nodes in a transportation network. Using these measures, they plan to improve the design of the network and devise plans for coping with the network failures. They used Singapore's mass rapid train - MRT dataset (given by Land Transport Authority (LTA) of Singapore) for their analysis.

3. DATASET

Data has been obtained from the websites of BTS [1] and FAA [2]. The nodes represent the airports and the edges represent the flights between these airports. There are 303 airports and 4115 edges in our dataset. In addition to the airline's data we have passenger data that has been used for link prediction. A snapshot of our dataset can be seen below:



ORIGIN_CITY_NAME	DEST_CITY_NAME	count	avg_distance
Aberdeen, SD	Minneapolis, MN	62	257
Abilene, TX	Dallas/Fort Worth, TX	211	158
Adak Island, AK	Anchorage, AK	9	1,192



Year	Atlanta	Chicago	Los Angeles
2005	42,402,653	36,720,005	29,372,272
2006	41,352,038	36,825,097	29,357,327
2015	49,340,732	36,351,272	36,305,668

Airline flight network database snapshot

Airline passenger database snapshot

In our directed graph, the edges connect the source and the destination of our flight and are weighted by the count of flights. The airline network shown in the cover page of our report has been generated using Tableau using the above dataset.

4. PROJECT DETAILS

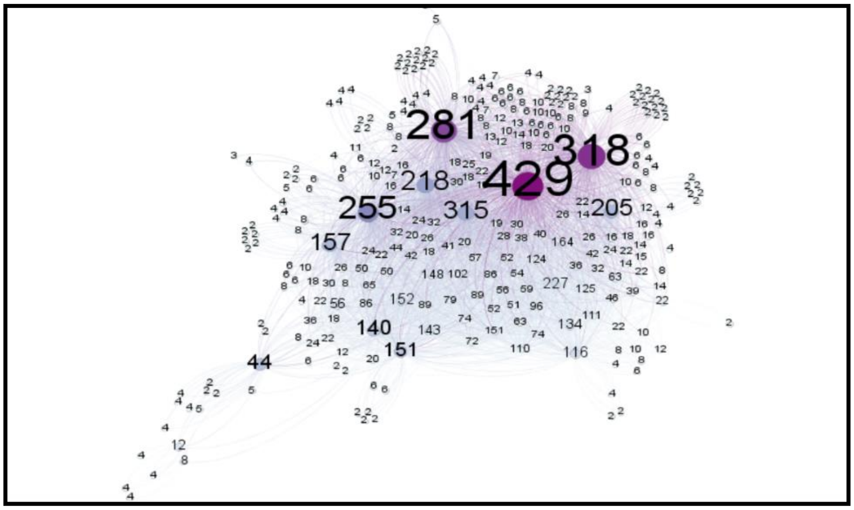
CENTRALITY MEASURES:

We begin to evaluate the important airports by calculating the centrality measures, hub and authority score and page rank measures.

Degree Centrality: The degree centrality of a node is the number of ties a node has.

$$C_{deg}(i) = |\{(i, j)|(i, j) \in E\}|$$

Degree indicates the importance of an airport with respect to number of inflow and outflow of flights.



Airport	Degree
Chicago	429
Atlanta	318
Houston	315
Dallas	281
Denver	255
New York	227

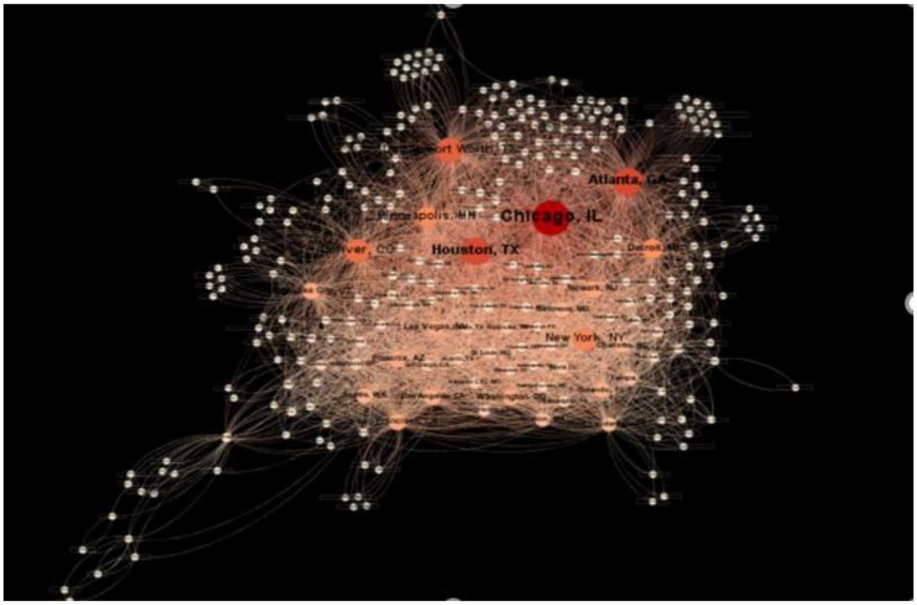
Fig3- Degree centrality plot

As you can see in the above figure 3, Chicago is the most important airport with the highest degree. It clearly states that most of the flight moves from and to Chicago followed by Atlanta and Houston.

Betweenness Centrality: The extent to which a node is part of transactions among other nodes can be studied using Freeman's (1978) betweenness measure.

$$C_{btw}(i) = \sum_{j \in V} \sum_{k \in V, k > j} \frac{g_{jk}(i)}{g_{jk}}$$

After analyzing the degree, we were curious to see how betweenness plays a role in identifying the significant airports. It is vital to analyze the shortest path connecting the airports which in turn will help us in evaluating the easily accessible airports.



Airport	Betweenness
Chicago	19110
Atlanta	16595
Dallas	14919
Denver	10437
Minnesota	8436
Houston	7385

Fig4- Betweenness centrality plot

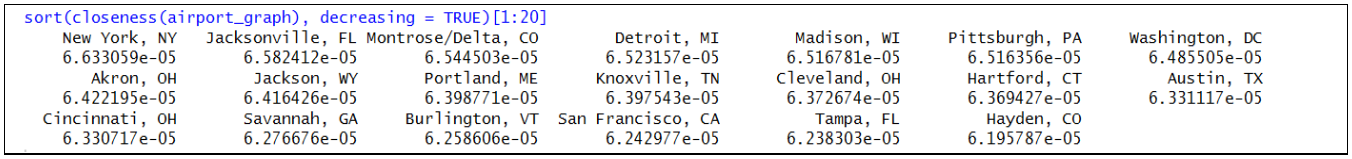
We see that there is an airport – Minnesota, which doesn't have a huge degree but has a high betweenness. Minnesota's airport is connected to various popular as well as unpopular destinations and we can reach those cities easily using Minnesota as a stop over.

Closeness Centrality: Closeness is defined as the inverse of farness, which in turn, is the sum of distances to all other nodes (Freeman, 1978). The intent behind this measure was to identify the nodes which could reach others quickly.

The limitation of closeness - the lack of applicability to networks with disconnected components (two nodes that belong to different components do not have a finite distance between them) didn't bother us since there is only one giant component in our network.

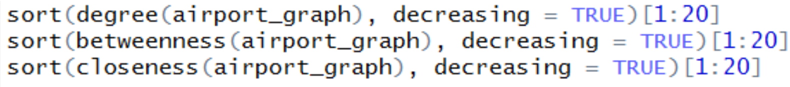
$$C_{cls}(i) = \frac{1}{d(i)} \quad \text{where } d(i) = \sum_{j \in V, j \neq i} d(i, j).$$

Closeness also indicates the ease of access to locations while travelling from one airport to another airport. The result of closeness is quite different from degree and betweenness. New York may not have the highest degree and betweenness but there are flights connecting New York to all other airports.



```
sort(closeness(airport_graph), decreasing = TRUE)[1:20]
    New York, NY   Jacksonville, FL Montrose/Delta, CO        Detroit, MI        Madison, WI     Pittsburgh, PA     Washington, DC 
    6.633059e-05       6.582412e-05       6.544503e-05       6.523157e-05       6.516781e-05       6.516356e-05       6.485505e-05 
      Akron, OH        Jackson, WY       Portland, ME      Knoxville, TN      Cleveland, OH       Hartford, CT         Austin, TX 
    6.422195e-05       6.416426e-05       6.398771e-05       6.397543e-05       6.372674e-05       6.369427e-05       6.331117e-05 
  Cincinnati, OH       Savannah, GA     Burlington, VT San Francisco, CA          Tampa, FL         Hayden, CO 
    6.330717e-05       6.276676e-05       6.258606e-05       6.242977e-05       6.238303e-05       6.195787e-05 
```

Code snippet for centrality measures:



```
sort(degree(airport_graph), decreasing = TRUE)[1:20]
sort(betweenness(airport_graph), decreasing = TRUE)[1:20]
sort(closeness(airport_graph), decreasing = TRUE)[1:20]
```

Page Rank: A method for computing a ranking, for every web page based on the graph of the web. PageRank also has applications in search, browsing, and traffic estimation.

```
$R_0 \leftarrow S$
loop:
    $R_{i+1} \leftarrow AR_i$
    $d \leftarrow ||R_i||_1 - ||R_{i+1}||_1$
    $R_{i+1} \leftarrow R_{i+1} + dE$
    $\delta \leftarrow ||R_{i+1} - R_i||_1$
while $\delta > \epsilon$
```

‘S’ can be any vector over Web pages

$$R'(u) = c \sum_{v \in B_u} \frac{R'(v)}{N_v} + cE(u)$$

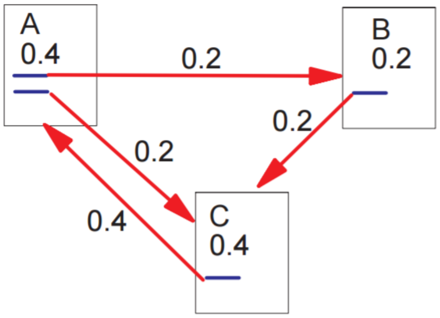
E(u) is some vector over the Web pages that corresponds to a source of rank.

PageRank of a set of Web pages is an assignment, R0

(Details can be found here [5])

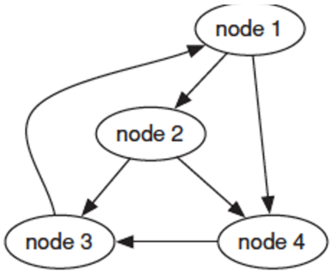
Page rank, hub and authority scores helped us determine the importance of an airport. An airport with high page rank means that we can arrive at these airports quickly using the sequence of random flights.

City	Page Rank
Chicago	0.0492
Houston	0.0414
New York	0.0380
Washington DC	0.0335
Atlanta	0.0299



A simplified page rank calculation mechanism [5]

Hubs and Authorities: A scheme in which, given a query, every web page is assigned *two* scores. One is called its *hub score* and the other its *authority score.* A good hub page is one that points to many good authorities; a good authority page is one that is pointed to by many good hub pages [4].



$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \\ 2 \\ 4 \end{bmatrix}$$

(Image source: Professor Ali Tafti’s lecture slides)

Using an adjacency matrix:

Update rule for *hubs*: $h_i \leftarrow M_{i1}\, a_1 + M_{i2} a_2 + \ldots + M_{in}\, a_n$

Update rule for *authorities*: $a_i \leftarrow M_{1i}\, h_1 + M_{2i} h_2 + \ldots + M_{ni}\, h_n$

As seen in figure 5 and 6 below, Chicago and Houston has high hub and authority score. Hub score represents that they offer many flights to many other large and important airports. Authority score represents that these airports share high amount of traffic with the airports of high hub score.

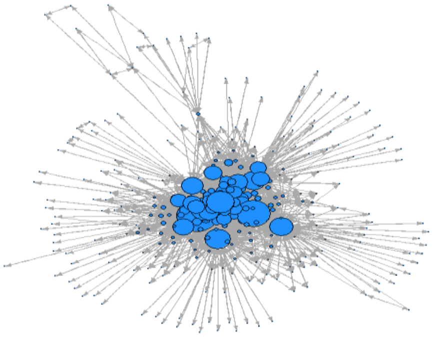


Fig5: Hub plot

City	Hub Score
Chicago	1
Houston	0.8359
New York	0.7356
Washington DC	0.6822
	0.6272

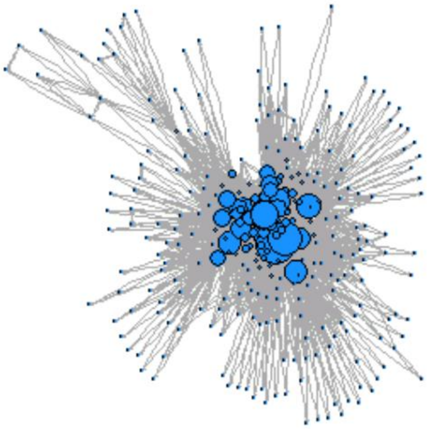
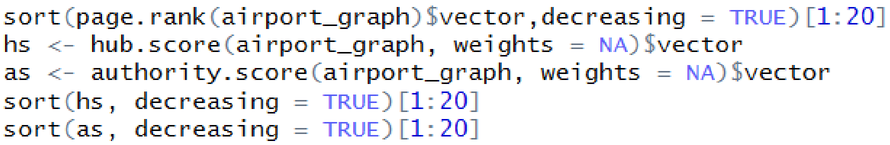


Fig6: Authority plot

City	Authority Score
Chicago	1
Houston	0.8413
New York	0.7327
Washington DC	0.6768
Atlanta	0.6279

Code snippet:



```
sort(page.rank(airport_graph)$vector,decreasing = TRUE)[1:20]
hs <- hub.score(airport_graph, weights = NA)$vector
as <- authority.score(airport_graph, weights = NA)$vector
sort(hs, decreasing = TRUE)[1:20]
sort(as, decreasing = TRUE)[1:20]
```

MODULARITY:

In the study of networks, modularity (networks) is a benefit function that measures the quality of a division of a network into groups or communities [9].

$$Q = \sum_{i=1}^{k} \left(e_{ii} - a_i^2\right)$$

a: probability a random edge would fall into module i
e: probability edge is in module I

After understanding the basics of airport network, we checked if there is any existence of homophily within the network. We can clearly see distinct communities of airports in the above Figure 7 below. The airports are well connected with each other within their region. However, we can see very few airports which are connected across other regions.

This suggests that there is a heavy movement of people due to business/leisure within the region whereas a comparatively lesser percentage of people move across the regions. We can see some trace of homophily where airports in the same region are connected more with each other than with airports of different region.

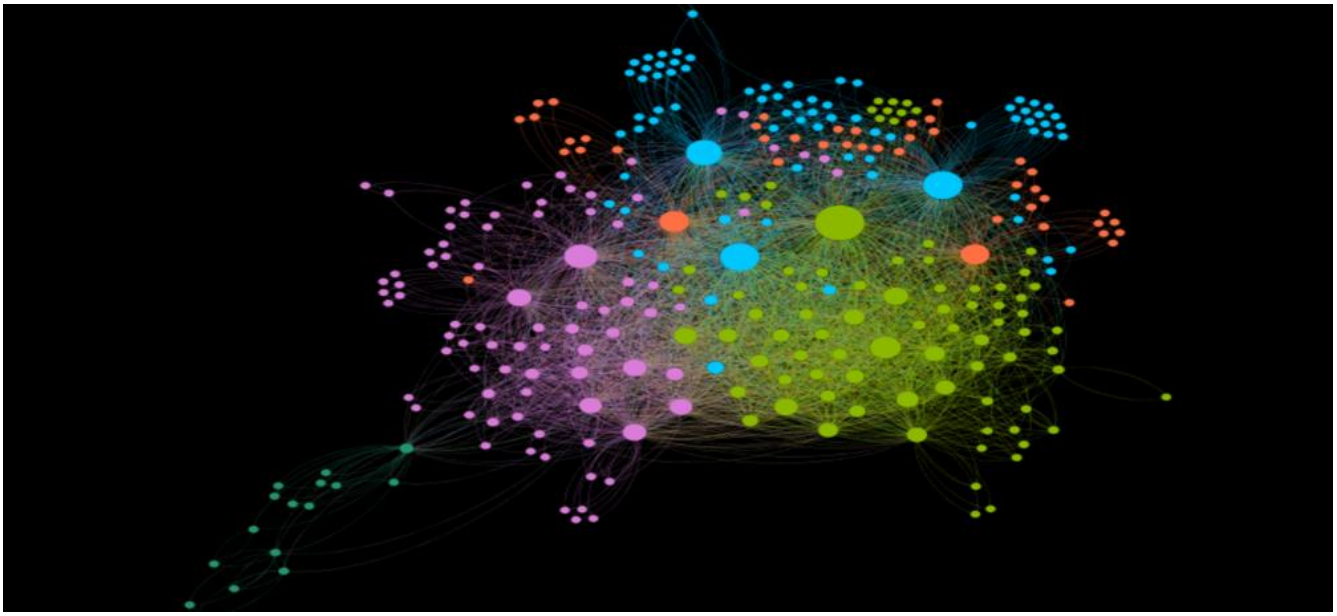


Fig7: Modularity plot using Gephi

TRIADIC CLOSURE:

Triadic closure is the property among three nodes A, B, and C, such that if a strong tie exists between A-B and A-C, there is a weak or strong tie between B-C. We see in our airline network how this property has been violated on several instances. A measure for the presence of triadic closure is *transitivity*.

$$T(G) = \frac{3\delta(G)}{\tau(G)}.$$

$\delta(i)$: the number of triangles that vertex 'i' is involved in

$\tau(i) = \binom{d_i}{2}$: the number of triples of vertex i

The existence of homophily urged us to check the presence of triadic closure in the network. We can see three nodes that are circled in yellow in the figure 8 below. There are flights connecting the airports of Detroit and Phoenix. We can also see flights connecting Phoenix and Honolulu but currently there are no flights from Honolulu to Detroit. We can say Triadic Closure is violated for these three destinations. Currently the customers from Detroit travel to Phoenix and then to Honolulu. As Honolulu is a leisure place, there are customers who are travelling from Detroit to Honolulu. Hence, we can have a potential flight between these two destinations.

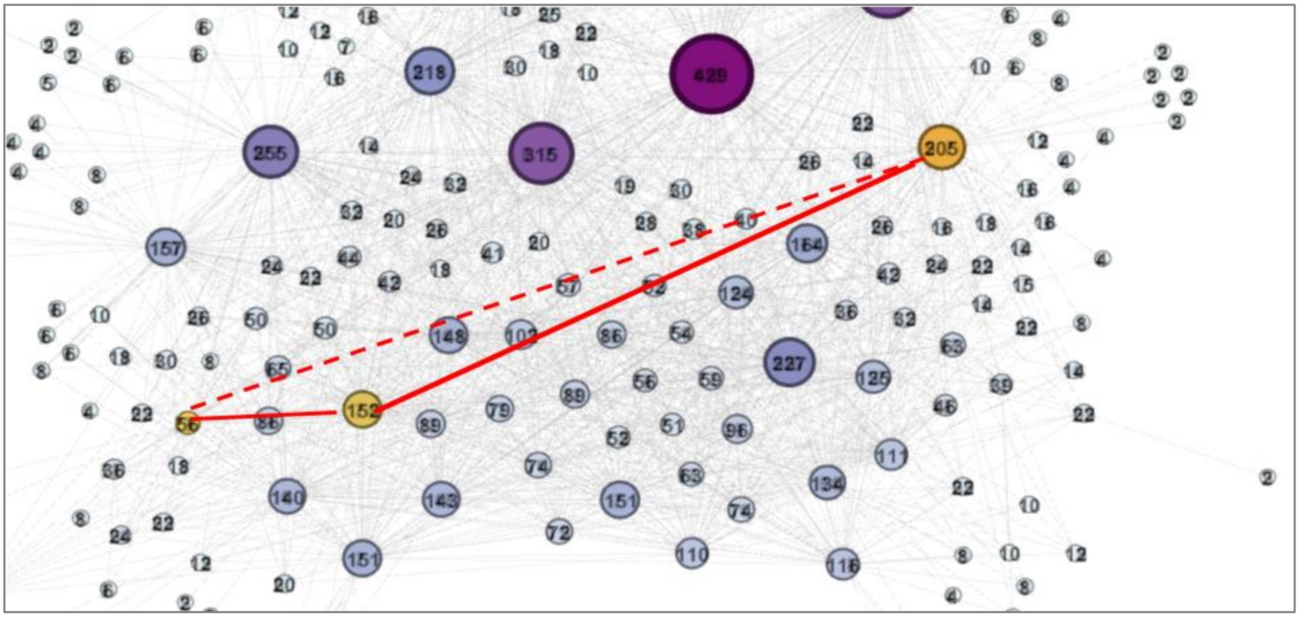


Fig8: Violation of Triadic Closure for 205(Detroit), 152(Phoenix) and 56(Honolulu)

AIRPORT FAILURE AND ITS IMPACT:

Small World Phenomenon: Since the average path length of our network is **2.38**, our network satisfies the small world phenomenon.

This means that every airport is on an average, 2 hops away from every other airport in the network. This is somewhat obvious too as our airports are largely connected with hubs which has connections with lot of other airports.

Natural calamity and terrorist attack generally tend to obstruct our movement. The fear of getting stuck in an airport due to unavoidable reasons advocated us to analyze the consequences of shut down of few airports. We started this by calculating the average length path for our network is 2.38. We took few of our important airports and started removing those airports, one at a time, from our network.

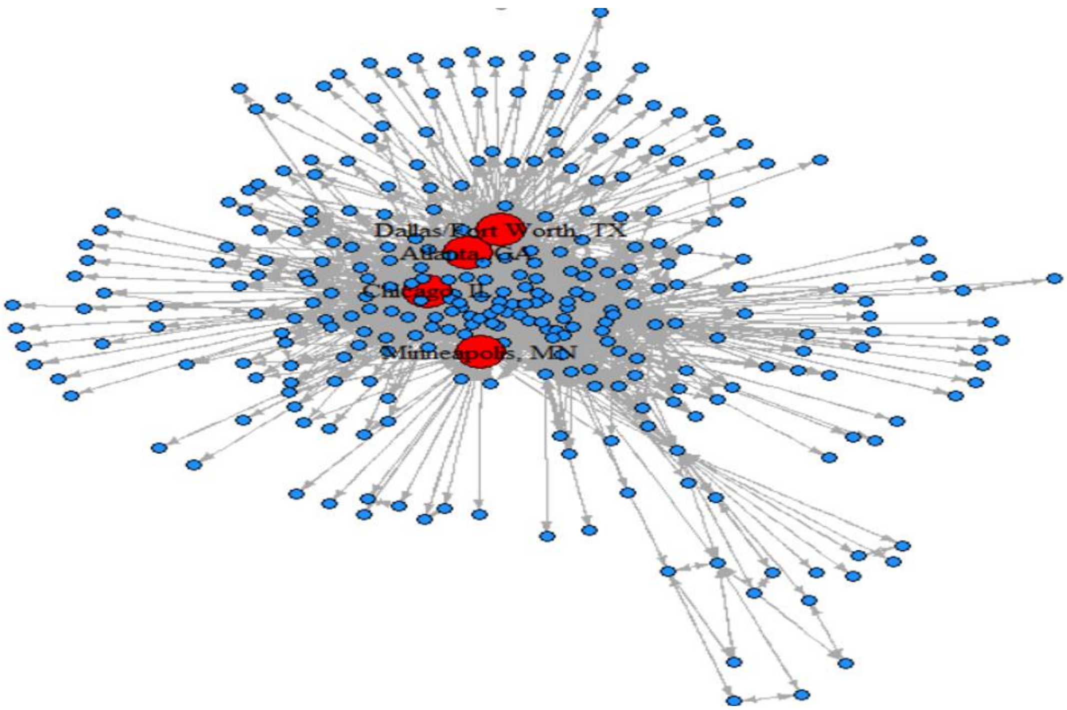
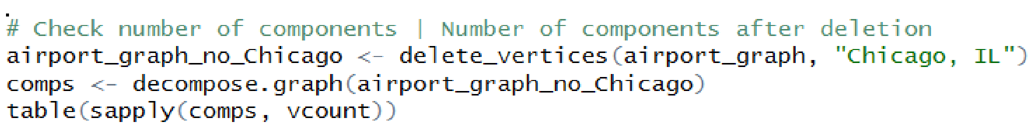


Fig9: Airport network – Nodes sized by their importance

After removing Chicago, we found that there will be 9 unreachable airports in our network and the average path length has increased from 2.38 to 2.49. If we remove Chicago from our network the flights will take 4.6 % more stops to reach from one destination to another destination.

Deleted Node	Unreachable Component	Average Path Length After Disconnecting	% Increase in Path Length
Chicago, IL	9	2.49	4.6%
Atlanta, GA	14	2.43	2.1%
Minneapolis, MN	8	2.39	0.42%
Dallas/Fort Worth, YX	13	2.40	0.84%

Code snippet:



```
# Check number of components | Number of components after deletion
airport_graph_no_Chicago <- delete_vertices(airport_graph, "Chicago, IL")
comps <- decompose.graph(airport_graph_no_Chicago)
table(sapply(comps, vcount))
```

NETWORK GROWTH:

In our analysis of the growth model of the Airport Network from year 2000 to year 2015, we considered the addition of airports through the years. In the analysis, we found that both the networks show a random growth behavior. This signifies that the nodes attach to each other in a random fashion over time, with each node equally likely to get ‘m’ links with equal probability. This means that, during the time of these 15 years, there is no change in the way the airports are added to the network. The behavior of attachment of nodes over time remains random. This can be interpreted rightly considering the network, that an airport (small or big) can be connected to another airport which could again be small or big. It is not necessary that the new airport that is born at any time ‘I’, will connect to a higher degree airport (more important) or a lower degree airport (less important).

This can be inferred from the alpha values obtained using the Ordinary Least Square (OLS) Regression on the Linear Model formed by the frequency and the degree distribution relationship (Log(1-F(d)) vs Log(Degree + 2*(alpha_0)*m/(1 - alpha_0))).

The alpha value plot for both the years can be seen in figure below. The alpha_0 and alpha_1 iteration was done on both the models of 2001 and 2015, and the best alpha_1 was selected based on the difference in the alpha values.

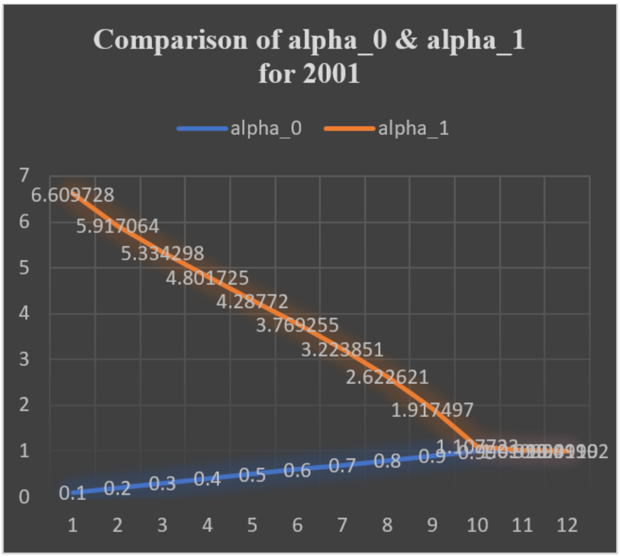


Fig10: Comparison of alpha_0 and alpha_1 for 2001

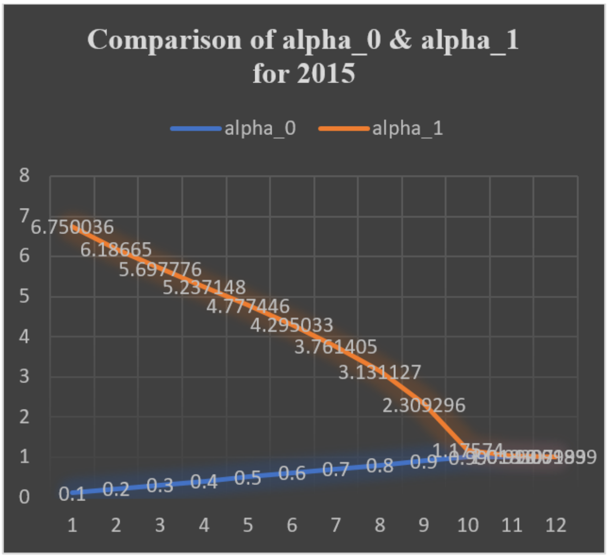


Fig11: Comparison of alpha_0 and alpha_1 for 2015

The *Figure10* above, represents the alpha values for the year 2001. It is seen that the alpha value converges around ‘1’. The best alpha_1 value: 1.0011. This value is approaching 1, which signifies the random network attachment behavior of the network.

Similarly, for the comparison of the alpha values for the 2015 network can be seen in *Figure 11*. This plot also shows that the alpha value converges at the end of the iteration of the alpha values on the linear model. The exact value of the alpha can be sorted out from the tabulated version of the alpha values. It comes out to be 1.001839.

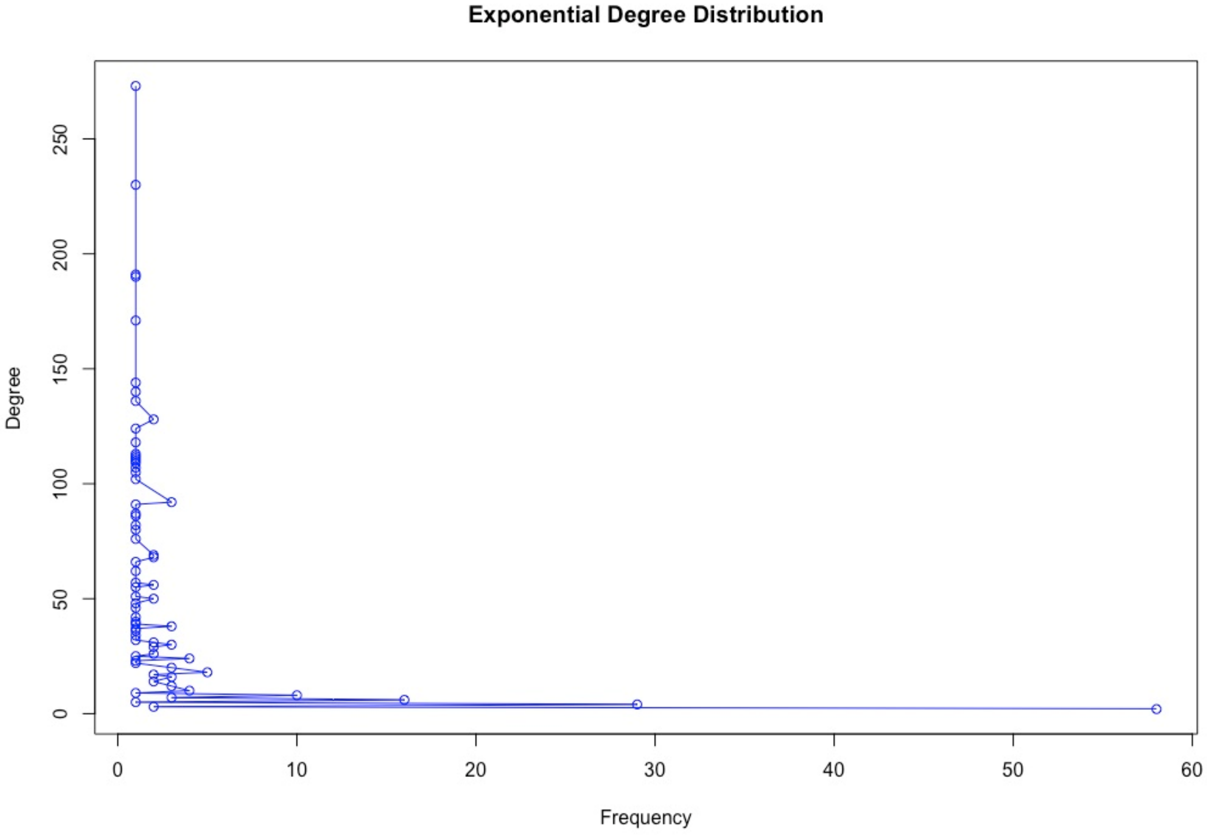
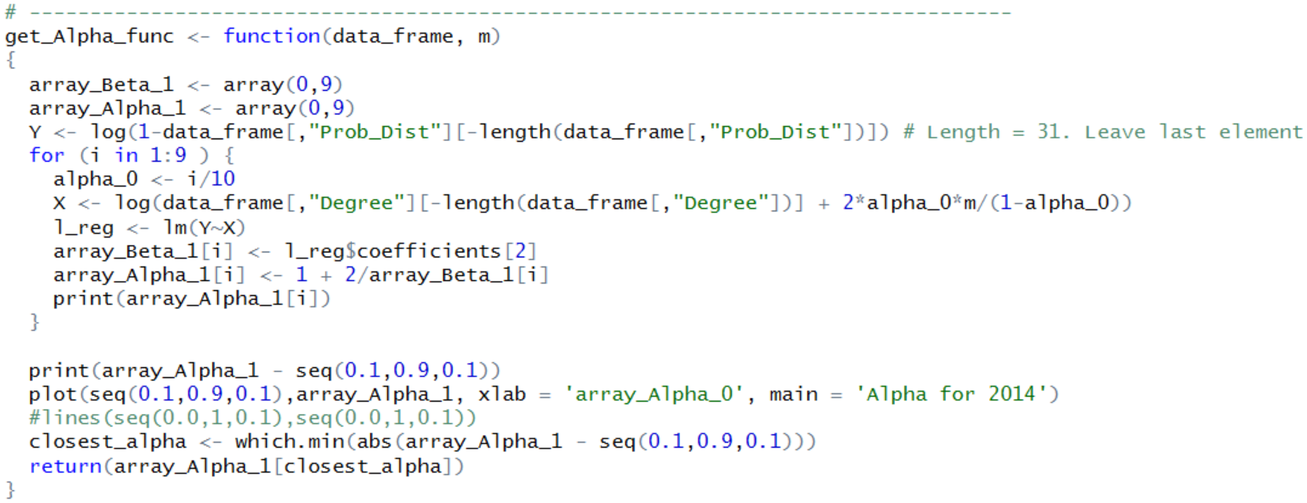


Fig12: Degree distribution of the airport network (2001) showing exponential decay of growth

The above *figure 12* shows a degree distribution plot (Degree vs Frequency) for the airport network of year 2001. It is evident from the plot that the airport network has a distribution which decays exponentially, which resembles the degree distribution of a random growth model of a network.

Code snippet for getting the value of alpha:



```
# ---------------------------------------------------------------------------------
get_Alpha_func <- function(data_frame, m)
{
  array_Beta_1 <- array(0,9)
  array_Alpha_1 <- array(0,9)
  Y <- log(1-data_frame[,"Prob_Dist"][-length(data_frame[,"Prob_Dist"])]) # Length = 31. Leave last element
  for (i in 1:9 ) {
    alpha_0 <- i/10
    X <- log(data_frame[,"Degree"][-length(data_frame[,"Degree"])] + 2*alpha_0*m/(1-alpha_0))
    l_reg <- lm(Y~X)
    array_Beta_1[i] <- l_reg$coefficients[2]
    array_Alpha_1[i] <- 1 + 2/array_Beta_1[i]
    print(array_Alpha_1[i])
  }

  print(array_Alpha_1 - seq(0.1,0.9,0.1))
  plot(seq(0.1,0.9,0.1),array_Alpha_1, xlab = 'array_Alpha_0', main = 'Alpha for 2014')
  #lines(seq(0.0,1,0.1),seq(0.0,1,0.1))
  closest_alpha <- which.min(abs(array_Alpha_1 - seq(0.1,0.9,0.1)))
  return(array_Alpha_1[closest_alpha])
}
```

OTHER NETWORK CHARACTERISTICS:

The plot in the below *figure 13* shows us a positive relationship between average betweenness and degree of nodes. This suggests a low neighborhood overlap. This can be underpinned with the fact that the airports which have high degree like O'Hare and JFK have a huge proportion of neighboring airports with comparatively small degree, ones such as Midway and Newark respectively.

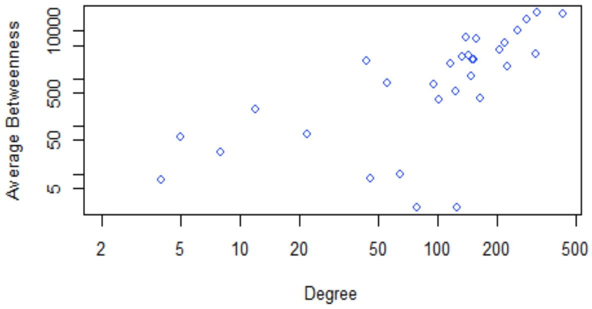


Fig13: Average Between-ness vs Degree

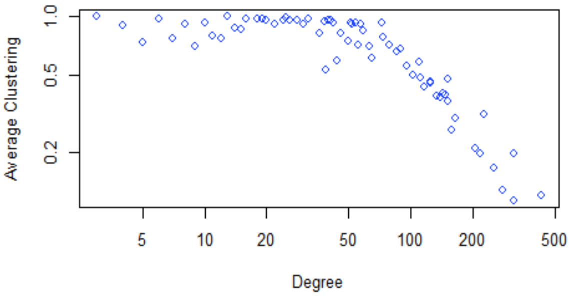


Fig14: Average Clustering vs Degree

In the above *figure 14*, a negative relationship can be seen between average clustering and degree of nodes. This is in-sync with the observation that any airport with high degree doesn't need to be clustered with other airports, however the airports with low degree should be clustered for better connectivity.

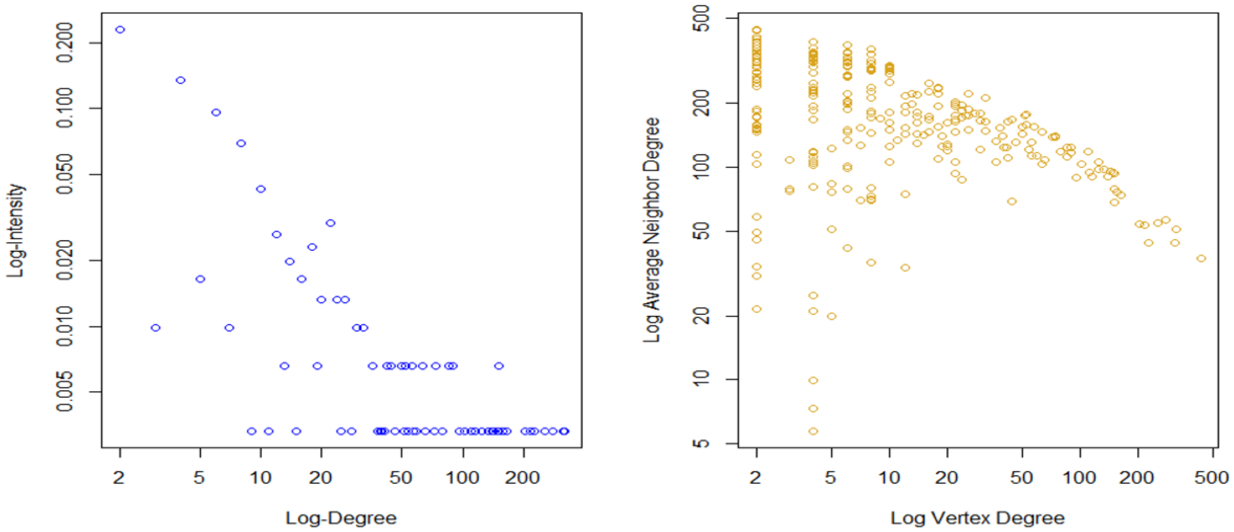


Fig15: The above plots represent the Log-Log Degree Distribution for the airport network

The *figure 15* shows the log-log plot for intensity vs degree distribution. It shows a negative relationship, which shows that there is a drop in intensity as the degree of a node increases for the airport network. Considering the airport network, it shows how the airports don't have interdependence on each other as their degree increases. For a lower degree airport like Midway International Airport, it might have interdependence on other airports. However, a high degree airport like Los Angeles International Airport, is not interdependent on any airport for its operations.

Also from *figure 15*, we can see that if vertex degree is high, then its neighbor's degree is low. It is evident from the graph that nodes with high degree collaborate with low degree, whereas nodes with low degree collaborate with both high as well as low degree. It is consistent with the fact that in the vicinity of a high degree airport, there is a rare possibility that another airport of high degree exists. However, there might be many airports in its neighborhood with low degree. Albeit the airports with low degree might collaborate both with low and high degree airports.

LINK PREDICTION:

Research has shown that it is possible to extract information about future interactions within a given network [10]. Given a snapshot of our airline network, we have tried to infer some new connections among its nodes that are probable to occur soon. David Nowell and Jon Klienberg in their paper mention several proximity measures leading to predictions that outperform chance by factors of 40 to 50, indicating that the network topology does indeed contain latent information from which we can infer about future interactions.

Several methods based on *node neighborhood* are defined that help us predict future interactions:

- Common neighbors: The number of common neighbors of two nodes 'x' and 'y' at time t indicates the probability that the two nodes will collaborate in the future.

  $$\text{score}(x, y) := |\Gamma(x) \cap \Gamma(y)|$$

  The common neighbors is a fairly simple approach and yet performs relatively well on accuracy.
- Jaccard's coefficient and Adamic/Adar: The similarity score based on a set of attributes of two nodes indicates the probability that the two nodes 'x' and 'y' will collaborate in future.

  $$\sum_{z\,:\,\text{feature shared by }x,y} \frac{1}{\log(\text{frequency}(z))}$$
- Preferential attachment: The probability that a new edge involved node 'x' is proportional to the current number of neighbors of x.

In addition to these, there are several other deterministic, probabilistic, feature construction and supervised learning methods used for link prediction. One such method described in the subsequent paragraph has been used by us for link prediction.

In this network, two airports are linked with flights if there is a statistically significant partial correlation in the number of flights between them. The passenger data of airports were collected for period of 10 years from 2005 to 2015. We fetched the data from the Federal Aviation Administration website [2]. Our dataset shows the number of passengers flying between two airports during this time.

To determine statistically significant links between two airports, following steps were followed:

1) Calculated the partial correlation coefficients between each node.
2) Compute the Fisher's transformation to approximate the bivariate distributions and to determine the confidence intervals that are used to obtain p-values.
3) Apply the Benjamini-Hochberg adjustment to control for the false discovery rate; and use a threshold of $p < 0.05$ to identify statistically significant partial correlations.
4) Use the calculations in the above steps to determine the edges among the nodes.
5) Construct the network of firms for whom degree centrality is at least 1.

Based on the above analysis, there were quite a few link predictions based on the correlation among the nodes. Some of the links that we predicted in the year 2005 were added by the year 2015. For example, our model rightly predicted a link between Santa Barbara and Chicago. Earlier, there were flights indirectly from Chicago to Santa Barbara via Los Angeles International Airport. Over the course of time, Santa Barbara has evolved as the industrial hub and now there is a direct flight connecting it with Chicago. Same scenario can be observed with a link prediction between Seattle and Chicago. There were no flights connecting Chicago directly to Seattle in 2005, however now we can see a link connecting them directly.

Our link prediction model just considers the passenger data. If we extend our scope to consider the cost of flight ticket and distance between two airports, we can find better correlations amongst the airports and predict links at a more granular level. This can be better understood taking an example of Panama, which is a tourism hub. As of now there is no flight connecting New York and Panama directly. However, there is a flight connecting the two airports via Tampa. However, considering the large number of tourists and the distance between Tampa and Panama via road (~400 miles), it could be inconvenient for the tourists. Also considering the flight between New York and Panama via Tampa which is very expensive, a link could well be predicted if distance and cost was considered as a factor.

5. CONCLUSION

In our project, we have used an airline network and have applied various centrality measures to find out the most important airports within the network. We have used degree centrality, betweenness centrality, page rank, hub and authority scores to find out the important nodes. From our analysis, Chicago's O'Hare, Atlanta, Houston, Dallas and Denver are the most important ones.

We have also analyzed the impact of an airport failure. When Chicago's O'Hare airport is hit by a snow storm and goes out of operation then there are nine airports that become disconnected from the main component. Also, the average travel time from one source to destination in this case goes up by 4.6%.

Going forward, in future we can increase the depth of research by using concepts of commuter flow centrality and delay centrality to find out the important airports. Using measure of delay centrality, we can evaluate the appropriate time to increase or decrease the number of flights and thereby control the passenger congestion.

By extending our work beyond the borders of the United States, we can develop a model that can help prevent spread of epidemics. By knowing the infected areas, we can figure out which airports to disconnect from the main component and thereby prevent the spread of epidemics.

A famous quote by Enrico Fermi says: If the result confirms the hypothesis, we have made a discovery. If the result is contrary to the hypothesis, we have still made a discovery!!

6. REFERENCES

1. http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/press_releases/airline_traffic_data.html
2. https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/
3. http://web.mit.edu/sheffi/www/selectedMedia/sheffi_urban_trans_networks.pdf
4. http://nlp.stanford.edu/IR-book/html/htmledition/hubs-and-authorities-1.html
5. http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf
6. https://www.cs.cornell.edu/home/kleinber/networks-book/
7. https://toreopsahl.com/tnet/weighted-networks/node-centrality/
8. http://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4160&context=sis_research
9. https://en.wikipedia.org/wiki/Modularity
10. https://www.cs.cornell.edu/home/kleinber/link-pred.pdf