Social Network Analysis - Project Report

Airways Transport Network

Team Members – Group14

Likhil Kumar Rachuri (lkr46@drexel.edu)

Vuthej Krishna Reddy (vv334@drexel.edu)

1 Abstract

This paper deals with the domestic airport network of United States of America. USA has the world's largest domestic airport network and the demand for air traffic is expected to increase 2 to 3 times in the next few decades. There are many previous studies which concentrated on the airport network, but they mostly included the Freight network into their consideration. In our case, we have restricted our scope for passenger flight network. This is to identify the most robust airport community and airport communities within the network which are easily prone to disruption. Understanding these facts are critical to identify the areas of improvement.

Each individual airport within the network is considered to be a node and all incoming/outgoing flights from one airport to another is assumed to form the connections, ties or edges. This is a directional network with number of incoming flights to a node is treated as in-degree of the node and number of outgoing flights from a node is considered to be out-degree.

Firstly, Clauset, Newman and Moore (CNM) method comprising of greedy modularity communities for faster identification of clusters which is best suited for the large scale networks is employed in this paper for network analysis. [1]

Secondly, Girvan-Newman Algorithm for identification of robust communities is employed and analyzed.

2 Literature Review

Previous work and research on topics similar to air traffic analysis includes Network characteristics of air traffic in the continental United States by Banavar Sridhar and Kapil Sheth [2003] from NASA [2]. This paper talks about the performance of future air traffic that can be anticipated by using the characteristics from the current air traffic system and includes future scenarios of air traffic network. Preliminary analysis made by the team found that a three-times growth in the overall air traffic network in US tends to result in a ten-times increase in the density of air traffic in certain parts of US. This also increases the scope for delays encountered with bad weather causing disruptions in air network monetarily. Hence the need for a better analysis to understand the area of improvements that can withstand the future issues efficiently.

Another study involves the effect of air traffic growth on the US air transportation system dealt by Tatsuya Kotegawa, Daniel DeLaurentis, Kimberly Noonan and Joseph Post in their research paper titled Impact of Commercial Airline Network Evolution on the US Air Transportation System [2011] [3]. This paper considers the US NAS as a dynamic network rather than a static network and coins a restructuring algorithm which is capable of estimating the gaps in future predictions more accurately than the Federal Aviation Administration of US currently employs.

Trends in the arrival and departure of US air traffic has been investigated by Li Shan-Mei, Xu Xiao-Hao and Men Ling-Hang in their paper Fluctuations in airport arrival and departure traffic through Network Analysis [2012] [4].

Another interesting topic dealt was the proposal of model for pricing of landing slots by Sungwook Hong, Patrick T Harker. One model with an exogenously determined allocation and the second with an endogenous allocation. [5]

A comparative study on airport connectivity in countries like China, Europe and US to understand the quality of service provided by these entities towards the passengers, describes that US network in a way is most precisely coordinated when we take into account the indirect connections. Unparallel level of service across the airport networks unbiased by the size and monetary value a particular airport adds to the network can be seen in Europe air network. Swiftness in commutation offered to the users by relatively smaller number of airports per inhabitant sums it up for the quickest travel network administered in China. [6]

3 Focus on Research Problem

Huge data is being generated with the increasing demand for air-traffic, allowing the need to analyze the data more deeply to come up with interesting facts and trends. These trends will further help the competitors to better understand the needs of the users and to provide better facilities and connectivity.

Demand can be forecasted by observing trends and patterns from the network built using past data. This puts the airline companies in a better position for supply and demand.

From our observations, we would like to unearth interesting facts and trends which suggests a better understanding of the working of Airways Transportation Network.

These include –

- Busiest Airport
- Least busy Airport
- Centrality - Influential Airports
- Eigenvector Centrality
- Group-based Community Detection
- Robust Communities

It is a known fact that the pandemic COVID-19 has shaken lives across the world and it also had profound impact on the economy of most of the countries. Post COVID-19, Governments across the globe must act swiftly and accurately in reviving their economies to avoid being absorbed into recessions.

Through our analysis on the robustness of communities within the airport network of USA, we will shed light on to the communities that are strongly connected or highly cohesive. Understanding this aspect will majorly benefit the Government of USA in taking necessary steps in reviving the airport network post COVID by focusing on strongly connected communities and be prepared for any such impacts obstructing the network connectivity.

4 Details of the Dataset

U.S. Department of Transportation provides interesting data through Bureau of Transportation Statistics website www.transports.bts.gov. This website includes information of all airline carriers operating within United States of America.

Dataset we are interested in provides information about the source and destination of a flight within United States of America for the year 2019.

Airports are treated as Nodes and a flight traveling from one airport (source) to another (destination) establishes a connection/tie between the nodes. i.e. Source and Destination.

Current dataset consists of 1,946 nodes with 263,846 ties. These connections are spanned across 313 different carriers operating under Federal Aviation Administration (FAA). Federal Aviation Administration is a department of transportation system which monitors the air traffic network in United States of America.

UNIQUE_CARRIER_NAME	ORIGIN_AIRPORT_ID	ORIGIN	DEST_AIRPORT_ID	DEST	MONTH
Empire Airlines Inc.	10140	ABQ	11413	DRO	3
Empire Airlines Inc.	10140	ABQ	11711	FMN	3
Empire Airlines Inc.	10194	AFW	12278	ICT	3
Empire Airlines Inc.	10194	AFW	12896	LBB	3
Empire Airlines Inc.	10194	AFW	13158	MAF	3
Empire Airlines Inc.	10194	AFW	15370	TUL	3
Empire Airlines Inc.	10299	ANC	10299	ANC	3
Empire Airlines Inc.	10299	ANC	11555	ENA	3
Empire Airlines Inc.	10299	ANC	11630	FAI	3
Empire Airlines Inc.	10299	ANC	12184	HOM	3
Empire Airlines Inc.	10666	BLI	11762	FRD	3
Empire Airlines Inc.	10666	BLI	14747	SEA	3
Empire Airlines Inc.	11050	CLM	14747	SEA	3
Empire Airlines Inc.	11198	CVO	14057	PDX	3

Above table displays few instances/rows of data from our current dataset.

Unique_Carrier_Name : This is a unique Carrier Name allotted to every individual airline. When the same name has been used by multiple carriers, a numeric suffix is used for earlier users. For example, ‘Air Caribbean’, ‘Air Caribbean (1)’.

Origin_Airport_ID : Origin Airport, Airport ID. An identification number assigned by U.S. Department of Transportation to identify a unique airport.

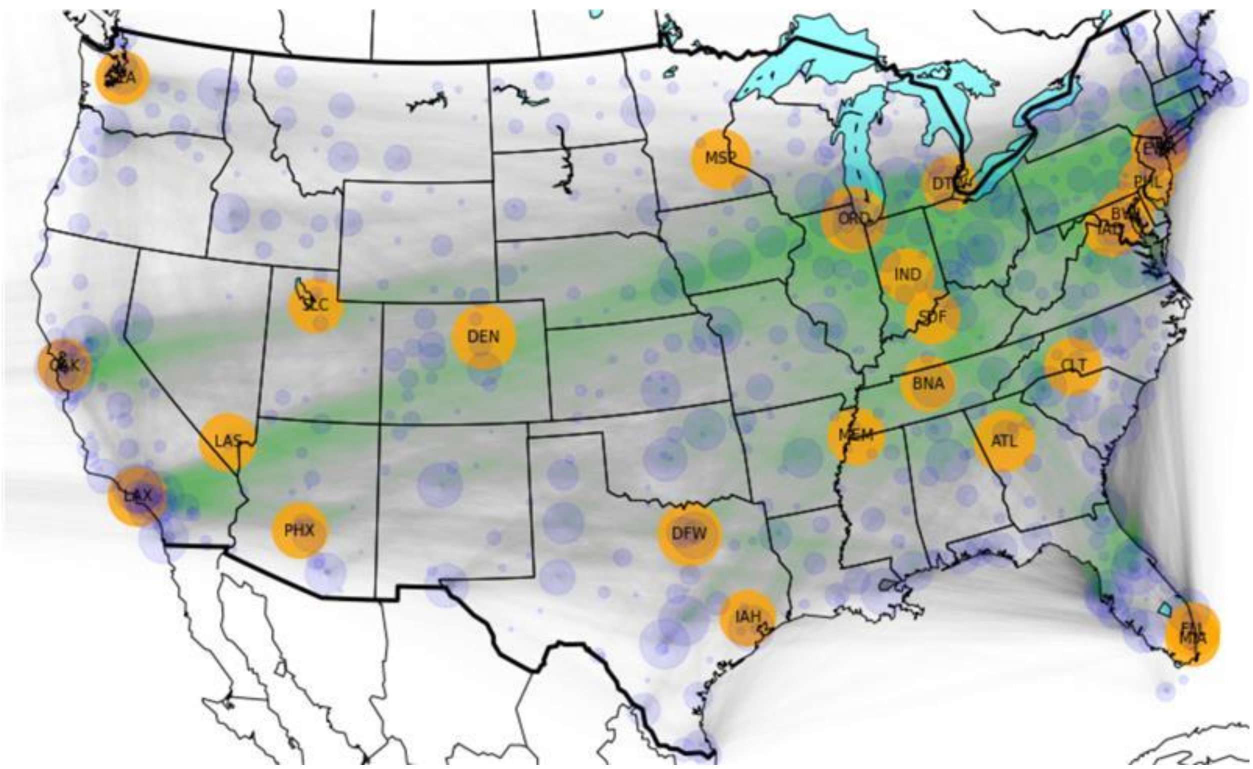
Origin : Origin Airport Code.

Dest_Airport_ID : Destination Airport, Airport ID. An identification number assigned by U.S Department of Transportation to identify a unique airport.

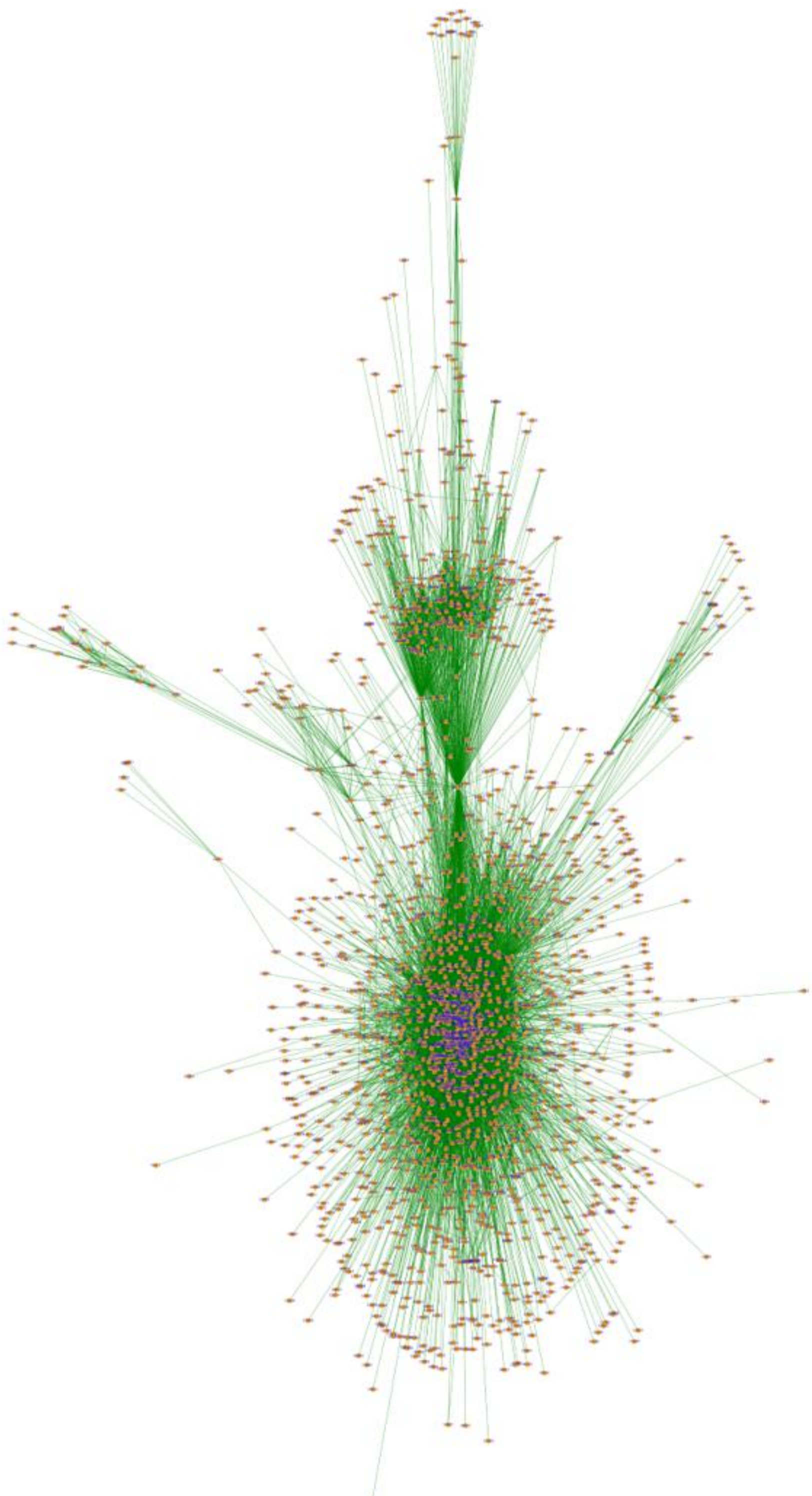
Dest : Destination Airport Code.

Month : This field provides the time period in months within a year.

Visualization-1 (Network built over geographical data using NetworkX)



Visualization-2 (Network built using Networkx without geomap)



<u>5 Applied Social Network Analysis Methods</u>

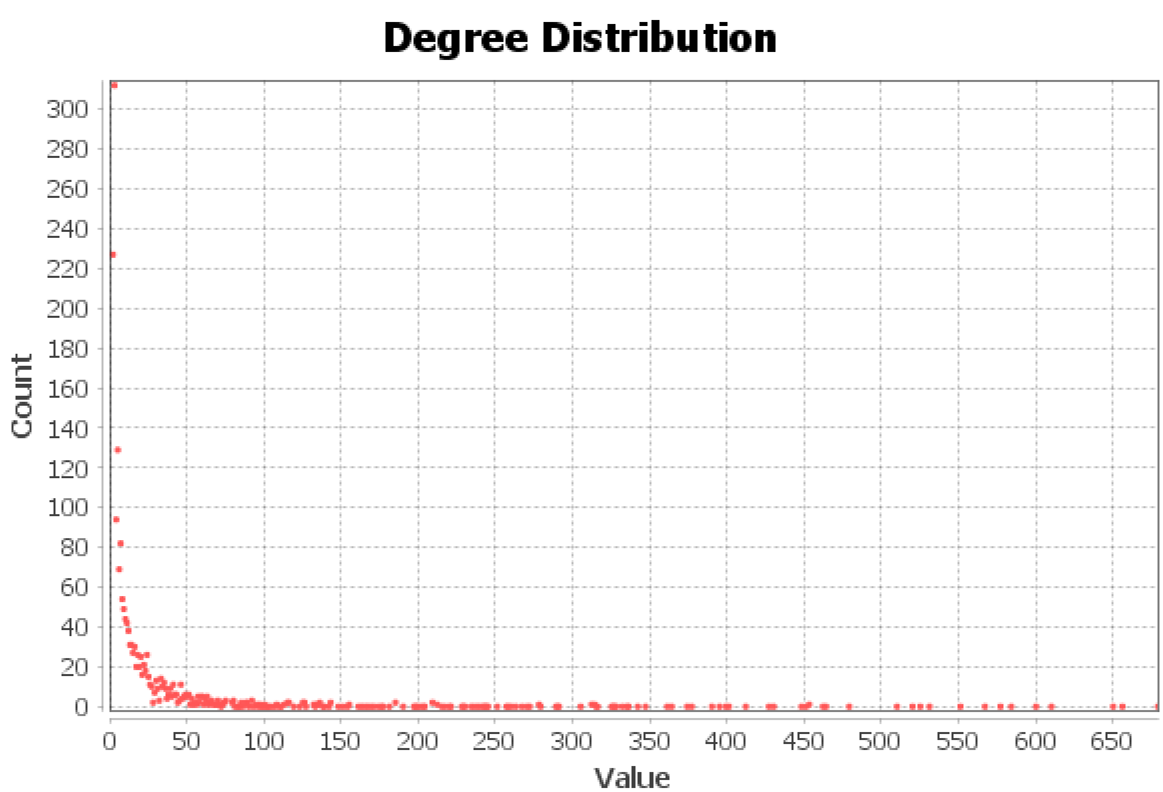
Centralities provide information pertaining to the influential nodes in the network. Influential nodes are the ones with wide range of connections within the network. In this paper, we will be mainly discussing about four major centralities. Namely, Degree Centrality, Betweenness Centrality, Closeness Centrality and Eigenvector Centrality.

Degree of a node is defined as the number of edges/links it has with the other nodes.

For Non-directional relations, the central actor is defined as the one involved in many ties. **'Degree Centrality'** for directed network takes into account, both the in-degree and out-degree of a particular node within the network. Degree centrality is one of the measurements to calculate the importance of the nodes. If the value is high, it indicates that the respective node is having more connections within the network, thus tends to be more important. It is more probable to be important node because information can be passed to many other nodes as number of links are high for the respective node.

The degree distribution plot follows the power-law distribution as indicated below. Airports with very high degree centrality are relatively low compared to the high number of airports with low degree centrality.

Average Degree identified in the network is 16.410.



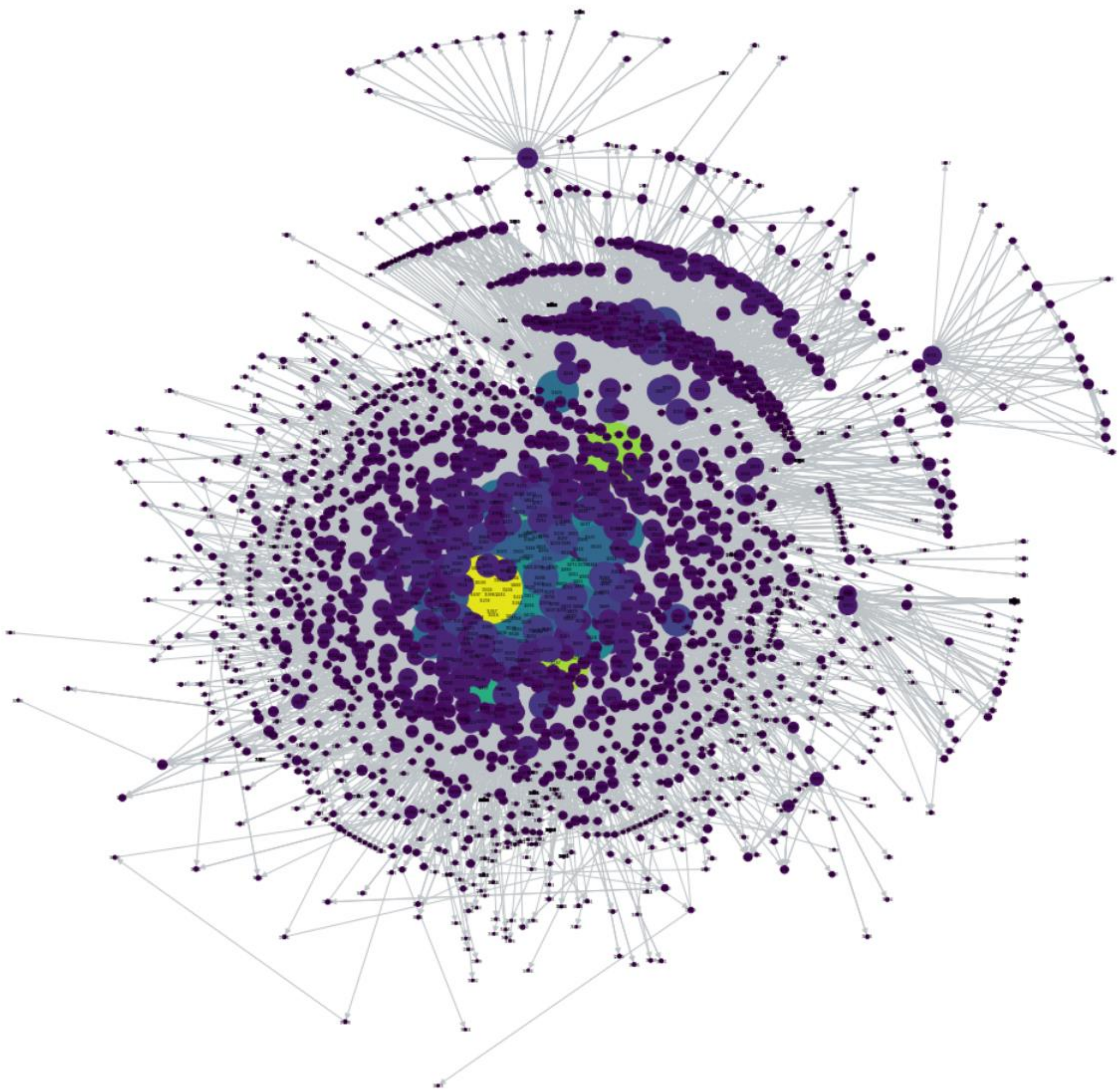


Figure 1 – Degree Centrality for the whole network

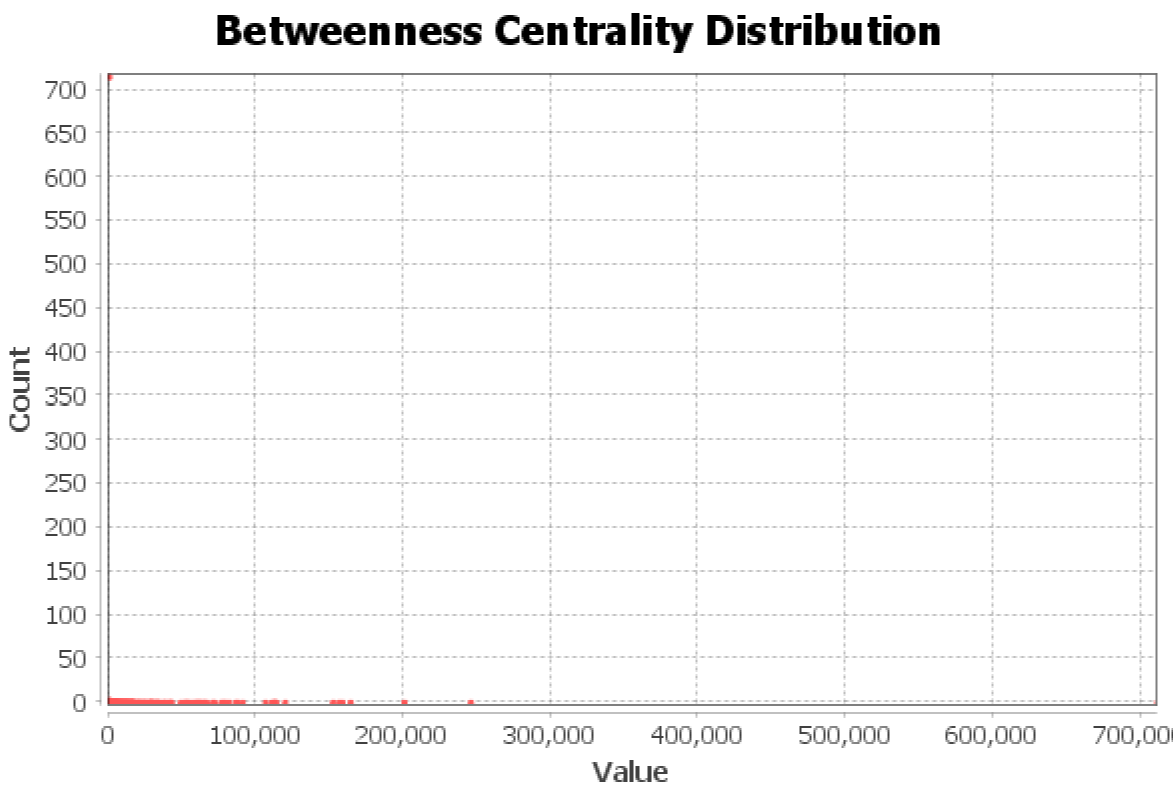
From above visualization we can observe that the degree centrality ranges from as little as 0.00051 to 0.34841. It was conscious effort to include the standardized values in order to have the provision for comparison with other networks.

O'Hare International Airport (ORD), Chicago shows to have the highest degree centrality of 0.34841. Con trarily Atlantic Municipal Airport (AIO), Atlantic has the lowest degree centrality of 0.00051.

For a network with 'g' nodes, **'Betweenness Centrality'** is the value obtained by the division of summation of number of geodesics linking two nodes (for all possible pairs except for the one node that we are calculating for) which contain our node that we are calculating for in the geodesic path, over the total number of geodesics that actually exist (for all possible pairs except for the one node that we are calculating for).

Standardized Betweenness Centrality for a network with 'g' nodes, is the value obtained by multiplying the 'Betweenness Centrality' with 2 and then dividing it by the product of 'g-1' and 'g-2').

Below displayed plot displays the betweenness centrality distribution.



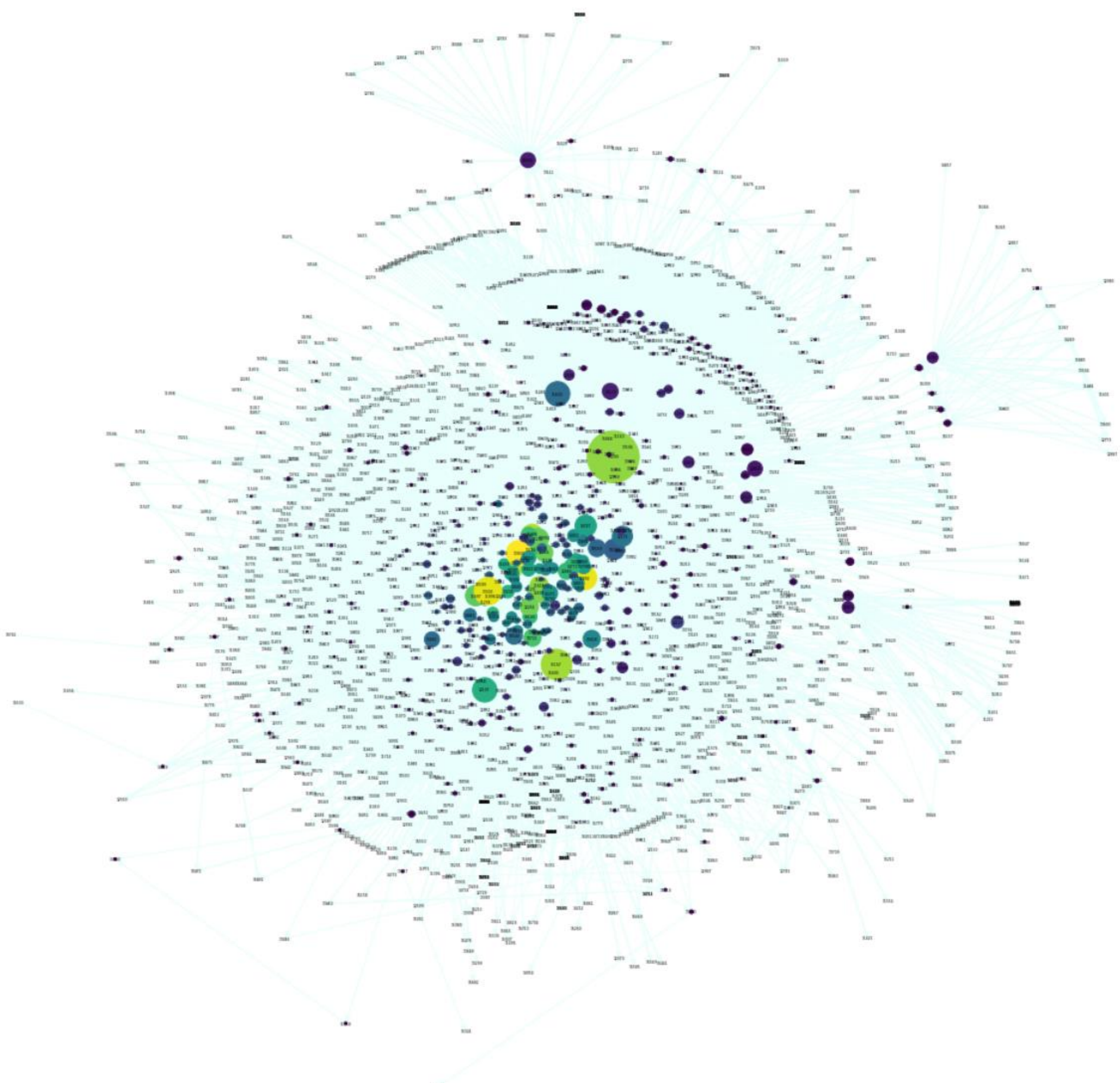


Figure 2 – Betweenness Centrality for the whole network

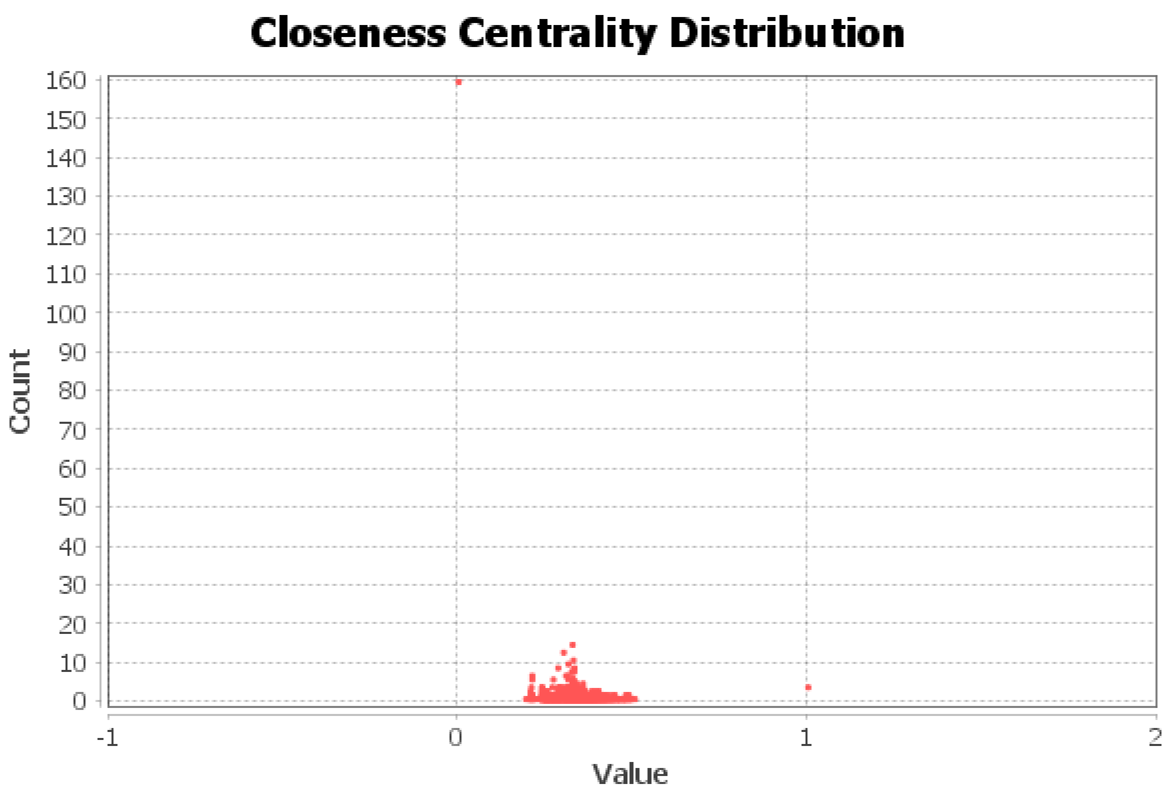
Ted Stevens Anchorage International Airport (ANC), Anchorage has the highest betweenness centrality of 0.18889.

Boca Chica Field (NQX), Key West has the lowest betweenness centrality of 0.

Considering the geographical position of both the above airports, we can observe that ANC which is connecting all the smaller airports in Alaska is acting as a bridge to the rest of the airports in other states. Had there been another major airport which has the capacity to take in the traffic as does the ANC, betweenness centrality would have been different. Coming to NQX, as this is located in the coastal area of West Florida it doesn't specifically hold a route that needs to be dependent on this airport. Thus the routes this particular airport is connected to, can also be reached from other airports. This is the main reason to observe the 0 betweenness centrality.

'Closeness Centrality' of a node is the reciprocal of the summation of geodesic distance from that node to the rest of the nodes in the network. Standardized Closeness Centrality for a network if 'g' nodes, is the product of 'g-1' with reciprocal of the summation of geodesic distance from that node to the rest of the nodes in the network. This parameter talks about how close or attached a particular node is to the rest of the network.

Below plot displays the closeness centrality distribution. Highest diameter observed within the network is 8 and an average path length of value 3.097. Closeness centrality for all the airports are almost equal and a cluster is observed.



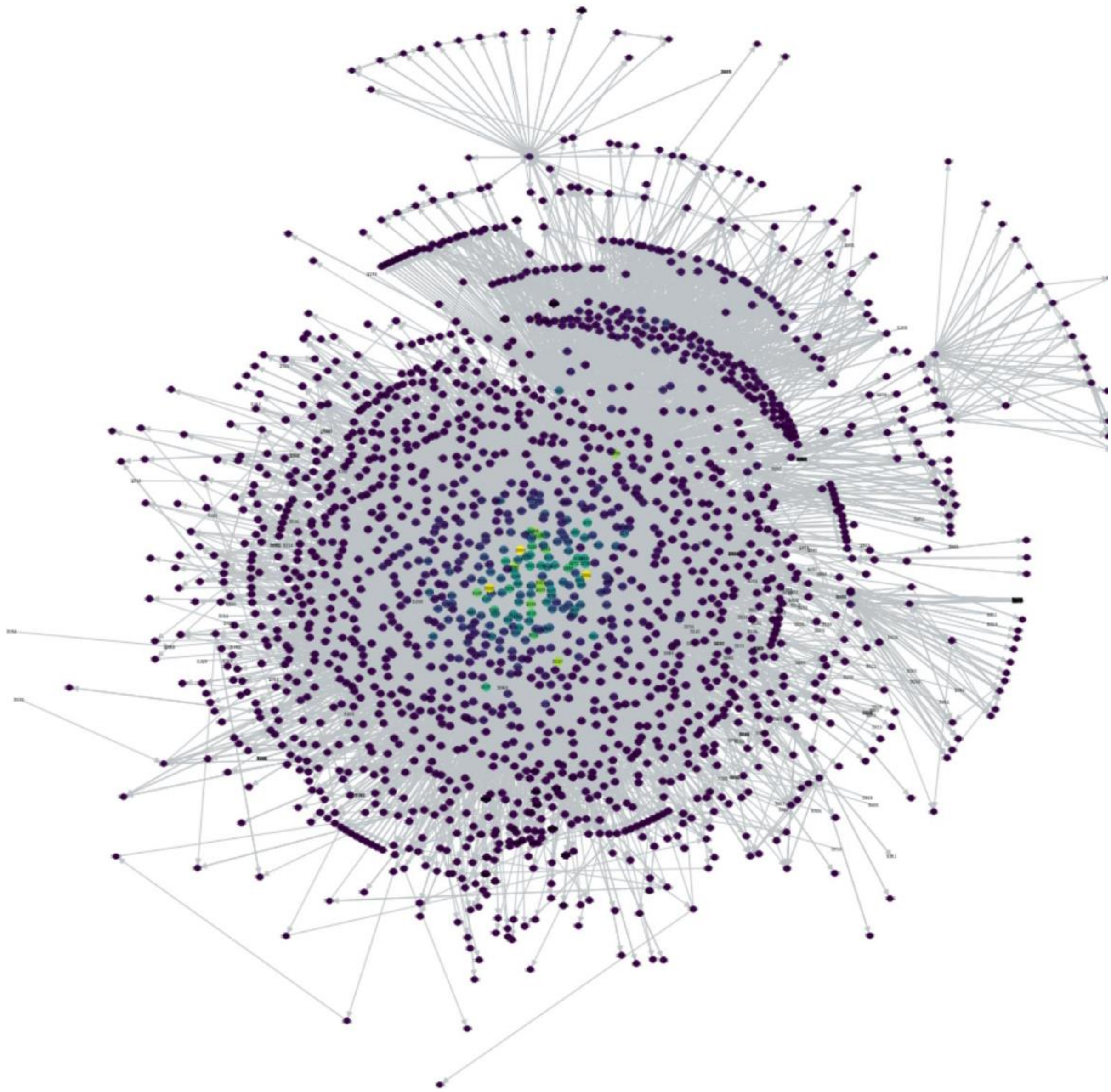
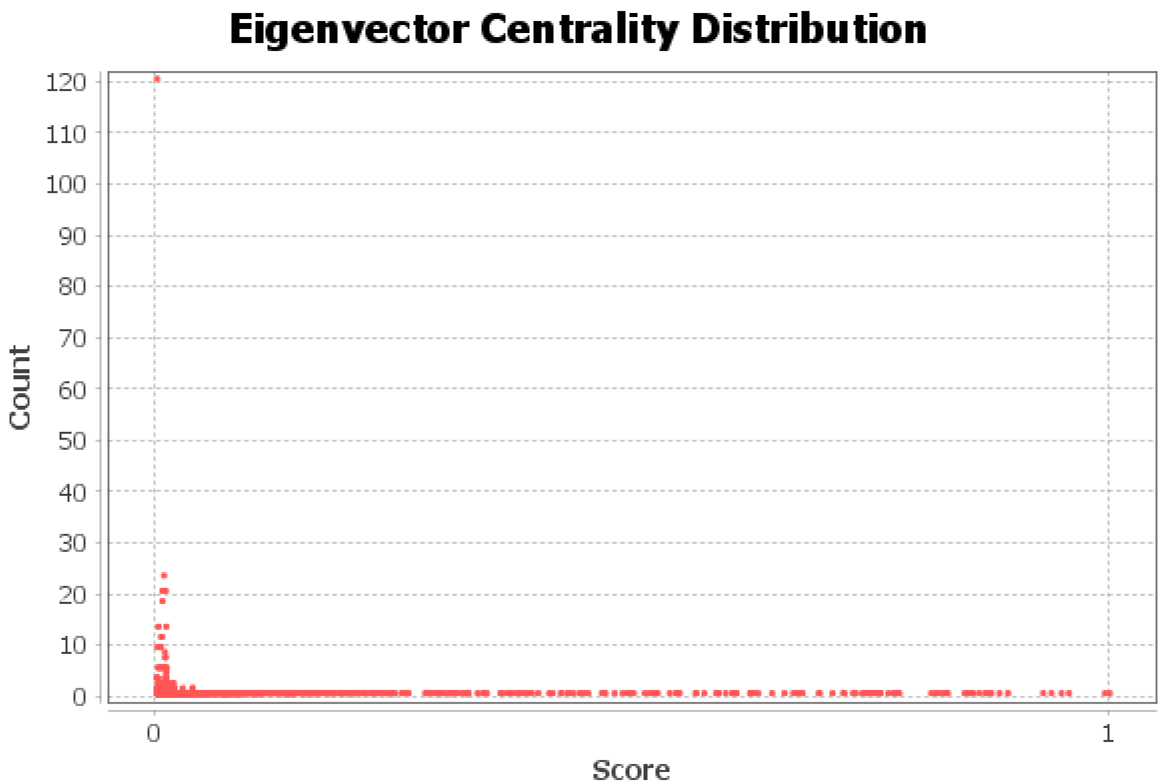


Figure 3 – Closeness Centrality for the whole network

O’Hare International Airport (ORD), Chicago displays the highest closeness centrality of 0.45658.
Boca Chica Field (NQX), Key West has the lowest closeness centrality of 0. ORD airport has the highest values for both degree and closeness centrality. This explains that this airport is connected to many other airports with the shortest distance. NQX has the least value (zero) for both closeness and betweenness centrality. Geographical location of NQX along with the type of connections it has towards other airports contributes towards this lowest centrality values.

‘**Eigenvector Centrality**’ takes into consideration the centrality of neighbors of a node in order to compute the eigenvector centrality of a node. This is a measure of influence a particular node has over the network. If a node has lot of in-degree, it automatically has high eigenvector centrality as many other nodes are pointing towards this node.

Below is the plot for Eigenvector Centrality Distribution. Number of iterations are 100. One major observation is that it looks almost similar to the Degree Centrality Distribution.



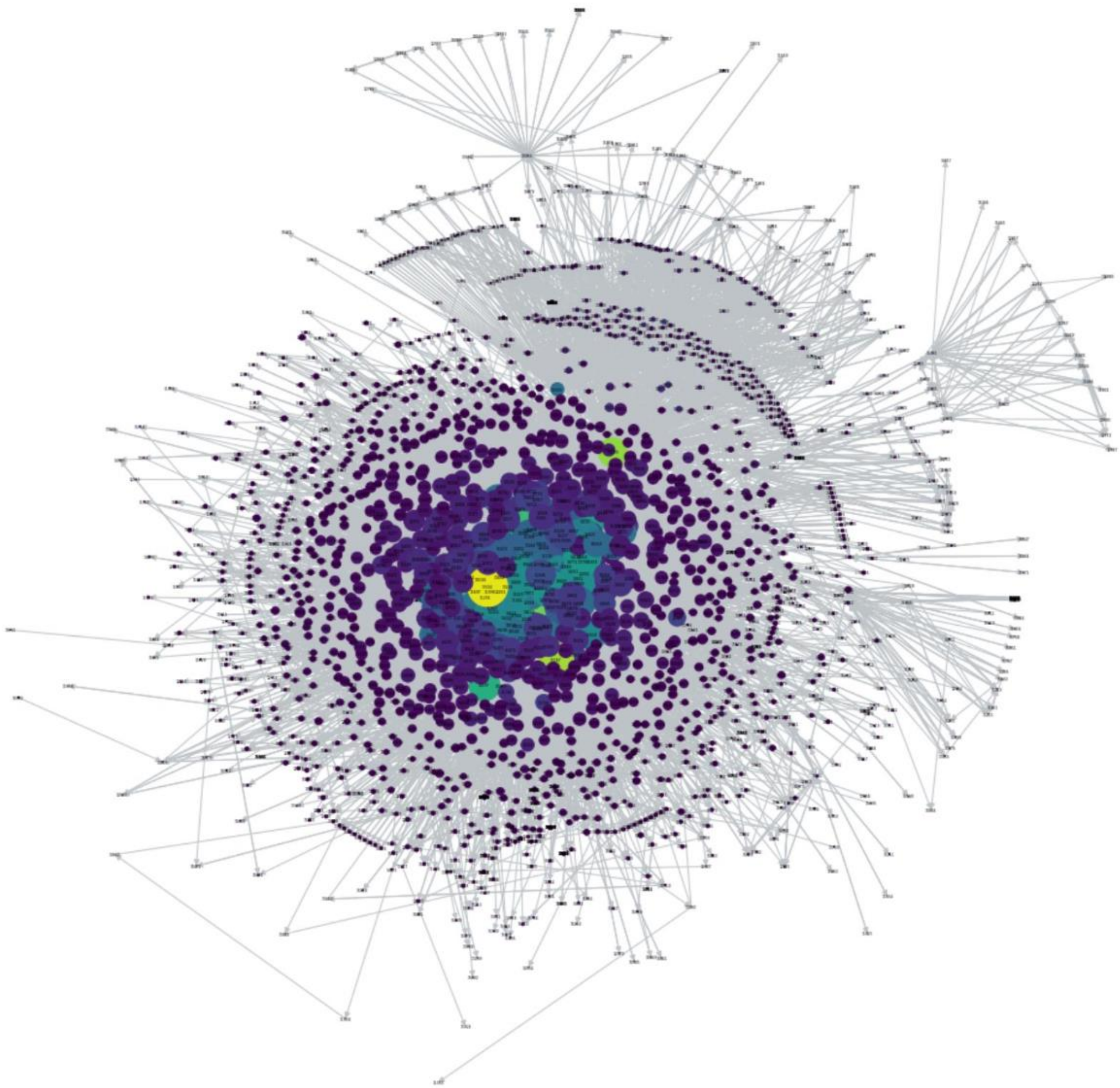


Figure 4 – Eigenvector Centrality for the whole network

O'Hare International Airport (ORD), Chicago has the highest eigenvector centrality of 0.12884. Boca Chica Field (NQX), Key West has the lowest eigenvector centrality of 0.

We can conclude that the ORD airport is having very high inbound connections thus proving to be a central node. High outbound connections can also be observed but these are relatively less high when compared to inbound traffic.

To understand the trends in the air traffic network, we can divide the traffic over the period of 12 months and can analyze each quarter. This helps to unearth intrinsic properties of the network which otherwise cannot be observed on the network as a whole. As the data we've collected is limited to first 10 months of 2019, we are only considering the first three quarters to understand the trends.

Quarter-1 (January to March)

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
ORD (0.39301)	LAX (0.45778)	ANC (0.22052)	ATL (0.14461)
A29 (0.00071)	ANB (0.0)	05A (0.0)	ANB (0.0)

Quarter-2 (April to June)

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
ORD (0.38496)	ORD (0.45475)	ANC (0.22234)	ORD (0.14338)
A27 (0.00065)	05A (0.0)	05A (0.0)	05A (0.0)

Quarter-3 (July to October)

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
ORD (0.35868)	ORD (0.44897)	ANC (0.19925)	ORD (0.14190)
BFB (0.00060)	A65 (0.0)	05A (0.0)	A65 (0.0)

Table 1 – Influential Airports from each Quarter of 2019

Even after splitting the data into 3 different quarters, O'Hare International Airport (ORD) seems to have the highest degree over the three different quarters. This is similar to what we've seen for the whole network. At any given point of time in the year, ORD seems to have more inbound and outbound flights.

Los Angeles International Airport (LAX) has the highest closeness centrality over the first quarter of the year followed by ORD for the next two quarters. It can be assumed that LAX can be seen as a go to destination over the month of January as we witness the New Years Eve. Hence, a greater number of airports are easily reachable to LAX than any other airport. However, over the next two quarters, ORD seems to have highest closeness centrality.

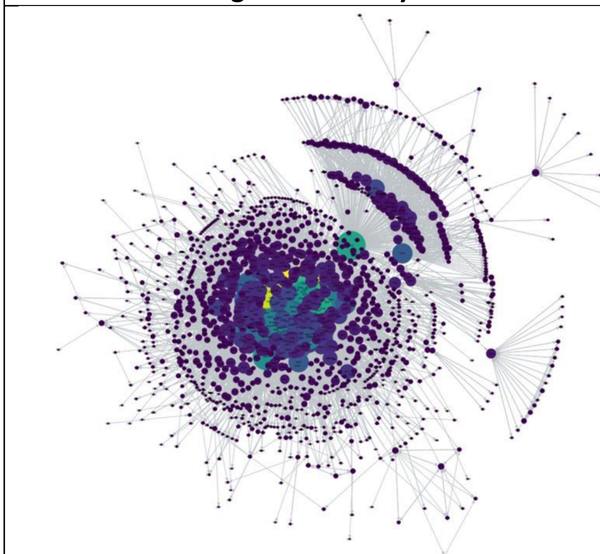
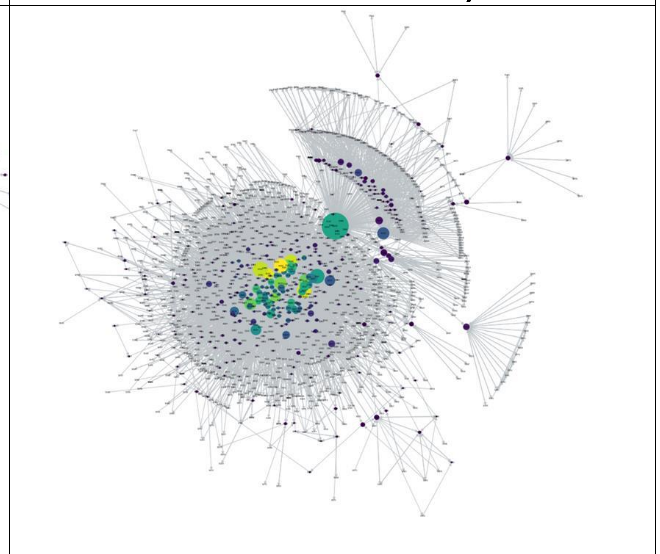
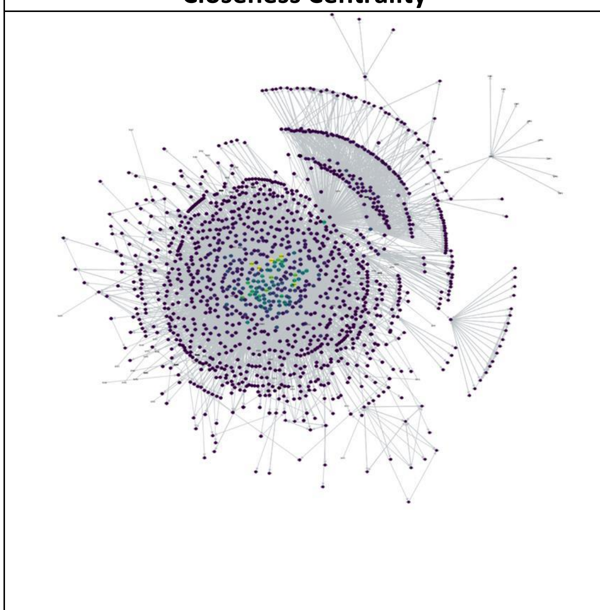
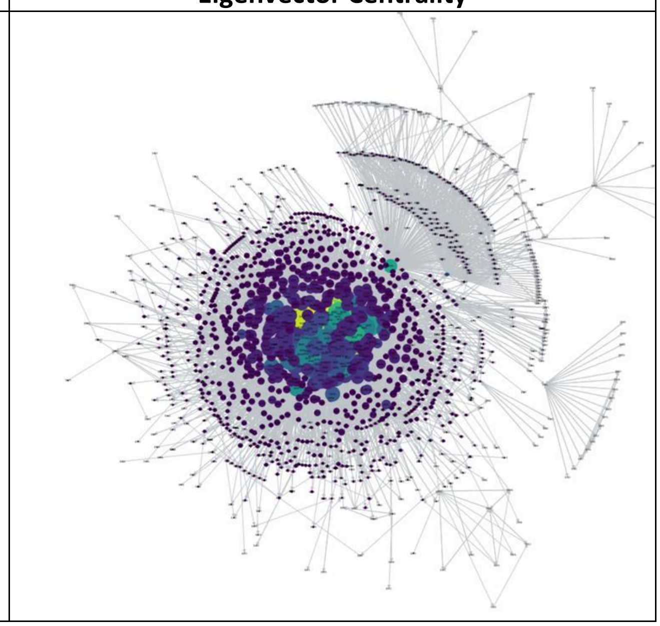
Throughout the year, ANC has the high betweenness centrality value proving it a major player in connecting parts of Alaska to the rest of the states in USA. Hence, ANC acts as a bridge between the local airports of Alaska and rest of the states in US to be connected and it also proves to be crucial.

Hartsfield-Jackson Atlanta International Airport (ATL) seems to be highly influential over the first quarter of the year then followed by ORD for the next two quarters. No specific insights can be deduced at this point with the current air traffic data as we can't inspect day to day trends rather only per 3 months.

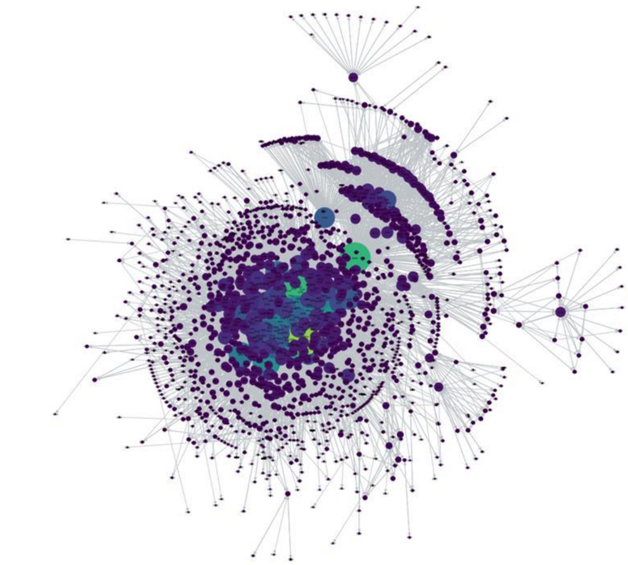
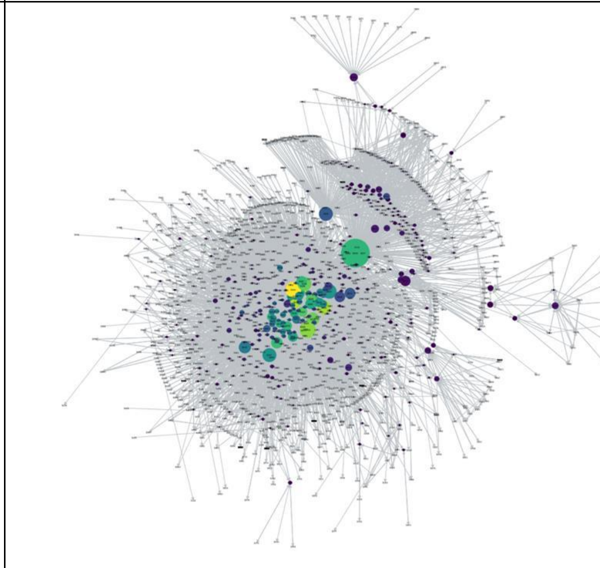
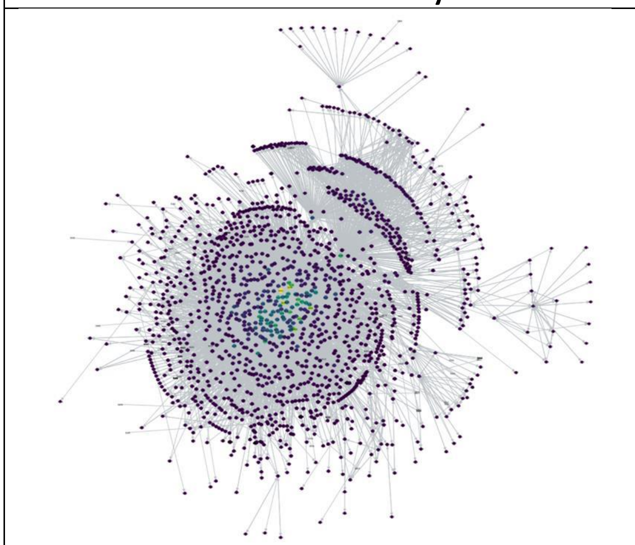
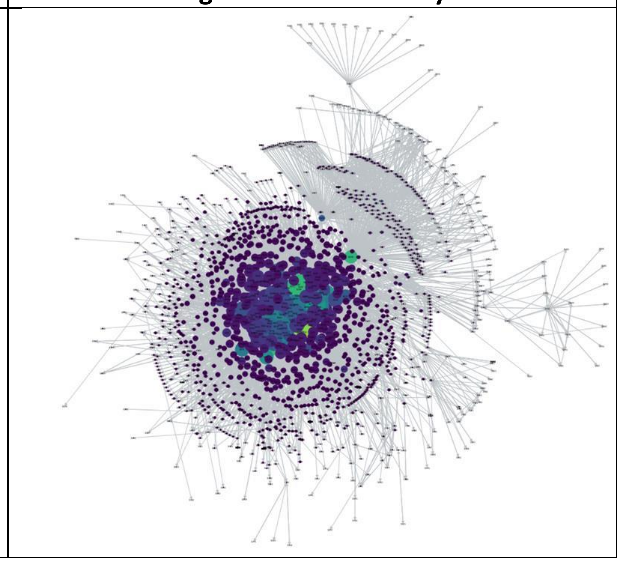
All in all, ORD as a single entity seems to be heavily influential across the network throughout the year 2019. It also possess the geographical advantage of being more centered and has the potential of acting as a point connecting the east coast and west coast.

5.1 Visualization for each quarter.

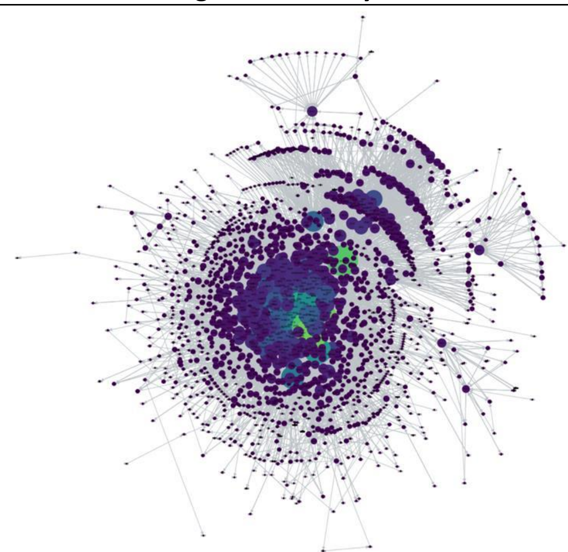
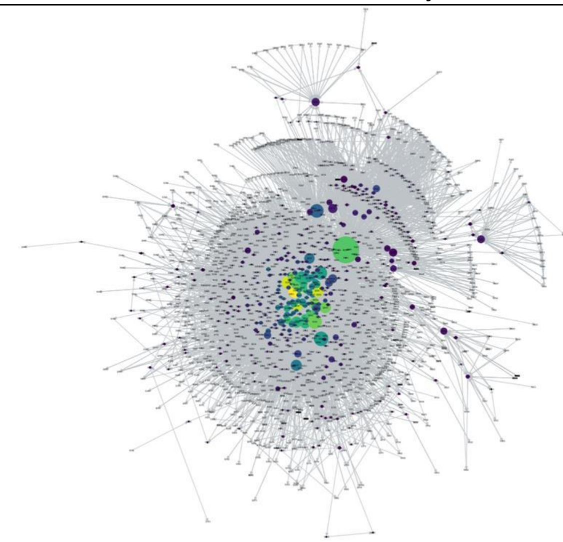
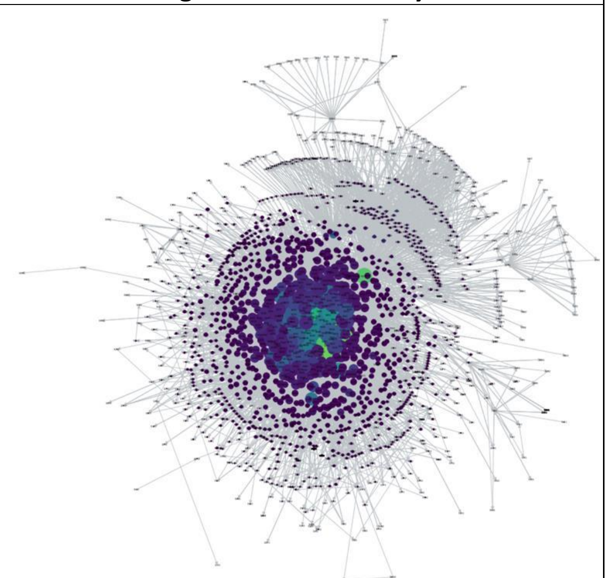
Quarter-1

Degree Centrality	Betweenness Centrality
	
Closeness Centrality	**Eigenvector Centrality**
	

Quarter-2

Degree Centrality	Betweenness Centrality
	
Closeness Centrality	**Eigenvector Centrality**
	

Quarter-3

Degree Centrality	Betweenness Centrality
	
Closeness Centrality	**Eigenvector Centrality**
	

5.2 Community Identification

By employing the Clauset, Nweman and Moore (CNM) algorithm with greedy modularity, we will identify the communities and then discuss about their robustness using the Modularity as measurement. We will then move on to the Girvan-Newman Algorithm for community detection.

5.2.1 Greedy Modularity Algorithm

Communities identified within the network can be evaluated based on the **Modularity** Measurement which is denoted by Q. It speaks about how well the clusters are obtained from the overall network. Ideally, nodes within the community should have more connections within it rather than having more connections outside of the community. Larger the value of Q, stronger is the community.

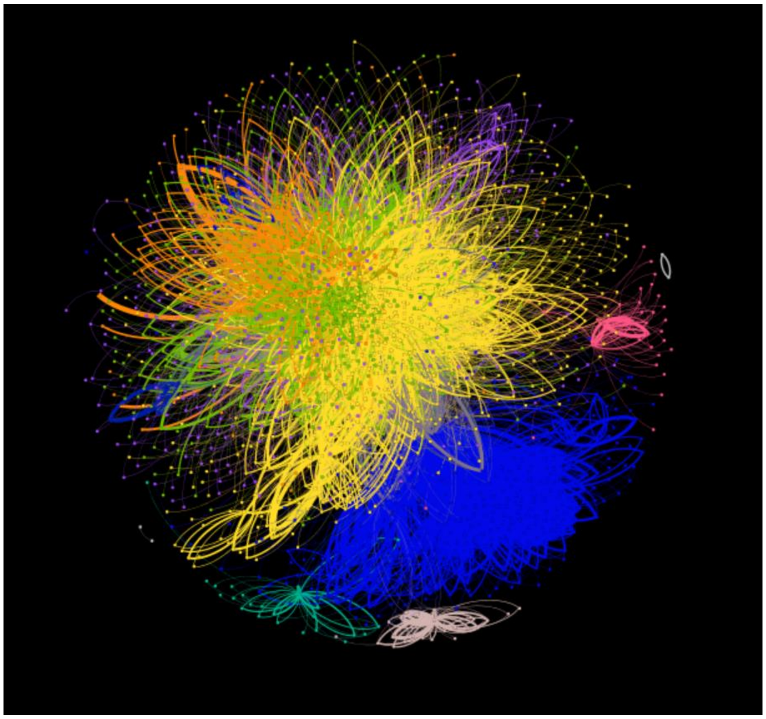
$$Q = \sum_{c_i \in C} \left[\frac{|E_{c_i}^{in}|}{|E|} - \left(\frac{2\,|E_{c_i}^{in}| + |E_{c_i}^{out}|}{2\,|E|} \right)^2 \right]$$

Where C is the set of all communities within the network iterated one by one by assigning them to Ci.

First part of the above equation deals with the ratio of connections formed by the nodes within the community to the overall connections in the network. Second part of the equation gives total outbound edges/connections from the nodes within the community. This modularity evaluates the strength of the communities.

Greedy Modularity algorithm initially assumes all the nodes within the network as individual communities and then calculates the modularity for all possible combinations. Among these, the one with highest modularity will be retained and the same process will be iterated throughout the network until all the identified communities have relatively high connections within the individual communities than the connections formed by the nodes within the community with nodes outside of that particular community.

Visualization of CNM Greedy Modularity Algorith,



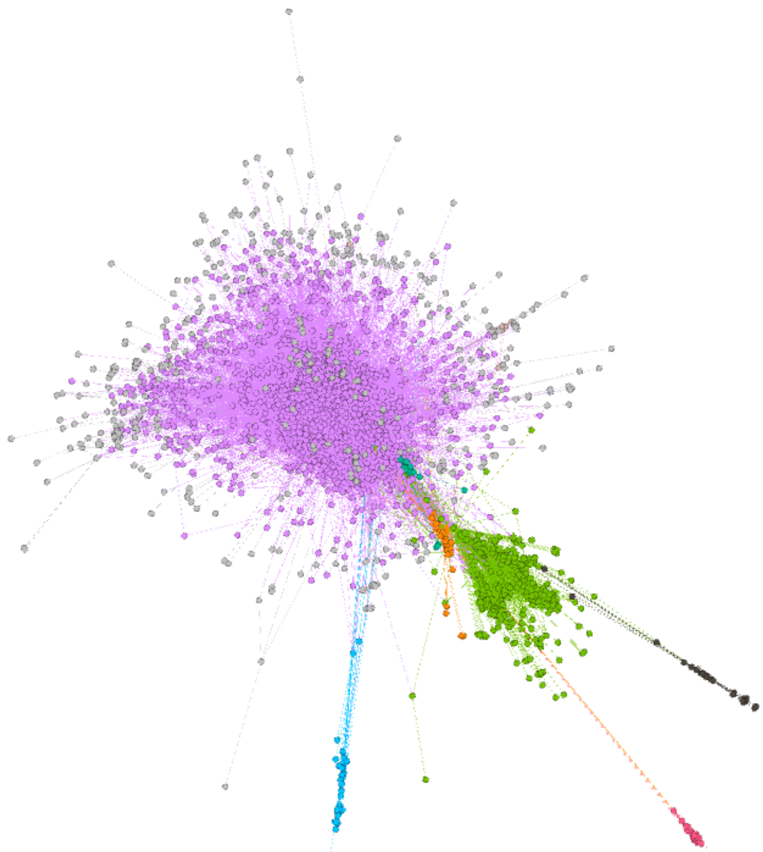
<u>5.2.2</u> <u>Girvan-Newman Algorithm</u>

Betweenness Centrality is a valued entity in this algorithm. At every iteration, this algorithm inspects the betweenness centrality of the nodes and identifies the one with highest betweenness centrality value which then removes it from the original graph.

After following similar approach for many iterations until the graph gets break down into cohesive subgroups or communities by using all of its edges from original graph. Modularity measure can then be employed to understand the robustness of the identified communities. [7]

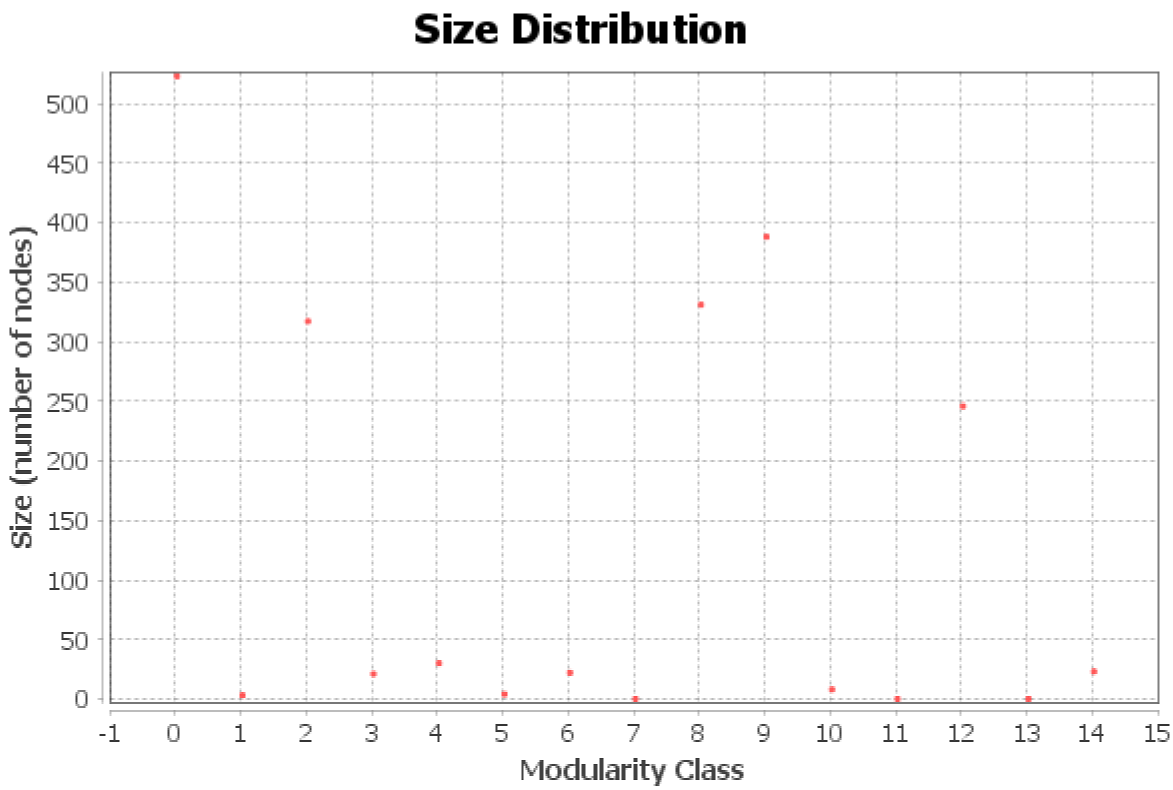
By employing this algorithm, 408 different communities were identified. Beyond this, further inclusion of groups into one to form a community resulted in less cohesiveness of the community. One major impediment in following this approach is with time parameter. Processing time for this algorithm is 15910.776 seconds which is a tremendous increase from 50.843 identified from Greedy Modularity Algorithm. Time consumed is 50.843 seconds to identify the communities which is relatively very less compared to Girvan-Newman Community Detection Algorithm.

Visualization of Girvan-Newman Algorithm



6 Results and Statistics

A total of 16 communities have been detected using the greedy modularity algorithm. Below is the plot displaying the number of actors in each community. Identified communities have a modularity value of 0.319 which is greater than 0.17908931 that we have obtained from Girvan-Newman Algorithm.



Below is the plot that shows the number of edges deleted in order to identify the communities with a good modularity value by following the Girvan-Newman algorithm. Modularity obtained is 0.179. Number of communities identified are 408.

6.1 Communities identified through Greedy Modularity algorithm.

Community-1

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
JFK (0.55766)	JFK (0.49365)	JFK (0.18178)	MIA (0.24678)
DME (0.00146)	NQX (0.0)	NQX (0.0)	NQX (0.0)

Community-2

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
DFW (0.69955)	DFW (0.54834)	LAX (0.07532)	DFW (0.13820)
ELP (0.00149)	AEY (0.0)	BRD (0.0)	AEY (0.0)

Community-3

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
ANC (1.01416)	ANC (0.57688)	ANC (0.50117)	ANC (0.30544)
WUX (0.00283)	WUX (0.0)	WUX (0.0)	WUX (0.0)

Community-4

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
ORD (0.55797)	ORD (0.36263)	ORD (0.31413)	ORD (0.44947)
SHV (0.00725)	OLS (0.0)	IAG (0.0)	OLS (0.0)

Community-5

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
SFO (1.34375)	SFO (0.81166)	SFO (0.34610)	SFO (0.53930)
YR1 (0.03125)	YPT (0.0)	BRD (0.0)	YPT (0.0)

Community-6

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
ATL (2.0)	ATL (0.95455)	ATL (0.65664)	ATL (0.54762)
SKL (0.04545)	LIJ (0.36446)	KWF (0.0)	LIJ (0.03172)

Community-7

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
BLD (1.90)	BLD (0.90909)	BLD (0.48241)	BLD (0.50263)
PGS (0.20)	AZ7 (0.0)	AZ1 (0.0)	AZ7 (0.0)

Community-8

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
SMO (1.0)	SMO (0.66667)	SMO (0.33333)	SMO (0.70754)
HAF (0.16667)	HAF (0.0)	HAF (0.0)	HAF (0.0)

Community-9

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
LAX (2.0)	LAX (1.0)	LAX (1.0)	LAX (0.70711)
APW (0.40)	APW (0.55556)	APW (0.0)	APW (0.31623)

Community-10

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
DEN (1.0)	PER (0.44444)	DEN (0.33333)	PER (0.70710)
PER (0.33333)	SC1 (0.0)	PER (0.0)	SC1 (0.0)

Community-11

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
LAL (1.50)	LAL (0.66667)	GYR (0.50)	LAL (1.0)
PGO (0.50)	PGO (0.0)	LAL (0.0)	PGO (0.0)

Community-12

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
SEA (1.0)	SEA (1.0)	SEA (0.0)	IPT (1.0)
YXL (0.5)	YXL (0.0)	YXL (0.0)	YXL (0.00086)

Community-13

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
MIA (1.0)	ASB (1.0)	MIA (0.0)	ASB (1.0)
COF (1.0)	COF (0.0)	COF (0.0)	COF (0.00141)

Community-14

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
EWR (1.0)	FTG (1.0)	EWR (0.0)	FTG (1.0)
WY1 (1.0)	WY1 (0.0)	WY1 (0.0)	WY1 (0.00141)

Community-15

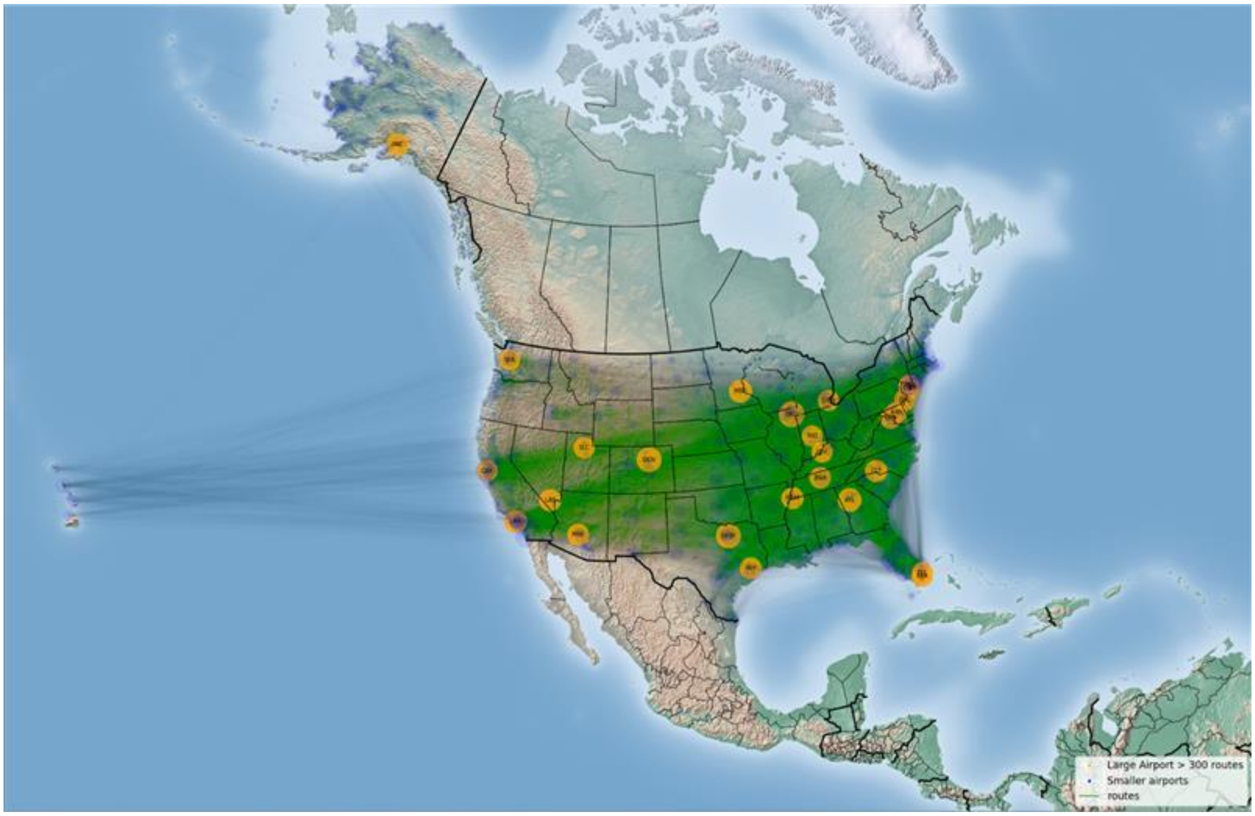
Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
HNL (2.0)	HNL (1.0)	SSB (0.0)	HNL (0.70711)
SSB (2.0)	SSB (1.0)	SSB (0.0)	SSB (0.70711)

Community-16

Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
PAE (2.0)	PAE (1.0)	TIK (0.0)	PAE (0.70711)
TIK (2.0)	TIK (1.0)	TIK (0.0)	TIK (0.70711)

Table 2 – Influential Airports in each Community

Visualization of influential airports



- Busiest Airport – *O'Hare International Airport*
- Least busy Airport – *Anniston Regional Airport*
- Centrality - Influential Airports – *Provided in Table 3 below*
- Group-based Community Detection – *Using CNM Greedy Modularity and Girvan-Newman*
- Robust Communities – *Provided in Table 2 above*

7 Discussion

Considering the ongoing COVID situation, operations in most of the airports have been shut down and will soon be reinstated as the situation betters. This reinstation of airports needs to be done effectively such that more connectivity is established even with less number of airports.

Our analysis helps identify these robust networks that delivers best possible connectivity.

Initially, influential airports were identified based on their centrality measures in the whole network. Then we have segregated the data into 3 different quarters and analysis has been carried out to identify the important airports.

Furthermore, we have applied both the algorithms of CNM Greedy Modularity and Girvan-Newman on the network to provide communities which were then analyzed to observe influential airports in each of the communities.

To achieve the goal of operating low number of airports but yet to attain maximum reach over the network, we need influential airports to be operational without any hurdles post COVID. So necessary measures can be implemented by the government authorities to revive the network starting with these influential airports.

It is also to be noted that the modularity of communities identified by Greedy Modularity Algorithm is very high when compared to the modularity of clusters identified through Girvan-Newman Algorithm.

Theoretically, Girvan-Newman approach is best suited for the network with less nodes and connections.

As ours being a moderately large network with 1946 nodes and over 264000 connections, it proved to be inefficient.

Coming back to CNM Greedy Modularity method, obtained clusters or communities are with a high modularity value suggesting their strength. Hence, we've chosen the influential airports from these communities as our final list of airports required to be operational in each community.

Community No.	Airport ID	Influential Airport Name
1	JFK	John F. Kennedy International Airport
2	DFW	Dallas Fort Worth International Airport
3	ANC	Ted Stevens Anchorage International Airport
4	ORD	O'Hare International Airport
5	SFO	San Francisco International Airport
6	ATL	Hartsfield-Jackson Atlanta International Airport
7	BLD	Bradley International Airport
8	SMO	Santa Monica Airport
9	LAX	Los Angeles International Airport
10	DEN	Denver International Airport
11	LAL	Lakeland Linder International Airport
12	SEA	Seattle-Tacoma International Airport
13	MIA	Miami International Airport
14	EWR	Newark Liberty International Airport
15	HNL	Daniel K. Inouye International Airport
16	PAE	Paine Field Airport

Table 3 – Influential Airports

References

[1] https://www.hindawi.com/journals/mpe/2014/502809/

[2] https://www.sciencedirect.com/science/article/pii/S1474667016410827

[3] http://www.atmseminar.org/seminarContent/seminar9/papers/143-Kotegawa-Final-Paper-4-18-11.pdf

[4] https://iopscience.iop.org/article/10.1088/1674-1056/21/8/088901/meta

[5] https://www.sciencedirect.com/science/article/abs/pii/0191261592900404

[6] https://www.sciencedirect.com/science/article/pii/S1366554509001094

[7] https://networkx.github.io/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.centrality.girvan_newman.html

[8] https://towardsdatascience.com/catching-that-flight-visualizing-social-network-with-networkx-and-basemap-ce4a0d2eaea6