CSE3021 – Social and Information Networks

AIRLINE ROUTES AND AIRPORT ANALYSIS

PROJECT REPORT

Submitted By:

19BCE0421	TANMAY BANSAL
19BCE0436	ATUL AGARWAL
19BCE0866	AADlTYA PAREEK

Slot: B2+TB2

Name of faculty: PROF. ARUN KUMAR S

SCHOOL OF COMPUTER SClENCE AND ENGlNEERlNG

CERTIFICATE

This is to certify that the project work entitled “AIRLINE ROUTES AND AIRPORT ANALYSIS” that is being submitted by Tanmay Bansal ,Atul Agarwal and Aaditya Pareek for Social and Information Networks (CSE3021) is a record of bonafide work done under my supervision. The contents of this Project work in full have neither been taken from any other source nor have been submitted for any other CAL course.

Place: Vellore

Signature of Students:

1. Tanmay Bansal
2. Atul Agarwal
3. Aaditya Pareek

Signature of faculty:

ACKNOWLEDGEMENT

We would like to express our gratitude to all those who have helped us in the successful completion of this project. Without their support, we would not have been able to achieve the goal of the project successfully. We would like to take this opportunity to thank our guide, Arun Kumar S and Suraj Shankarlal Meghwani, for their constant support, guidance and mentorship without which it would have certainly been difficult to complete the project on time. We would like to thank our Dean, who provided us with the facilities required and conductive conditions for the project. Finally, we would like to express our sincere gratitude to VIT University, which provided us with a platform to hone our skills.

ABSTRACT

We intend to use the freely available dataset from openflights.org to visualize and analyse airplane routes throughout the world. This will help us implement social network analysis concepts to determine the most influential and busy airports in the world. These conclusions can be made through the airport's degree centrality, betweenness centrality, eigenvalue and closeness centrality. By analysing the countries' clustering coefficient and average shortest path analysis conclusions can be made about the safety of the airline network in the country.

We plan to use **NetworkX** and **Python** to implement our project.
NetworkX is a python library that can be used to create, edit and study complex networks. It will provide us with the ability to convert our datasets into graphs and to make centrality measure calculations on it.
Python is the core programming language whose libraries will be critical in reading and visualizing our data and network.

Upon creation of the network we will use inbuilt functions of NetworkX which enable us to make centrality measure calculations on them.

AIM

The aim is to calculate all centrality measures of the graph we generate and to compare these results with the population density of the corresponding countries and to find if there exists an inverse or direct relation with the importance of the airport and the population density of the country.

OBJECTIVES

The main objectives of this project were to:

1. implement Social and Information Network Concepts to the airline routes dataset.
2. calculate the most important and busy airports in the world
3. to find the next potential and upcoming important airports
4. to compare our results with the population densities and find the corresponding conclusions

LITERATURE SURVEY

<table>
<tr><td colspan="2">Analysis and visualization of airport network to strengthen the economy</td></tr>
<tr><td colspan="2">Overview:
The world’s eminent airports are directly or indirectly connected to many other airports. Every airport is considered as a node and the route can be considered as an edge connecting them. The work analyzes the USA airport network using different centrality measures of social network analysis. The centrality measures calculated on airport network help in identification of certain characteristics of the airports. Some of the characteristics are like the busiest airport and the airports which influence trade, alternate path, fastest route, nearest airports, etc.
The characteristics help to find the designated airports meant for improving the economy. The results of this paper say about the prominent communication and connections among the airports in the U.S.A.</td></tr>
<tr><td>Conclusion:
The project dealt completely with an American Airspace Dataset. It used UCINET to compute its centrality measure results and concluded that the Chicago O’hare Airport was the most vital airport in the USA. It has the highest degree, Betweenness, closeness, Eigenvector, reach and information centrality which shows that Chicago is one of the designated cities in the U.S.A and the airport Chicago</td><td>Authors:
Saleena, P.
P. Swetha
D. Radha</td></tr>
</table>

O'hare International is the most influential. Analysis of such datasets on different centrality measures helps in identifying the importance of the airports in various aspects.

Tools Used:
UCINET 6, NetDraw

Saleena, P., P. Swetha, and D. Radha. "Analysis and visualization of airport network to strengthen the economy." *International Journal of Engineering & Technology* (2018): 708-713.

A Network Science-Based Performance Improvement Model for the Airline Industry Using NetworkX

Overview:
In this paper, the connectivity and profits of an airline company were analyzed with the theoretical approach and by proposing a model to increase the performance of the above parameters. In this paper, multi-airlines have been considered.

Conclusion:
The models for Indigo, SpiceJet and Air Asia were considered and analysed. In this research, complex multigraph airlines were analyzed by using the graph analytics technique for the optimum solution. Standard parameters like edges, nodes, degree, clustering, and shortest path on indigo, spice jet, and AirAsia airline systems were also compared.

Authors:
Kollu Vishnu
V. R. Amiripalli
Shanmuk S

Tools Used:
Python, NetworkX

Kollu, Vishnu VR, et al. "A Network Science-Based Performance Improvement Model for the Airline Industry Using NetworkX." *International Journal of Sensors Wireless Communications and Control* 11.7 (2021): 768-773.

<table>
<tr><td colspan="3">EXPLORATION OF GLOBAL AIR TRANSPORT NETWORK USING SOCIAL NETWORK ANALYSIS</td></tr>
<tr><td colspan="3">Overview:
By calculating the clustering coefficient and the average shortest path, the team aimed to confirm that the world airport network has the characteristics of a small-world network. The objective of this project was to better understand the characteristics and patterns of air transport around the world. The team used various measures of social network analysis to arrive at our output, which included a comparison of regional airport networks, their importance in the network, and influence airports have on the world airport network.</td></tr>
<tr><td colspan="2">Conclusion:
The team drew up results using multiple Social Network Analysis measures on the dataset. They provided visual representation of how the results of these measures can be used to derive inferences regarding importance of global airports in different aspects.</td><td>Authors:
Prabhakar
Nikhilesh
L. Jani Anbarasi</td></tr>
<tr><td>Tools Used:
Python, NetworkX</td><td colspan="2">Prabhakar, Nikhilesh, and L. Jani Anbarasi. "Exploration of the global air transport network using social network analysis." Social Network Analysis and Mining 11.1 (2021): 1-12.</td></tr>
</table>

<table>
<tr><td colspan="3">A complex network approach towards modeling and analysis of the Australian Airport Network</td></tr>
<tr><td colspan="3">Overview:
In this paper, the team modeled Australia's civil domestic airport infrastructure as a network and analyzed the resulting network structure and its features using network tools. This case study identifies and investigates network measures, such as the degree distribution, characteristics path length, clustering coefficient and centrality measure as well as the correlations among them to understand the topology of an airport network.</td></tr>
<tr><td colspan="2">Conclusion:
The paper concluded that the characteristics of the Australian Airport Network was similar to that of a small world network wherein the average shortest path between two nodes increased rather gradually with increase in size of the network. It also concluded that the most connected airports were not the most central nodes as found in the case of Sydney and Brisbane.</td><td>Authors:
Hossain,
Md Murad,
Sameer Alam</td></tr>
<tr><td>Tools Used:
Not specified</td><td colspan="2">Hossain, Md Murad, and Sameer Alam. "A complex network approach towards modeling and analysis of the Australian Airport Network." Journal of Air Transport Management 60 (2017): 1-9.</td></tr>
</table>

Analysis of Airport Network in Pakistan Utilizing Complex Network Approach

Overview:
The research team gathered data from Civil Aviation Authority of Pakistan (CAA) and conducted complex network analysis on the network constructed. The data gave them a distribution that represented the routes of airlines over the period of seven days. They aimed to compare these results with results from other countries which shared similar topologies. It also aimed to find the categories of networks that the Airport Network of Pakistan would fit in.

Conclusion:
The paper concluded that Pakistan’s network was a small world network which sufficed a power law. It found that three airports held controlled over 80% of Pakistan’s Airport Network.

Authors:
Malik,
Hafiz Abid Mahmood,
Nadeem Mahmood,
Mir Hammal Usman,
Kashif Rziwan,
Faiza Abid.

Tools Used:
Not specified

Malik, Hafiz Abid Mahmood, Nadeem Mahmood, Mir Hammal Usman, Kashif Rziwan, and Faiza Abid. "Analysis of airport network in Pakistan utilizing complex network approach." *network* 10, no. 1 (2019).

Identifying the influential nodes via eigen-centrality from the differences and similarities of structure

Overview:
One of the most important problems in complex network is the identification of the influential nodes. For this purpose, the use of differences and similarities of structure to enrich the centrality method in complex networks is proposed. The centrality method called ECDS centrality used is the eigen-centrality which is based on the Jaccard similarities between the two random nodes.

Conclusion:
Traditional eigenvector centrality only considers the number of the neighbors of the node, but the new eigenvector centrality (ECDS) determined by the differences and similarities can identify the influential nodes more accurately than the traditional eigenvector centrality. In this paper, by taking into account the differences and

Authors:
Lin-Feng Zhong
Ming-Sheng Shang
Xiao-Long Chen

<table>
<tr><td colspan="2">similarities of structure between the two random nodes, we present a new eigen-centrality (ECDS) to identify the node spreading influence. The ECDS centrality considers the differences and similarities of structure which are determined by the Jaccard similarity of the random two nodes.</td><td></td></tr>
<tr><td>Tools Used:
Theoretical Analysis</td><td colspan="2">Zhong, Lin-Feng, et al. "Identifying the influential nodes via eigen-centrality from the differences and similarities of structure." Physica A: Statistical Mechanics and its Applications 510.C (2018): 77-82.</td></tr>
</table>

<table>
<tr><td colspan="3">Analysis of the airport network of India as a complex weighted network</td></tr>
<tr><td colspan="3">Overview:
This paper talks about Airport Network of India (ANI) which represents India’s domestic civil aviation infrastructure as a complex network. It finds that ANI, a network of domestic airports connected by air links, is a small-world network characterized by a truncated power-law degree distribution and has a signature of hierarchy. It investigates ANI as a weighted network to explore its various properties and compare them with their topological counterparts.</td></tr>
<tr><td colspan="2">Conclusion:
It shows that ANI, despite being small in size, has complex dynamics similar to those of bigger air transportation networks.ANI, whose topology has a signature of hierarchy, has small-world network features and is characterized by a truncated scale-free degree distribution. Analysis of weighted ANI reveals clearer picture of the network dynamics. ANI is expected to grow at a rapid speed with addition of airports, several low-cost air services [16], and importantly by increase in strength and complexity of interactions.</td><td>Authors:
Ganesh Bagler</td></tr>
<tr><td>Tools Used:
unknown</td><td colspan="2">Bagler, Ganesh. "Analysis of the airport network of India as a complex weighted network." Physica A: Statistical Mechanics and its Applications 387.12 (2018): 2972-2980.</td></tr>
</table>

Exploring the network structure and nodal centrality of China’s air transport network: A complex network approach	
Overview: This paper uses a complex network approach to examine the network structure and nodal centrality of individual cities in the air transport network of China (ATNC). Measures for overall network structure include degree distribution, average path length and clustering coefficient.Centrality metrics for individual cities are degree, closeness and betweenness, representing a node’s location advantage as being directly connected to others, being accessible to others, and being the intermediary between others,	
Conclusion: This paper has used complex network theory to examine the overall structure of China’s air transport network and the centrality of individual cities. The air transport network of China (ATNC) has small-world characteristics, but is not a scale-free network. Like the US, the ATNC is largely disassortative. In conclusion, the research shows that the rapid development of the air transport network in China has produced a distinctive pattern. In recent years, new and small airports in China are inclined to supply direct links to the top hubs and so bypass the regional ones, resulting in underdeveloped regional centers.	Authors: Jiaoe Wang Huihui Mo Fahui Wang Fengjun Jin
Tools Used: unknown	Wang, Jiaoe, et al. "Exploring the network structure and nodal centrality of China’s air transport network: A complex network approach." *Journal of Transport Geography* 19.4 (2011): 712-721.

<table>
<tr><td colspan="3">Analysis of the Air Transport Network Characteristics of Major Airports</td></tr>
<tr><td colspan="3">Overview:
The world’s major airports are directly connected to hundreds of airports without intermediate routes. This connectivity can be described as the network in which the airport becomes a node and the route becomes a connection line. This paper analyzes the air transport network of 1,060 airports using the social network analysis (SNA) methodology. This paper aids in the understanding of air transport networks and logistics connectivity in inter-city and inter-country transport</td></tr>
<tr><td colspan="2">Conclusion: This paper identified the airports with high centrality in the network. They integrated the result from the airport analysis into the national level and identified the characteristics of major countries. Next they compared the individual networks of the United States and China, which have the largest portion in the total network. Lastly, the differences between degree and betweenness centrality in the network were elucidated.</td><td>Authors:
Min Geun SONG
Gi Tae YEO</td></tr>
<tr><td>Tools Used:
unknown</td><td colspan="2">Song, Min Geun, and Gi Tae Yeo. "Analysis of the air transport network characteristics of major airports." The Asian Journal of Shipping and Logistics 33.3 (2017): 117-125.</td></tr>
</table>

Empirical analysis of airport network and critical airports - China

Overview:

In this paper, the authors present methods to investigate network properties and to identify critical airports in the network. A novel network model is proposed with airports as nodes and the correlations between traffic flow of airports as edges. Spectral clustering algorithms are developed to classify airports. Spatial distribution characteristics and intraclass correlation of different categories of airports are carefully analyzed. The analyses based on the fluctuation trend of distance-correlation and power spectrum of time series are performed to examine the self-organized criticality of the network. The results indicate that there is one category of airports which dominates the self-organized critical state of the network. Six airports in this category are found to be the most important ones in the Chinese air transport network. The flight delay occurring in these six airports can propagate to the other airports, having a huge impact on the operation characteristics of the entire network.

Conclusion:

A novel aviation network is constructed with airports as nodes and the correlations between traffic flow of airport pairs as edges. All airports are classified into 6 categories by the proposed spectral clustering algorithm. Among the 6 categories, airports in C6 are of the highest intraclass correlation, and they have quite similar features in terms of traffic characteristics and geographical locations

Authors:

Cong Wei,Hu Minghua, Dong Bin, Wang Yanjun,Feng Cheng

Tools Used:
Not specified

Cong, Wei; Hu, Minghua; Dong, Bin; Wang, Yanjun; Feng, Cheng (2016). *Empirical analysis of airport network and critical airports. Chinese Journal of Aeronautics, 29(2), 512–519*

METHODOLOGY

First we import the datasets from openflights.org and process it. The initial dataset is too complex and gives inaccurate results for centrality measures calculations. To make the results more accurate we refine our data.

The route dataset was refined by creating a new column called count. Here we maintained the count of how many times a particular directed route appeared in our data and dropped the duplicates.

Next we created a directed graph using NetworkX where the size of each node was directly proportional to how many times the airport appeared in the dataset.

Next we used NetworkX tools to calculate various centrality measures on our graph and generated a list of all top 10 measures such as most influential, most busy and most important route. Finally we compare all the lists with the population density of the corresponding countries and draw our conclusions.

NOVELTY

- In all the research papers we reviewed the teams either used a small dataset restricted to countries or even when they used a global dataset they stopped after computing the centrality measures.
- We have added the calculation of the next potential hubs and comparisons with the population densities of the countries.

CODE

```
!apt-get install libgeos-3.5.0
!apt-get install libgeos-dev
!pip install https://github.com/matplotlib/basemap/archive/master.zip

import numpy as np
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
import statistics as stat
from mpl_toolkits.basemap import Basemap

airports = pd.read_csv("sample_data/airports-extended.csv", names=["Airport
ID", "Airport Name", "City", "Country", "IATA", "ICAO", "Latitude",
"Longitude", "Altitude", "Timezone", "DST", "Tz database", "Type",
"Source"])
airports.head()

routes = pd.read_csv("sample_data/routes.csv")
routes.head()

population = pd.read_csv("sample_data/population_by_country_2020 (1).csv")
population.head()
routes.drop_duplicates(subset=["Source airport", "Destination airport"]
,keep = False, inplace = True)
routes_graph = nx.from_pandas_edgelist(routes, source="Source airport",
target="Destination airport"



#Finding out the basic characteristics of the Graph.
g_nodes = nx.number_of_nodes(routes_graph)
g_edges = nx.number_of_edges(routes_graph)
g_density = nx.density(routes_graph)
g_number_of_connected_components =
nx.number_connected_components(routes_graph)
g_connected_components = nx.connected_components(routes_graph)
degrees = [v for k, v in routes_graph.degree()]
minimum = min(degrees)
maximum = max(degrees)
avg = stat.mean(degrees)
stdev = stat.stdev(degrees)
#Printing the basics of graph
print("Number of nodes : ",g_nodes)
```

```
print("Number of edges : ",g_edges)
print("Density : ",g_density)
print("Number of Connected components : ",g_number_of_connected_components)
print("Minimum Degree: ",minimum)
print("Maximum Degree: ",maximum)
print("Average Degree: ",avg)
print("Standard Deviation: ",stdev)
routes = pd.read_csv("sample_data/routes.csv")
routes['Source airport ID'] = pd.to_numeric(routes['Source airport
ID'].astype(str), 'coerce')
routes['Destination airport ID'] = pd.to_numeric(routes['Destination airport
ID'].astype(str), 'coerce')
routes = routes.dropna(subset=["Source airport ID", "Destination airport
ID"])
airport_new = airports[['Airport Name','Latitude', 'Longitude', 'IATA',
'ICAO']]
airport_indexs = airport_new.index.values
routes_new = routes[(routes['Source airport ID'].isin(airport_indexs)) &
                         (routes['Destination airport
ID'].isin(airport_indexs))]
routes_new =  pd.DataFrame(routes_new.groupby(['Source airport',
'Destination airport']).size().reset_index(name='counts'))
routes_counts = routes_new['Source
airport'].append(routes_new.loc[routes_new['Source airport'] !=
routes_new['Destination airport'], 'Destination airport']).value_counts()
routes_counts = pd.DataFrame({'IATA': routes_counts.index, 'total_flight':
routes_counts})
pos_data = routes_counts.merge(airport_new, on = 'IATA')
graph_new = nx.from_pandas_edgelist(routes_new, source = 'Source airport',
target = 'Destination airport', edge_attr = 'counts',create_using =
nx.DiGraph())

fig = plt.figure(figsize=(80, 60), edgecolor='w')
m = Basemap(projection='cea',llcrnrlat=-90,urcrnrlat=90,\
            llcrnrlon=-180,urcrnrlon=180,resolution='c',suppress_ticks =
True)
mx, my = m(pos_data['Longitude'].values, pos_data['Latitude'].values)
pos = {}
m.drawcoastlines()
m.fillcontinents(color='white')
# draw parallels and meridians.
m.drawparallels(np.arange(-90.,91.,30.))
m.drawmeridians(np.arange(-180.,181.,60.))
```

```
m.drawmapboundary(fill_color='lightgrey')
for count, elem in enumerate (pos_data['IATA']):
  pos[elem] = (mx[count], my[count])
nx.draw_networkx_nodes(G = graph_new, pos = pos, nodelist =
graph_new.nodes(), node_color = 'r', alpha = 0.8, node_size =
[routes_counts['total_flight'][s]*3 for s in graph_new.nodes()])

fig = plt.figure(figsize=(80, 60), edgecolor='w')
nx.draw_networkx_nodes(G = graph_new, pos = pos, nodelist =
graph_new.nodes(), node_color = 'r', alpha = 0.8, node_size =
[routes_counts['total_flight'][s]*3 for s in graph_new.nodes()])
nx.draw_networkx_edges(G = graph_new, pos = pos, edge_color='g', width =
routes_new['counts']*0.75, alpha=0.2, arrows = False)
m = Basemap(projection='cea',llcrnrlat=-90,urcrnrlat=90,\
            llcrnrlon=-180,urcrnrlon=180,resolution='c',suppress_ticks =
True)
m.drawcoastlines()
m.fillcontinents(color='white')
# draw parallels and meridians.
m.drawparallels(np.arange(-90.,91.,30.))
m.drawmeridians(np.arange(-180.,181.,60.))
m.drawmapboundary(fill_color='lightgrey')
plt.show()

hubs_dict = nx.degree_centrality(graph_new)
top10hubs = {k: hubs_dict[k] for k in list(hubs_dict)[:10]}
print("\nThe top 10 hubs are: ")
k=0
data = []
for w in top10hubs:
  country_name = airports.loc[airports["IATA"] == w]["Country"].values[0]
  data.append([w, airports.loc[airports['IATA'] == w]['Airport
Name'].values[0], country_name, population.loc[population['Country (or
dependency)'] == country_name]['Density (P/Km²)'].values[0],
population.loc[population['Country (or dependency)'] ==
country_name]['Population (2020)'].values[0]])
  k+=1
df = pd.DataFrame(data, columns = ['IATA', 'Airport Name', 'Country (or
dependency)', 'Density (P/Km²) of Country', 'Population (2020) of Country'])
df


print("\nThe top 10 most influential airports are: ")
```

```
most_influential = nx.degree_centrality(graph_new)
k=0
data = []
for w in sorted(most_influential, key = most_influential.get, reverse =
True):
  if(k<10):
    country_name = airports.loc[airports["IATA"] == w]["Country"].values[0]
    data.append([w, most_influential[w], airports.loc[airports['IATA'] ==
w]['Airport Name'].values[0], country_name,
population.loc[population['Country (or dependency)'] ==
country_name]['Density (P/Km²)'].values[0],
population.loc[population['Country (or dependency)'] ==
country_name]['Population (2020)'].values[0]])
    k+=1
df = pd.DataFrame(data, columns = ['IATA', 'Degree Centrality', 'Airport
Name', 'Country (or dependency)', 'Density (P/Km²) of Country', 'Population
(2020) of Country'])
df


#finding the most busiest hubs
busy = []
most_busiest = nx.eigenvector_centrality(graph_new)
k = 0
data = []
print("\nThe top 10 busiest airports are: ")
for w in sorted(most_busiest, key=most_busiest.get, reverse=True):
  if(k < 10):
    country_name = airports.loc[airports["IATA"] == w]["Country"].values[0]
    data.append([w, most_busiest[w], airports.loc[airports['IATA'] ==
w]['Airport Name'].values[0], country_name,
population.loc[population['Country (or dependency)'] ==
country_name]['Density (P/Km²)'].values[0],
population.loc[population['Country (or dependency)'] ==
country_name]['Population (2020)'].values[0]])
  if(k<30):
    busy.append(w)
    k+=1
df = pd.DataFrame(data, columns = ['IATA', 'Eigenvector Centrality','Airport
Name', 'Country (or dependency)', 'Density (P/Km²) of Country', 'Population
(2020) of Country'])
df
```

```
#Finding the best connecting hubs
best_con = []
best_connector = nx.betweenness_centrality(graph_new)
k=0
data = []
print("\nThe top 10 best connecting airports are: ")
for w in sorted(best_connector, key=best_connector.get, reverse=True):
  if(k<10):
    country_name = airports.loc[airports["IATA"] == w]["Country"].values[0]
    data.append([w, best_connector[w], airports.loc[airports['IATA'] ==
w]['Airport Name'].values[0], country_name,
population.loc[population['Country (or dependency)'] ==
country_name]['Density (P/Km²)'].values[0],
population.loc[population['Country (or dependency)'] ==
country_name]['Population (2020)'].values[0]])
  if(k<30):
    k+=1
    best_con.append(w)
df = pd.DataFrame(data, columns = ['IATA', 'Betweenness Centrality','Airport
Name', 'Country (or dependency)', 'Density (P/Km²) of Country', 'Population
(2020) of Country'])
df


# Printing the next potential hubs for the airlines
# If the airports are not the hubs, but they do exist in most busiest and
most important ports.
print("\nThe next potential hubs could be: ")
k=0
data = []
for w in busy:
  if w in best_con:
    if w not in top10hubs:
      country_name = airports.loc[airports["IATA"] ==
w]["Country"].values[0]
      data.append([w, airports.loc[airports['IATA'] == w]['Airport
Name'].values[0], country_name, population.loc[population['Country (or
dependency)'] == country_name]['Density (P/Km²)'].values[0],
population.loc[population['Country (or dependency)'] ==
country_name]['Population (2020)'].values[0]])
      k+=1
df = pd.DataFrame(data, columns = ['IATA', 'Airport Name', 'Country (or
```

```
dependency)', 'Density (P/Km²) of Country', 'Population (2020) of Country'])
df


print("\nThe top 10 used routes are: ")
most_important_edge = nx.edge_betweenness_centrality(graph_new)
k = 0
data = []
for w in sorted(most_important_edge, key = most_important_edge.get, reverse
= True):
  if(k<10):
    data.append([w[0], airports.loc[airports['IATA'] == w[0]]['Airport
Name'].values[0], w[1], airports.loc[airports['IATA'] == w[1]]['Airport
Name'].values[0], most_important_edge[w]])
    k+=1
df = pd.DataFrame(data, columns = ['IATA Source Airport', 'Source Airport
Name', 'IATA Destination Airport', 'Source Destination Name',  'Edge
Betweenness Centrality'])
df
```

RESULTS

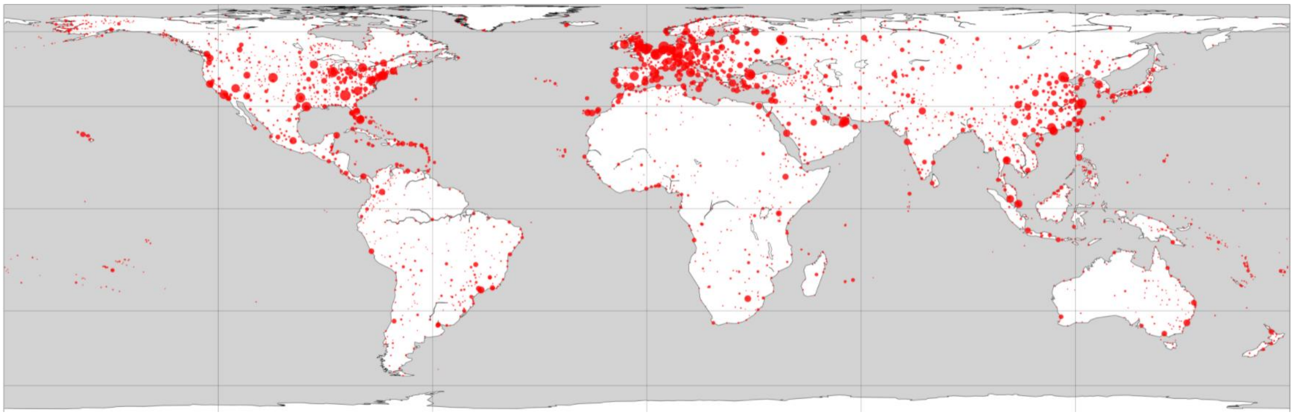

Fig. 1: Airports as nodes of graph

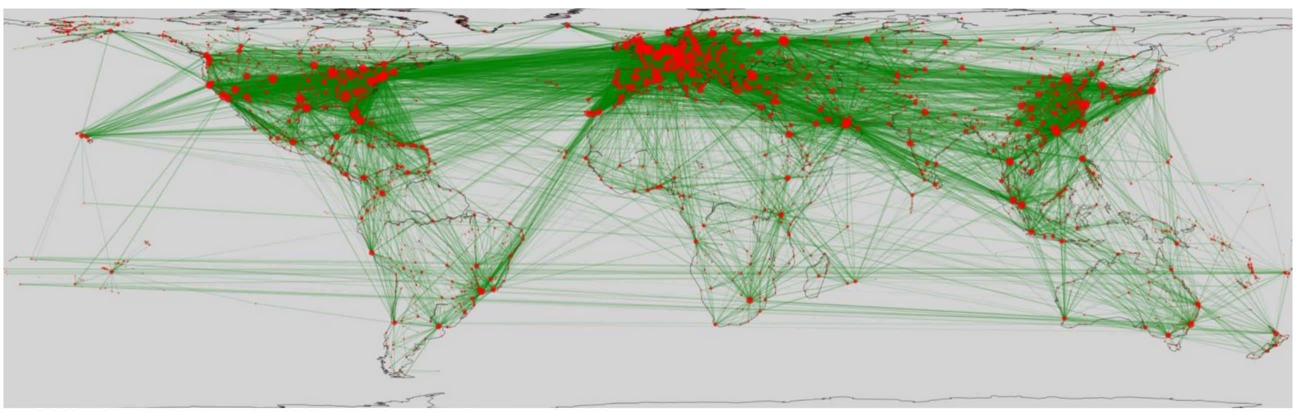

Fig 2: Complete Graph on World Map

The top 10 most influential airports are:

	IATA	Degree Centrality	Airport Name	Country (or dependency)	Density (P/Km²) of Country	Population (2020) of Country
0	FRA	0.143288	Frankfurt am Main International Airport	Germany	240	83830972
1	CDG	0.141176	Charles de Gaulle International Airport	France	119	65298930
2	AMS	0.139065	Amsterdam Airport Schiphol	Netherlands	508	17141544
3	IST	0.136048	Atatürk International Airport	Turkey	110	84495243
4	ATL	0.130618	Hartsfield Jackson Atlanta International Airport	United States	36	331341050
5	PEK	0.123680	Beijing Capital International Airport	China	153	1440297825
6	ORD	0.122775	Chicago O'Hare International Airport	United States	36	331341050
7	MUC	0.114027	Munich International Airport	Germany	240	83830972
8	DME	0.112519	Domodedovo International Airport	Russia	9	145945524
9	DFW	0.112217	Dallas Fort Worth International Airport	United States	36	331341050

Fig 3: 10 Most Influential Airports

The top 10 busiest airports are:

	IATA	Eigenvector Centrality	Airport Name	Country (or dependency)	Density (P/Km²) of Country	Population (2020) of Country
0	AMS	0.166481	Amsterdam Airport Schiphol	Netherlands	508	17141544
1	FRA	0.165915	Frankfurt am Main International Airport	Germany	240	83830972
2	CDG	0.159405	Charles de Gaulle International Airport	France	119	65298930
3	MUC	0.149393	Munich International Airport	Germany	240	83830972
4	LHR	0.136390	London Heathrow Airport	United Kingdom	281	67948282
5	FCO	0.135904	Leonardo da Vinci–Fiumicino Airport	Italy	206	60446035
6	BCN	0.129996	Barcelona International Airport	Spain	94	46757980
7	IST	0.128651	Atatürk International Airport	Turkey	110	84495243
8	ZRH	0.126310	Zürich Airport	Switzerland	219	8665615
9	MAD	0.123602	Adolfo Suárez Madrid–Barajas Airport	Spain	94	46757980

Fig 4: Most Busy Airports

The top 10 best connecting airports are:

	IATA	Betweenness Centrality	Airport Name	Country (or dependency)	Density (P/Km²) of Country	Population (2020) of Country
0	ANC	0.071284	Ted Stevens Anchorage International Airport	United States	36	331341050
1	LAX	0.066839	Los Angeles International Airport	United States	36	331341050
2	DXB	0.062511	Dubai International Airport	United Arab Emirates	118	9910892
3	CDG	0.062332	Charles de Gaulle International Airport	France	119	65298930
4	FRA	0.052634	Frankfurt am Main International Airport	Germany	240	83830972
5	PEK	0.049477	Beijing Capital International Airport	China	153	1440297825
6	ORD	0.047840	Chicago O'Hare International Airport	United States	36	331341050
7	AMS	0.044159	Amsterdam Airport Schiphol	Netherlands	508	17141544
8	YYZ	0.043803	Lester B. Pearson International Airport	Canada	4	37799407
9	SEA	0.043680	Seattle Tacoma International Airport	United States	36	331341050

Fig 5: Most Connected Airports

The top 10 used routes are:

	IATA Source Airport	Source Airport Name	IATA Destination Airport	Source Destination Name	Edge Betweenness Centrality
0	LAX	Los Angeles International Airport	ANC	Ted Stevens Anchorage International Airport	0.010718
1	ANC	Ted Stevens Anchorage International Airport	LAX	Los Angeles International Airport	0.010013
2	BET	Bethel Airport	ANC	Ted Stevens Anchorage International Airport	0.009434
3	ANC	Ted Stevens Anchorage International Airport	BET	Bethel Airport	0.009382
4	ORD	Chicago O'Hare International Airport	ANC	Ted Stevens Anchorage International Airport	0.008970
5	ANC	Ted Stevens Anchorage International Airport	ORD	Chicago O'Hare International Airport	0.008419
6	FAI	Fairbanks International Airport	SEA	Seattle Tacoma International Airport	0.007036
7	SEA	Seattle Tacoma International Airport	FAI	Fairbanks International Airport	0.006965
8	YQT	Thunder Bay Airport	YYZ	Lester B. Pearson International Airport	0.005748
9	YYZ	Lester B. Pearson International Airport	YQT	Thunder Bay Airport	0.005457

Fig 6: Most Used Routes

The next potential hubs could be:

	IATA	Airport Name	Country (or dependency)	Density (P/Km²) of Country	Population (2020) of Country
0	AMS	Amsterdam Airport Schiphol	Netherlands	508	17141544
1	FRA	Frankfurt am Main International Airport	Germany	240	83830972
2	LHR	London Heathrow Airport	United Kingdom	281	67948282
3	MAD	Adolfo Suárez Madrid–Barajas Airport	Spain	94	46757980
4	JFK	John F Kennedy International Airport	United States	36	331341050
5	DXB	Dubai International Airport	United Arab Emirates	118	9910892
6	ORD	Chicago O'Hare International Airport	United States	36	331341050
7	YYZ	Lester B. Pearson International Airport	Canada	4	37799407

Fig 7: Important Hubs

CONCLUSION

The visualization and implementation of the Social and Information Network concepts helped us understand how the global air traffic is dependent on certain central airports. We also concluded that the countries with very high population densities have much less traffic and central/important airports than countries with lower population densities. Through this project we have calculated and listed out the most central, influential and busy airports in the world.