# Topic: Network Analytics

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: Prajay B. Urkude Batch ID: 16092021**

**Topic: Network Analyti****cs**

**Problem Statement: -**

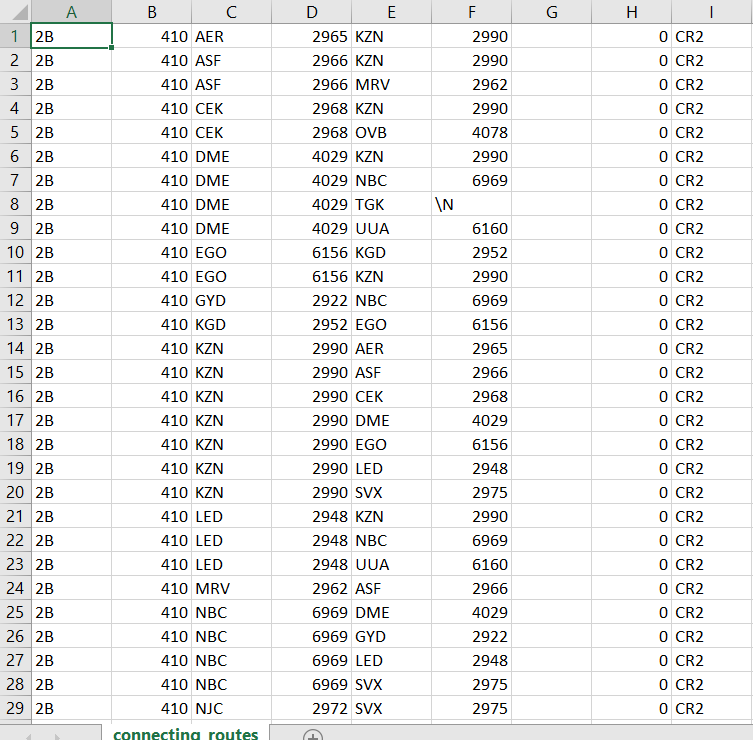
There are two datasets consisting of information for the connecting routes and flight halt. Create network analytics models on both the datasets separately and measure degree centrality, degree of closeness centrality, and degree of in-between centrality.

* Create a network using edge list matrix (directed only**).**
* Columns to be used ***python:***

Flight\_halt=c("ID","Name","City","Country","IATA\_FAA","ICAO","Latitude","Longitude","Altitude","Time","DST","Tz database time")

connecting routes=c ("flights", " ID", "main Airport”, “main Airport ID", "Destination ","Destination ID","haults","machinary")

**connecting routes**



**Ans:**

**Business Objectives:**

To find the busiest airport by calculating measure degree centrality, degree of closeness centrality, and degree of in-between centrality.

**Business Constraints:**

The traffic of airplanes at the particular airport.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name Of Feature** | **Description** | **Type** | **Relevance** |
| Flights | Name of flight | Qualitative, Nominal | Irrelevant for the analysis |
| ID | ID of flight | Quantitative, Nominal | Irrelevant for the analysis |
| Main Airport | Name of the airport | Qualitative, nominal | Relevant for the analysis |
| Main Airport ID | Id of the Airports | Quantitative, Nominal | Irrelevant for the analysis |
| Destination | Destination airport | Qualitative, nominal | Relevant for the analysis |
| Destination Id | Id of the destination airport | Quantitative, Nominal | Irrelevant for the analysis |
| Haults | Hault of the airplane at the airport | Quantitative, Nominal | Irrelevant for the analysis |
| machinery | Machinery in the airplane | Qualitative/ Quantitative, Nominal | Irrelevant for the analysis |

1. Importing libraries like pandas, matplotlib, numpy and networkx.

Networkx is a Python Package for the creation, manipulation and study of the structure, dynamics and functions of complex networks.

2. Loading the data and check for columns, missing values, datatypes.

Here we have taken 1st 1000 rows for the analysis.

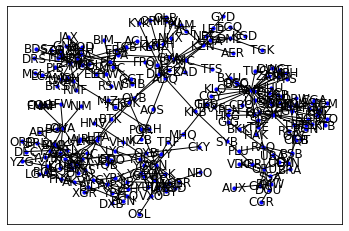
3. Creating the empty graph for no of nodes and no. of edges. graph stores nodes and edges with optimal data or attributes.

4. Takes source and destination nodes from the data frame by using from\_pandas\_edgelist variable.

5. nx.info gives the no. of nodes and edges and average degree from the source and the destinations of the datasets.

6. calculate the degree of centrality and we found that 4I7 airport has the highest degree of centrality.

7. Draw the graph which shows the connections routes of one airport to another.



8. Calculating the closeness centrality and we found that MNL has highest closeness centrality.

9. Calculate the betweenness Centrality and we found that MNL airport has the highest betweenness centrality.

10. Calculate the Eigenvector Centrality and we found that FMM airport has the highest Eigenvector Centrality

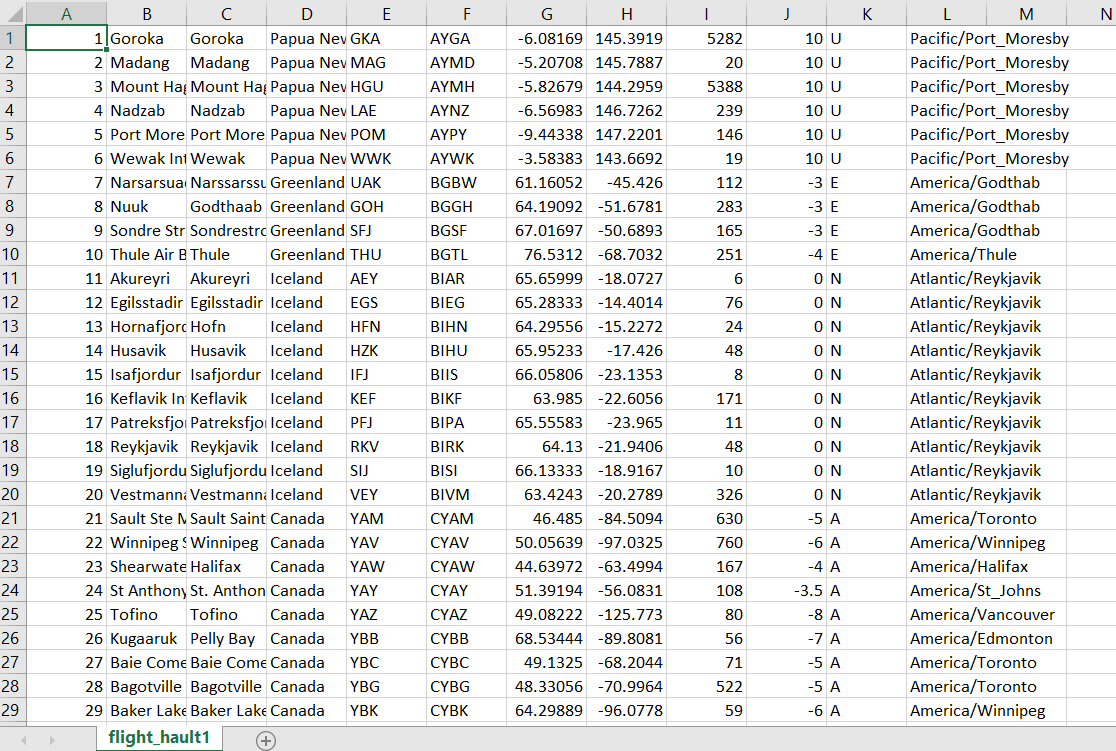
11.Calculate the cluster coefficient and we found that MNL airport has the highest cluster coefficient.

12. calculate average clustering and it is found 0.144

From the above calculation we found that MNL airport has the highest degree of centrality, closeness centrality, betweenness centrality, and cluster coefficient so this the very important and busiest airport and most connected airport to another airport. As FMM airport has the highest Eigen values it also has important it is connecting to the airports which further connected to the another major airports.



**Flight\_hault1**



**Ans:**

**Business Objectives:**

To find the busiest airport by calculating measure degree centrality, degree of closeness centrality, and degree of in-between centrality.

**Business Constraints:**

The traffic of airplanes at the particular airport.

Flight\_halt=c("ID","Name","City","Country","IATA\_FAA","ICAO","Latitude","Longitude","Altitude","Time","DST","Tz database time")

|  |  |  |  |
| --- | --- | --- | --- |
| **Name Of Feature** | **Description** | **Type** | **Relevance** |
| Id | ID of flight | Qualitative, Nominal | Irrelevant for the analysis |
| Name | Name of flight | Quantitative, Nominal | Irrelevant for the analysis |
| City | Name of the city where airport is available | Qualitative, nominal | Relevant for the analysis |
| Country | Name of country | Quantitative, Nominal | Irrelevant for the analysis |
| IATA\_FAA | It has the name for the airport | Qualitative, nominal | Relevant for the analysis |
| ICAO | It contains the destination airport | Quantitative, Nominal | Irrelevant for the analysis |
| Latitude | Value of the latitude of the airports | Quantitative, Nominal | Irrelevant for the analysis |
| Longitude | Value of the longitude of the airports | Qualitative/ Quantitative, Nominal | Irrelevant for the analysis |
| Altitude | Altitude of the airports |  |  |
| Time | Time of hault of the airplane |  |  |
| DST |  |  |  |
| Tz Database Time |  |  |  |

1. Importing libraries like pandas, matplotlib, numpy and networkx.

Networkx is a Python Package for the creation, manipulation and study of the structure, dynamics and functions of complex networks.

2. Loading the data and check for columns, adding names to the column, check missing values, datatypes.

Here we have taken 1st 1000 rows for the analysis. Here we see that in IATA\_FAA there is 103 missing values so we do mean imputation

for this column.

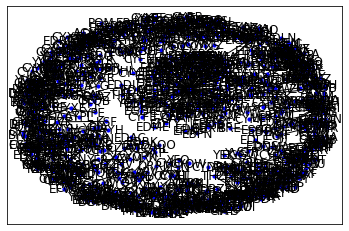
3. Creating the empty graph for no of nodes and no. of edges. graph stores nodes and edges with optimal data or attributes.

4. Takes source and destination nodes from the data frame by using from\_pandas\_edgelist variable.

5. nx.info gives the no. of nodes and edges and average degree from the source and the destinations of the datasets.

6. calculate the degree of centrality and we found that **MNL** airport has the highest degree of centrality.

7. Draw the graph which shows the connections routes of one airport to another



8. Calculating the closeness centrality and we found that 4I7 has highest closeness centrality.

9. Calculate the betweenness Centrality and we found that 4I7 airport has the highest betweenness centrality.

10. Calculate the Eigenvector Centrality and we found that AEE airport has the highest Eigenvector Centrality

From the above calculation we found that 4I7 airport has the highest degree of centrality, closeness centrality, betweenness centrality, and cluster coefficient so this the very important and busiest airport and we have do to another airport through this airport. AEE airport has the highest Eigen values it also has important it is connecting to the airports which further connected to the another major airports.

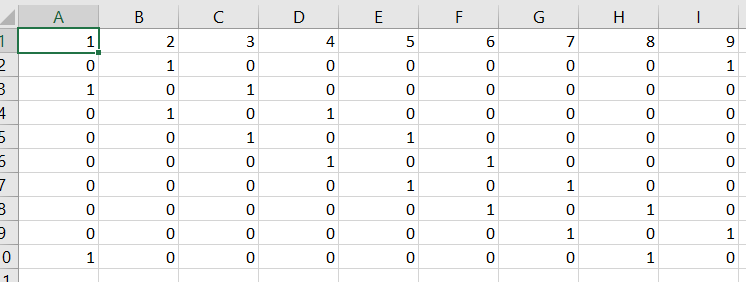
**Problem statement**

There are three datasets given (Facebook, Instagram, and LinkedIn). Construct and visualize the following networks:

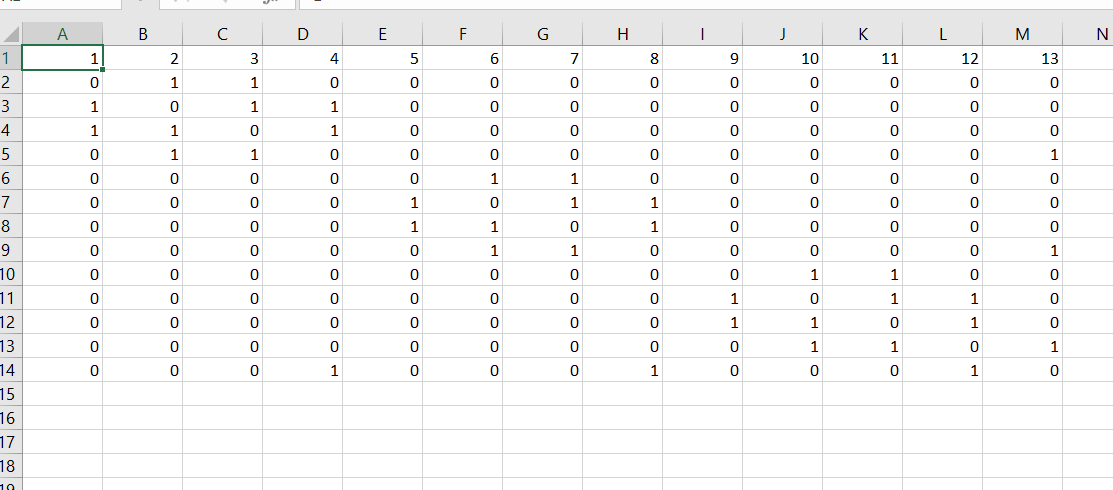
* circular network for Facebook
* star network for Instagram
* star network for LinkedIn

Create a network using an adjacency matrix (undirected only). The snapshots of those datasets are given below:

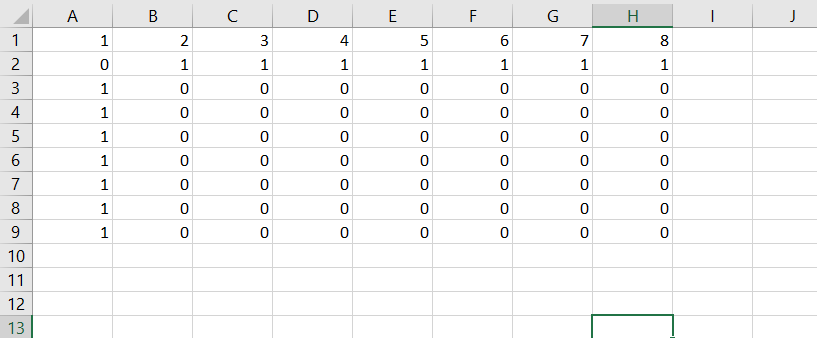
**Facebook**

****

**Instagram**

****

**LinkedIn**



Ans:-

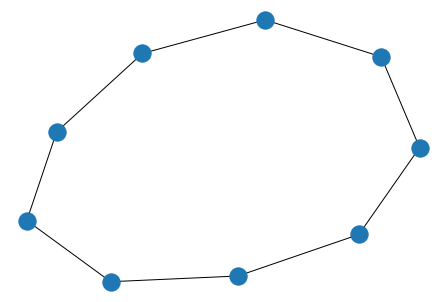
1. Importing libraries like pandas, matplotlib, numpy and networkx.

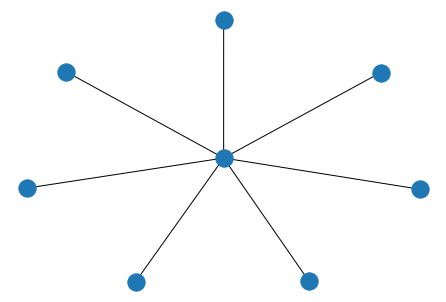
Networkx is a Python Package for the creation, manipulation and study of the structure, dynamics and functions of complex networks.

2. Loading the data and check for columns, missing values, datatypes.

3. Convert the data into matrix form by using matrix function of numpy package and then draw the graph.

a) Circular Network for Facebook



b) star network for Instagram

c) Star network for LinkedIn

