# 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: \_\_\_Jaidev Chhabria\_\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_DSWEMON 290521\_\_\_\_\_\_\_

**Topic: Network Analytics**

**Grading Guidelines:**

**1. An assignment submission is considered complete only when correct and executable code(s) are submitted along with the documentation explaining the method and results. Failing to submit either of those will be considered an invalid submission and will not be considered for evaluation.**

**2. Assignments submitted after the deadline will affect your grades.**

**Grading:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ans** | **Date** |  |  | **Ans** | **Date** |
| Correct | On time | A | 100 |  |  |
| 80% & above | On time | B | 85 | Correct | Late |
| 50% & above | On time | C | 75 | 80% & above | Late |
| 50% & below | On time | D | 65 | 50% & above | Late |
|  |  | E | 55 | 50% & below |  |
| Copied/No Submission |  | F | 45 |  |  |

* **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline.
* **Grade B: (>= 80 and < 90):** 
  + When assignments are submitted on time but less than 80% of problems are completed.

(OR)

* + All assignments are submitted after the deadline.
* **Grade C: (>= 70 and < 80):** 
  + When assignments are submitted on time but less than 50% of the problems are completed.

(OR)

* + Less than 80% of problems in the assignments are submitted after the deadline.
* **Grade D: (>= 60 and < 70):**
  + Assignments submitted after the deadline and with 50% or less problems.
* **Grade E: (>= 50 and < 60):** 
  + Less than 30% of problems in the assignments are submitted after the deadline.

(OR)

* + Less than 30% of problems in the assignments are submitted before the deadline.
* **Grade F: (< 50):** No submission (or) malpractice.

**Hints:**

1. **Business Problem**
   1. **What is the business objective?**
   2. **Are there any constraints?**
2. **Work on each feature of the dataset to create a data dictionary as displayed in the below image:**



**2.1 Make a table as shown above and provide information about the features such as its data type and its relevance to the model building. And if not relevant, provide reasons and a description of the feature.**

1. **Data Cleaning**
2. **Model Building**
   1. **Perform network analytics on the given datasets.**
   2. **Briefly explain the model output in the documentation.**
3. **Write about the benefits/impact of the solution - in what way does the business (client) benefit from the solution provided?**

With directed model we can see Average clustering, coefficient, centrality, betweenness, degree etc. Business can check the least value between routes and plan their business strategies accordingly.

**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.

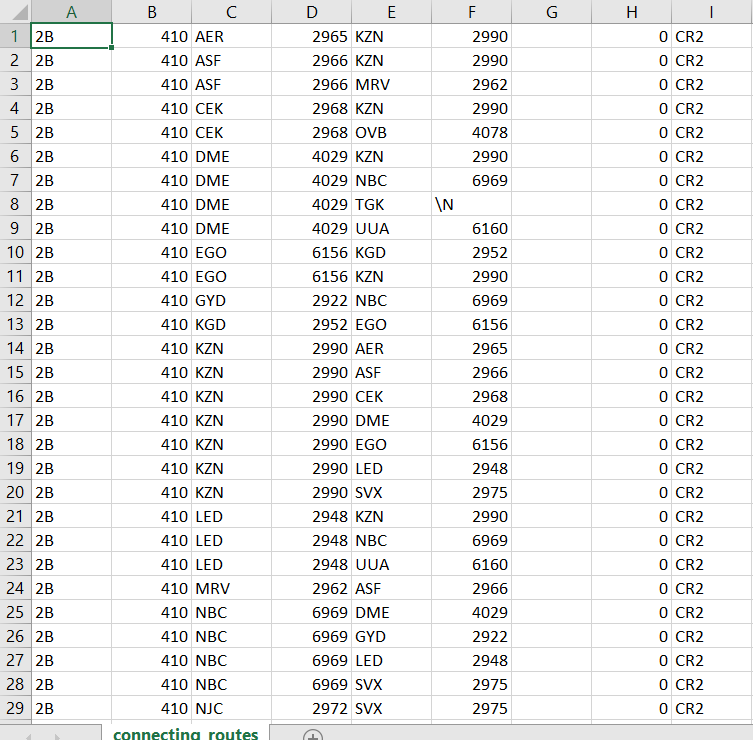
* Create a network using edge list matrix (directed only**).**
* Columns to be used ***in R*:**

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")

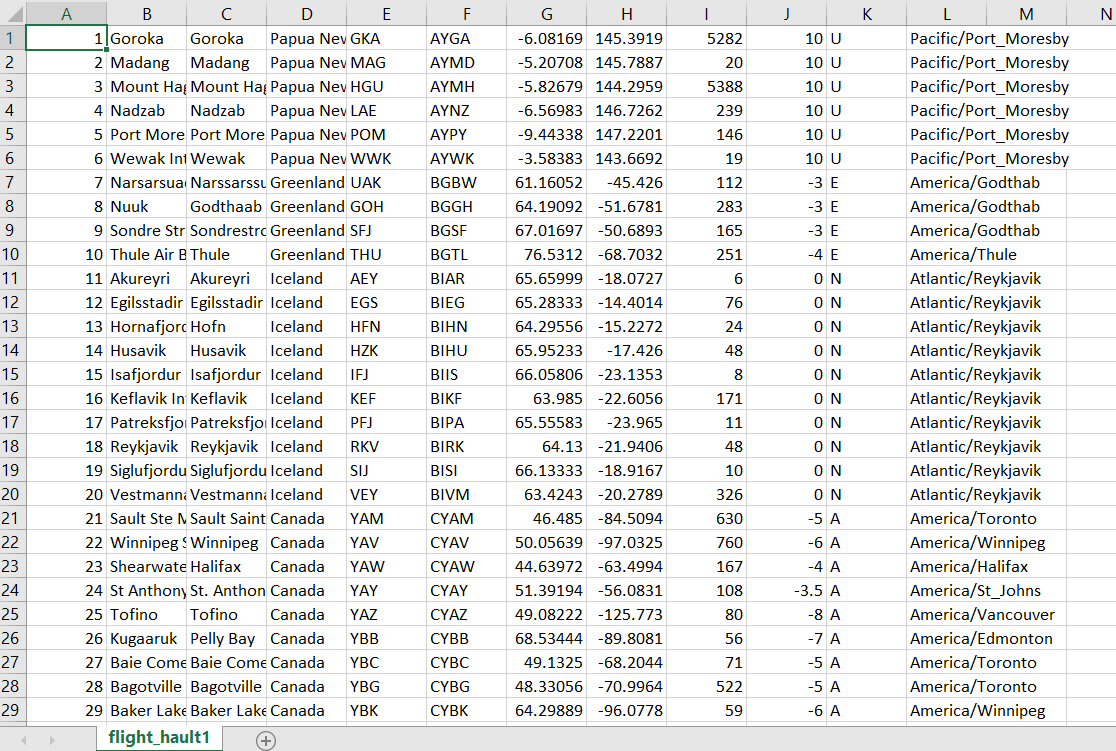


**connecting routes**





**Flight\_hault1**



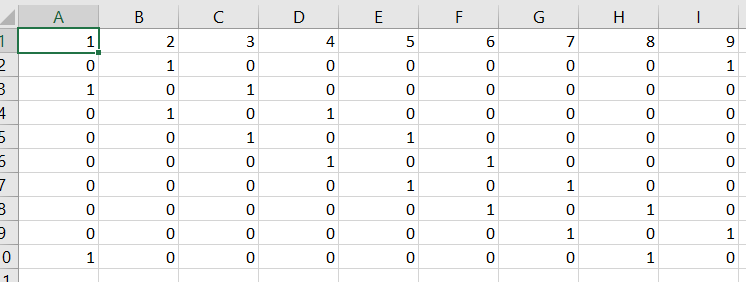
**Problem statement**

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

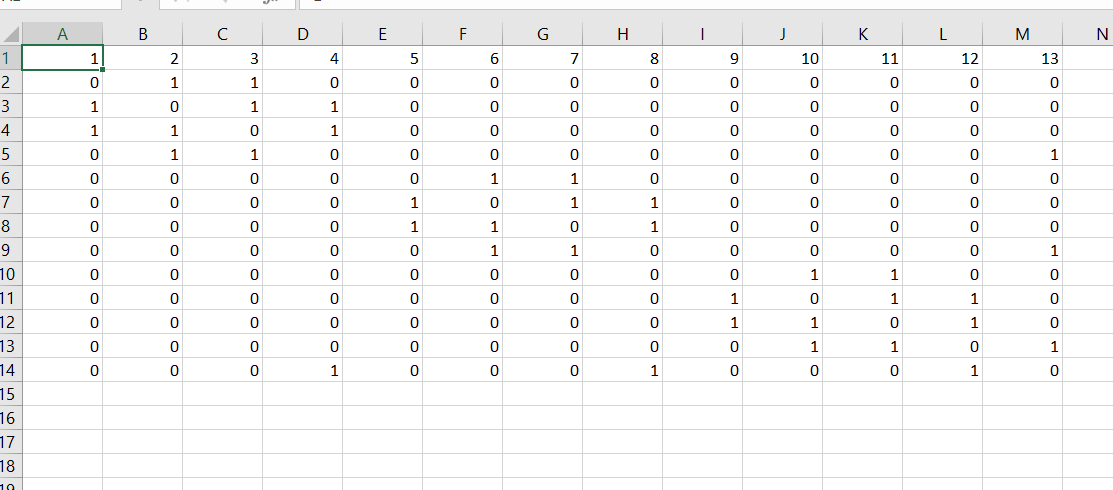
* circular network for Facebook
* star network for Instagram
* star network for LinkedIn

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

**Facebook**

****

**Instagram**

****

**LinkedIn**

