**DASC 5300 Project 2 Report**

1. **Overview**

This project provided a lot of hands-on knowledge about programming and mainly the analysis involved in a project. I started the project by going through the project description and files provided. Then I started preprocessing the data by cleaning it first and then filtering out my assigned airlines from the list. I created the graph structure and extracted the characteristics.

I then analysed the graph to find top 5 nodes with most connections, manually and with algorithms. Analysis on this data revealed that the airlines I was working with was Swiss International Airlines. Further I did my own analysis and predicted the next hub for this airline.

1. **File Descriptions**
2. **DASC5300\_Project2.ipynb**: This is the python notebook file created containing code for the analysis done on the data set.
3. **Network\_summary.py:** Python file to find max, min, avg, stddev
4. **Routes\_v3.txt**: txt file of unfiltered data.
5. **My\_airline.csv**: Csv file to output filtered data
6. **Deg\_stats.csv**: graph statistics output to csv file
7. **My\_airline\_top\_hubs**: Csv file to output top hubs by different algorithms
8. **\_\_deg\_dist.png:** visualisation of degree distribution
9. **AL54\_graph.png:** Visualisation of graph data structure
10. **Division of Labour**

Since I was the only one in the team, I completed the project alone. I divided the work among myself, that is, by creating milestones and assigning time to each section. I was able to complete the project in timely manner, by dividing the work into milestones.

1. **Problems Encountered**

My major obstacle was removing entries for return flight. I spent a lot of time on removing these entries. I finally figured that these entries were nothing but duplicates in reverse order. So, I sorted columns “Source\_code” and “Destination\_code” by rows and then removed the duplicate entries.

1. **Analysis**
2. **Graph Construction:**

Chart, radar chart

Description automatically generatedI used a dictionary to map the ‘Source\_code” and “Destination\_code” column to a list of number starting from 1 assigning each unique entries a unique number. I then used the Networkx module to create my graph data structure directly from the dataframe. I further visualized the data structure using Networkx, coloring top nodes (with connections more than **average degree**) with “Red” and rest with green “Green” and giving each node weight according to their connections.

Figure1. Graph of AL54 routes

1. **Graph Characteristics:**

I used networkx functions and functions provided by Mr. Enamul to get the graph characteristics for my airlines, like number of nodes, edges, density, etc.

Table1.1 Graph Characteristics

|  |  |
| --- | --- |
| **Characteristics** | **Value** |
| Number of nodes | 103 |
| Number of edges | 145 |
| Density | 0.0276 |
| Number of connected components | 1 |
| Connected component | [103] |
| Diameter | 5 |
| Minimum Degree | 1 |
| Maximum Degree | 94 |
| Average Degree | 2.8155 |
| Standard Deviation | 9.7123 |

* We can see that the number of nodes are only 103 and edges are 145, which is a small number. Leading us to believe that it is a mid to small scale airline.
* Minimum degree is 1 and maximum is 94, that would probably be the node which will be the hub of this airline.
* Chart, histogram

  Description automatically generatedAverage degree is 2.8155 and standard deviation is a big number 9.7123. This tells us that there are very few nodes with large number of connections and rest are nodes with single or two connections.

Figure1. Distribution of the degrees in the graph

1. **Finding Top k Hubs:**

I then manually analysed the graph to find top 5 nodes with most connections. I further used the degree centrality, closeness centrality and betweenness centrality algo to find top hubs. These are listed below:

Table1.2 Top 5 hubs of AL54 by different algorithms

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Hubs** | **Manual Analysis** | **Degree Centrality** | **Closeness Centrality** | **Betweenness Centrality** | **Degrees** |
| First | ZRH | ZRH | ZRH | ZRH | 94 |
| Second | GVA | GVA | GVA | GVA | 35 |
| Third | BSL | BSL | BSL | BSL | 9 |
| Fourth | BRU | BRU | BRU | BRU | 6 |
| Fifth | FRA | FRA | FRA | FRA | 4 |

* We can see that for this graph all the algorithms yield the same top 5 nodes.
* Also, after the top 2 hubs the degree drops drastically.
* For the sake of analysis, we will consider all five hubs, even though BSL has almost 4 times fewer connections than GVA.

1. **Identification of Airline:**

* Zurich (ZRH) is the top hub for this airline. Doing a quick reverse search of airlines and their hubs on the internet [1] tells us that ZRH is the hub of only one airline, that is, Swiss International Airlines.
* Swiss International has 102 destination which matches our number of nodes 102+1(main hub included)
* This also matches our characteristics analysis, Swiss International is a small airline compared to giants like Emirates, AA etc who have fleet size of 252 or more and have destinations more than 150 [2].

1. **Identification of the Next Hub**

* From section 5.iii we have three contenders for the next hub since all the **other nodes** have too few connections and investing in them would require the airlines to create new connections and that **would require a lot of investment**.
* We consider BSL, BRU and FRA for our analysis since ZRH and GVA are already hubs.
* Coming to Frankfurt (FRA), Brussels (BRU) and Basel (BSL), we see that BSL has **higher number of connections**. Therefore, converting it into a hub would require lesser investment than BSL. Though the difference in connections between these three is quite small so this point does not yield us anything useful.
* Out of all three airports, FRA is the **biggest airport** having 2 terminals and third one in construction. BSL which has only 1 terminal and BRU has two piers [3][4][5].
* Geographically, FRA is **closer to other European cities** across central European countries, like Czechia, Denmark, Poland, Sweden, Netherlands, etc. BRU is towards the west where the only country close to it is UK and Ireland. Lastly, BSL is close to Zurich and Bern, Zurich being a hub itself would **defeat the purposing** of making BSL a hub. Therefore, flights from FRA will have to travel less distance on average to a given European destination [6].
* All these points brings us to the conclusion that FRA would be the best choice to be the next hub for Swiss International Airlines.

**Appendix A:**

**References**

1. Major airlines of the world, accessed on Oct 20, 2021, <https://www.nationsonline.org/oneworld/major_airlines.htm>
2. Emirates (Airlines), accessed on Oct 21, 2021, <https://en.wikipedia.org/wiki/Emirates_(airline)>
3. EuroAirport Basel Mulhouse Freiburg, accessed on Oct 21, 2021, <https://en.wikipedia.org/wiki/EuroAirport_Basel_Mulhouse_Freiburg>
4. Frankfurt Airport, accessed on Oct 21, 2021, <https://en.wikipedia.org/wiki/Frankfurt_Airport>
5. Brussels Airport, accessed on Oct 21, 2021, <https://en.wikipedia.org/wiki/Brussels_Airport>
6. Google Maps, accessed on Oct 21, 2021, <https://www.google.com/maps/place/Frankfurt+Airport/@50.0379326,-0.4026919,5z/data=!4m5!3m4!1s0x47bd0b745442f979:0x8a74d84d21f8f331!8m2!3d50.0379326!4d8.5621518>