Global Air Transportation Network Graph Analytics

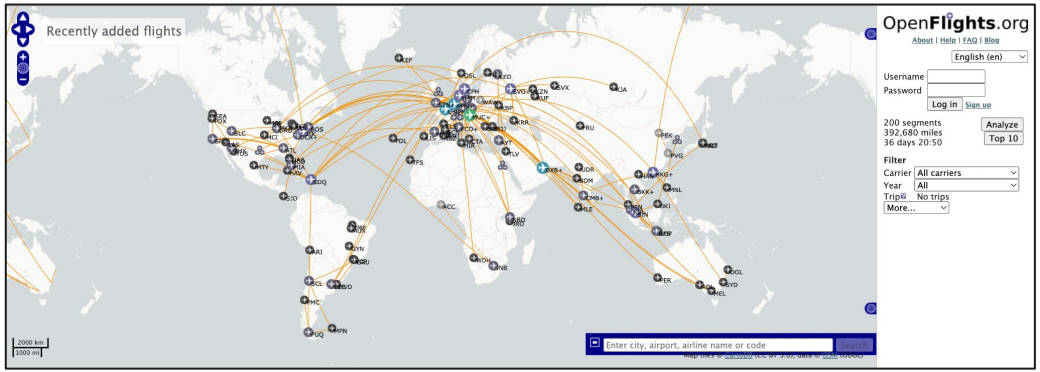
By Akhil Devarasetty, Jasmine Camacho, and Russell Tapia

Purpose of the Project

- Question:
  - Which airport would have the biggest impact on the other airports when it is shut down
- Example use cases:
  - Prioritize quarantine in case of outbreaks
  - Resource allocation in order to keep transportation working
  - Economical seasonality (busy period and non-busy periods). Useful for planning trips.

About the Dataset

- Airports, Airlines and Routes is a comprehensive air travel dataset
  - Including: airports, airlines, routes and airplanes
- Contains over 10,000 data points
- Compiled by OpenFlights
- Airports -> nodes (7698)
- Routes -> edges (66772)



About the Software

- Analysis was done using Gephi
  - An open-source network analysis and graph visualization software
- Allows CSV imports to create graphical visualizations
- Capable of running analytical algorithms in order to get statistics on graphs

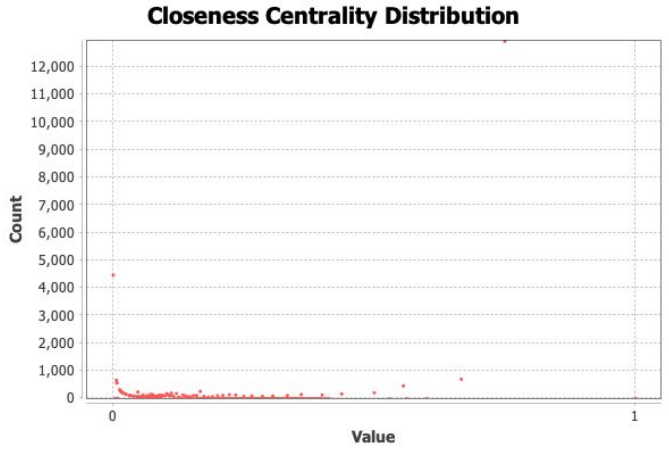
Process

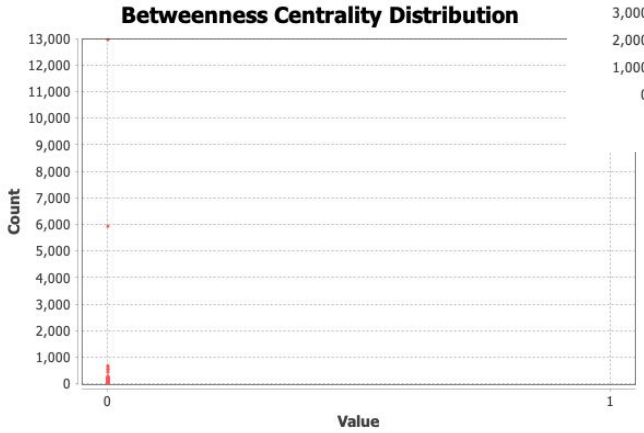
- Clean the dataset
  - Understand the dataset
  - Create proper header name for columns (Gephi requires ‘Source’ and ‘Target’ for edge table)
  - Verify correct data type per column
  - Remove ‘Null’ values and duplicates
- Analyze graph
  - Run network algorithms
  - Filter for degree

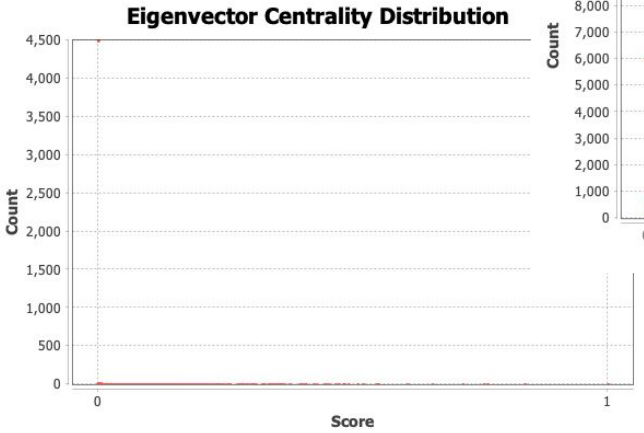
Results

Red = low weight
Yellow = medium weight
Blue = high weight

Results contd.







Explaining Results

- **Closeness Centrality** - Conceicao do Araguaia Airport, Larsen Bay Airport, Kongolo Airport, Eros Airport, Unalaska Airport, Victoria Harbour Seaplane Base, Nerlerit Inaat Airport, Vancouver Harbour Water Aerodrome, Ittoqqortoormiit Heliport
    - Valued at 1
    - How central a node is
- **Betweenness Centrality** - Los Angeles International Airport
    - Valued at 0.01455
    - The extent a node lies in the path between two nodes
    - Think of connecting flights
- **Eigenvector Centrality** - Hartsfield Jackson Atlanta International Airport
    - Valued at 1
    - Influence of a node to the nodes next to it
    - Higher score if nearby nodes have more connections