

University of Bologna

Social Network Analysis

Open Flights

Analysis of Airplane Routes based on Airports Data

Brajucha Filippo 0001130613

Leonardo Dessì

Gianmarco Gabrielli

Simone Rinaldi

1 Introduction

The idea is to develop the analysis of the global network of airports and the air routes connecting them.

We founded a data source on the OpenFlights website ([link](#) and [github-link](#)), which allow us to construct a graph representing this network. In the graph, the nodes represent the airports, while the edges represent the direct air routes between two different airports.

We weight the edges according to the number of flights or connections that exist between the various airports.

The main objective of the project is to identify the most relevant and strategic airports on a network level. We identified the airports that serve as major transit nodes and subsequently test the hypothesis that these airports are critical for the overall connectivity of the network.

2 Dataset

2.1 Dataset Description: OpenFlights Air Route Network

The dataset used for this project is derived from OpenFlights, a public resource for airline, airport, and route data. Specifically, the dataset is an **edgelist** format file that describes the global air transportation network. This dataset is particularly valuable for studying the structure and dynamics of global air travel.

2.1.1 Source

- **URL:** OpenFlights Dataset
- **Provider:** The dataset is hosted on Tore Opsahl's website, a notable source for network analysis datasets.

2.1.2 Format and Structure

The dataset is formatted as an edgelist, where each row represents a direct connection (route) between two airports. An edgelist is a common representation of a graph in network analysis, detailing the edges (connections) between nodes (airports). Each line contains the following key information:

1. **Source Node (Airport):** A numerical ID representing the airport where the route originates.
2. **Destination Node (Airport):** A numerical ID representing the airport where the route ends.
3. **Weight (Number):** An additional column representing the weight of the edge, often indicating frequency or capacity.

Each node number is linked to a list (`openflights_airports.txt`) where is possible to find all the info about the airport (for example the country, the city, the code, the coordinates etc).

2.2 Key Characteristics

- **Nodes:** Airports, each uniquely identified by an ID.
- **Edges:** Direct flight routes connecting pairs of airports.
- **Network Type:**
 - *Directed:* Routes are directional, from source airport to destination airport.
 - *Weighted:* Depending on the data, edges might include weights for analyzing traffic intensity or other attributes.

3 Applications

This dataset is ideal for exploring various network science topics, including:

- **Degree Centrality:** Identifying major hubs in the air transportation network.
- **Clustering:** Finding regional clusters of airports.
- **Shortest Path Analysis:** Examining the efficiency of global air connectivity.
- **Community Detection:** Identifying clusters of airports with dense internal connections.

4 Limitations and Considerations

- **Static Data:** The dataset represents a snapshot in time and does not capture temporal variations in routes.
- **Coverage:** The dataset may not include all airlines or routes, as coverage depends on OpenFlights' data collection.
- **Data Cleaning:** It may require preprocessing to handle missing data, self-loops, or duplicate edges.

5 Relevance to the Project

The OpenFlights edgelist provides a comprehensive representation of global air-line routes, making it a powerful tool for analyzing the structure of air transportation networks. By leveraging this dataset, the project aims to explore metrics such as centrality, connectivity, and clustering within the global airport network, contributing to the broader understanding of network dynamics in transportation.

6 Validity and Reliability