Airports Network Dynamics and Passenger Flow

Lucia Colin Cosano
Email: lcolincosano@hawk.iit.edu

May 2024

This paper presents a comprehensive analysis of airline network dynamics and passenger flow using datasets obtained from the Bureau of Transportation Statistics (BTS) of the US Department of Transportation. The primary objective is to investigate the intricate relationships within the air transportation system, focusing on aspects such as airport connectivity, seasonal variations in passenger traffic, and airline performance. The project employs a network modeling approach, constructing two distinct network models: Airport Network and Airline Network. These models represent the air transportation system as directed and weighted networks, facilitating the exploration of connectivity patterns and passenger movements. The analysis aims to identify major routes and hubs, understand seasonal travel patterns, and assess the performance of different airlines. Data collection, cleaning, transformation, and feature engineering processes are thoroughly described to ensure the integrity and quality of the analysis. Insights derived from the study are expected to contribute to a deeper understanding of the dynamics of the air transportation network and inform decision-making processes within the aviation industry.

Keywords: *Airline network, Passenger flow, Airport connectivity, Seasonal variations, Network modeling, Data analysis, Feature engineering.*

The global aviation industry plays a pivotal role in facilitating economic growth, cultural exchange, and international connectivity. As air travel continues to witness steady growth, understanding the underlying dynamics of airline networks and passenger flow becomes increasingly crucial for policymakers, industry stakeholders, and researchers alike. This paper presents a comprehensive study aimed at analyzing the intricate interplay between airline network dynamics and passenger movements within the United States air transportation system.

The foundation of this study lies in the utilization of rich and extensive datasets provided by the Bureau of Transportation Statistics (BTS), a vital resource under the U.S. Department of Transportation.

At the heart of the approach is the adoption of network modeling techniques, which offer a powerful framework for studying complex systems characterized by interconnectedness and dynamic interactions. Two distinct network models have been constructed: the Airport Network and the Airline Network. These models provide a holistic representation of the air transportation system, offering insights into airport connectivity, airline operations, and the flow of passengers between origins and destinations.

The Airport Network encapsulates the spatial relationships between airports, depicting the network of routes traversed by flights. By treating airports as nodes and flight routes as edges, this model enables to identify major routes, hubs, and patterns of connectivity within the network. Similarly, the Airline Network sheds light on the strategic interactions among airlines, mapping out the relationships between carriers based on shared route segments and passenger volumes.

Through rigorous analysis and exploration of these network models, we aim to achieve several objectives. Firstly, we seek to identify major routes and hubs within the air transportation system, shedding light on critical nodes of connectivity and traffic concentration. Secondly, we endeavor to uncover seasonal variations in passenger traffic, discerning patterns of travel behavior that fluctuate throughout the year. Lastly, we strive to evaluate the performance of different airlines, assessing their market presence, network connectivity, and operational efficiency.

The findings of this study hold significant implications for various stakeholders within the aviation industry. From policymakers shaping aviation policy to airlines optimizing their route networks, the insights derived from our analysis can inform strategic decision-making processes and drive innovation within the sector. By unraveling the intricate dynamics of airline networks and passenger flow, we contribute to a deeper understanding of the air transportation system, paving the way for a more efficient, resilient, and interconnected aviation ecosystem.

The dataset utilized in this study was obtained from the Bureau of Transportation Statistics (BTS), an authoritative source of transportation data within the United States Department of Transportation. The BTS provides a comprehensive repository of datasets encompassing various modes of transportation, including air travel. The specific dataset employed in our analysis comprises detailed information on airline flights, encompassing attributes such as passenger counts, flight distances, airline details, airport identifiers, and temporal information.

The decision to utilize the BTS dataset was driven by several factors. Firstly, the BTS datasets are widely recognized for their reliability, comprehensiveness, and accessibility, making them an ideal choice for conducting rigorous analysis within the aviation domain. Moreover, the dataset offers a rich array of variables that capture key aspects of airline operations and passenger movements, aligning closely with the objectives of our study.

The dataset utilized in this study encompasses eleven features, each offering unique insights into various aspects of airline operations and passenger movements within the air transportation system. The main features are as follows:

- Passengers: this feature represents the number of passengers on a specific flight departure. It serves as a fundamental indicator of passenger demand and traffic volume on individual flights, enabling the analysis of load factors and passenger distribution across different routes and airlines.
- Distance: the distance of the flight route is captured by this feature, providing crucial information about the spatial dimensions of airline operations. Distance serves as a key determinant of flight duration, fuel consumption, and operational costs, influencing route planning and scheduling decisions.
- Unique_carrier: this variable serves as a unique identifier for the airline operating the flight. It enables the categorization and segmentation of flight data based on carrier attributes, facilitating the analysis of airline-specific trends, performance metrics.
- Unique_carrier_name: the name of the airline operating the flight is represented by this feature. It provides descriptive information about the carrier, enhancing the interpretability and contextual understanding of the dataset.
- Origin_airport_ID: This feature comprises a unique identifier for the origin airport of the flight. It enables the geospatial mapping of flight routes and facilitates the analysis of airport connectivity patterns, hub operations, and regional air traffic flows.
- Origin: The name of the origin airport is captured by this variable, offering descriptive information about the airport's location and identity.
- Dest_airport_ID: Similar to the origin airport identifier, this feature represents a unique identifier for the destination airport of the flight. It enables the analysis of airport pairs, route structures, and network connectivity.
- Destination: The name of the destination airport is provided by this variable, offering descriptive information similar to the origin airport name.
- Year: This feature denotes the year of the flight departure. It enables temporal analysis, allowing the exploration of trends, seasonality, and long-term patterns in airline operations and passenger behavior.
- Quarter: the quarter of the year (e.g., Q1, Q2) facilitates the segmentation of data into discrete time intervals, enabling the analysis of seasonal variations and quarterly trends in passenger traffic.
- Month: it refers to the month of the year, facilitating the analysis of monthly trends, seasonal peaks, and periodic fluctuations in airline operations.

The approach to data collection and dataset selection has remained consistent with the initial plan outlined in the data description deliverable.

Upon initial inspection, the dataset was found to be free from missing cells, ensuring completeness. However, the presence of duplicate rows, accounting for 0.6% of total observations, posed a challenge to data integrity.

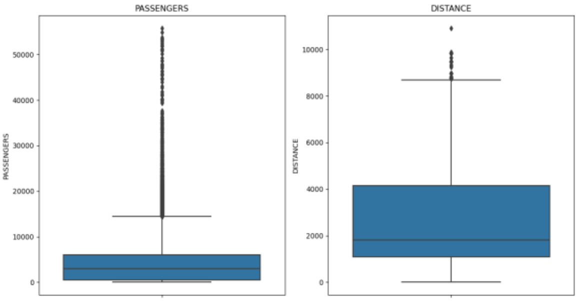
The subsequent phase of the project focused on data cleaning and transformation to prepare the dataset for analysis. This involved handling missing values, unique values, duplicate rows, and outlier detection.

The *passengers* feature exhibited null values, with approximately 19.92% of its entries being zero, indicating null entries. Rows with null values were removed to maintain data integrity and ensure accurate analysis.

Related to unique values, the *year* feature contained only one unique value, rendering it redundant for analysis purposes. Due to this fact, the feature was deemed irrelevant and subsequently removed from the dataset to focus on features contributing meaningful insights.

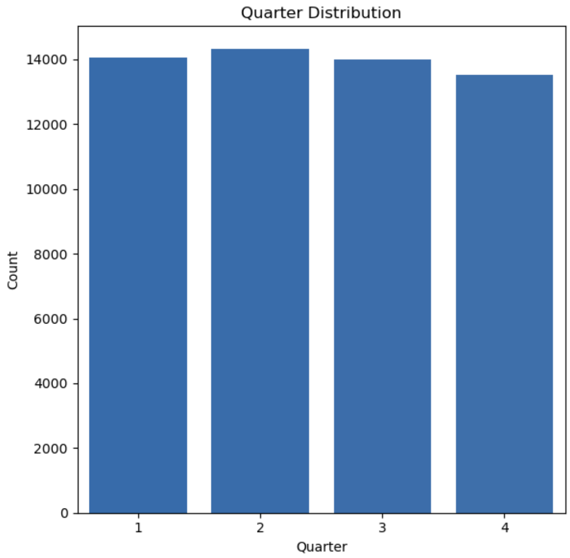
Outlier detection was conducted using visualizations, particularly box plots, which aided in identifying outliers within numerical features such as passenger counts and distances traveled. Subsequently, outliers were carefully identified and addressed to mitigate their potential impact on analysis results.

Figure 1. Outliers detection



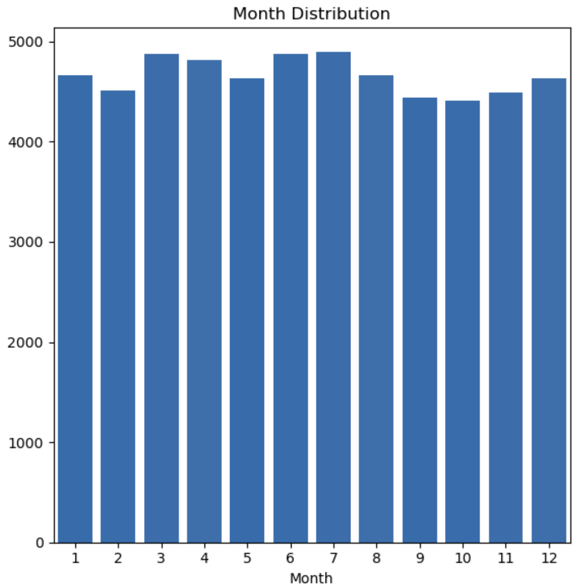
Moving beyond numerical attributes, relevant categorical features have been explored, particularly Quarter and Month distributions.

Figure 2. Quarter Distribution



The Quarter Distribution chart illustrates the distribution of data across quarters, each representing a three-month period within a year. Uniformity in bar heights across quarters implies a consistent pattern in business activity throughout the year, without significant seasonal variations.

Figure 3. Month Distribution



Similarly, the Month distribution chart showcases the distribution of data across months. Consistency in bar heights indicates stable business activity throughout the year, with subtle variations within each quarter suggesting potential peak seasons or specific events.

To enrich the analysis, a new feature has been created: *average distance per passenger*. By calculating the average distance traveled per passenger for each flight, this feature provides valuable insights into the efficiency of passenger travel, shedding light on flight dynamics and passenger behavior.

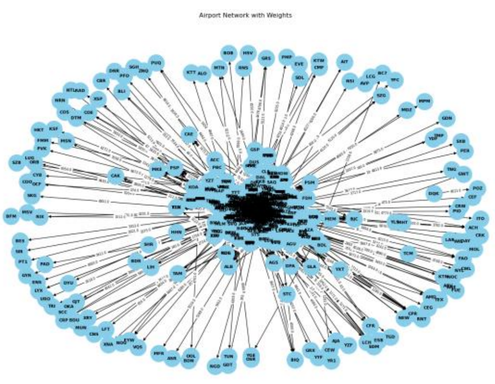
Entering the subsequent phase of the project, the focus shifts towards the intricate construction and visualization of the Airport Network. This network encapsulates the spatial relationships between airports, delineating the web of flight routes traversed by commercial airlines.

Each airport serves as a node within the network, interconnected by edges representing the direct flight routes between them. The presence of an edge signifies the existence of a direct air link between two airports, with the weight of the edge indicative of the distance between the connected airports.

Related to the Airline Network, this network elucidates the strategic relationships between airlines, mapping out the flow of passengers between origin and destination airports. Each airline is represented as a node within the network, interconnected by edges representing shared route segments between airports.

.In the pursuit of understanding the intricate dynamics of airport and airline, a formidable challenge arose in visually representing the vast array of connections among airports. Initially, a comprehensive network model was constructed using all available samples from the dataset. However, upon visualizing the resulting graph, significant complexity was encountered, obscuring any discernible patterns or insights.

Figure 4. Airport network



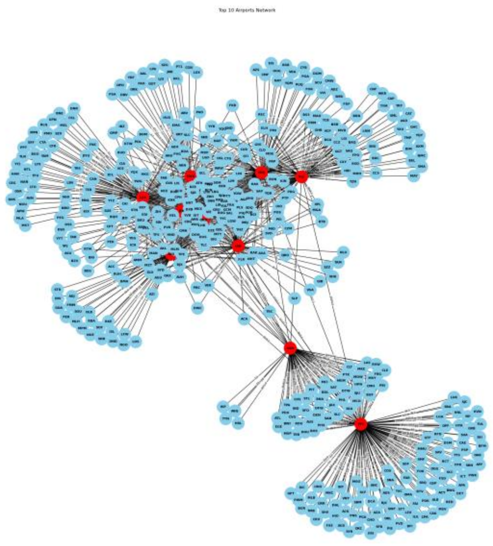
To address this challenge a random sample comprising 10% of the available data was used. This was driven by the objectives of reducing computational complexity and providing a representative subset for modeling purposes. Despite the rationale behind this approach, challenges were encountered as the randomness introduced by the sampling obscured meaningful patterns within the network.

To address this challenge, an iterative approach was adopted to devise a more focused method aimed at enhancing the interpretability of the network visualization. Recognizing the pivotal role of certain airports in shaping connectivity and passenger flows, the analysis was narrowed down to the top 10 airports, characterized by their high volume of flight activity and strategic importance.

The method commenced with the extraction of airport frequency data from the dataset, followed by the identification of the top 10 airports based on their prominence in flight volume. Subsequently, the dataset was filtered to include only data pertaining to these top airports, streamlining the complexity of the network while retaining its fundamental structure and insights.

Utilizing the NetworkX library in Python, a weighted graph was constructed to encapsulate connectivity between the top airports, with each airport serving as a node and flight routes as edges. Importantly, the graph was augmented with additional information on passenger counts, transforming it into a visualization of passenger flow between key nodes.

Figure 5. Top 10 Airport network



The top 10 airports—John F. Kennedy International Airport (JFK), Miami International Airport (MIA), Los Angeles International Airport (LAX), Toronto Pearson International Airport (YYZ), Newark Liberty International Airport (EWR), O'Hare International Airport (ORD), George Bush Intercontinental Airport (IAH), Fort Lauderdale-Hollywood International Airport (FLL), Cancún International Airport (CUN), and Washington Dulles International Airport (IAD)—emerge as critical nodes within the air transportation system, each playing a pivotal role in facilitating domestic and international travel.

John F. Kennedy International Airport in New York City serves as a major gateway for transatlantic flights, connecting the United States with Europe and beyond. Miami International Airport serves as a key hub for flights to and from Latin America, positioning it as a vital link for travelers to and from the region. Los Angeles International Airport is a prominent hub on the West Coast, serving as a gateway to the Asia-Pacific region and a major center for domestic flights. Toronto Pearson International Airport is Canada's busiest airport, serving as a critical hub for both domestic and international travel. Newark Liberty International Airport in New Jersey complements 'JFK' as a key gateway to the New York metropolitan area, offering a range of domestic and international flights. O'Hare International Airport in Chicago is one of the busiest airports in the world, serving as a major hub for both domestic and international flights. George Bush Intercontinental Airport in Houston is a key hub for flights to and from Latin America, as well as serving domestic routes across the United States. Fort Lauderdale-Hollywood International Airport in Florida is a popular gateway for travelers to the Caribbean and Latin America, as well as serving domestic routes within the United States. Cancún International Airport in Mexico is a major hub for tourism in the Caribbean, attracting travelers from around the world to its pristine beaches and vibrant culture. Washington Dulles International Airport serves as a key hub for international flights to and from the Washington, D.C. metropolitan area, connecting the capital region with destinations across the globe.

These top 10 airports are characterized by their strategic locations, extensive route networks, and state-of-the-art facilities, making them indispensable components of the global air transportation network.

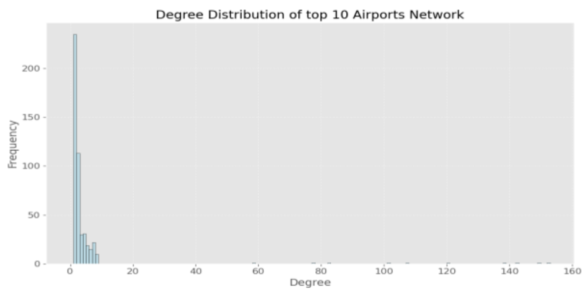
Related to the main network measures, the degree distribution of the Airport Network reveals a pattern characterized by a large number of nodes with relatively low degrees and a small number of nodes with exceptionally high degrees.

Nodes with low degrees represent airports with fewer direct connections to other airports within the network. These airports may serve as regional hubs or destinations with limited direct flight routes, contributing to the abundance of nodes with low degrees in the distribution.

Conversely, nodes with high degrees denote airports that serve as major hubs or central connecting points within the network. These airports enjoy extensive connectivity, with numerous direct flight routes to a wide range of destinations. While they comprise only a fraction of the total nodes in the network, their prominence is significant in facilitating efficient passenger flows and route accessibility across the air transportation system.

This skewed degree distribution underscores the hierarchical nature of the Airport Network, with a small number of major hubs playing a disproportionately influential role in shaping overall network connectivity and accessibility. Understanding this distribution is crucial for analyzing the network's resilience to disruptions, identifying critical nodes, and optimizing route planning and scheduling strategies within the aviation industry.

Figure 6. Degree distribution



Related to the airport network and despite potentil desired insights gleaned from the analysis of the Airport Network, it's important to acknowledge its limitations in providing comprehensive understanding. While it offers insights into spatial relationships and connectivity patterns among airports, it may not capture the nuanced dynamics of airline operations and passenger movements. Recognizing this, a new network has not been developed using the new strategy. This decision reflects the need for a more nuanced approach that considers additional factors beyond airport connectivity, such as airline operations, passenger flows, and strategic interactions among carriers. By adopting a more holistic approach, we aim to gain deeper insights into the intricate dynamics of the air transportation system, paving the way for more informed decision-making and innovation within the aviation industry.

Stepping into the seasonal analysis, monthly passenger flows between airports have been represented.The graphical representation reveals fluctuations in major hub airports depending on the month, indicative of seasonal variations in passenger traffic. Despite these temporal shifts, the visualization underscores the persistent prominence of certain airports as key hubs throughout the year, suggesting a degree of stability amidst changing travel patterns.

Figure 7. Airport network for January

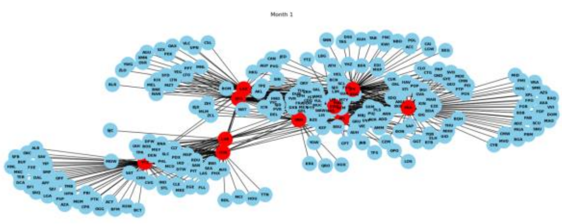
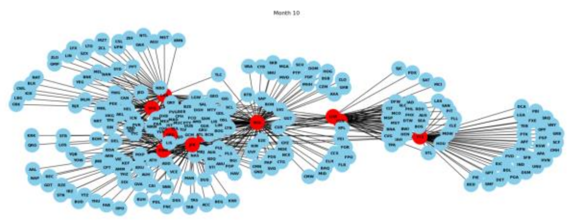


Figure 8. Airport network for December



In conclusion, this comprehensive study has shed light on the intricate dynamics of airline networks and passenger movements within the United States air transportation system. By leveraging rich datasets provided by the Bureau of Transportation Statistics and employing network modeling techniques, we have uncovered valuable insights into airport connectivity and seasonal variations in passenger traffic.

Throughout the project, challenges have been encountered and critical decisions have been made to ensure the integrity and relevance of our analysis. The utilization of rigorous data cleaning and transformation techniques, coupled with the strategic focus on key airports, enabled us to distill complex network structures into actionable insights.

Reflecting on the project, it provided a valuable learning experience, highlighting the importance of visualization techniques, and iterative problem-solving. Moving forward, there are areas for improvement, such as exploring more sophisticated modeling approaches to capture the nuanced dynamics of airline operations and passenger behaviors.

Future work could involve delving deeper into the strategic interactions among airlines, incorporating real-time data sources for more dynamic analyses, and exploring the impact of external factors such as weather patterns and geopolitical events on air transportation systems.

Overall, this project has deepened our understanding of the global aviation industry and its complex network dynamics. It has underscored the importance of interdisciplinary approaches and continuous learning in addressing the evolving challenges and opportunities within the aviation sector.

References

[1] “*Bureau of Transportation Statistic”,* March 20, 2024. https://www.bts.gov/.

[2] *“Software for Complex Networks”* , Apr 15, 2024.https://networkx.org/documentation/stable/index.html

[3] “*New York Airport - JFK Airport: The Main NYC Airport*”. https://new-york-jfk-airport.com

[4] *"Miami International Airport"*, Apr 28. https://www.miamiandbeaches.com/l/transportation/miami-international-airport/

[5] "*Air Transport Network", Apr 27.* https://en.wikipedia.org/wiki/Air_Transport_Network

[6] *"Empirical analysis of airport network and critical airports",* Apr 28 https://www.sciencedirect.com/science/article/pii/S1000936116000376

[7] *Uso básico de NetworkX, Apr 15.* https://madi.nekomath.com/P5/NetworkX.html