UCI Paul Merage School of Business

Leadership for a Digitally Driven World™

# BANA 277 Customer and Social Analytics Project

# Airline Routes: Social Network Analysis and its Business Implications

## Team 12A

Bijay Gautam, Brigitte Harijanto, Conrad Ning, Suyog Sanklapur, Yilu Chen
Date of Completion: 03/10/2024

# Table of Contents

**Project Background**

Let's dive into the world of airline routes and airports—not just as dots and lines on a map, but as a buzzing network that pretty much holds up the globe's economy, tourism, and our plans for next spring break. This paper isn't just another academic chore; it's a journey into understanding how all these flights crisscrossing the sky impact everything from our holiday plans to the world's carbon footprint. We're talking network effects of airline routes, and we're breaking it down with some cool network analysis to see what's really going on.

***Why We Care***

Airports and their spiderwebs of routes do more than just get us from A to B. They are the playgrounds of economic showdowns and technological leaps. Through this paper, we're going to peel back the layers of this complex system to figure out how things work. In an attempt to understand more about the importance of this, we will explore how network effects affect businesses, governments, and airlines.

***Questions to be Answered***

Using some smart network analysis tricks, we'll dive into questions like: How does being in the center of this web benefit an airport or a region? What happens when the network gets shaken up by new tech, political affairs, or pandemics? And how do all these moving parts affect us, the travelers?

***What's Next***

By mixing theories with real-world data from Kaggle, this exploration aims to shed some light on the ripple effects that airline routes have on our world. We will be using GEPHI to graph out our network and have places related to each of our members as a clearer example, whilst making a neater picture so that we can provide insights that could help make better decisions for those running the show in the air travel game, and maybe, just maybe, making flying a better experience for all of us. So buckle up, and let's take this flight together through the interconnected world of air travel, uncovering insights and truths about a network that's as fascinating as it is vital to our way of life. Here's to hoping we land some useful knowledge and maybe contribute a bit to making the future of flying brighter.

## Project Data Description & Exploratory Data Analysis

### *Data Description*

The dataset is collected from Kaggle titled "Global Air Transportation Network"[1] encompasses a comprehensive collection of data on worldwide air travel routes, showcasing the intricacies of the global air transportation network. It consists of four data files, "airlines.csv," "airplanes.csv," "airports.csv," and "routes.csv." With the objective to explore relationships using network analysis, this project utilizes the "routes.csv" file, which serves as the focal point for analyzing the connectivity between airports across different countries and airlines. This file delineates various attributes of air routes, including the airline ID, the source and destination airports by their unique codes, and additional details like equipment used. The dataset's scope covers a vast array of airlines, both global and regional, providing a detailed picture of international and domestic flight paths.

Analyzing this dataset allows for a deeper understanding of the global connectivity provided by air travel, highlighting major hubs and the routes that link cities across continents. It offers valuable insights into the network's density, the significance of certain airports in global traffic, and the diversity of aircraft deployed on different routes. Researchers and analysts can utilize this information for a multitude of applications, ranging from optimizing flight schedules and analyzing airline market share to studying the environmental impact of air travel and planning infrastructure development in response to air traffic demands. The "routes.csv" file stands as a crucial resource for comprehending the dynamics of global air travel and its implications on worldwide connectivity and economic interactions.

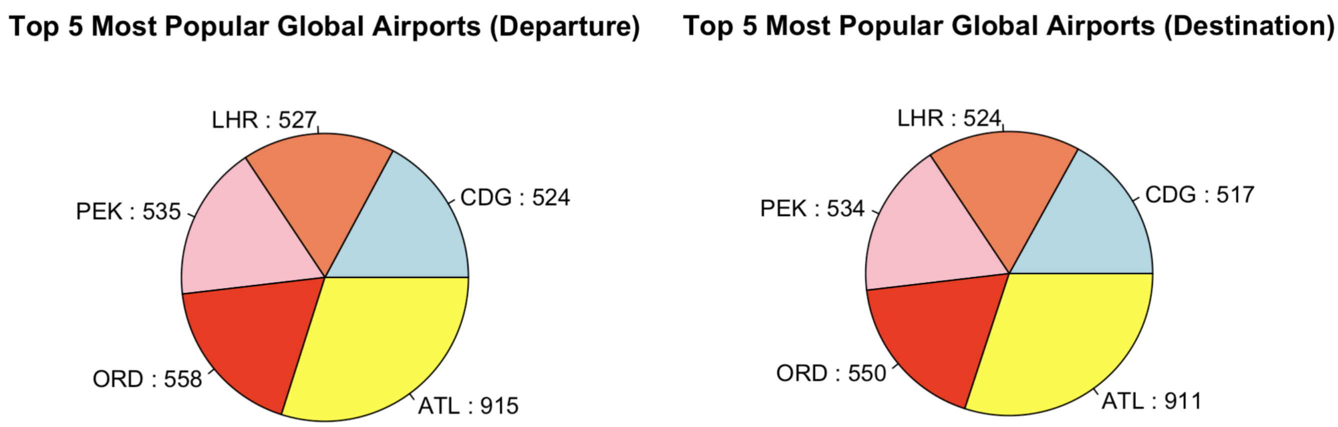
### *Exploratory Data Analysis (EDA)*

Within the "routes.csv" dataset, we have over 67,000 rows and 10 variables. For network analysis, we work with two variables, source and destination or source and target nodes. Before we dive deeper into network analysis and visualizations, we also conduct an exploratory data analysis (EDA) and compute some descriptive statistics that provides a comprehensive understanding of the data.

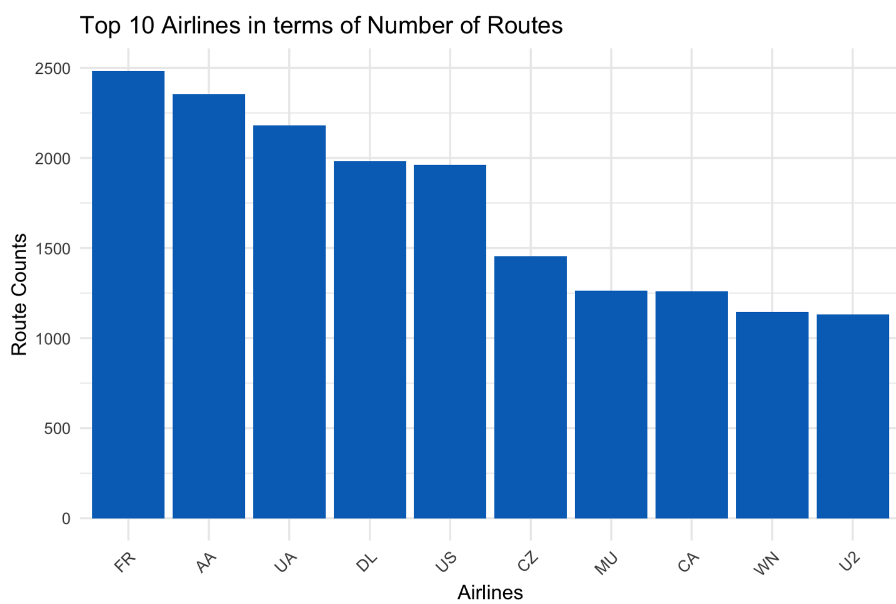
According to Graph 1, we acknowledge that Europe and North America are popular departures and destinations for travelers around the world. For Europe, Paris (CDG) and London (LHR) are popular international destinations for travelers. In North America, specifically the US, Atlanta (ATL) and Chicago (ORD) have the highest number of plane routes. Beijing is the only Asian country that is ranked within the Top 5.

---

[1] Woebkenberg , Tyler. "Global Air Transportation Network." Kaggle, 6 Dec. 2022, www.kaggle.com/datasets/thedevastator/global-air-transportation-network-mapping-the-wo/data?select=routes.csv.




*Graph 1: Top 5 Most Popular Global Airports (for both departure and destination airport)*




*Graph 2: Top 10 Airlines in relations to Number of Routes*

Based on the results from Graph 1, we are expected to have airlines that are based from these countries. According to Graph 2, we can see that US airlines such as American Airlines, United Airlines, and Delta Airlines are ranked 2nd, 3rd, and 4th places. RyanAir ranks the top place, which is an Irish airline that offers cheap flight tickets in Europe. China Southern Airlines (CZ) ranks 5th place. Overall, the origin of the airlines corresponds with the popularity of the airports.

**Methodology**

***Data Visualization using Gephi***

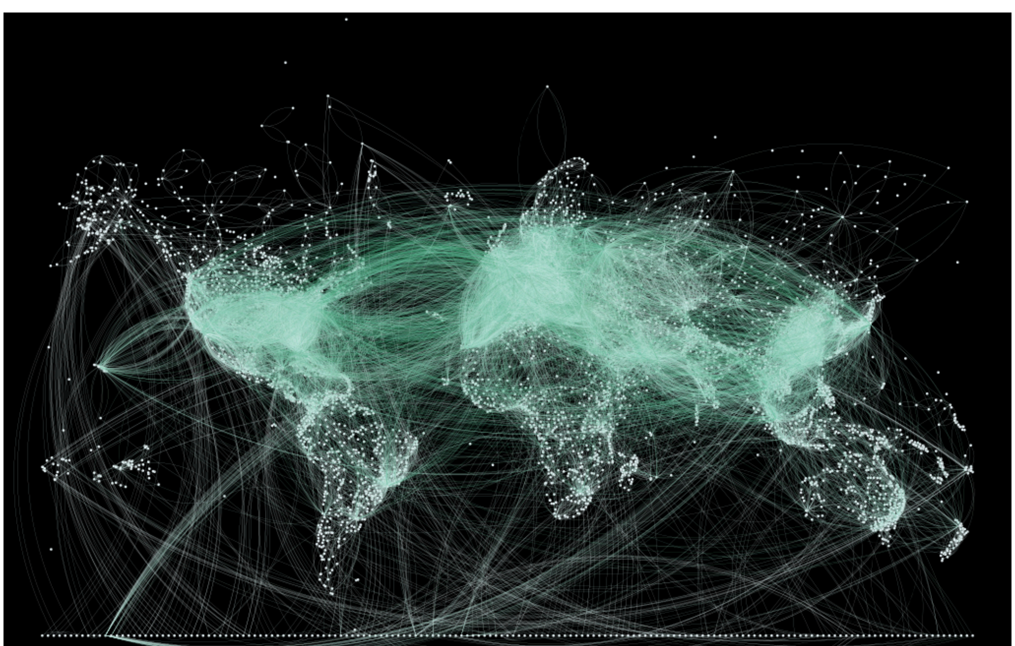
We used the Gephi platform for conducting our network analysis. This involved organizing our data into two distinct sheets: one for nodes and the other for edges.

1. Nodes Sheet: In this sheet, we recorded the following attributes:
    - ID: Used the International Air Transport Association (IATA) airport codes to uniquely identify airports.
    - Label: Full name of the airport, facilitating easier comprehension, particularly in cases where IATA codes are not familiar.
    - Latitude and Longitude: Geographic coordinates of the airports were used to accurately position the nodes within the network for geographic visualization.
2. Edges Sheet: This sheet contained the following information:
    - Source: Identifies the airports from which a flight is about to take off.
    - Target: Identifies the airport at which the flight is about to land.

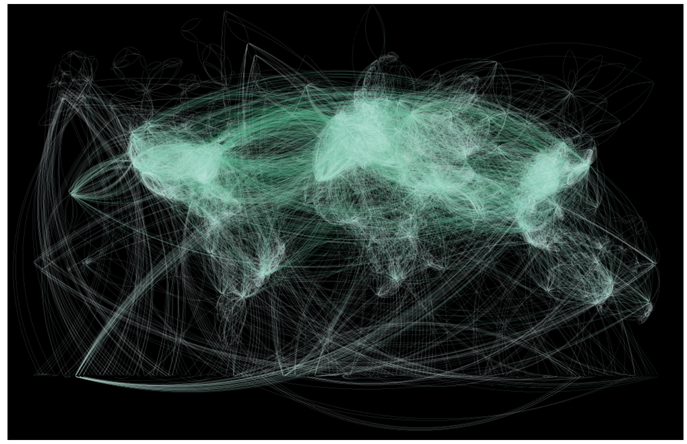
**Parameters & Procedures**

Type: We selected the directed relationship type to represent the one-way flow of flights from one airport to another.

Upon uploading the data onto the Gephi platform, we initiated the graphing process by navigating to the overview section. Since the data already had geographic coordinates for the nodes, we chose the Geo Layout algorithm and specified the longitude and latitude for accurate mapping of airports. Additionally, we selected the Mercator projection to represent the 3D surface of the earth in a 2D plane. Following this, we executed the Geo Layout algorithm, resulting in a visually appealing snapshot of the global airport network.



In the images below, the nodes and edges have been divided separately to provide a clearer visualization.

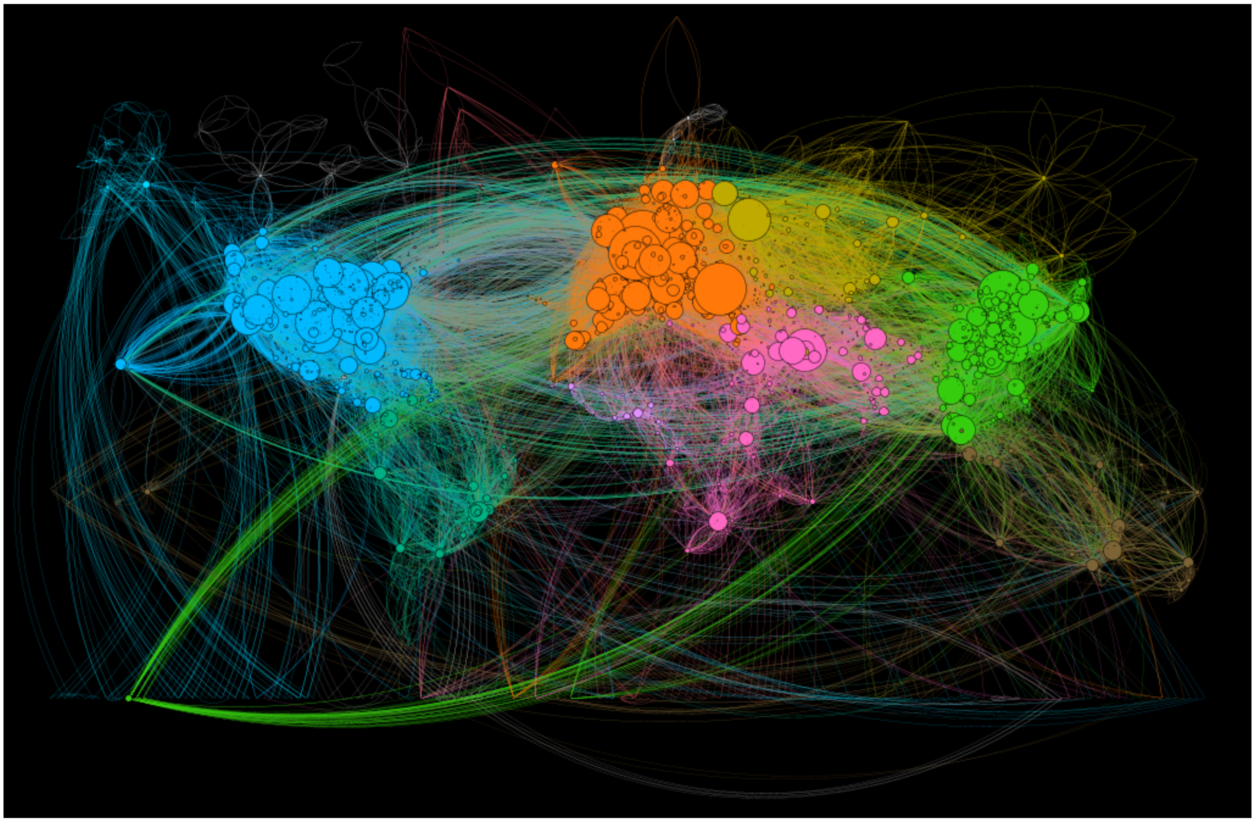


Next, we proceeded to the overview section. We focused on nodes and selected the ranking feature to better understand global airport connectivity. By selecting the "degree" ranking, we adjusted the size of nodes based on their degree centrality, with node sizes ranging from a minimum of 1 to maximum of 150. As a result, airports with higher degrees were represented with larger nodes, while airports with lower degrees were represented with smaller nodes.

To further explore the structure of the airport network, we applied the community detection modularity algorithm. We applied randomization to avoid bias and edge weights were enabled to give importance to the number of flights leaving from the origin airports and arriving at the destination airports. Then, we set the resolution parameter to 0.92 to properly align the cluster with the nature of the travel patterns around the world. After running the algorithms, the result showed that the modularity score was .658. This suggests that the network exhibits a relatively strong community structure.

The research revealed a network with a strong community structure. Therefore, by applying the modularity class to both nodes and edges, we were able to identify communities within our network. This method helped with the identification of seven major communities within the network, each represented with a unique color for clarity and distinction.

By implementing these steps, we were able to effectively analyze the global airport network, identifying key nodes and community structures essential for understanding air travel patterns and connectivity.



**Network Analysis**

**Diameter** : As the directed graph is not strongly connected and includes infinite path lengths, using python we found out diameter for the largest strongly connected component has a value of 14 which implies a more intricate,extended connectivity pattern.

**Density** : Density of the directed graph is 0.0032. This indicates that the graph is sparse and approximately 0.32% of all possible routes are realized in the airline network.

**Degree Centrality** : Frankfurt airport, with a degree centrality value of 0.1393, has the highest connectivity in the airline routes network. This implies that Frankfurt Airport serves as a major hub with direct connections to a substantial proportion of other airports.

**Closeness Centrality** : Frankfurt Airport with value of 0.3924 and Charles de Gaulle Airport,Paris with 0.3900 are the top two airports with the highest closeness centrality.This shows that these are more central airports and play crucial roles in facilitating efficient connectivity within the overall airport network. These airports serve as key hubs with shorter average travel distances to other airports.

**Betweenness Centrality** : Ted Stevens International Airport, Alaska has the highest betweenness centrality value of 0.070. This indicates it plays a critical role in connecting different routes within the airline network and serves as a key bridge or an intermediary, influencing efficient routing.

**Eigenvector Centrality :** Amsterdam airport, with the highest eigenvector centrality score of 0.1659, implies that it is not only well-connected but also connected to other influential airports, reinforcing its importance in the overall network structure.

**Hub and Authority scores :** Frankfurt Airport, with the highest hub score of 0.0080, holds a crucial position as a major hub in the airline network. This implies that Frankfurt Airport has a significant number of direct connections to other airports, indicating its importance in facilitating efficient and direct air travel routes.

Amsterdam Airport Schiphol,possessing the highest authority score of 0.0080, holds a position of influence as a crucial destination in the airline network. This implies that Amsterdam Airport Schiphol is frequently the target of incoming flights, highlighting its significance as a major arrival point.

**Clustering Coefficient** : The average clustering coefficient for the graph is 0.4692 which indicates a substantial level of local clustering or connectivity within the network.
There are 708 nodes in total which possess a maximum clustering coefficient value of 1.These nodes in the graph form tightly connected clusters and likely act as central points within their local communities.

**Reciprocity** : The graph has a reciprocity score of 0.9756 which indicates a high level of bidirectional connections within the directed graph.This bidirectional nature of routes can contribute to more efficient and flexible travel options for passengers, allowing for convenient round-trip flights.

**Business Implications**
Airports with a high degree of centrality, like FRA (Frankfurt), CDG (Charles de Gaulle), and AMS (Amsterdam Schiphol), are significant hubs in the global air network. These airports are crucial for planning as they represent central nodes where airlines can connect passengers to various destinations. Additionally, these hubs are likely to require more resources, including staffing, security, and infrastructure, to manage the higher volume of passengers and flights efficiently.

The closeness centrality of FRA shows its efficiency in the network, suggesting that it has more direct flights from other airports. This metric is helpful for planning and marketing, as airports with high closeness centrality can be promoted as ideal transfer points for passengers looking for the shortest travel times to their final destinations. ANC (Anchorage) has the highest betweenness centrality, indicating it serves as a bridge within the network, often appearing on the shortest path between other airports.

ATL (Atlanta), ORD (Chicago), and DFW (Dallas) have the most in-degree centrality in the United States which means they have the largest number of incoming flights in the United States. ATL and ORD also have the most out-degree centrality, meaning they have the largest number of outgoing flights in the United States. Airports like these can implement better security checkpoint technologies and processes, such as X-ray machines that don't require removing electronics from bags and more connectivity from check-in to departure gates to help streamline the departure experience.

Airports are critical for efficient network flow and may require additional resources to manage transit traffic. AMS, with the highest eigenvector centrality, is influential within the network, suggesting it is well-connected to other well-connected hubs. For airlines, this implies a strategic positioning for partnerships and code-sharing agreements. For the airport, this means prioritizing service quality and capacity to maintain its influential position. Airports like ANC and LAX (Los Angeles), which are important for connections, need to prioritize efficient turnaround times for aircraft and passenger transfers.

**Operational Considerations**

During peak periods like Christmas and summer holidays, airports need to prepare for extra traffic and congestion. Airports should make backup strategies to bring in more workers, optimize terminal operations, and ensure smooth passenger flow. This could include increasing operating hours, enhancing queueing systems, and adding crowd management measures. For high centrality airports like FRA, CDG, and AMS that are operating close to or at capacity, they can consider expanding the airport infrastructures like additional terminals, runways, and more ground transportation connections. Airlines could consider changing flight schedules, increasing frequency on high-demand routes, and allocating resources more efficiently.

ATL (Atlanta), ORD (Chicago), and DFW (Dallas) with high in-degree centrality, suggest that they are popular destinations. These airports can implement more advanced passenger processing systems, such as automatic gates and biometrics for identification to expedite the immigration process. Because of the high volume of incoming passengers, faster baggage handling systems and processes can significantly speed up the flow of passengers out of the airports since they don't have to wait for a long time for baggage claims. Better design of terminal layout that separates arrival and departure passengers, close proximity from arrival gates to immigration and baggage claim, clear and real time multi-language signage to direct passengers could also help optimizing the operations for high in-degree airports.

**Efficiency**
With an increased volume of travelers during peak seasons, airports and airlines also need to prioritize delivering high-quality service and ensuring a positive passenger experience. Airports can use mobile and digital platforms to provide passengers with real-time updates on flight status, gate changes, and other relevant information, which can reduce stress and improve the overall travel experience. Additionally, both airlines and airports can introduce passenger feedback systems that can help identify areas for improvement and address concerns promptly.

**Social Implications**
Airports with high centrality signal the connectivity of the region it serves and the importance of the city it is in the region itself. From the network analysis, airports with high centrality attributes are usually located in metropolis areas with substantial populations. Airports can act as the gateway to cultural exchange, economic opportunities, education and knowledge sharing, health care access, tourism resources growth, and even equity and inclusiveness. They allow workers to move easily between areas to better collaborate and tourists to experience different parts of the world. Patients can seek health care from places with more developed medical infrastructures and personnel. Special airports like ANC serve as hubs for global logistics where shipping companies can utilize geographic convenience to optimize the costs and efficiencies of their operations.

**Recommendations for Airports**
1. **Resource Optimization**: For high centrality airports, they can invest in advanced systems that can predict passenger flow and optimize security checkpoints, as well as gate allocations, to reduce congestion and improve efficiency.
2. **Infrastructure Investment**: These central hubs can consider expanding the airport infrastructures like additional terminals, runways, and more ground transportation connections.
3. **Strategic Positioning**: Airports with high closeness centrality should market themselves as prime transfer points in international travel, highlighting their accessibility and convenience for connecting flights.
4. **Transit-Focused Facilities**: Airports with high betweenness centrality can establish facilities and services for transit passengers or cargo, such as quicker security checks and buggies throughout the airports for passengers and off-shore cargo handling to streamline logistics.
5. **Inter-Airport Collaboration**: Airports with high eigenvector centrality can develop strategic partnerships with other hubs to enhance their network's efficiency.
6. **Sustainability Initiatives**: Adopt sustainable techniques to reduce the environmental impact of heavy traffic, which might also be used as a marketing tool for environmentally concerned passengers.

**Recommendations for Airlines**

1. **Network Planning**: Use centrality data to optimize route planning, focusing on high-centrality airports for hub operations and ensuring that aircraft schedules match passenger demand for connectivity.
2. **Code-Sharing and Alliances**: Develop code-sharing agreements and join global airline alliances, such as Oneworld, SkyTeam, and Star Alliance, to maximize the benefits of high centrality and eigenvector scores, allowing for a more extensive network with seamless passenger experiences.
3. **Dynamic Capacity Management**: Use predictive analytics to adjust capacity on routes connecting to high-centrality airports, ensuring that demand is met without over-saturating the market.
4. **Customer Experience**: Focus on enhancing the customer experience, particularly at high betweenness airports, by offering services such as expedited streamlined transfers, lounge access, and real-time flight information.
5. **Pricing Strategies**: Implement dynamic pricing strategies that reflect the value of efficient connectivity offered at high-centrality airports while also providing competitive fares on routes to and from these hubs.
6. **Marketing Campaigns**: Create targeted marketing campaigns that highlight the advantages of the airline's network, mainly the ease of connections at high centrality and high betweenness airports.

**Conclusion**

The dynamic nature of air travel serves as a unifying force, connecting continents and travelers worldwide. Research on airline routes and airports has offered valuable insights into the world of the global air traffic network. Through careful examination using network analysis, we have identified key airports and routes that have shaped the aviation industry. This insight has provided a deeper understanding of how airlines and airports can adapt and thrive within this complex network. By leveraging these insights, airlines and airports can make better strategic decisions to enhance operations and to provide superior service to travelers. The research empowers businesses to optimize resources to improve revenue, customer service, and make data-driven decisions.