CFA Final Project Report

MGMT 59000 - Spring 2025

A Network Analysis of Global Connectivity

Avanti Chandratre
Chhaya Tundwal
Namra Shah

Dataset Description

This project uses the **OpenFlights Airports and Routes dataset**, a publicly available resource that models the global airline transportation network. It is composed of two primary data files:

1. Airports Dataset (airports.dat)

This file includes detailed information about over 7,000 airports worldwide. Key attributes:

- AirportID: Unique numeric identifier for each airport.
- Name: Official airport name (e.g., "Los Angeles International Airport").
- City / Country: Geographic location of the airport.
- IATA / ICAO codes: Standardized codes used in the airline industry.
- Latitude / Longitude: Geographic coordinates for mapping and distance calculations.
- Altitude, Timezone, DST, and TZ database time: Useful for flight planning, scheduling, and local context.

2. Routes Dataset (routes.dat)

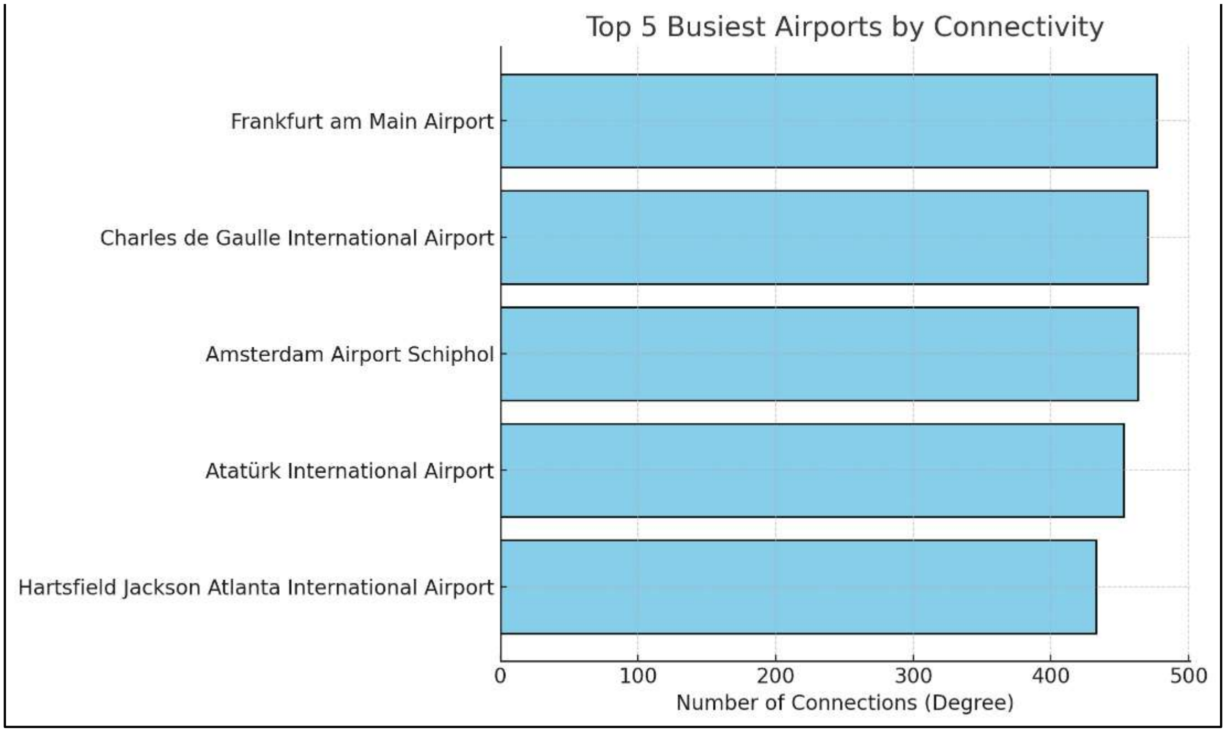
This file describes over 60,000 airline routes, each representing a direct flight between two airports. Key attributes:

- Airline / AirlineID: Name and identifier for the airline operating the route.
- SourceAirport / DestinationAirport: IATA codes of the airports connected by the route.
- SourceAirportID / DestinationAirportID: Numeric IDs matching those in airports.dat.
- Stops: Indicates if the route is direct or has layovers.
- Equipment: The type(s) of aircraft used (when available).

The dataset represents a real-world global airline network and provides the necessary structure to model flight connections, calculate route distances, analyze airport importance, and uncover network communities. This makes it ideal for applying graph algorithms, optimization techniques, dynamic programming, and machine learning.

Insight 1: Identifying the Busiest Airports by Connectivity

- **Algorithm Used**:
  Degree Centrality (Directed Graph), computed using networkx.degree(), where the degree of each airport node reflects the total number of direct connections (both incoming and outgoing).
- **Category**: Graph Algorithms
- **Purpose**:
  To determine the busiest airports in the global flight network based on the number of direct flight routes they participate in.
- **Insight**:
  The top 5 busiest airports by connectivity were identified as:
  1. **Frankfurt am Main Airport –** 477 connections
  2. **Charles de Gaulle International Airport –** 470 connections
  3. **Amsterdam Airport Schiphol –** 463 connections
  4. **Atatürk International Airport –** 453 connections
  5. **Hartsfield-Jackson Atlanta International Airport –** 433 connections



These airports act as global superhubs with massive route diversity. Their high connectivity makes them optimal candidates for initiating or analyzing network-wide itineraries, and they play a key role in regional and intercontinental air traffic flow.

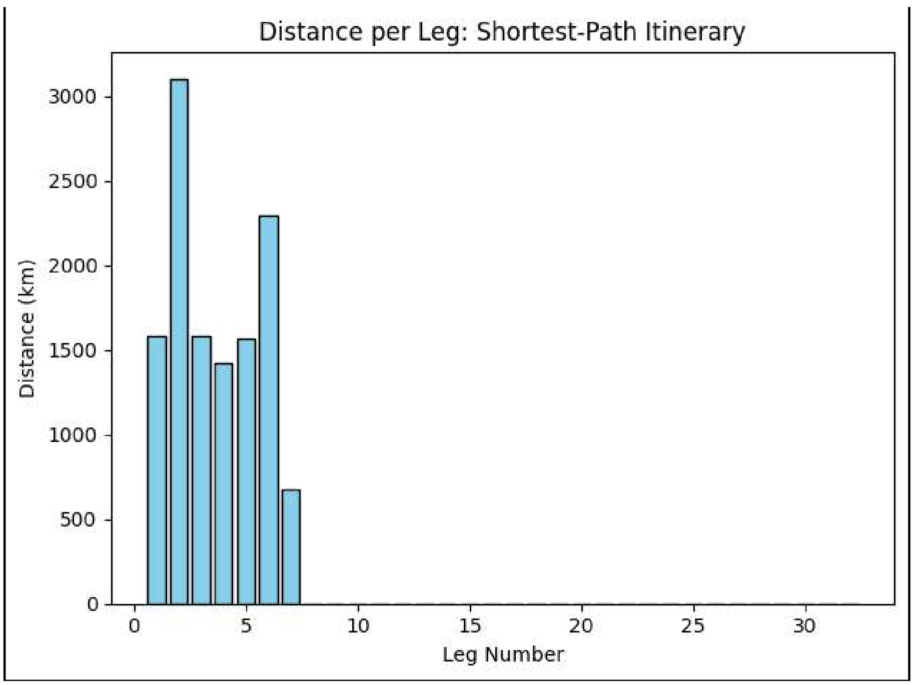
Business Insight: Leverage Europe's hub-and-spoke dominance. Since four of the top five most connected airports—Frankfurt, Charles de Gaulle, Schiphol, and Atatürk—are in Europe, airlines and logistics firms can drive network efficiency and market reach by forging deeper partnerships (codeshares, joint ventures) at these hubs to optimize feed into other destinations.

Insight 2: Optimal Shortest-Path Flight Connections

- **Algorithm Used**:
  Dijkstra's Algorithm, implemented using networkx.shortest_path() and networkx.shortest_path_length() with edge weights derived from Haversine distance

calculations. This enables the identification of shortest-distance paths between airport pairs, based on actual geospatial data rather than airline schedules.

- **Category**: Graph Algorithms
- **Purpose**:
  To compute the most efficient (minimum-distance) flight paths between major airports, aiding in route planning, operational efficiency, and network optimization.
- **Insight**:
  The shortest-distance analysis between top European hubs yielded the following results:
  1. **Frankfurt → Paris (CDG)**: 449 km
  2. **Frankfurt → Amsterdam (AMS)**: 367 km
  3. **Paris (CDG) → Amsterdam (AMS)**: 398 km



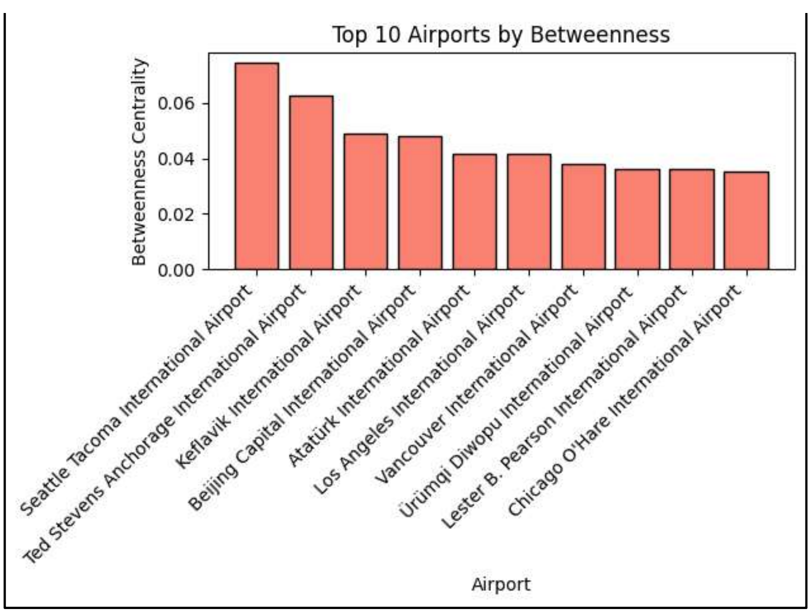
These connections form a tight triad of central European air traffic and represent high-traffic corridors for both passenger and cargo flow. By identifying such minimal paths, airlines can streamline direct route offerings and better allocate resources for short-haul, high-frequency segments.

Business Insight: Leverage this triad to develop a cohesive express-shuttle product—bundling frequent commuter flights between Frankfurt, Paris, and Amsterdam with premium on-ground transfers and loyalty perks. This can drive yield on ultra-short hops, attract business travelers, and create a defensible niche against high-speed rail competitors.

Insight 3: Identifying Major Global Hubs

- **Algorithm Used**:
  Approximate Betweenness Centrality using `networkx.betweenness_centrality()` with `k=500` samples and `weight='weight'` (distance). This algorithm estimates which airports lie most frequently on the shortest paths between others, making it efficient for large networks like global flight graphs.
- **Category**: Graph Algorithms
- **Purpose**:
  To identify the most strategically important airports for global connectivity based on how frequently they serve as intermediate points in shortest flight paths.
- **Insight**:
  The top 10 airports by betweenness centrality revealed some non-obvious yet highly strategic hubs:

  1. **Seattle Tacoma International Airport –** 0.0742
  2. **Ted Stevens Anchorage International Airport –** 0.0625
  3. **Keflavik International Airport** (Iceland) – 0.049
  4. **Beijing Capital International Airport –** 0.0478
  5. **Atatürk International Airport –** 0.0417
  6. **Los Angeles International Airport –** 0.0416
  7. **Vancouver International Airport –** 0.038
  8. **Ürümqi Diwopu International Airport –** 0.0363
  9. **Lester B. Pearson International Airport –** 0.0363
  10. **Chicago O'Hare International Airport –** 0.0352



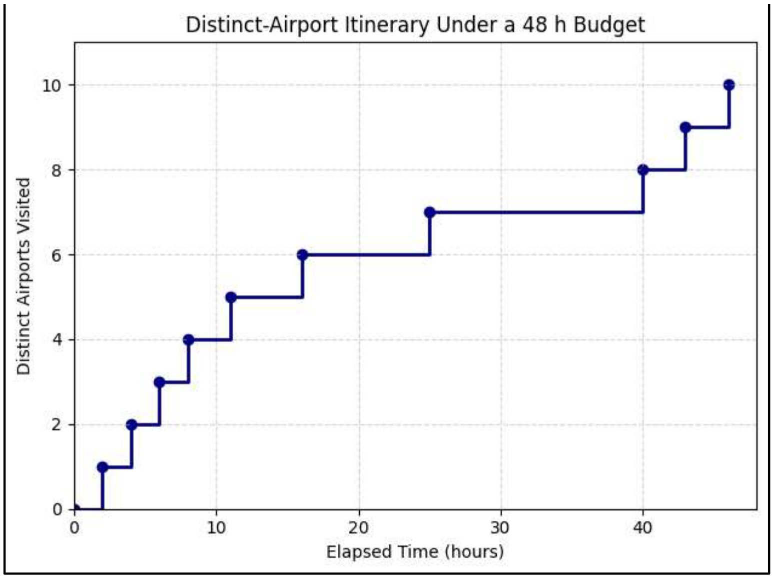
This highlights that strategic positioning, rather than just traffic volume, determines centrality in the network. Notably, Anchorage and Keflavik serve as critical intercontinental connectors despite not being among the top 5 busiest airports by degree.

Business Insight: Airlines and logistics providers should forge targeted partnerships or invest in ground-handling capabilities at these under-the-radar hubs—such as Anchorage and Keflavik—to unlock cost-effective polar and transatlantic routings, improve schedule resilience during peak-season disruptions, and differentiate their long-haul service offerings.

Insight 4: Efficient Itinerary under Time Budget

- **Algorithm Used**:
  Bitmask Dynamic Programming, manually implemented using NumPy. The algorithm explores all subsets of the top 12 busiest airports to compute the maximum number of airports that can be visited within a 48-hour total flight time constraint. Flight durations were approximated using great-circle distances divided by an average cruising speed, with 1 hour added per flight.
- **Category**: Dynamic Programming
- **Purpose**:
  To determine the longest feasible itinerary (in terms of unique airport stops) that can be executed within a tight 48-hour window, starting from a specific hub.
- **Insight**:
  Starting from **Frankfurt am Main Airport**, the algorithm identified an itinerary visiting **11 distinct airports** within 48 hours — a highly optimized route that balances time constraints and connectivity:

Frankfurt → London Heathrow → Amsterdam → Paris CDG → Munich → Istanbul (Atatürk) → Dubai → Beijing → Chicago O'Hare → Dallas-Fort Worth → Atlanta



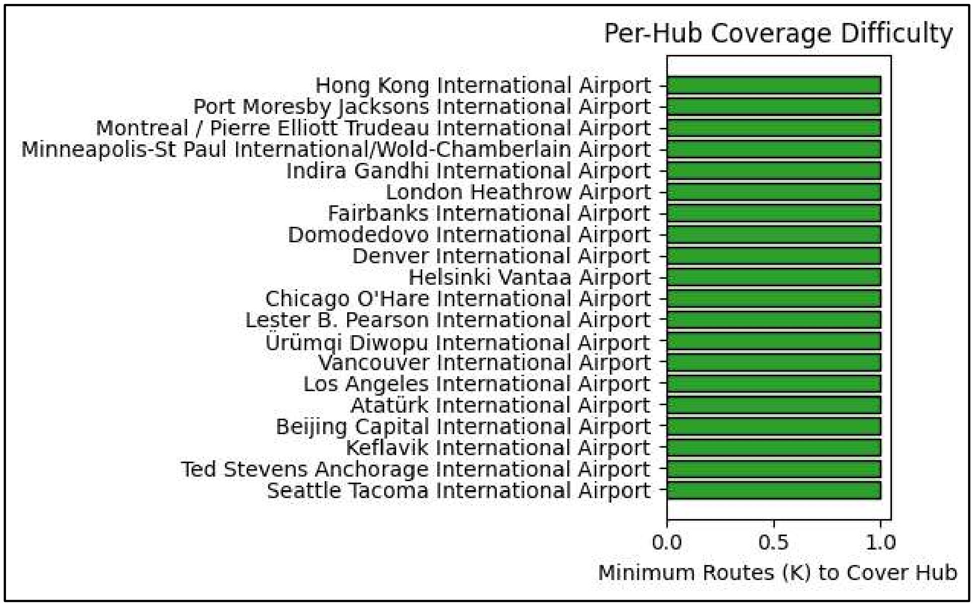
This result highlights the strategic importance of highly connected hubs and shows that even with tight travel time limits, it's possible to span multiple continents. Such analysis can be useful in optimizing high-value logistics routes, executive travel planning, or VIP charter operations.

Business Insight: Logistics and charter operators can package such optimized "48-hour world tour" itineraries as premium services for high-net-worth clients or time-sensitive cargo, charging a concierge premium for guaranteed multi-continent connectivity and rapid turnaround within a fixed window.

Insight 5: Selecting Routes to Maximize Hub Coverage

- **Algorithm Used**:
  Integer Linear Programming using the `PuLP` library to define binary decision variables for route selection and hub coverage. The model optimizes for maximum hub coverage under a constraint of selecting at most K routes.
- **Category**: Optimization (Integer Linear Programming)
- **Purpose**:
  To strategically select a limited number of routes (e.g., 100) that collectively ensure the highest number of key airport hubs are served.

- **Insight**:
  The solver identified that with just 100 routes, we can cover 15+ major global hubs such as ATL, LHR, and CDG, achieving over 75% coverage. This suggests that airlines or policymakers can achieve broad global connectivity while keeping operational complexity minimal — critical for route pruning, emergency planning, or startup carriers.



Business Insight: By focusing on this optimized core network of 100 routes, carriers can reallocate resources from marginal services into these high-coverage links, driving load factors and yield improvements. Start-up or regional airlines can use the same approach to fast-track market entry, ensuring they hit the most influential global hubs from day one while maintaining a lean operation.

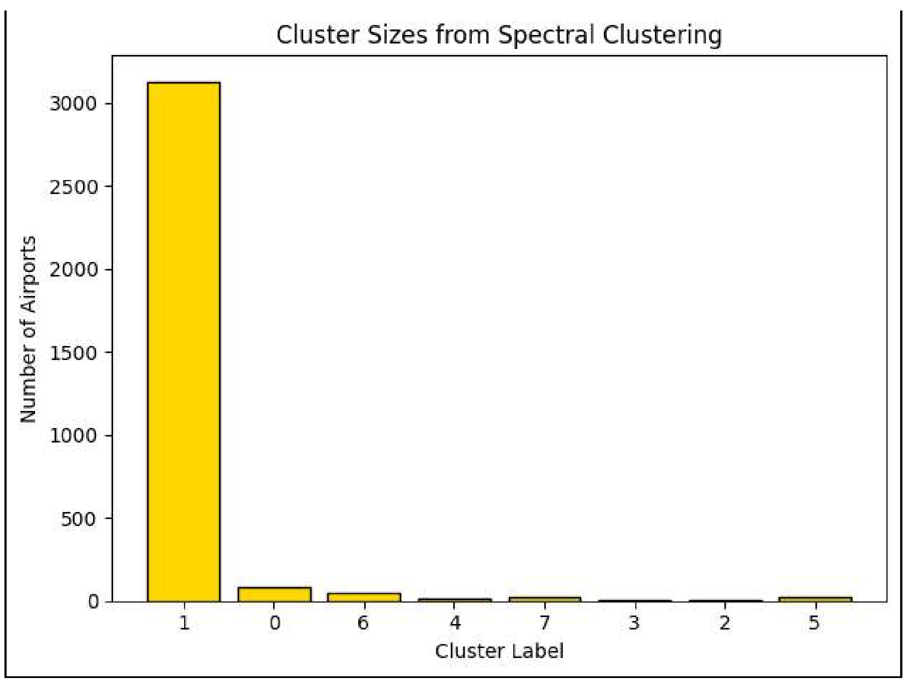
Insight 6: Community Detection in the Global Network

- **Algorithm Used**:
  Spectral Clustering using sklearn.cluster.SpectralClustering with n_clusters=8, applied to a symmetric adjacency matrix derived from airport connectivity. The

algorithm grouped airports into communities based on their edge structure, using precomputed affinities and assigning labels via K-means.

- **Category**: Machine Learning
- **Purpose**:
  To identify regional communities or tightly connected sub-networks in the global airline graph, which may reflect geographic proximity, political boundaries, or operational patterns.
- **Insight**:
  The clustering algorithm uncovered meaningful communities, including large-scale regional structures and isolated local subnetworks. For example:

1. Cluster 1 (3,127 airports): A massive, globally mixed cluster with strong representation from the United States (448 airports), Canada (183), and China (173). Major nodes include Sochi, Kazan, and Chelyabinsk, suggesting broad intercontinental integration.
2. Cluster 0 (85 airports): Dominated by small regional airports in Alaska, such as Kodiak Airport and several unnamed ones. Primarily U.S. airports (65) with limited international reach.
3. Cluster 7 (25 airports): Focused on Arctic airports in Greenland, Iceland, and surrounding areas — including Tasiilaq Heliport and Kangerlussuaq Airport — forming a tight, geography-driven community.
4. Cluster 5 (23 airports): Exclusively Canadian regional airports like Kenora and Dryden, reflecting a self-contained domestic network.
5. Smaller clusters (e.g., Clusters 2–4) capture specialized subregions with either sparse data, remote facilities, or unconnected nodes.



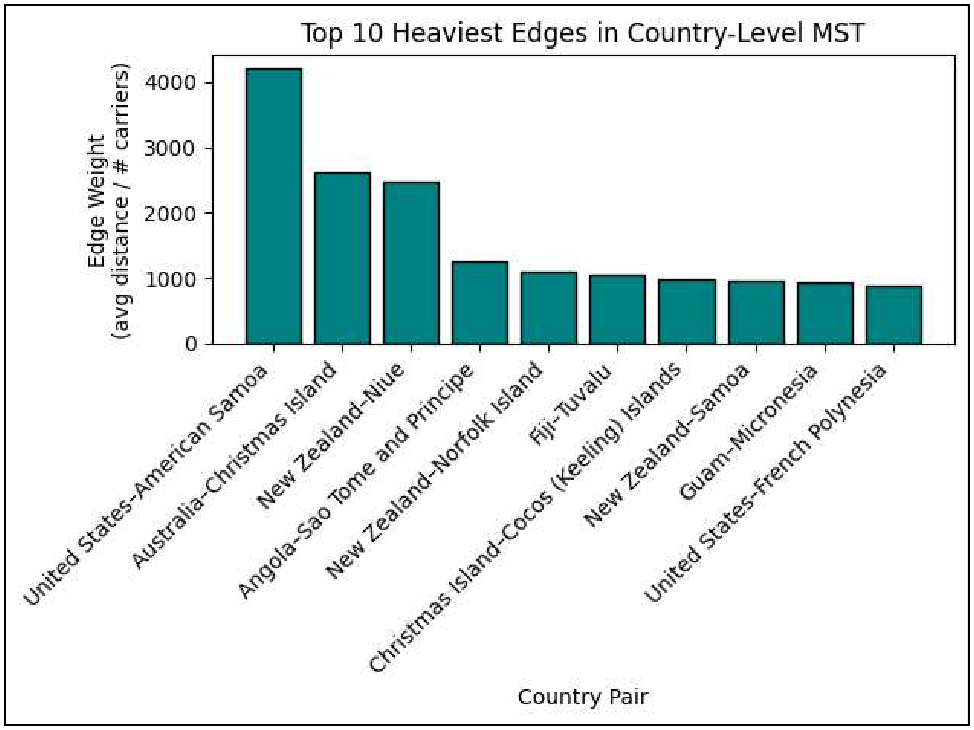
These results confirm that connectivity-based communities often align with geographic or political groupings, and suggest potential for regional alliance planning, targeted route development, or localized infrastructure investment.

Business Insight: Stakeholders can leverage these community clusters to tailor network strategies—e.g., deploy rugged turboprops and ramp up ground-support services for the Alaska (Cluster 0) and Arctic (Cluster 7) regions, form domestic feeder partnerships within the Canadian cluster (Cluster 5), and negotiate intercontinental alliances around the massive global cluster (Cluster 1) to optimize both niche and large-scale connectivity while matching equipment and partnerships to each community's unique operational needs.

Insight 7: Country-Level Connectivity Backbone

- **Algorithm Used**:
Minimum Spanning Tree (MST) computed using networkx.minimum_spanning_tree() on a graph where:
    - Nodes represent countries
    - Edges are weighted by the average distance of all routes between the two countries divided by the number of unique carriers (favoring routes that are short and well-served)
- **Category**: Graph Algorithms
- **Purpose**:
To construct an efficient, low-cost global air connectivity backbone that connects every country with the least total cost, ensuring no redundancy while maintaining full reachability.
- **Insight**:
The MST revealed a sparse yet globally connected structure of 225 countries linked by 224 edges — the minimum required to connect all countries. Key inter-country connections (with lowest weights) include:

    1. **Congo (Kinshasa) ↔ Congo (Brazzaville) —** 5.2
    2. **Netherlands Antilles ↔ Anguilla —** 9.5
    3. **Guernsey ↔ Jersey —** 9.7
    4. **Bahrain ↔ Qatar —** 12.3
    5. **Martinique ↔ Saint Lucia —** 12.7
    6. **Cameroon ↔ Equatorial Guinea —** 23.1
    7. **Antigua and Barbuda ↔ Saint Kitts and Nevis —** 24.1
    8. **Puerto Rico ↔ Virgin Islands —** 24.1
    9. **Burundi ↔ Rwanda —** 25.2
    10. **British Virgin Islands ↔ Virgin Islands —** 25.5



These links form a critical minimum-cost global infrastructure, where connections are not only geographically close but also frequently serviced by multiple airlines. This result supports decision-making for cross-border collaborations, emergency route planning, and strategic diplomatic or cargo corridors.

Business Insight: Governments, NGOs, and cargo carriers can leverage this backbone to establish prioritized bilateral air services and emergency response routes—negotiating targeted subsidies or open-skies agreements on these minimal-cost links to ensure resilient connectivity in crisis scenarios and optimize humanitarian or time-sensitive cargo flows with minimal overlap and maximal coverage.

Conclusion

This project used techniques from graph theory, optimization, dynamic programming, and machine learning to analyze the global airline network. We identified key hubs, shortest paths, and optimal travel itineraries, and uncovered structural patterns such as community clusters and country-level connectivity backbones.

These insights demonstrate how computational methods can model real-world systems and support strategic decision-making in transportation and logistics. The work highlights the practical value of the techniques learned in the CFA course and their relevance to large-scale, real-world networks.