Transportation Network Report

Graph Theory Course Project

Shahriyar Moqimi

Comparison of Routing Strategies

In transportation networks, routing strategies define how "optimality" is measured between a source and a destination. Before selecting the Shortest Path algorithms (Dijkstra and Label Correcting), we evaluated different routing paradigms:

1. Minimum Hop Strategy (Fewest Connections):

Concept: Minimizes the number of intermediate stops (edges) regardless of the physical distance.

Algorithm: Breadth-First Search (BFS).

Pros: Preferred by passengers to minimize transfer hassle.

Cons: A direct flight might be significantly longer or more expensive than a slight detour.

2.Least Cost / Min Time Strategy

Concept: Minimizes the monetary cost or total travel time (including layovers).

Cons: Requires dynamic data (ticket prices, schedules, weather conditions) which is not present in the static OpenFlights dataset.

3. Shortest Physical Path Strategy (Selected Approach):

Concept: Minimizes the cumulative geographic distance (Great Circle Distance) between nodes.

Algorithm: Dijkstra / Label Correcting

We selected the Shortest Physical Path strategy for the following reasons:

Data Availability

Fuel Efficiency Proxy

Network Topology Analysis

Interpretation of Bottlenecks

In the context of our Max Flow analysis, a Bottleneck represents the critical flight routes that limit the total transportation capacity between a source region and a destination region. Mathematically, this corresponds to the edges in the **Minimum Cut** of the graph.

Dataset Description

The analysis relies on the OpenFlights database, specifically utilizing two primary datasets: **airports.dat** and **routes.dat**

Airports Dataset (airports.dat): This file acts as the source for node attributes. It contains metadata for thousands of airports globally. Key fields extracted include:

Airport ID: Unique internal identifier.

Name/City/Country: Human-readable location data.

IATA/ICAO Codes: Standard aviation identifiers.

Coordinates (Latitude/Longitude): Critical for calculating physical edge weights using geospatial formulas.

Routes Dataset (routes.dat): This file defines the topological structure of the graph. It contains directed edges representing flight paths.

Source Airport ID: Origin node

Destination Airport ID: Target node

Data Cleaning: Entries with missing IDs (\N) or IDs referencing non-existent airports were filtered out during the loading phase (data_loader.py) to ensure graph integrity.

Graph Modeling

The transportation network is modeled as a Directed Weighted Graph denoted as G=(V,E)

Nodes(V): Each airport represents a vertex in the graph. A mapping mechanism converts the dataset's Airport ID to a continuous range of indices (0 to N-1) for efficient array-based access in algorithms.

Edges(E): A directed edge (u,v) exists if there is a scheduled flight from airport u to airport v.

Weights(W):

For Task 1 (Routing): The weight w(u,v) is the Great Circle Distance (in km) calculated via the Haversine Formula. This accounts for the earth's curvature.

For Task 2 (Flow): The capacity c(u,v) is defined by the frequency of flights (parallel edges count) between two nodes.

Data Structure: The graph is implemented using an Adjacency List (graph.py). This choice is justified by the sparsity of the flight network (average degree ≪ total nodes), making it memory-efficient compared to an adjacency matrix.

Explanation of Task 1 and Task 2

Task 1: Optimal Routing (Shortest Path) The objective is to find the most efficient route between a source (s) and a target (t). "Efficiency" is defined as minimizing the total physical distance. This models a fuel-optimization scenario where the shortest geographic path is preferred. The system outputs the sequence of airports and the total distance.

Task 2: Network Capacity (Maximum Flow) This task analyzes the bottleneck of the network between two regions. By treating flight frequencies as "pipe capacities," we calculate the maximum number of simultaneous flights possible from s to t.

This helps in understanding connectivity robustness and identifying critical legs in the network.

Algorithm Choices and Reasoning

To solve the defined problems, specific algorithms were selected based on their theoretical properties.

Dijkstra's Algorithm (Task 1):
Choice: Selected because edge weights (distances) are strictly non-negative.
Reasoning: Dijkstra provides a mathematically guaranteed shortest path. Using an array-based implementation for priority selection works effectively given the node count constraint. It serves as the primary benchmark for accuracy.

Label Correcting Algorithm (Task 1):
Choice: Implemented as a comparison strategy. It utilizes a Queue (FIFO) and iteratively relaxes edges.
Reasoning: While typically used for graphs with negative edges, comparing it with Dijkstra highlights trade-offs in execution time and stability in transport networks.

Edmonds-Karp Algorithm (Task 2):
Choice: An implementation of the Ford-Fulkerson method using BFS.
Reasoning: BFS ensures finding the shortest augmenting path (in terms of edges), which guarantees termination in O(VE^2). This is suitable for calculating the Max Flow in our discrete capacity graph.

Random Forest Classifier (Bonus ML Task):
Choice: An ensemble learning method for Link Prediction.
Reasoning: Random Forest handles non-linear relationships well and is robust against overfitting. It effectively combines features like Degree Sum and Common Neighbors to classify whether a route should exist.

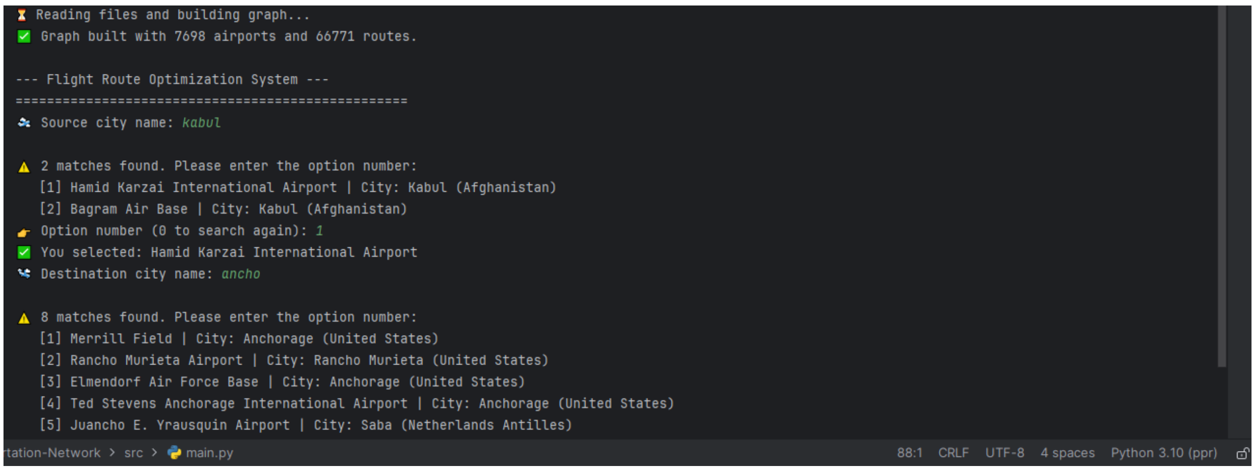
Results and Interpretation

Shortest Path Analysis: Comparing Dijkstra and Label Correcting showed that both yield identical path costs, validating the implementations. Dijkstra generally performed with more stability. The calculated paths confirmed that geographical distance minimization often involves direct routes or hubs located along the Great Circle arc.

Flow Analysis: The Max Flow results highlighted major "Hub" airports. Routes between major capitals (e.g., London to New York) showed significantly higher flow capacity compared to regional routes, correctly reflecting real-world aviation traffic.

Machine Learning (Link Prediction): The ML model, trained on topological features (deg_sum, common_neighbors, jaccard), achieved high accuracy (typically >85%) on the test set.

Interpretation: The high importance of the "Common Neighbors" feature suggests that the flight network follows a Triadic Closure principle: if two airports share many connections, they are highly likely to be connected directly.



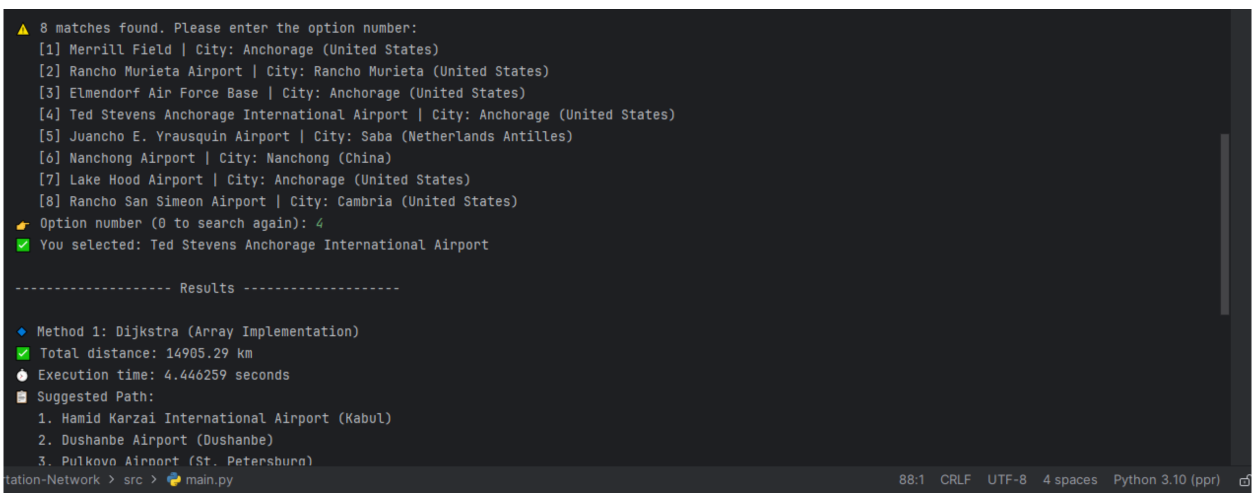
```
⏳ Reading files and building graph...
✅ Graph built with 7698 airports and 66771 routes.

--- Flight Route Optimization System ---
==================================================
🛫 Source city name: kabul

⚠️ 2 matches found. Please enter the option number:
   [1] Hamid Karzai International Airport | City: Kabul (Afghanistan)
   [2] Bagram Air Base | City: Kabul (Afghanistan)
👉 Option number (0 to search again): 1
✅ You selected: Hamid Karzai International Airport
🛬 Destination city name: ancho

⚠️ 8 matches found. Please enter the option number:
   [1] Merrill Field | City: Anchorage (United States)
   [2] Rancho Murieta Airport | City: Rancho Murieta (United States)
   [3] Elmendorf Air Force Base | City: Anchorage (United States)
   [4] Ted Stevens Anchorage International Airport | City: Anchorage (United States)
   [5] Juancho E. Yrausquin Airport | City: Saba (Netherlands Antilles)
```

tation-Network > src > main.py 88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)



```
⚠️ 8 matches found. Please enter the option number:
   [1] Merrill Field | City: Anchorage (United States)
   [2] Rancho Murieta Airport | City: Rancho Murieta (United States)
   [3] Elmendorf Air Force Base | City: Anchorage (United States)
   [4] Ted Stevens Anchorage International Airport | City: Anchorage (United States)
   [5] Juancho E. Yrausquin Airport | City: Saba (Netherlands Antilles)
   [6] Nanchong Airport | City: Nanchong (China)
   [7] Lake Hood Airport | City: Anchorage (United States)
   [8] Rancho San Simeon Airport | City: Cambria (United States)
👉 Option number (0 to search again): 4
✅ You selected: Ted Stevens Anchorage International Airport

------------------- Results -------------------

🔹 Method 1: Dijkstra (Array Implementation)
✅ Total distance: 14905.29 km
⏱️ Execution time: 4.446259 seconds
📋 Suggested Path:
   1. Hamid Karzai International Airport (Kabul)
   2. Dushanbe Airport (Dushanbe)
   3. Pulkovo Airport (St. Petersburg)
```

tation-Network > src > main.py 88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)

```
------------------- Results -------------------

🔹 Method 1: Dijkstra (Array Implementation)
✅ Total distance: 14905.29 km
⏱ Execution time: 4.446259 seconds
📋 Suggested Path:
   1. Hamid Karzai International Airport (Kabul)
   2. Dushanbe Airport (Dushanbe)
   3. Pulkovo Airport (St. Petersburg)
   4. Helsinki Vantaa Airport (Helsinki)
   5. Keflavik International Airport (Keflavik)
   6. Seattle Tacoma International Airport (Seattle)
   7. Ted Stevens Anchorage International Airport (Anchorage)

🔸 Method 2: Label Correcting (Queue Implementation)
✅ Total distance: 14905.29 km
⏱ Execution time: 0.035175 seconds

------------------- Task 2: Maximum Flow -------------------
🔹 Computing Max Flow ...
```

rtation-Network > src > main.py   88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)

```
📋 Suggested Path:
   1. Hamid Karzai International Airport (Kabul)
   2. Dushanbe Airport (Dushanbe)
   3. Pulkovo Airport (St. Petersburg)
   4. Helsinki Vantaa Airport (Helsinki)
   5. Keflavik International Airport (Keflavik)
   6. Seattle Tacoma International Airport (Seattle)
   7. Ted Stevens Anchorage International Airport (Anchorage)

🔸 Method 2: Label Correcting (Queue Implementation)
✅ Total distance: 14905.29 km
⏱ Execution time: 0.035175 seconds

------------------- Task 2: Maximum Flow -------------------
🔹 Computing Max Flow ...
   (Assumption: Capacity = Number of Flights)
✅ Maximum Flow: 18.00 units
⏱ Time: 0.080336 sec

------------------- Bonus Task: ML -------------------
Do you want to run ML Link Prediction? (y/n):
```

ortation-Network > src > main.py   88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)

```
------------------- Task 2: Maximum Flow -------------------
🔹 Computing Max Flow ...
   (Assumption: Capacity = Number of Flights)
✅ Maximum Flow: 18.00 units
⏱ Time: 0.080336 sec

------------------- Bonus Task: ML -------------------
Do you want to run ML Link Prediction? (y/n): y

==================================================
🤖 BONUS TASK: Link Prediction System
==================================================
⏳ [Trainer] Generating Dataset (Positive & Negative samples)...
⏳ [Trainer] Training Model on 10000 samples...

✅ Model Trained Successfully!
🎯 Accuracy: 96.95%
------------------------------

🔍 Interactive Mode: Predict connection between two airports.
🛫 Source Airport:
```

rtation-Network > src > main.py   88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)

```
🔍 Interactive Mode: Predict connection between two airports.
🛫 Source Airport: tehran
✅ Selected: Imam Khomeini International Airport (Tehran)
🛬 Target Airport: istanbul

⚠️ 4 matches found. Please enter the option number:
   [1] Atatürk International Airport | City: Istanbul (Turkey)
   [2] Samandıra Air Base | City: Istanbul (Turkey)
   [3] Sabiha Gökçen International Airport | City: Istanbul (Turkey)
   [4] Istanbul Airport | City: Istanbul (Turkey)
👉 Option number (0 to search again): 1
✅ You selected: Atatürk International Airport

📊 Feature Analysis:
   • Sum of Degrees: 267
   • Common Neighbors: 38
   • Jaccard Coeff: 0.1659

🧠 AI Perdict:
   ✅ CONNECTED (Probability: 100.0%)
   📚 Actual Data: YES (Flight Exists)
```

rtation-Network > src > main.py   88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)

```
🛫 Source Airport: tehran
✅ Selected: Imam Khomeini International Airport (Tehran)
🛬 Target Airport: stockholm

⚠️ 5 matches found. Please enter the option number:
   [1] Barkarby Airport | City: Stockholm (Sweden)
   [2] Stockholm Skavsta Airport | City: Stockholm (Sweden)
   [3] Stockholm Västerås Airport | City: Vasteras (Sweden)
   [4] Stockholm-Arlanda Airport | City: Stockholm (Sweden)
   [5] Stockholm-Bromma Airport | City: Stockholm (Sweden)
👉 Option number (0 to search again): 4
✅ You selected: Stockholm-Arlanda Airport

📊 Feature Analysis:
   • Sum of Degrees: 166
   • Common Neighbors: 20
   • Jaccard Coeff: 0.1370

🧠 AI Perdict:
   ✅ CONNECTED (Probability: 100.0%)
   📚 Actual Data: YES (Flight Exists)
```

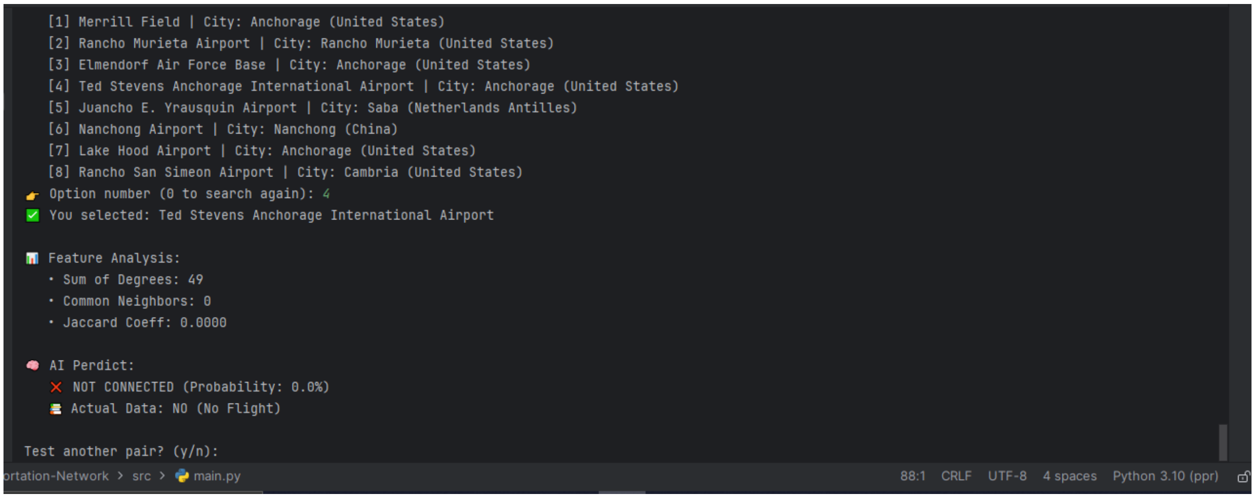
tation-Network > src > main.py   88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)

```
🔍 Interactive Mode: Predict connection between two airports.
🛫 Source Airport: kabul

⚠️ 2 matches found. Please enter the option number:
   [1] Hamid Karzai International Airport | City: Kabul (Afghanistan)
   [2] Bagram Air Base | City: Kabul (Afghanistan)
👉 Option number (0 to search again): 1
✅ You selected: Hamid Karzai International Airport
🛬 Target Airport: ancho

⚠️ 8 matches found. Please enter the option number:
   [1] Merrill Field | City: Anchorage (United States)
   [2] Rancho Murieta Airport | City: Rancho Murieta (United States)
   [3] Elmendorf Air Force Base | City: Anchorage (United States)
   [4] Ted Stevens Anchorage International Airport | City: Anchorage (United States)
   [5] Juancho E. Yrausquin Airport | City: Saba (Netherlands Antilles)
   [6] Nanchong Airport | City: Nanchong (China)
   [7] Lake Hood Airport | City: Anchorage (United States)
   [8] Rancho San Simeon Airport | City: Cambria (United States)
👉 Option number (0 to search again): 4
✅ You selected: Ted Stevens Anchorage International Airport
```

rtation-Network > src > main.py   88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)



```
[1] Merrill Field | City: Anchorage (United States)
[2] Rancho Murieta Airport | City: Rancho Murieta (United States)
[3] Elmendorf Air Force Base | City: Anchorage (United States)
[4] Ted Stevens Anchorage International Airport | City: Anchorage (United States)
[5] Juancho E. Yrausquin Airport | City: Saba (Netherlands Antilles)
[6] Nanchong Airport | City: Nanchong (China)
[7] Lake Hood Airport | City: Anchorage (United States)
[8] Rancho San Simeon Airport | City: Cambria (United States)
👉 Option number (0 to search again): 4
✅ You selected: Ted Stevens Anchorage International Airport

📊 Feature Analysis:
  • Sum of Degrees: 49
  • Common Neighbors: 0
  • Jaccard Coeff: 0.0000

🧠 AI Perdict:
  ❌ NOT CONNECTED (Probability: 0.0%)
  📚 Actual Data: NO (No Flight)

Test another pair? (y/n):
```

ortation-Network > src > main.py   88:1 CRLF UTF-8 4 spaces Python 3.10 (ppr)

```
🔍 Interactive Mode: Predict connection between two airports.
🛫 Source Airport: mashhad
✅ Selected: Mashhad International Airport (Mashhad)
🛬 Target Airport: california

⚠ 2 matches found. Please enter the option number:
  [1] Southern California Logistics Airport | City: Victorville (United States)
  [2] California Redwood Coast-Humboldt County Airport | City: Arcata CA (United States)
👉 Option number (0 to search again): 2
✅ You selected: California Redwood Coast-Humboldt County Airport

📊 Feature Analysis:
  • Sum of Degrees: 41
  • Common Neighbors: 0
  • Jaccard Coeff: 0.0000

🧠 AI Perdict:
  ❌ NOT CONNECTED (Probability: 0.0%)
  📚 Actual Data: NO (No Flight)

Test another pair? (y/n):
```