Project Unit One
Subject: Social Networks Analysis

Student:
Luis Fernando Monterrubio Cota
Teacher:
Didier Gamboa

Abstract

In this study, we examined the USAir97 network, which maps out the U.S. air transportation system, to better understand its structure and key components. By applying various centrality measures such as Degree, Eigenvector, Katz, PageRank, Betweenness, and Closeness Centralities, we were able to identify the most influential airports in the network. Betweenness Centrality was especially effective in identifying crucial hubs that handle the majority of connections, thus pinpointing potential vulnerabilities and areas of resilience within the network.[1] We also analyzed the degree distribution of the network and visualized it using a histogram, scatter plot, and cumulative degree distribution. This revealed a scale-free distribution, indicating that a few airports have significantly higher connectivity compared to the rest. Our findings suggest that the USAir97 network has a complex yet well-organized structure, with key airports playing vital roles in maintaining overall connectivity. The identified communities are essential for strategic planning and enhancing the robustness of the air transportation system. [?].

June 25, 2024

Networks Theory

I. INTRODUCTION

The USAir97 network maps out the U.S. air transportation system, with airports as nodes and flights as edges connecting them. This network is crucial as it illustrates the structure and dynamics of air travel across the U.S., showcasing key hubs and connections that enable the movement of people and goods.[4] By analyzing this network, we can gain valuable insights into optimizing routes, improving resilience against disruptions, and enhancing overall efficiency.

Researchers have utilized this dataset in numerous studies to explore complex network properties such as robustness, centrality measures, and community detection. These studies typically find that the air transportation network is highly interconnected, with a few airports serving as critical hubs that maintain global connectivity. Such findings highlight the importance of these key hubs and the potential impact of their failure on the entire network.

The USAir97 network is classified as a unidirected, non-planar network. It is unidirected because the edges represent direct flights between airports, meaning connections go both ways. It is non-planar due to the high number of connections, making it impossible to draw the network on a plane without the edges crossing.
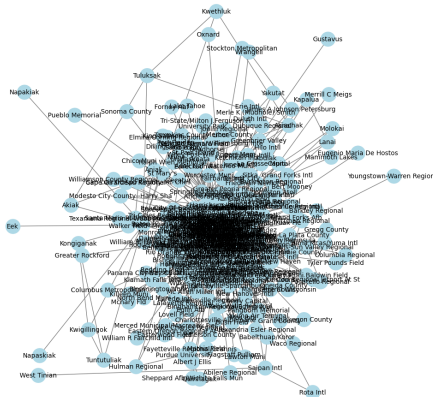

Fig. 1. Network

II. NETWORK CHARACTERISTICS

The analysis of the USAir97 network reveals several key characteristics that highlight the structure and dynamics of the U.S. air transportation system. Here, we present the measurements and interpret their significance based on graph theory concepts.

```
1  Size of the Network: 332
2  Number of Links: 2126
3  Average Path Length: 2.74
4  Clustering Coefficient: 0.63
5  Average Distance: 2.74
6  Diameter: 6
7  Radius: 3
8  Eccentricity: {'Wiley Post−Will Rogers Mem':
      5, 'Deadhorse': 5, 'Ralph Wien Memorial':
      5}
9  Periphery: ['Tuluksak', 'Akiachak', 'Akiak']
10 Center: ['Honolulu Intl']
```

- **Size of the Network**: With 332 airports, the network is extensive, indicating a large number of destinations within the U.S.
- **Number of Links**: There are 2126 direct flights, showing a high level of connectivity which ensures multiple routes between airports, enhancing the network's robustness.
- **Average Path Length**: At 2.74, this short average path length indicates that, on average, an airport can be reached from any other airport in under three flights. This reflects a small-world property, ensuring efficient travel across the network.
- **Clustering Coefficient**: The value of 0.63 signifies a highly clustered network. Airports tend to form tightly-knit groups, which means local connectivity is strong, promoting regional travel efficiency.
- **Average Distance**: Similar to the average path length, this reinforces the idea of a well-connected network where distances between nodes are generally short.
- **Diameter**: The diameter of 6 means the longest shortest path between any two airports is just six flights, emphasizing the network's compactness.
- **Radius**: The radius of 3 indicates the minimum eccentricity of any node, reflecting that some airports are very central, with short paths to all other nodes.
- **Eccentricity**: Values range from 3 to 6. Airports with lower eccentricity (e.g., 'Honolulu' with 3) are more central, meaning they are closer to all other airports. In contrast, peripheral airports have higher eccentricity values, indicating they are farther from the network's center.
- **Periphery**: Airports like 'Tuluksak' and 'Akiachak' with the highest eccentricity values (6) are on the network's periphery, meaning they have the longest shortest paths to other airports.
- **Center**: Central nodes such as 'Honolulu' are crucial for network connectivity, as they have the shortest maximum distance to all other nodes,

making them key hubs for efficient travel.

III. CENTRALITY MEASURES

In this section, we analyze the USAir97 network using various centrality measures to identify the most influential airports. The centrality measures provide insights into different aspects of network importance, such as connectivity, influence, and control over information flow. Here, we focus on Degree Centrality, Eigenvector Centrality, Katz Centrality, PageRank, Betweenness Centrality, and Closeness Centrality.[2]

Chicago O'Hare Intl, Dallas/Fort Worth Intl, and The William B Hartsfield Atlan consistently appear at the top of most centrality measures, highlighting their overall importance in the network. Chicago O'Hare Intl, in particular, stands out as the most influential airport, ranking highest in Degree, Eigenvector, PageRank, Betweenness, and Closeness Centralities.

These key hubs are critical for the network's robustness and efficiency. Disruptions at these airports could have significant impacts on the overall connectivity and functionality of the air transportation system. Therefore, maintaining the operational stability of these major hubs is essential for the resilience of the network.

IV. DEGREE DISTRIBUTION

The degree distribution provides insights into the connectivity of nodes within the network. In the context of the USAir97 network, it helps us understand how flight connections are distributed among airports.
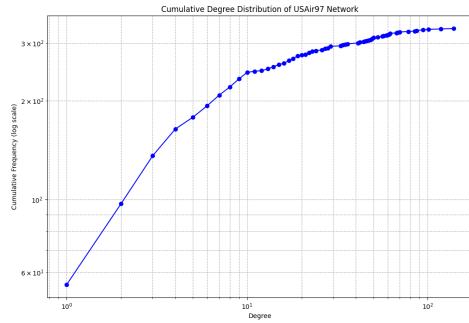

Fig. 2. Cumulative degree distribution

The cumulative degree distribution plot indicates that the USAir97 network has a small number of highly connected nodes and a large number of sparsely connected nodes. This pattern is consistent with many real-world networks, where connectivity is unevenly distributed. This distribution highlights the importance of key airports in maintaining network connectivity and the potential impact of their failure on the overall network structure.

V. COMMUNITY DETECTION

The community detection analysis helps us understand the structure of the USAir97 network by identifying groups of airports that are more densely connected to each other than to the rest of the network. This information can be used to improve the efficiency and robustness of the air transportation system.

A. Modularity

Modularity is a measure of the strength of division of a network into communities. High modularity indicates dense connections within communities but sparse connections between them.[3]
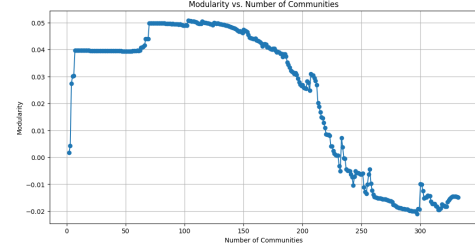

Fig. 3. Modularity of the network

The highest modularity value (approximately 0.05) indicates the optimal number of communities. In this case, the optimal number of communities is around 103.

As the number of communities increases beyond the optimal point, modularity decreases, indicating that the community structure becomes less meaningful with too many small communities.

B. Community Detection Visualization

The visualization shows a clear core-periphery structure, with a dense central core of highly interconnected airports and a sparser periphery of less connected airports. Each color represents a different community, and the layout helps us see the clustering of airports within the same community.
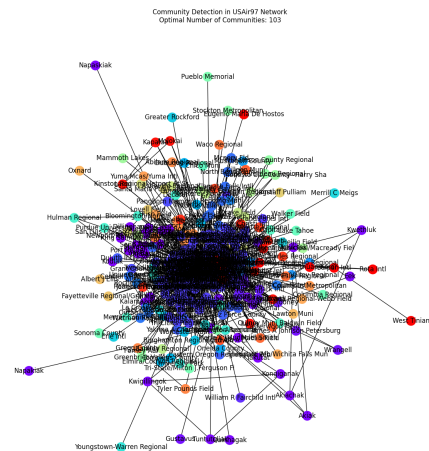

Fig. 4. USAir97 network with nodes colored

VI. CONCLUSION

Analyzing the USAir97 network sheds light on the U.S. air transportation system. Airports like Chicago O'Hare, Dallas/Fort Worth, and Atlanta's Hartsfield-Jackson are vital for keeping the network connected, as shown by their high centrality scores. If these hubs were disrupted, it would have a major impact on the entire network.

The network's degree distribution suggests it's resilient to random failures but vulnerable to targeted attacks on major hubs. Community detection identified 103 optimal communities, revealing a core-periphery structure. Enhancing connections in less connected communities can boost the network's overall efficiency and resilience.

These insights are crucial for strategic planning, emphasizing the need to focus on key hubs, improve connectivity, and prepare for potential disruptions. Future studies could examine how the network changes over time and how external factors like weather or economic shifts affect it. This analysis helps make informed decisions to optimize and secure the U.S. air transportation network, ensuring it remains robust and efficient.

REFERENCES

[1] Wright, G., Yasar, K. (2022, December 12). social networking. WhatIs.

[2] Kenton, W. (2022, February 11). Social Networking Service (SNS): characteristics and risks. Investopedia. https://www.investopedia.com/terms/s/social-networking-service-sns.asp

[3] Bello, M. (2023, June 10). Online communities vs social networks: What's the difference? Amity. https://www.amity.co/blog/online-communities-vs-social-networks-whats-the-difference.

[4] Disney, A. (2023, August 14). Social network analysis: Understanding centrality measures. Cambridge Intelligence. https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/