SNA U1 - project - Networks Theory

Luis David Martínez Gutiérrez
Data Engineering
Universidad Politécnica de Yucatán
Km. 4.5. Carretera Mérida — Tetiz
Tablaje Catastral 4448. CP 97357
Ucú, Yucatán. México
Email: 2109099@upy.edu.mx

Didider Omar Gamboa Ángulo
Universidad Politécnica de Yucatán
Km. 4.5. Carretera Mérida — Tetiz
Tablaje Catastral 4448. CP 97357
Ucú, Yucatán. México
Email: didier.gamboa@upy.edu.mx

Abstract

This analysis of the US Airlines network involved a comprehensive examination of its structure and characteristics. The network was visualized, and key metrics such as the size, number of links, average path length, and clustering coefficient were measured. Centrality measures including degree, betweenness, closeness, and eigenvector centrality were calculated to identify influential airports. The degree distribution was plotted to understand the connectivity pattern. Community detection using the Girvan-Newman method revealed clusters within the network. The findings underscore the network's centralized nature, with critical hubs playing a significant role in connectivity, and provide insights for optimizing network design and operations.

I. Introduction

Description

The chosen network represents the US Airlines network, where nodes represent airports and edges represent direct flights between these airports. This analysis is significant as it provides insights into the connectivity and efficiency of the airline network, which can inform decisions on infrastructure development, route optimization, and emergency response planning.

Importance

Understanding the structure of the airline network can help in optimizing flight routes, improving network resilience, and enhancing customer experience. This dataset is often used in studies related to transportation, logistics, and network theory to analyze and improve the efficiency and robustness of airline networks.

Dataset Usage and General Results

The dataset is commonly used in network analysis studies to examine the structural properties of airline networks, such as connectivity, centrality, and community structure. Typical results include identifying key hubs, understanding the distribution of connections, and detecting communities within the network.

Network Classification

The US Airlines network is classified as a directed network because the connections (flights) between nodes (airports) have a direction (from source to destination). It is not a bipartite network as it doesn't involve two distinct sets of nodes. It can also be planar as it can be drawn on a plane without any edges crossing, although this is not always practical.

UNIVERSIDAD POLITÉCNICA DE YUCATÁN BIS Universities

SNA U1 - project - Networks Theory

II. Network Characteristics.

A. *Size of the Network*

The size of the network refers to the total number of nodes, which in this case are the airports. This metric gives us an idea of the scale of the network.

B. *Number of Links*

The number of links (edges) represents the total number of direct flights between airports. This helps us understand the connectivity within the network.

C. *Average Path Length*

The average path length is the average number of steps along the shortest paths for all possible pairs of network nodes. It provides insight into the efficiency of the network in terms of how easily passengers can travel between airports.

D. *Clustering Coefficient*

The clustering coefficient measures the degree to which nodes in a network tend to cluster together. It indicates the likelihood that two airports connected to the same airport are also connected to each other. A high clustering coefficient implies a network with tightly knit groups.

E. *Additional Distance Metrics*

Average Distance: Similar to the average path length, it represents the average number of steps along the shortest paths in the network.
Diameter: The longest shortest path between any two nodes in the network. It provides insight into the network's size in terms of reachability.
Eccentricity: The greatest distance between a node and any other node. It helps identify the nodes that are the farthest apart in the network.
Radius: The minimum eccentricity of any node. It represents the shortest maximum distance from any node to all other nodes.
Periphery: The set of nodes with the highest eccentricity. These nodes are the farthest from all other nodes in the network.
Center: The set of nodes with the lowest eccentricity. These nodes are the most central in the network.

F. *CODE 1: Spring Layout with Default Parameters*

Graph Layout: The nodes are positioned using the spring layout algorithm with default parameters. This layout simulates a physical system where nodes repel each other, and edges act like springs pulling nodes together, leading to a visually appealing representation.

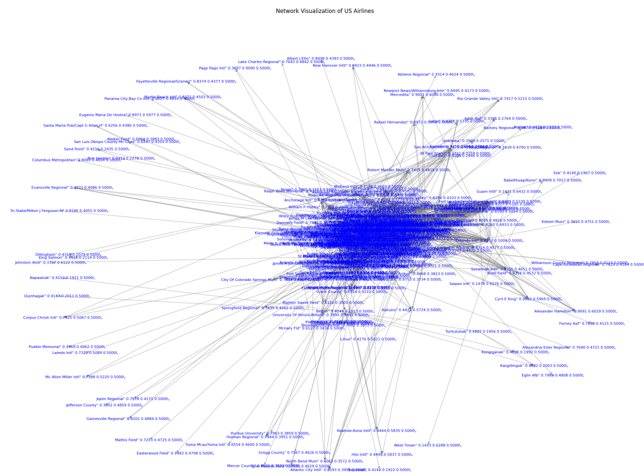

Fig. 1. Spring Layout with Default Parameters

Colors and Labels: Nodes are light blue, edges are gray, and labels are blue.
Visualization: This layout generally results in a more aesthetically pleasing visualization where clusters and hubs are more visible, although it might be a bit cramped in densely connected regions.

G. *CODE 2: Circular Layout*

Graph Layout: The nodes are positioned using a circular layout. This positions all nodes on a circle, spreading them out evenly around the circumference.
Colors and Labels: Nodes are blue, edges are light gray, and labels are red.
Visualization: The circular layout helps to evenly distribute nodes around a central point, making it easier to see individual nodes and their connections, but it can make it harder to see the overall structure and clusters within the network.

H. *CODE 3: Spring Layout with Adjusted Parameters*

Graph Layout: The nodes are positioned using the spring layout algorithm with an adjusted k parameter (0.3). This parameter controls the optimal distance between nodes, resulting in a more spread-out graph.
Colors and Labels: Nodes are blue, edges are light gray, and labels are red.
Visualization: Adjusting the k parameter helps to reduce the cramping of nodes, providing a clearer view of individual nodes and connections, and revealing the network's structure more effectively.

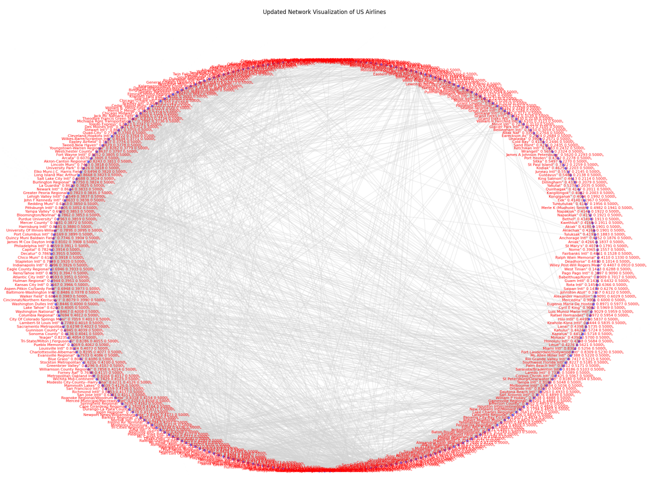

Fig. 2. Circular Layout

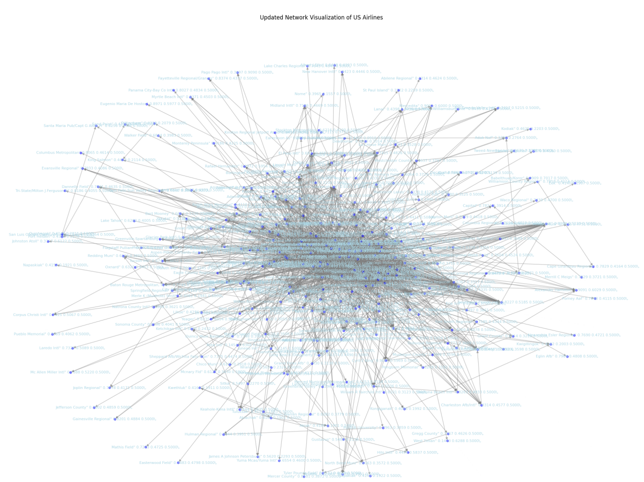

Fig. 3. Spring Layout with Adjusted Parameters

III. How the Different Graphs Help in Visualization

A. CODE 1 (Default Spring Layout)

Strengths: Good for visualizing the overall structure and clustering within the network.
Use Case: Ideal when you want to see the general layout and identify clusters and hubs.

B. CODE 2 (Circular Layout)

Strengths: Excellent for identifying individual nodes and their direct connections due to even spacing.
Use Case: Useful when the focus is on the connections of individual nodes rather than the overall network structure.

C. CODE 3 (Adjusted Spring Layout)

Strengths: Provides a clearer view of individual nodes and connections by reducing node overlap.
Use Case: Best for detailed analysis of node connections and network structure without the clutter seen in the default spring layout.

Each layout offers unique advantages depending on the specific aspects of the network you wish to analyze. The spring layout (default and adjusted) is better for overall structural insights, while the circular layout excels in clarity of individual node connections.

D. How the Different Graphs Help in Visualization

E. CODE 1 (Default Spring Layout)

Strengths: Good for visualizing the overall structure and clustering within the network.
Use Case: Ideal when you want to see the general layout and identify clusters and hubs.

F. CODE 2 (Circular Layout)

Strengths: Excellent for identifying individual nodes and their direct connections due to even spacing.
Use Case: Useful when the focus is on the connections of individual nodes rather than the overall network structure.

G. CODE 3 (Adjusted Spring Layout)

Strengths: Provides a clearer view of individual nodes and connections by reducing node overlap.
Use Case: Best for detailed analysis of node connections and network structure without the clutter seen in the default spring layout.

Each layout offers unique advantages depending on the specific aspects of the network you wish to analyze. The spring layout (default and adjusted) is better for overall structural insights, while the circular layout excels in clarity of individual node connections.

IV. Centrality Measures.

Centrality measures help identify the most important nodes in a network based on different criteria. For the US Airlines network, the following centrality measures are considered:

A. *Degree Centrality*

Degree centrality measures the number of direct connections a node has. Airports with high degree centrality are major hubs with many direct flights.

B. *Betweenness Centrality*

Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes. Airports with high betweenness centrality play a critical role in connecting different parts of the network.

C. *Closeness Centrality*

Closeness centrality reflects how close a node is to all other nodes. Airports with high closeness centrality can reach other airports more quickly and efficiently.

D. Eigenvector Centrality

Eigenvector centrality measures the influence of a node in the network. Airports with high eigenvector centrality are connected to other highly influential airports.

E. Observations

Degree Centrality: Indicates which airports are the main hubs. These hubs are critical for network efficiency and passenger traffic.
Betweenness Centrality: Highlights airports that are crucial for maintaining connectivity between different parts of the network. These airports are key for route optimization and emergency response.
Closeness Centrality: Identifies airports that can reach others quickly, which is important for efficient travel times.
Eigenvector Centrality: Reflects the overall influence of airports in the network. Airports with high eigenvector centrality are integral for maintaining network stability.

	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	Name
118	0.419940	0.048395	0.144455	3.368532e-58	Chicago O'hare Intl" 0.7914 0.3710 0.5000\,
261	0.356495	0.043513	0.373403	7.949398e-19	Dallas/Fort Worth Intl" 0.7449 0.4578 0.5000\,
255	0.305136	0.022892	0.337740	1.402415e-21	The William B Hartsfield Atlan" 0.8091 0.4506 ...
152	0.283988	0.022363	0.170822	4.940028e-49	Pittsburgh Intll" 0.8305 0.3852 0.5000\,
182	0.283988	0.026587	0.229017	1.035827e-34	Lambert-St Louis Intl" 0.7789 0.4018 0.5000\,
230	0.262840	0.018125	0.275244	1.301940e-26	Charlotte/Douglas Intl" 0.8268 0.4356 0.5000\,
166	0.256798	0.018262	0.204353	3.999268e-43	Stapleton Intl" 0.7049 0.3920 0.5000\,
67	0.235650	0.021579	0.068324	8.220781e-70	Minneapolis-St Paul Intl/Wold-" 0.7644 0.3432 ...
112	0.211480	0.006950	0.099924	1.000477e-59	Detroit Metropolitan Wayne Cou" 0.8146 0.3687 ...
201	0.205438	0.023056	0.255905	2.073334e-33	San Francisco Intl" 0.6159 0.4126 0.5000\,

Fig. 4. Spring Layout with Default Parameters

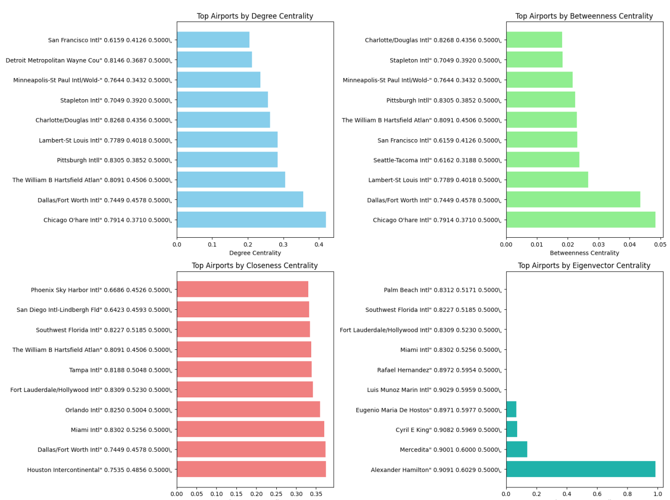

Fig. 5. Spring Layout with Default Parameters

V. Degree Distribution.

The degree distribution of a network shows how the connections (or edges) are distributed among the nodes (or vertices). It provides insights into the overall structure and robustness of the network.

Observations
The degree distribution can reveal whether the network follows a scale-free distribution, where a few nodes (airports) have a very high degree, while most nodes have a low degree. A heavy-tailed distribution indicates the presence of hubs, which are critical for the network's connectivity and resilience.

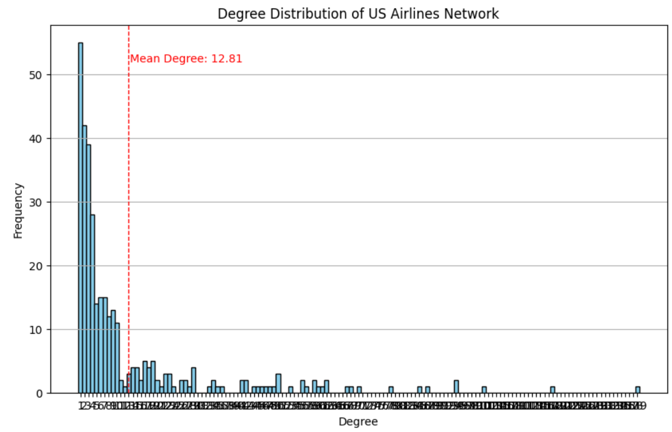

Fig. 6. Spring Layout with Default Parameters

VI. Community Detection.

Community detection identifies groups within the network where nodes are more densely connected to each other than to the rest of the network. This helps understand the modular structure of the network.

Communities within the airline network can indicate regional clusters or airline alliances. Understanding these communities helps in optimizing route planning and identifying critical connections that maintain network integrity. Communities may correspond to geographical regions or operational divisions within the airline industry, highlighting how different parts of the network are organized and interact.

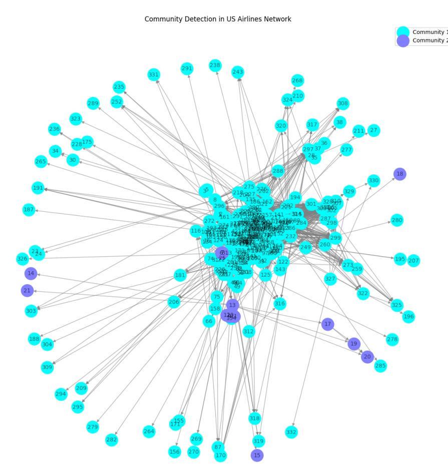

Fig. 7. Spring Layout with Default Parameters

VII. Conclusion

Based on the analysis, the US Airlines network exhibits a highly centralized structure with a few key hubs playing crucial roles in connectivity. The degree distribution suggests a hub-and-spoke model, which is common in airline networks. Centrality measures highlight the importance of specific airports in maintaining network efficiency and resilience. Community detection reveals regional clusters that can inform strategic planning and operational improvements.

REFERENCES

[1] Hagberg, A., Schult, D., Swart, P. (2008). NetworkX: Network Analysis with Python. Retrieved from https://networkx.github.io/documentation/stable/

[2] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science Engineering, 9(3), 90-95. Retrieved from https://matplotlib.org/stable/contents.html

[3] Girvan, M., Newman, M. E. J. (2002). Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12), 7821-7826. Retrieved from https://www.pnas.org/content/99/12/7821

[4] Stanford large network dataset collection. (n.d.). Retrieved from https://snap.stanford.edu/data/

[5] Pajek datasets. (n.d.). Retrieved from http://vlado.fmf.uni-lj.si/pub/networks/data/