

The Architecture of Flight: A Network Analysis of the US Airport System

Troy Altus

2026-05-01

Table of contents

1	Abstract	1
2	Introduction	2
3	Data	2
3.1	Source and Acquisition	2
3.2	Dataset Limitations	3
4	Methods	3
4.1	Graph Construction	3
4.2	Degree Distribution and Scale-Free Test	3
4.3	Small-World Coefficient	3
4.4	Centrality Measures	4
4.5	Robustness Analysis	4
5	Results	4
5.1	Network Overview	4
5.2	Geographic Map	5
5.3	Network Topology	5
5.4	Degree Distribution	6
5.5	Hub Rankings by Centrality	6
5.6	Small-World Properties	6
5.7	Robustness Analysis	6
6	Discussion	7
7	Conclusion	7
	References	8

1 Abstract

Every time a flight is delayed in Atlanta, the ripple spreads to Pittsburgh, Phoenix, and Portland. This is not a coincidence — it is topology. The US domestic air network is a textbook example

of a scale-free, small-world graph: a system shaped not by central planning but by the relentless arithmetic of preferential attachment, where large hubs attract more routes because they already have routes. This paper applies graph-theoretic analysis to the US domestic airport network using the OpenFlights dataset (OpenFlights.org 2024), quantifying degree distribution, betweenness centrality, clustering behavior, and robustness under both random failure and targeted attack. The findings confirm that a small number of hub airports act as disproportionate load-bearers, and that the network’s tolerance of random failure masks a striking vulnerability to coordinated disruption of its most central nodes.

2 Introduction

There is a particular brand of airport misery that frequent travelers know well: the connecting flight that exists only because some planner, decades ago, decided that Chicago should be the midpoint between everywhere and everywhere else. The hub-and-spoke model of air travel is not merely a scheduling convenience — it is an emergent property of network economics, and it leaves a distinctive fingerprint in the topology of the system it creates.

Modern network science offers a precise vocabulary for this fingerprint. The foundational insight of Watts and Strogatz (1998) is that many real-world systems — from social networks to neural circuits to power grids — occupy a middle ground between the rigid regularity of a lattice and the pure randomness of an Erdős–Rényi graph. These *small-world* networks exhibit high local clustering (my friends tend to know each other) while maintaining surprisingly short average path lengths (everyone is nonetheless a few handshakes from everyone else). A year later, Barabási and Albert (1999) showed that many of these same networks follow a power-law degree distribution — a signature of *scale-free* structure, where a small number of nodes accumulate a vastly disproportionate number of connections.

The airport network is a canonical instance of this class. Guimerà et al. (2005) analyzed the worldwide air transportation network and found that it exhibits both scale-free degree distributions and small-world properties, with hub cities playing roles far beyond their geographic or economic size. The present analysis restricts attention to the US domestic network, where data quality is high and policy implications are direct: understanding which airports are structurally critical is the first step toward designing a system that fails gracefully.

Three questions organize this paper:

1. **Topology:** Does the US airport network follow a power-law degree distribution, and does it exhibit small-world properties?
2. **Centrality:** Which airports are the most structurally critical, and by what measure?
3. **Robustness:** How does the network respond to random node failure versus targeted removal of its most central nodes?

3 Data

3.1 Source and Acquisition

Route and airport data are drawn from the OpenFlights dataset (OpenFlights.org 2024), a curated open-access database of global airline routes and airport coordinates maintained by OpenFlights.org. Two files are used:

- **airports.dat** — 7,698 airports worldwide with IATA code, name, city, country, latitude, and longitude.
- **routes.dat** — 67,663 routes covering 3,321 airports and 548 airlines.

Analysis is restricted to US domestic nonstop routes: both origin and destination must be US airports (country = “United States”), and routes with intermediate stops are excluded.

3.2 Dataset Limitations

OpenFlights represents scheduled routes as reported by airlines and may not reflect real-time or seasonal schedules. Routes are unweighted — passenger volume, frequency, and seat capacity are not encoded. All edges are treated as undirected, collapsing the distinction between origin and destination. For weighted or directed analyses, the BTS T-100 Domestic Segment database (Bureau of Transportation Statistics 2024) provides monthly passenger and departure counts at the route level and is the recommended complement to this dataset.

4 Methods

4.1 Graph Construction

The airport network is modeled as a simple undirected graph $G = (V, E)$, where each vertex $v \in V$ represents a US airport (identified by IATA code) and each edge $(u, v) \in E$ represents the existence of at least one nonstop domestic route between airports u and v . Multiple airlines operating the same route are collapsed to a single edge.

All path-length and clustering metrics are computed on the largest connected component (LCC) to avoid undefined values in disconnected subgraphs.

4.2 Degree Distribution and Scale-Free Test

The degree of a node $k_i = |\{j : (i, j) \in E\}|$ counts its direct connections. For a scale-free network, the degree distribution follows a power law:

$$P(k) \sim k^{-\gamma}$$

where γ is the scaling exponent. Following Clauset et al. (2009), we fit γ by ordinary least squares regression on the log-log complementary cumulative distribution function (CCDF). A value $2 < \gamma < 3$ is characteristic of most empirically observed scale-free networks.

4.3 Small-World Coefficient

Watts and Strogatz (1998) characterize a small-world network by comparing two statistics against null models drawn from equivalent Erdős–Rényi random graphs:

- **Clustering coefficient** C — probability that two neighbors of a node are also connected:

$$C = \frac{1}{n} \sum_i \frac{2t_i}{k_i(k_i - 1)}$$

where t_i is the number of triangles through node i .

- **Average shortest path length** L — mean geodesic distance between all pairs.

For an equivalent random graph with n nodes and mean degree $\langle k \rangle$:

$$C_{\text{rand}} \approx \frac{\langle k \rangle}{n}, \quad L_{\text{rand}} \approx \frac{\ln n}{\ln \langle k \rangle}$$

The small-world coefficient is:

$$\sigma = \frac{C/C_{\text{rand}}}{L/L_{\text{rand}}}$$

A value $\sigma \gg 1$ indicates small-world structure.

4.4 Centrality Measures

Two centrality measures are computed:

Degree centrality normalizes degree by the maximum possible:

$$C_D(i) = \frac{k_i}{n-1}$$

Betweenness centrality counts the fraction of all-pairs shortest paths that pass through node i :

$$C_B(i) = \frac{1}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from s to t and $\sigma_{st}(i)$ is the number passing through i . Betweenness centrality identifies *brokers* — airports that sit on many critical paths even if they are not the largest hubs by degree.

4.5 Robustness Analysis

Node removal experiments follow Albert et al. (2000). Two strategies are compared:

- **Random failure** — nodes removed in a uniformly random order (averaged over 10 trials)
- **Targeted attack** — nodes removed in descending order of betweenness centrality (recomputed after each removal step for accuracy; approximated here with a static pre-computed ranking for computational tractability)

After each removal step, the size of the largest connected component (LCC) as a fraction of the original network is recorded.

5 Results

5.1 Network Overview

Table 1: US domestic airport network summary statistics

Table 1	
Metric	Value
Nodes (airports)	541
Edges (routes)	2,780
Mean degree k	10.28
Max degree	153
Clustering coefficient C	0.4944
C (random graph equivalent)	0.0190
Avg. shortest path L	3.200
L (random graph equivalent)	2.701
Small-world coefficient	21.97
Connected components	3
LCC fraction of total	98.5%

The network is a sparse graph with a mean degree far below the theoretical maximum, yet exhibits a clustering coefficient substantially higher than an equivalent random graph. The small-world coefficient $\sigma \gg 1$ confirms small-world structure: the network is simultaneously locally dense (neighbors of airports tend to also serve each other) and globally compact (any two airports are reachable in just a few hops).

5.2 Geographic Map

Unable to display output for mime type(s): text/html

(a) US airport network. Node size scales with degree; color encodes betweenness centrality. Hover for airport details.

Unable to display output for mime type(s): text/html

(b)

Figure 1

The map makes the hub-and-spoke architecture immediately visible. A handful of large-degree airports — concentrated in the South, Midwest, and East — dominate the network, while smaller regional airports appear as peripheral spokes. Note that high degree and high betweenness do not always coincide: some mid-tier airports serve as critical bridges between regional clusters without themselves being large hubs.

5.3 Network Topology

The force-directed layout strips away geography to reveal pure network structure. Hub airports — those with many routes — migrate to the center, pulled there by the gravity of their connections. Regional clusters are visible as dense local neighborhoods; the long edges crossing from one cluster to another represent the high-betweenness airports that serve as critical bridges. Remove those bridges, and the clusters become islands.

5.4 Degree Distribution

Unable to display output for mime type(s): text/html

Figure 2: Degree distribution of the US airport network on a log-log scale. A power-law fit to the complementary CDF is overlaid.

The degree distribution follows an approximate power law with exponent $\gamma \approx 1.26$, consistent with the range $2 < \gamma < 3$ characteristic of scale-free networks (Barabási and Albert 1999). The log-log CCDF is approximately linear across more than a decade of degree values. The deviation at high degree is expected: the finite size of the network and the physical constraint of airport capacity impose natural cutoffs on how many routes any single hub can support.

5.5 Hub Rankings by Centrality

Unable to display output for mime type(s): text/html

Figure 3: Top 25 airports by betweenness centrality (left) and degree (right). Airports are labeled by IATA code; hover for full name.

The two rankings tell different stories. Degree centrality captures raw connectivity — the largest hubs by sheer number of destinations served. Betweenness centrality identifies *structural brokers*: airports that sit at the intersection of many shortest paths through the network, regardless of how many direct routes they operate. An airport can rank highly on betweenness without appearing in the degree top-ten if it serves as the primary gateway between two otherwise loosely connected regional clusters.

5.6 Small-World Properties

Table 2: Small-world metrics compared to an equivalent random graph

Table 2			
Metric	Airport Network	Random Graph (ER)	Ratio C/C_rand or L/L_rand
Clustering coefficient C	0.4944	0.0190	26.0×
Avg. path length L	3.200	2.701	1.18×
Ratio to random	—	—	= 21.97

The clustering coefficient of the airport network is many times larger than that of an equivalent random graph, while the average path length is close to the random-graph value. This combination — high clustering, short paths — is the defining signature of small-world structure (Watts and Strogatz 1998). In plain terms: airports tend to cluster into regional communities (carriers concentrate service in geographic hubs), yet the global network is so well-connected that even remote airports are just a few hops from any other.

5.7 Robustness Analysis

The divergence between the two curves is the central empirical result of this section. Under random failure — analogous to individual airport closures from weather, mechanical failure, or

Unable to display output for mime type(s): text/html

Figure 4: Network robustness under random failure and targeted attack. The y-axis shows the largest connected component (LCC) as a fraction of the original network size as nodes are progressively removed.

labor action — the network degrades gracefully: the LCC shrinks slowly even as a substantial fraction of airports are removed, because the probability of randomly hitting a critical hub is low. Under targeted attack — the removal of airports in descending order of betweenness centrality — the network fragments far more rapidly. The gap between the two curves quantifies the structural risk embedded in the hub-and-spoke design.

This result is not unique to air travel. Albert et al. (2000) demonstrated the same pattern in the internet topology and in metabolic networks: scale-free graphs are simultaneously robust to random failure and fragile to deliberate attack, precisely because the hubs that absorb random perturbations are also the nodes whose removal does the most damage.

6 Discussion

The US airport network is a textbook instance of emergent complexity. No single designer decreed that Atlanta should be the most critical node in the network; that distinction emerged from decades of airline route decisions, each individually rational, collectively producing a system with a small number of overloaded load-bearers and a long tail of peripheral spoke airports. The power-law degree distribution is not a quirk of aviation — it is what preferential attachment looks like at scale.

Several implications follow from the robustness analysis.

Resilience planning should prioritize by betweenness, not by size. The largest airports by passenger volume are not identical to the most structurally critical airports by betweenness centrality. A disruption event that closes a high-betweenness, moderate-degree airport may fragment the network more severely than closing a high-volume hub. Emergency response planning that treats all large airports equally is systematically misaligned with network topology.

Edge redundancy matters more than node redundancy. The robustness curves suggest that adding more airports (nodes) provides diminishing returns if those airports simply attach as additional spokes to existing hubs. Adding routes between mid-tier airports — increasing edge density in the periphery — would reduce the structural load on the high-betweenness brokers and flatten the targeted-attack curve.

The OpenFlights dataset underestimates redundancy. Routes are unweighted and binary: a single weekly service and a 20-flights-per-day shuttle appear identically. A weighted analysis using BTS T-100 passenger data (Bureau of Transportation Statistics 2024) would produce a more operationally accurate picture, likely showing that the effective network is even more concentrated on a handful of heavily-trafficked routes.

7 Conclusion

The US domestic airport network is simultaneously one of the most studied and most misunderstood complex systems in American infrastructure. Its everyday dysfunction — the missed

connection, the cascading delay, the overbooked hub — is not a failure of management; it is a direct consequence of its topology. A scale-free, small-world graph optimizes for average-case efficiency at the cost of worst-case fragility.

This analysis has quantified that trade-off. The degree distribution confirms scale-free structure. The small-world coefficient confirms that geographic clustering does not prevent global reachability. And the robustness curves make visible what operational experience has long suggested: the network is surprisingly robust to the random disruptions that fill the nightly news, and surprisingly vulnerable to the coordinated failures that resilience planners lose sleep over.

Future work in this series will apply comparable methods to the US power grid and internet autonomous system topology, extending the comparison across infrastructure domains and asking whether the hub-and-spoke architectural pattern is universal or domain-specific.

References

- Albert, Réka, Hawoong Jeong, and Albert-László Barabási. 2000. “Error and Attack Tolerance of Complex Networks.” *Nature* 406 (6794): 378–82. <https://doi.org/10.1038/35019019>.
- Barabási, Albert-László, and Réka Albert. 1999. “Emergence of Scaling in Random Networks.” *Science* 286 (5439): 509–12. <https://doi.org/10.1126/science.286.5439.509>.
- Bureau of Transportation Statistics. 2024. “T-100 Domestic Segment Data.” *US Department of Transportation*. <https://www.transtats.bts.gov/>.
- Clauset, Aaron, Cosma Rohilla Shalizi, and Mark EJ Newman. 2009. “Power-Law Distributions in Empirical Data.” *SIAM Review* 51 (4): 661–703. <https://doi.org/10.1137/070710111>.
- Guimerà, Roger, Stefano Mossa, Adrian Turttschi, and Luis A Nunes Amaral. 2005. “The Worldwide Air Transportation Network: Anomalous Centrality, Community Structure, and Cities’ Global Roles.” *Proceedings of the National Academy of Sciences* 102 (22): 7794–99. <https://doi.org/10.1073/pnas.0407994102>.
- OpenFlights.org. 2024. *OpenFlights: Airport and Airline Data*. <https://openflights.org/data.html>.
- Watts, Duncan J, and Steven H Strogatz. 1998. “Collective Dynamics of ‘Small-World’ Networks.” *Nature* 393 (6684): 440–42. <https://doi.org/10.1038/30918>.