Discovering Connectivity Patterns and Resilience in the Global Air Transportation Network

6-Step Workflow for Complex Network Analysis of the Global ATN

Jasleen Minhas
COMP 6917 – Complex Networks

December 2025

Abstract

This project applies the 6-step workflow from the course to the Global Air Transport Network (ATN). The network is modeled as a directed graph built from the OpenFlights airport and route data, where airports are nodes and direct routes are directed edges. The airport and route data were obtained from OpenFlights.org. The final graph contains **6 072 airports** and **37 042 routes**. Despite being extremely sparse, the network has an average shortest path length of only **3.97** flights and a clustering coefficient around **0.26**, indicating a clear small–world pattern. The degree distribution is strongly right–skewed and well approximated by a power law in the tail (exponent $\gamma \approx 1.48$), so a small set of hubs dominates global connectivity. Louvain community detection reveals about twenty geographically coherent communities with high modularity ($Q \approx 0.65$). Finally, resilience simulations show that the network is robust to random failures but fragile under targeted attacks on high–centrality airports and under regional (community–level) disruptions.

Introduction and Research Questions

Modern air travel behaves like a single, tightly coupled system. A snowstorm in Frankfurt or a closure of Chinese airspace quickly produces delays on other continents. This behaviour is an expression of the underlying network structure rather than a set of isolated events.

The project is driven by the following questions:

Q1: Is the ATN *scale–free* or at least strongly hub–dominated?

Q2: Does it show the *small–world* effect (short paths and high clustering)?

Q3: Which airports play the most important roles as hubs, bridges, and navigators?

Q4: Does the network naturally split into geographically meaningful communities?

Q5: How resilient is the network to random failures and to targeted or regional attacks?

The rest of the report is organised along the 6-step workflow specified in the project regulation: dataset identification, graph construction, metric selection, algorithmic implementation, knowledge discovery, and final understanding.

Step 1: Identifying a Dataset of Interest

The aim of Step 1 is to choose a real-world system that can naturally be represented as a network and that raises interesting structural and resilience questions.

For this project, I selected the **Global Air Transport Network** using the **OpenFlights** dataset:

- `airports.dat`: 7 698 airports with names, IATA codes, cities, countries, latitude and longitude.
- `routes.dat`: 67 663 routes specifying airline, source airport and destination airport.

This dataset is suitable because:

- it is large enough ($\sim 7\,000$ airports, $\sim 68\,000$ routes) to exhibit complex behaviour;
- it is global and safety-critical, so resilience questions are meaningful;
- airport codes and coordinates make it easy to visualise and interpret.

Step 2: Constructing a Proper Graph Model

In Step 2, I transform the raw data into a clean, well-defined graph.

Data cleaning

The following preprocessing steps are applied:

1. Remove airports without a valid three–letter IATA code or with missing coordinates.
2. Keep only routes whose source and destination both appear in the cleaned airport list.
3. Remove self–loops (routes from an airport to itself).
4. Collapse parallel routes between the same ordered pair of airports into a single directed edge.

Graph definition

The resulting directed graph $G = (V, E)$ is defined as:

- V: $|V| = 6\,072$ airports (IATA codes).
- E: $|E| = 37\,042$ directed edges, each indicating at least one scheduled route from u to v.

Each node stores attributes (name, country, latitude, longitude) for mapping and interpretation.

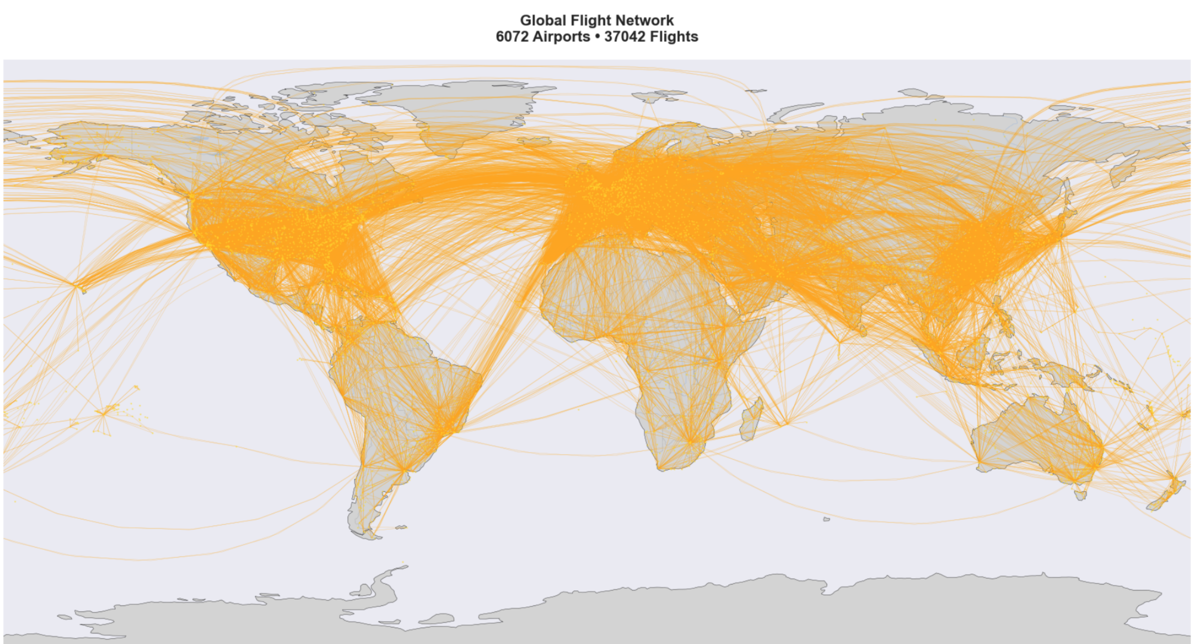

Figure 1: Global flight network: **6 072 airports** and **37 042 directed routes**. Edge density is low but connectivity between major regions is very high.

A global visualisation of all routes is shown in Figure 1. Dense bands connect North America and Europe, Europe and Asia, and parts of Asia with the Middle East and Australia, while some regions in Africa and the Pacific appear more sparsely connected.

Step 3: Determining a Set of Metrics

Step 3 is about deciding which quantitative metrics are needed to answer the research questions.

Global structure metrics

To test for small–world behaviour and core–periphery structure:

Number of nodes and edges; Average in/out/total degree; Average shortest path length and diameter (on the largest strongly connected component); Clustering coefficient (transitivity and average clustering); Reciprocity and Assortativity; Size of the largest strongly and weakly connected components.

Node-level importance metrics

To identify key airports:

- Degree centrality (hubs by flight count);
- Eigenvector and Katz centrality (importance by influential neighbours);
- PageRank (important destinations);
- Betweenness (bridges);
- Closeness (navigators).

Community and resilience metrics

To understand regional structure and robustness:

- Louvain modularity and community assignments;
- Fraction of nodes in each community and their main hubs;
- Size of the largest strongly connected component during:
  - Random node removal,
  - Targeted removal by degree,
  - Targeted removal by betweenness,
  - Progressive removal of entire communities.

Step 4: Implementing Algorithms to Compute Metrics

All computations are implemented in Python using `NetworkX`, `SciPy`, and `NetworKit` in a single Jupyter notebook (`source_code.ipynb`). The main algorithmic components are:

- **Graph construction**: parsing CSV files and adding nodes/edges with attributes.
- **Global metrics**: using NetworkX functions for degree statistics, clustering, path lengths and components on the largest strongly connected component.
- **Degree distribution and power-law fit**: computing degree histograms; plotting them in log–log scale; fitting tail exponents to estimate γ.
- **Centrality algorithms**:
  - degree centrality via simple counting;
  - eigenvector, Katz, and PageRank via power iteration;
  - betweenness via repeated shortest paths;
  - closeness via average shortest path length from each node.

- **Community detection**: Louvain (PLM implementation in NetworKit) to obtain community assignments and modularity.
- **Resilience simulations**: iterative node or community removal with recomputation of the largest strongly connected component at each step, storing the resulting size trajectory.

All intermediate outputs (tables and plots) are saved and a subset is included in this report.

Step 5: Discovering Knowledge from the Metrics

In Step 5, I use the computed metrics to extract concrete findings about structure, key actors, communities and resilience.

5.1 Global structure and small–world behaviour

Table 1 summarises the key structural metrics.

Table 1: Key network metrics and interpretations for the Global Air Transport Network.

Metric	Value	Interpretation
Nodes N	6 072	Total number of active airports.
Edges $\|E\|$	37 042	Unique directed (one–way) routes.
Avg_in_deg	6.0	Each airport receives flights from about 6 others.
Avg_out_deg	6.0	Each airport offers direct flights to about 6 destinations.
Avg_deg	12.0	Typical airport is directly connected to about 12 others.
Density μ	0.001	Only $\approx 0.1\%$ of possible directed pairs have a flight.
Avg. short path L	3.967	Any two airports are about **4** flights apart on average.
Transitivity ζ	0.249	About 25% of possible triangles exist.
Reciprocity R	0.978	Nearly all routes are bidirectional.
Assortativity r	-0.016	Slightly disassortative (hubs link to smaller airports).
Avg. clustering	0.2617	Strong local clustering.
Largest_scc	3 190 nodes	Core strongly connected backbone.
Largest_wcc	3 231 nodes	Largest weakly connected component.
Diameter d	12	Longest shortest path inside the LSCC.

Even though the density is extremely low, the average path length is about 4 and clustering is moderate (~ 0.26). This is a classic small–world pattern: long-range efficiency with local redundancy. Figure 2 shows that slightly more than half of airports form a dense core (LSCC) carrying almost all traffic.

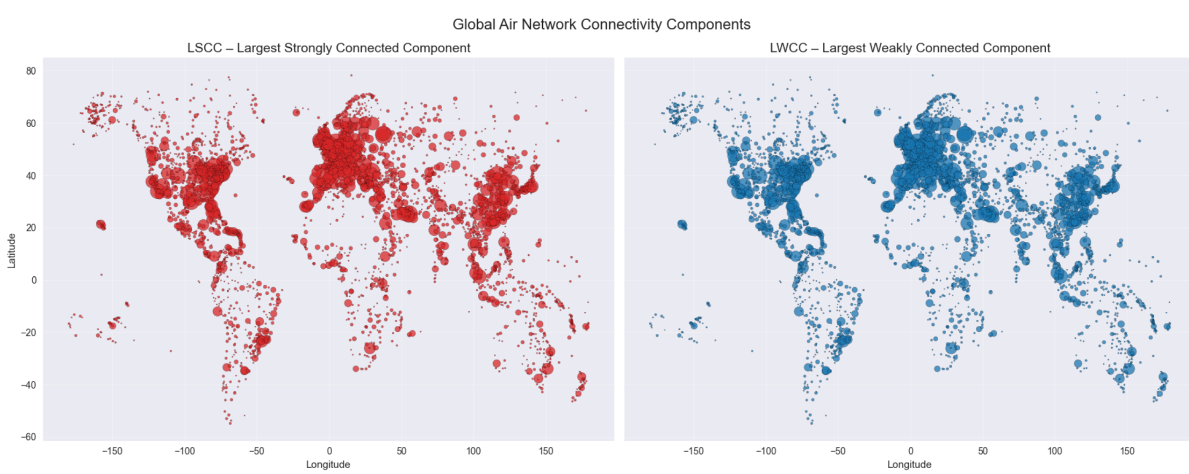

Figure 2: Largest strongly connected component (LSCC, left) and largest weakly connected component (LWCC, right). The LSCC contains **3 190 airports** and almost all routes, forming a dense global backbone.

5.2 Degree distribution and scale–free behaviour

Figure 3 shows the degree distribution on a log scale. Most airports have very few connections, while a small number have hundreds (e.g. Frankfurt with ~ 477 total connections).

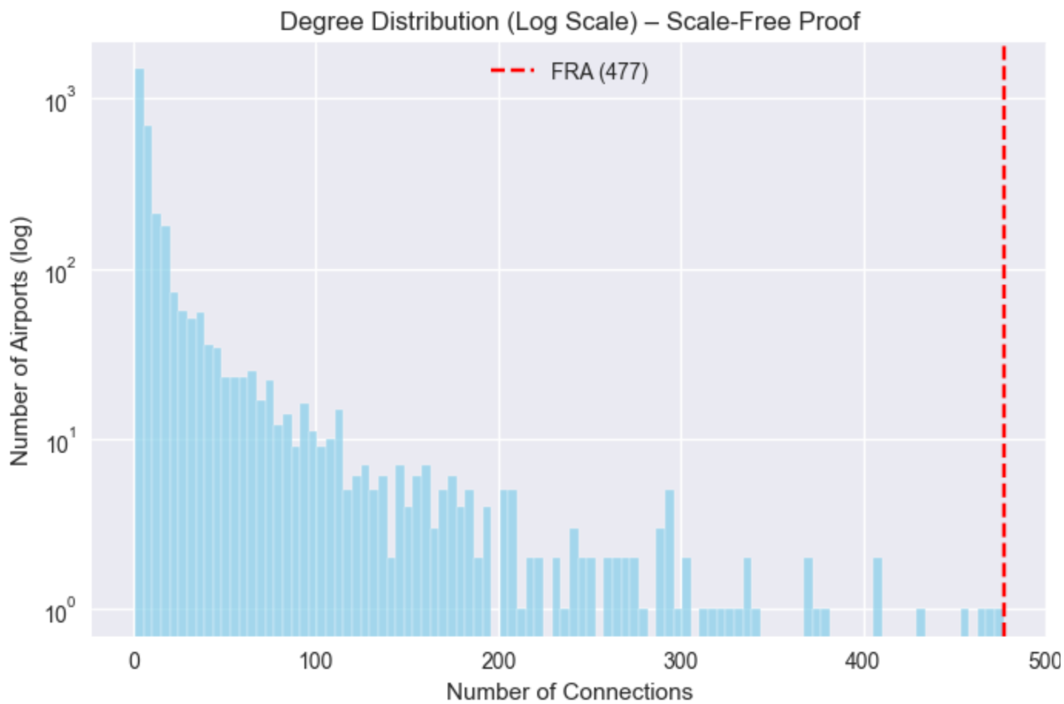

Figure 3: Total degree distribution on a log scale. The tall bar on the left corresponds to many small airports; the long tail reflects a small set of extremely well–connected hubs.

The in- and out-degree distributions in Figure 4 follow a near-linear trend in the tail on log–log axes. A power-law fit suggests $\gamma \approx 1.48$.

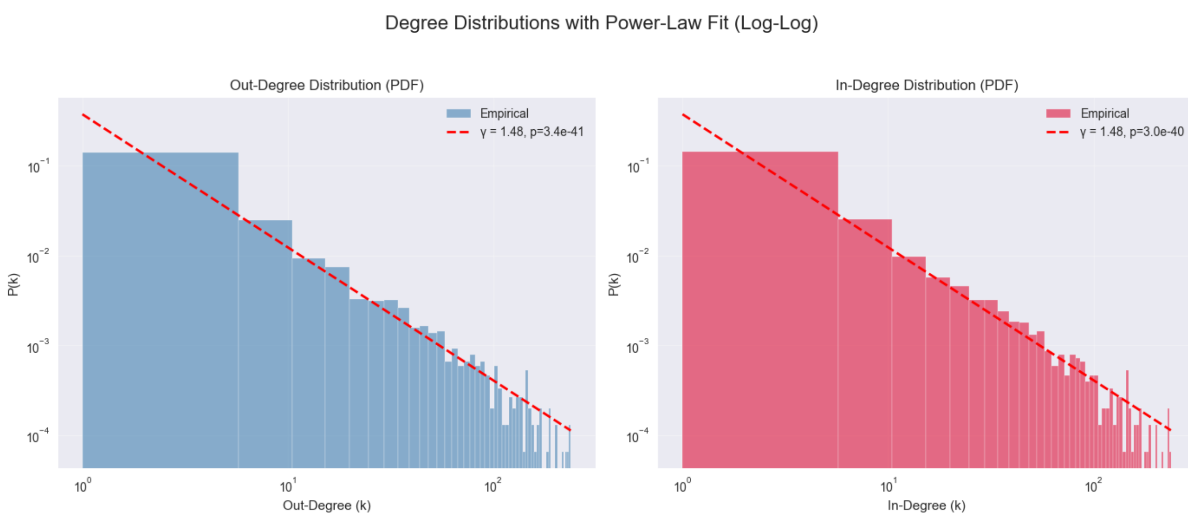

Figure 4: In– and out–degree distributions on log–log axes with power–law fits. The tail is close to linear, indicating a heavy–tailed, almost scale–free structure with exponent $\gamma \approx 1.48$.

This heavy tail shows a hub–dominated architecture: many airports with single-digit degree and a few megahubs with hundreds of connections.

5.3 Centrality: different notions of important airports

Centrality metrics tell different stories about importance. Figure 5 shows the top airports under various centrality measures.

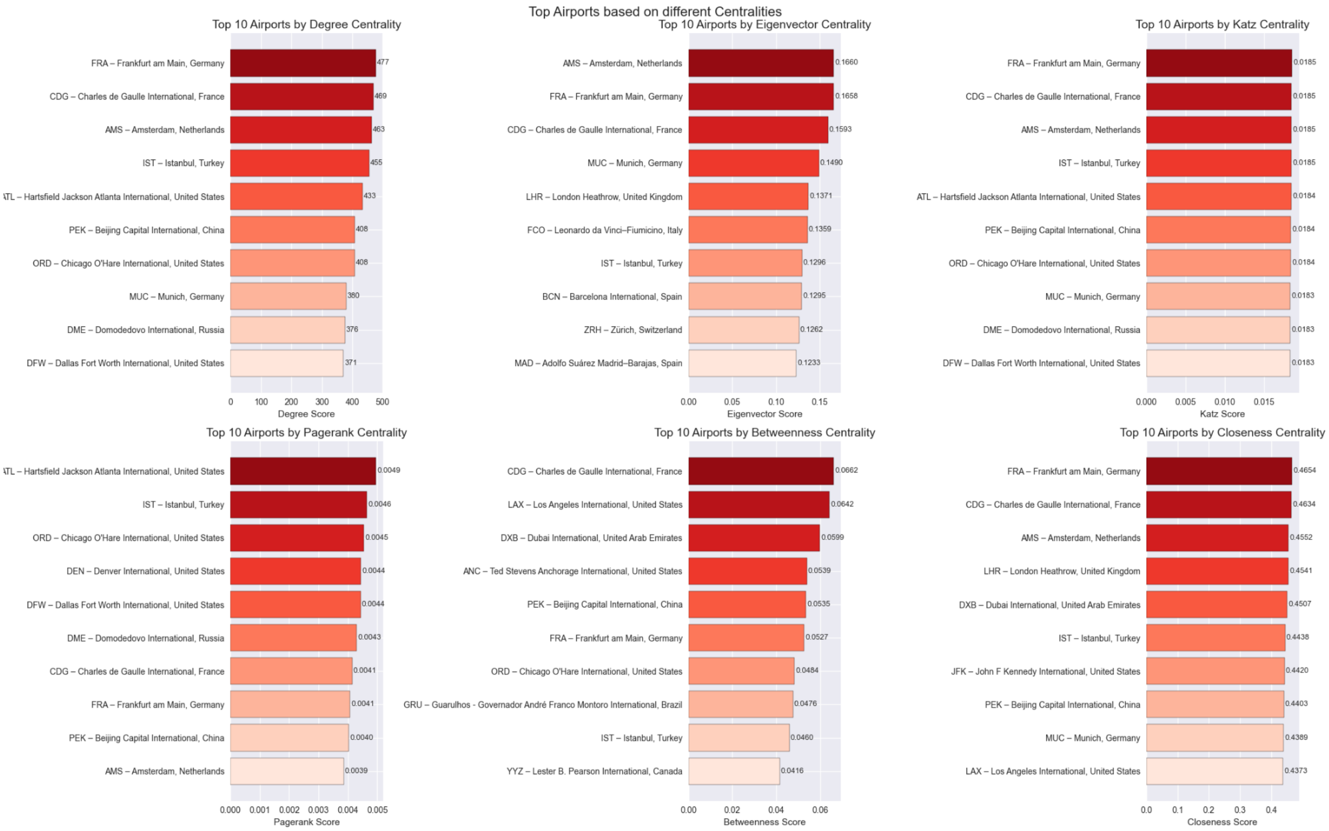

Figure 5: Top airports under different centrality measures. Frankfurt (FRA), Charles de Gaulle (CDG), Amsterdam (AMS), Istanbul (IST), Atlanta (ATL), Beijing (PEK) and Dubai (DXB) appear repeatedly, but with different roles.

Key patterns:

- **Megahubs (degree)**: FRA, CDG and AMS provide very large numbers of direct connections.
- **Bridges (betweenness)**: CDG, LAX and DXB sit on many shortest paths, acting as inter-regional gateways; ANC and PEK also act as bridges.
- **Navigators (closeness)**: European hubs such as FRA, CDG, AMS and LHR minimise average distance to the rest of the network.
- **Influential destinations (PageRank)**: ATL, IST and ORD are repeatedly visited when following random flights, making them central from a passenger-flow perspective.

5.4 Communities and geographic organisation

Louvain community detection yields about 20 communities with modularity $Q \approx 0.65$. Figure 6 shows these communities on the world map.

The communities align closely with continents and regional markets:

- a large North American community (ATL as hub),
- a dense European community (CDG, FRA, AMS),
- an East Asian community (PEK),

- distinct communities for South America (BOG), Africa (JNB), Arctic (ANC, YZF) and Pacific islands (PPT, POM).

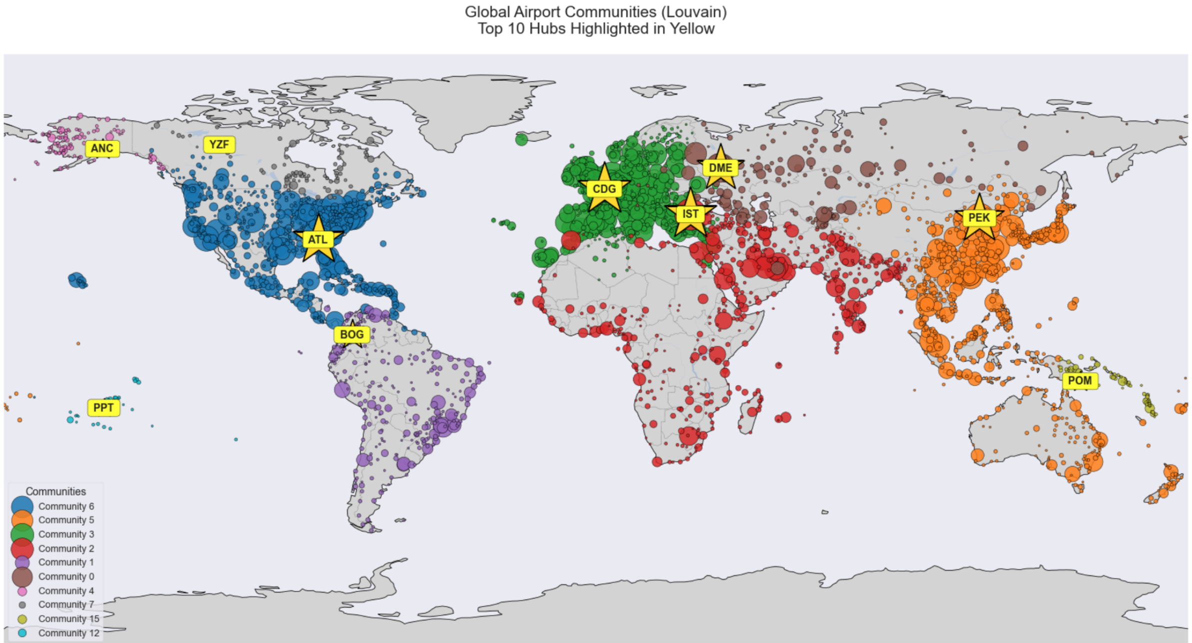

Figure 6: Louvain communities in the LSCC. Colours denote communities; yellow stars mark top hubs such as ATL (North America), CDG/DME (Europe), IST (Eurasia), PEK (East Asia), BOG (South America), ANC/YZF (Arctic), PPT/POM (Pacific).

5.5 Resilience under failures and attacks

Resilience experiments remove airports (or communities) and track the size of the largest strongly connected component.

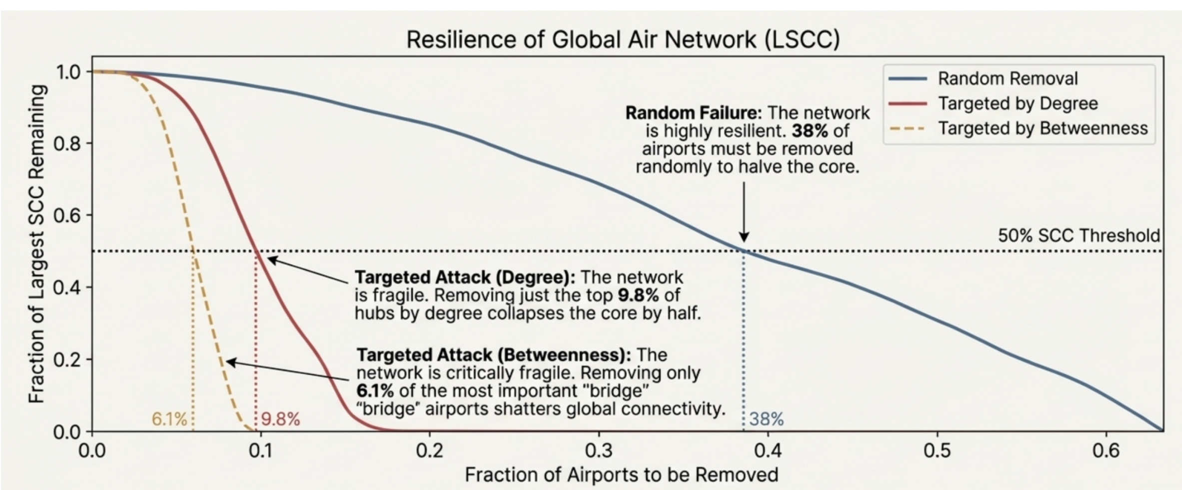

Figure 7: Resilience of the LSCC under different node removal strategies. Vertical lines mark where the LSCC shrinks to 50% of its original size.

Quantitative thresholds:

- **Random failure**: about **38%** of airports must be removed at random to halve the LSCC, indicating strong robustness to random local outages.

- **Degree-targeted attack**: removing only **9.8%** of the highest-degree airports halves the LSCC.
- **Betweenness-targeted attack**: removing around **6.1%** of the highest-betweenness airports fragments the core even faster.

Regional (community-level) disruptions show a different vulnerability (Figure 8):

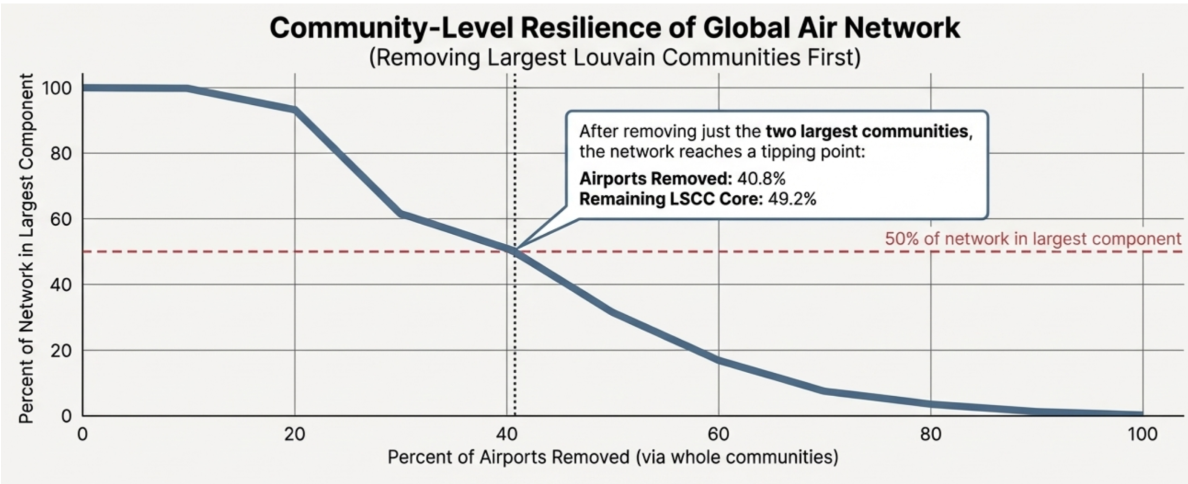

Figure 8: Community-level resilience: removing entire Louvain communities. After removing just the two largest communities, about **40.8%** of airports are gone and the largest component falls to **49.2%** of its original size.

After removing only two large communities, the largest component falls below half the network. This shows that the failure of a few regions (e.g. Europe and parts of North America or East Asia) can cripple global connectivity.

Step 6: Generating Understanding of the Problem

Finally, Step 6 synthesises the results into a coherent understanding of the ATN.

Scale-free and hubs. The heavy-tailed degree distributions with exponent $\gamma \approx 1.48$ confirm that the ATN is strongly hub–dominated. Many small airports depend on a few megahubs, which explains why disruptions at those hubs have global effects.

Small-world efficiency. The combination of low density (0.001), short paths (≈ 4 hops) and moderate clustering (≈ 0.26) shows that the ATN is a small–world network. Air travel is therefore both efficient globally and redundant locally.

Diverse roles of key airports. Centrality metrics reveal that "importance" is multi-dimensional. Some airports are megahubs (high degree), some are bridges (high betweenness), some are navigators (high closeness) and some are influential destinations (high PageRank). Operational planning should recognise these different roles instead of focusing on a single ranking.

Regional building blocks. Louvain communities with high modularity ($Q \approx 0.65$) indicate that the ATN is organised into geographically meaningful regions, each centred around one or a few hubs. These regions are the building blocks of the global system.

Resilience and vulnerability. The ATN is robust to random failures but fragile under targeted hub or bridge removal and under regional shutdowns. This dual behaviour is typical of hub–dominated networks and has clear implications for resilience planning, redundancy strategies and crisis management.

Overall, applying the 6-step workflow to the Global Air Transport Network shows how complex network analysis can turn raw route tables into a structured understanding of efficiency, inequality, regional organisation and vulnerability in a critical global infrastructure.

References

[1] OpenFlights. *OpenFlights Airport, Airline and Route Data*. Available at `https://openflights.org/data.php`.

[2] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[3] A.-L. Barabási. *Network Science*. Cambridge University Press, 2016.