BY: JASLEEN MINHAS

01 Dec, 2025

The Hidden Architecture of Global Air Travel

A Network Science Investigation into Connectivity and Resilience

Agenda Overview

Why Study the Air Transport Network?

Have you experienced the **ripple effect** of a single airport's failure?

A storm at a major hub like Frankfurt or Paris can cause delays and cancellations worldwide.

- **This isn't random; it's a feature of the network's structure.**

Problem Statement!

- The global air system is a **critical**, large-scale network **defined by a few mega-hubs** and thousands of **smaller** airports.

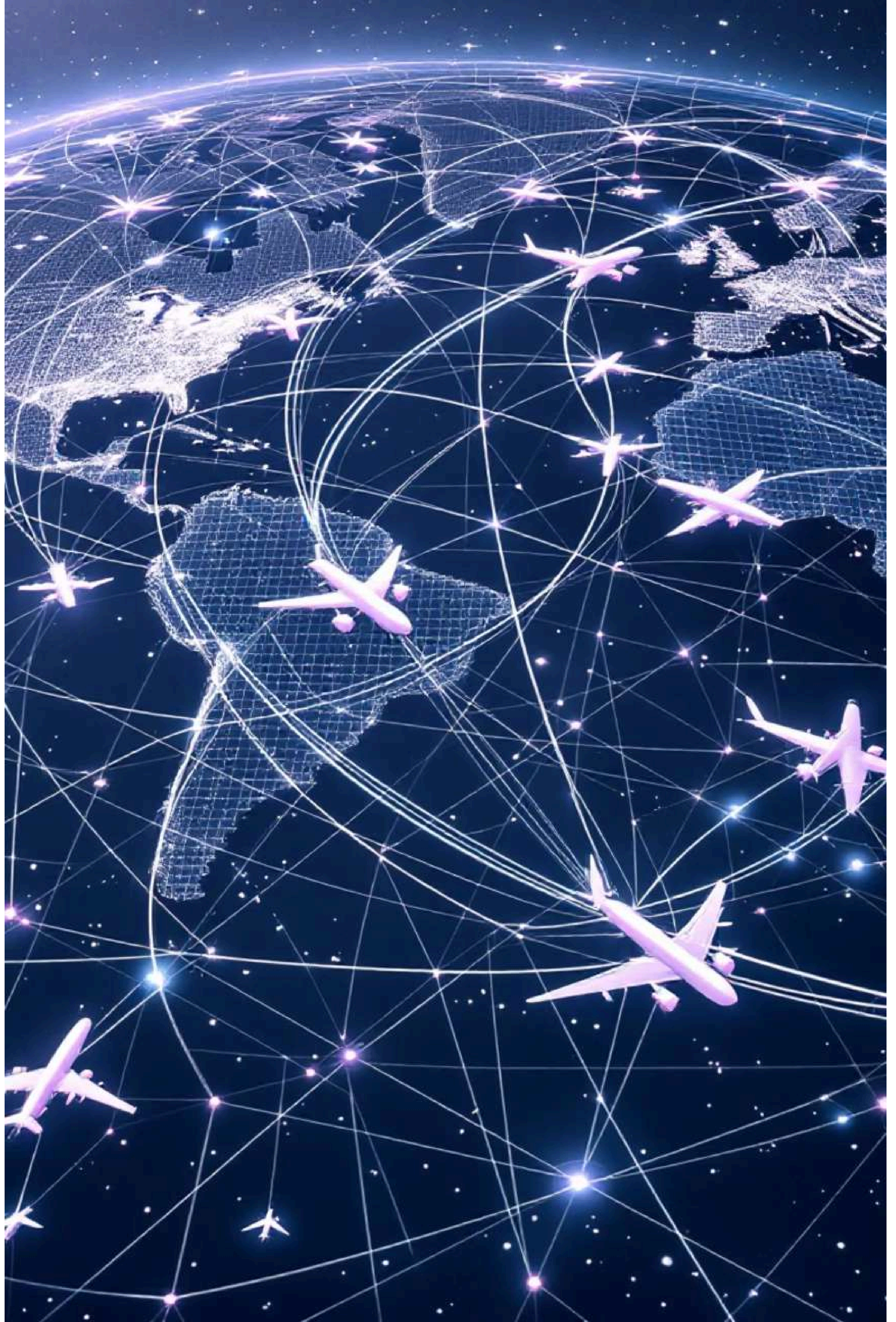
- Our goal is to **move beyond intuition** and **quantify this structure.**

- This investigation uses **complex network analysis** to reveal how the **world's air travel system is wired,** which airports are **critical to its function,** and just how **vulnerable it is to disruption.**

Importance?

This work leads to the foundation of:

- Improved resilience planning
- Strategic infrastructure investment
- Crisis management & recovery strategies



The Five Core Questions of the Investigation

1. **Structure:** Is the network “scale-free”, dominated by a few massive hubs?
2. **Efficiency:** Does it exhibit the "small-world" property of short paths despite the huge size of the Globe?
3. **Key Players:** Which airports are the true hubs and bridges that hold the network together?
4. **Territories:** Does the network naturally partition into regional communities?
5. **Fragility/Resilience:** How does the system collapse under random failures versus targeted attacks on its most critical components?

Dataset Used

OpenFlights: An Open-Source Atlas of Global Aviation Data

A suite of clean, comprehensive, and interconnected databases for developers, researchers, and enthusiasts.

Why OpenFlights Dataset?

Real-world Relevance

This dataset maps critical global infrastructure, offering insights into economy, tourism, logistics, and disease spread that are meaningful beyond theoretical models.

Natural Complex Network

With airports as nodes and routes as directed edges, it's perfectly structured for studying hubs, small-world effects, community detection, and network resilience.

Algorithmic Tractability

At ~6,000 nodes and ~37,000 edges, it's large enough to reveal complex structures, yet manageable for advanced algorithms to run efficiently on a laptop.

Rich Geographic Context

Detailed attributes like country, city, and coordinates enable powerful world-map visualisations and geographic interpretation of network communities.

Open & Reproducible

Being a public dataset and a common benchmark in network science ensures easy reproducibility of results and fosters collaborative research.

Graph Construction

I utilised **OpenFlights dataset,** which initially contains approximately 7,698 airport entries and 67,663 flight.

Raw Data Overview

- 7,698 airport entries
- 67,663 flight routes

- Airport attributes: name, city, country, lat/long
- Route attributes: source to destination airport

Preprocessing & Cleaning

- Dropped routes with invalid source/destination IATA codes.
- Removed self-loops (A to A).
- Treated multiple parallel routes as one edge for structural analysis.

Final Directed Graph Model

- Nodes (Airports): 6,072
- Edges (Routes): 37,042

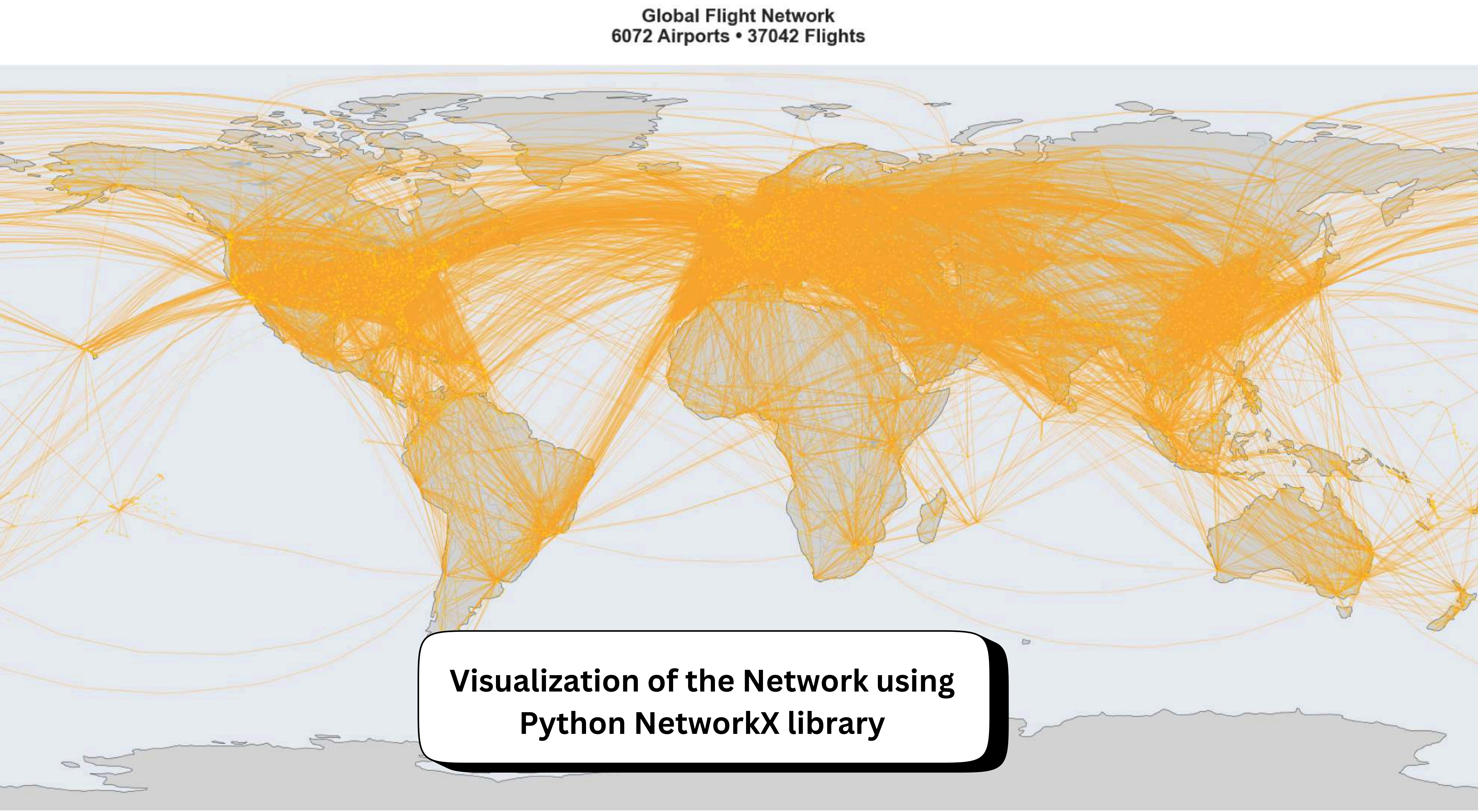
- A directed graph where nodes are airports (IATA codes) and a directed edge represents a scheduled route.

Formally, the resulting graph is defined as:

$$G=(V,E),\ |V|=6072,\ |E|=37042,$$

where each edge signifies a **directed airport-to-airport route.**

This yields a well-defined, consistent directed graph for robust analysis.



Visualization of the Network using Python NetworkX library

Basic Network Properties

Before deep diving, I computed fundamental network properties to understand its global structure.

Average Degree

In-degree ≈ 6, Out-degree ≈ 6. Total degree ≈ 12.

Density

Approx. 0.001, indicating an extremely sparse network.

Avg. Shortest Path

Approx. 3.97. On average, any two core airports are just ~ 4 flights apart, demonstrating incredible global efficiency.

Clustering Coefficient

Around 0.26, suggesting high clustering.

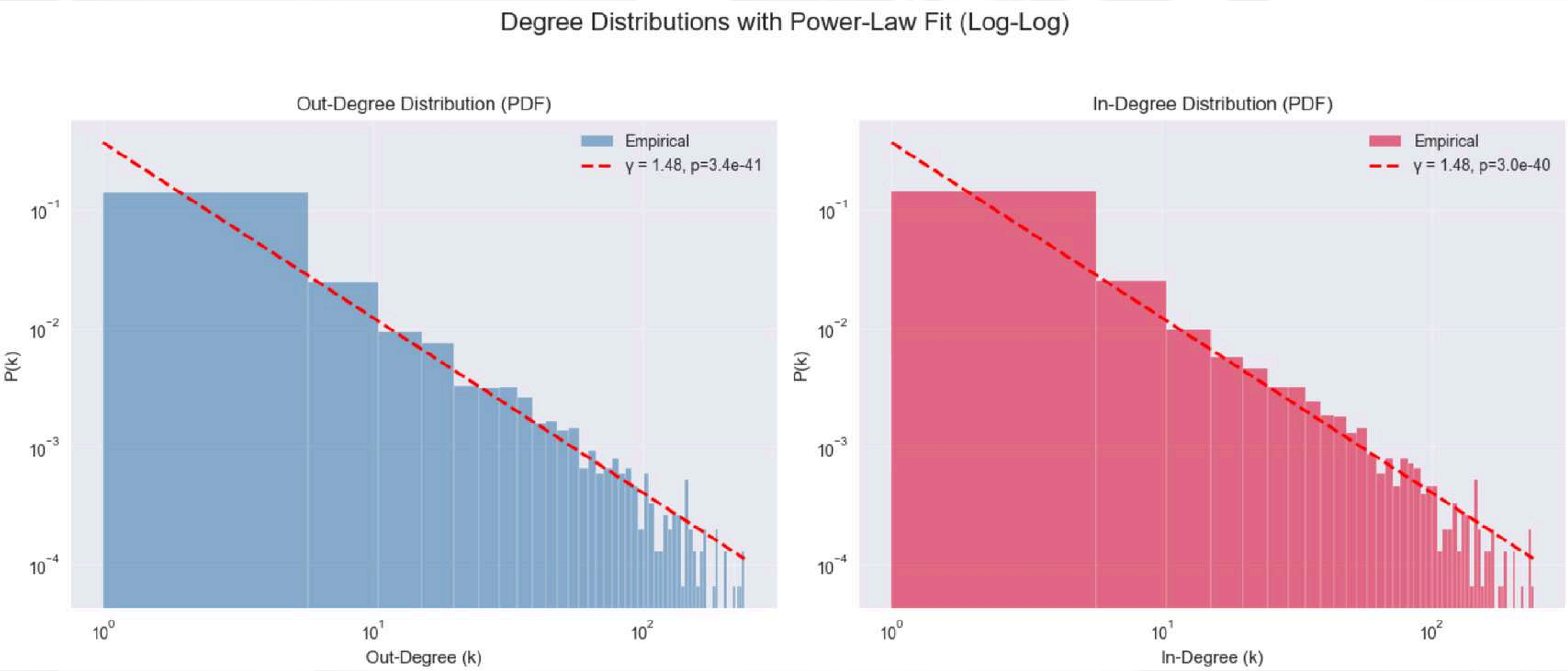
Reciprocity

Approx. 0.978, meaning most routes are bidirectional.

Assortativity

Slightly negative (−0.016), major hubs tend to connect with smaller airports, hinting at a mild hub–spoke pattern.

Degree Distribution and Scale-Free Behavior



Insight:

- The degree distribution is heavily right-skewed: most airports have very few connections, while a small number of hubs have hundreds.
- The linear relationship on the log-log plot is the classic signature of a heavy-tailed, near scale-free network.

Algorithms for Analysis: Uncovering Insights

Centrality Measures

Betweenness, PageRank, and closeness centrality to identify influential airports.

Connectivity Analysis

Metrics like the **Largest Strongly Connected component (LSCC)** to understand network structure.

Community Detection

Louvain method to identify functional clusters or modules within the global network.

Robustness Simulation

Simulating node removal (random vs. targeted) and observing the impact on network connectivity.

Algorithms for Analysis: Uncovering Insights

Centrality Measures

Betweenness, PageRank, and closeness centrality to identify influential airports.

Connectivity Analysis

Metrics like the Largest Strongly Connected component (LSCC) to understand network structure.

Community Detection

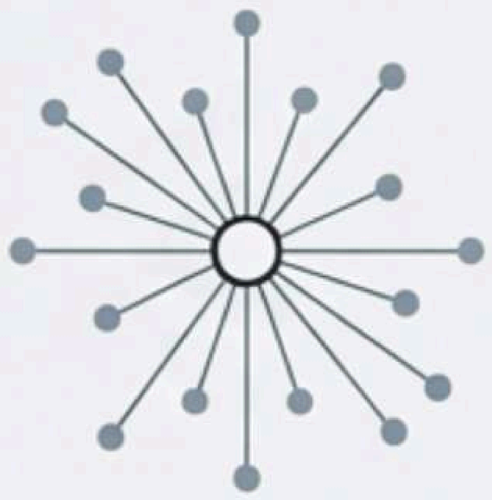
Louvain method to identify functional clusters or modules within the global network.

Robustness Simulation

Simulating node removal (random vs. targeted) and observing the impact on network connectivity.

Identifying Global Hubs: Centrality Measures

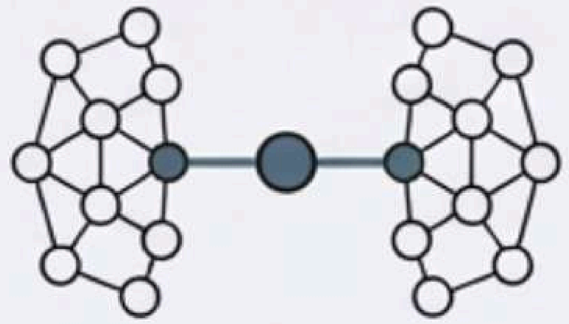
Identifying the Network's Vital Organs: Beyond Flight Counts



Degree Centrality

The "Popular Kid."

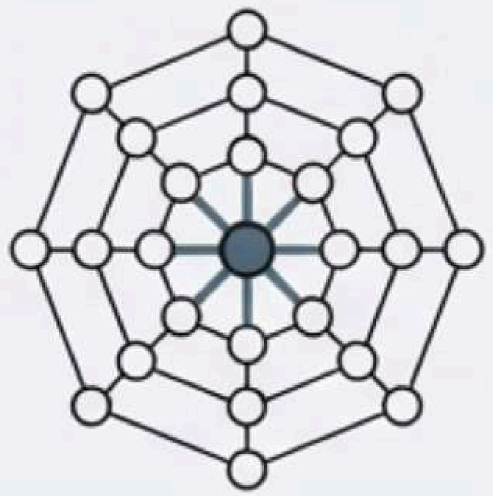
Measures the number of direct connections.



Betweenness Centrality

The "Bridge."

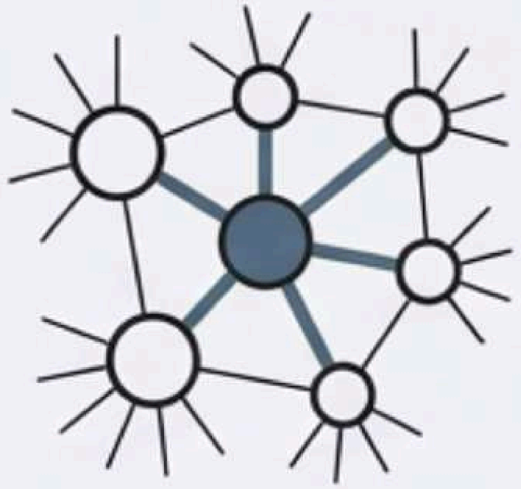
Measures how often an airport lies on the shortest path between others. Critical for connecting disparate regions.



Closeness Centrality

The "Geographic Center."

Measures how quickly an airport can reach all others. Essential for rapid global access.

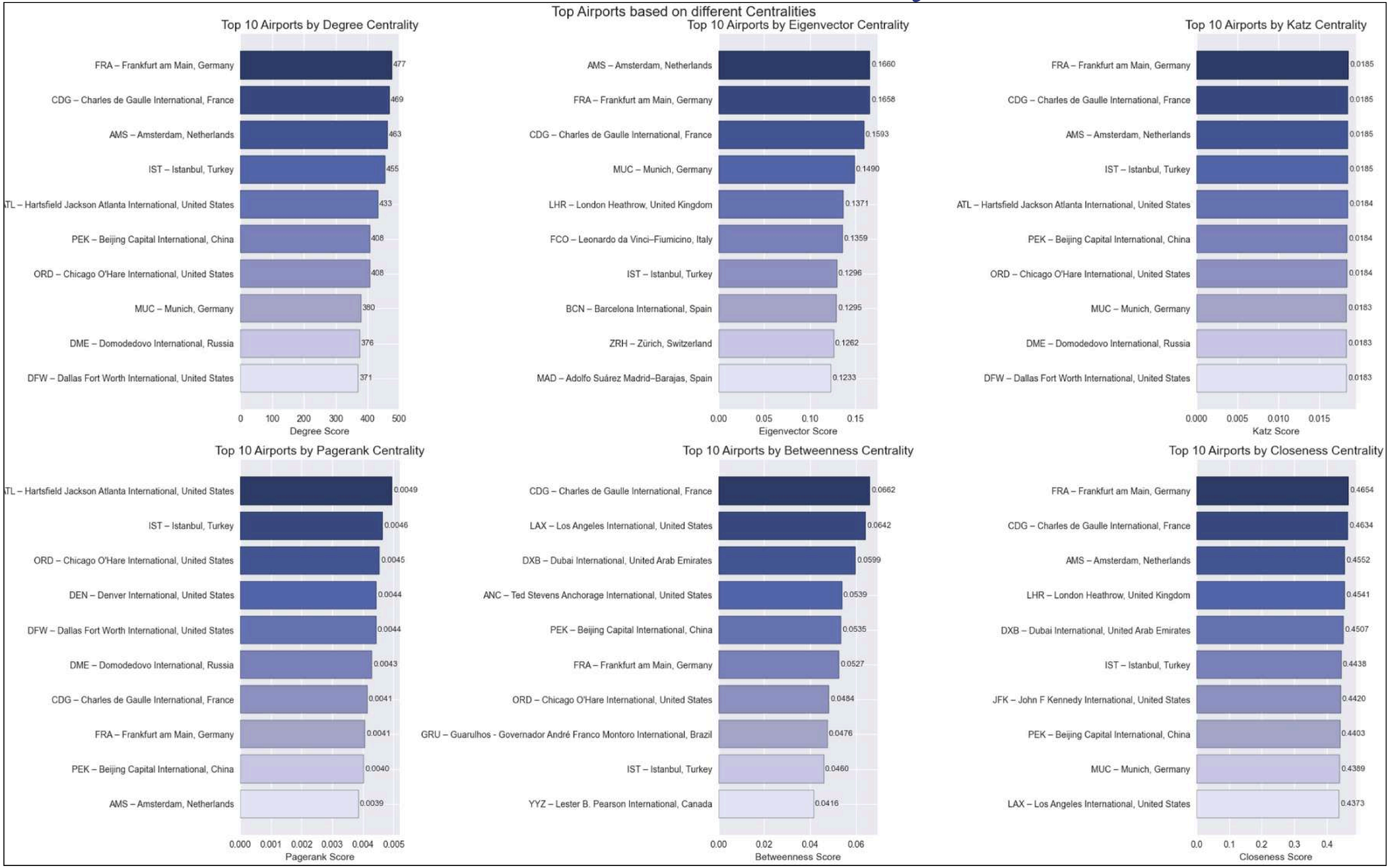


PageRank

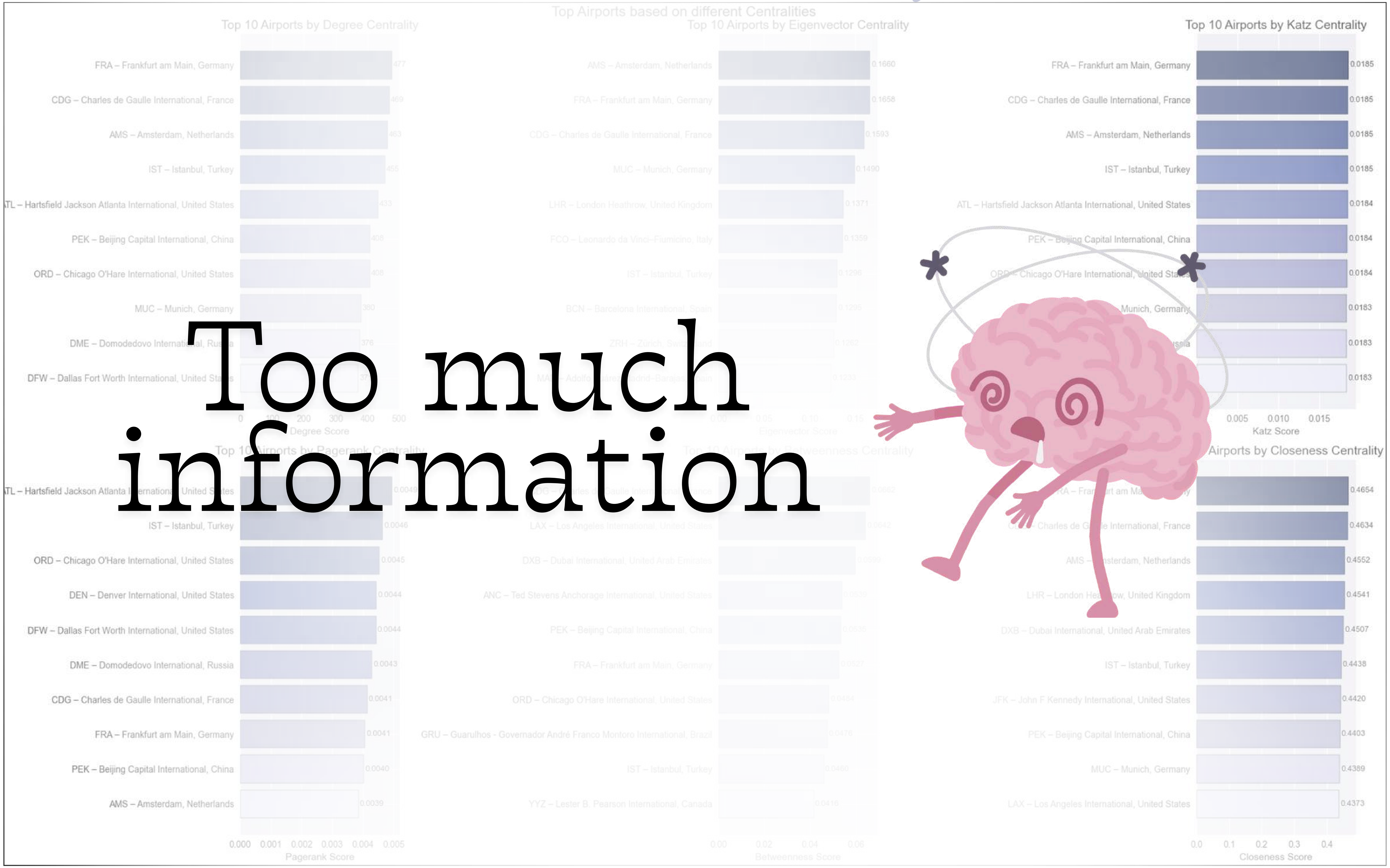
The "Influencer."

Measures importance based on the importance of its connections. A high PageRank airport is a hub connected to other hubs.

Global Hubs based on Centrality Measures



Global Hubs based on Centrality Measures



Too much information

Different Centraility Telling, Different stories

Airport	**Megahub** *(Degree)*	**Influencer** *(Eigenvector/Katz)*	**Destination** *(PageRank)*	**Keystone** *(Betweenness)*	**Navigator** *(Closeness)*
FRA	●	●			●
CDG	●	●		●	●
AMS	●	●			
ATL	●		●		
DXB				●	●
IST	●	●	●		●
LHR					●
ANC				●	

So, Which is the World's Most Important Airport?

Answer Depends on the Mission

For maximum direct options... the **Megahubs** like Frankfurt (FRA) are your target.

For strategic global transfers... rely on the **Keystones** like Dubai (DXB) or Anchorage (ANC).

For access to the US market... the center of gravity is the **Destination,** Atlanta (ATL).

For the most efficient global reach... start at a **Navigator** like Paris (CDG).

Algorithms for Analysis: Uncovering Insights

Centrality Measures

Betweenness, PageRank, and closeness centrality to identify influential airports.

Connectivity Analysis

Metrics like the **Largest Strongly Connected component (LSCC)** to understand network structure.

Community Detection

the global network.

Robustness Simulation

on network connectivity.

Connectivity Analysis: Connected Components

The network's reachability isn't uniform.

Why LSCC is important:

The Largest Strongly Connected Component (LSCC) is the core where every airport can reach every other via directed flights.

Insight:

- **52% of airport carry 99.75% routes.**

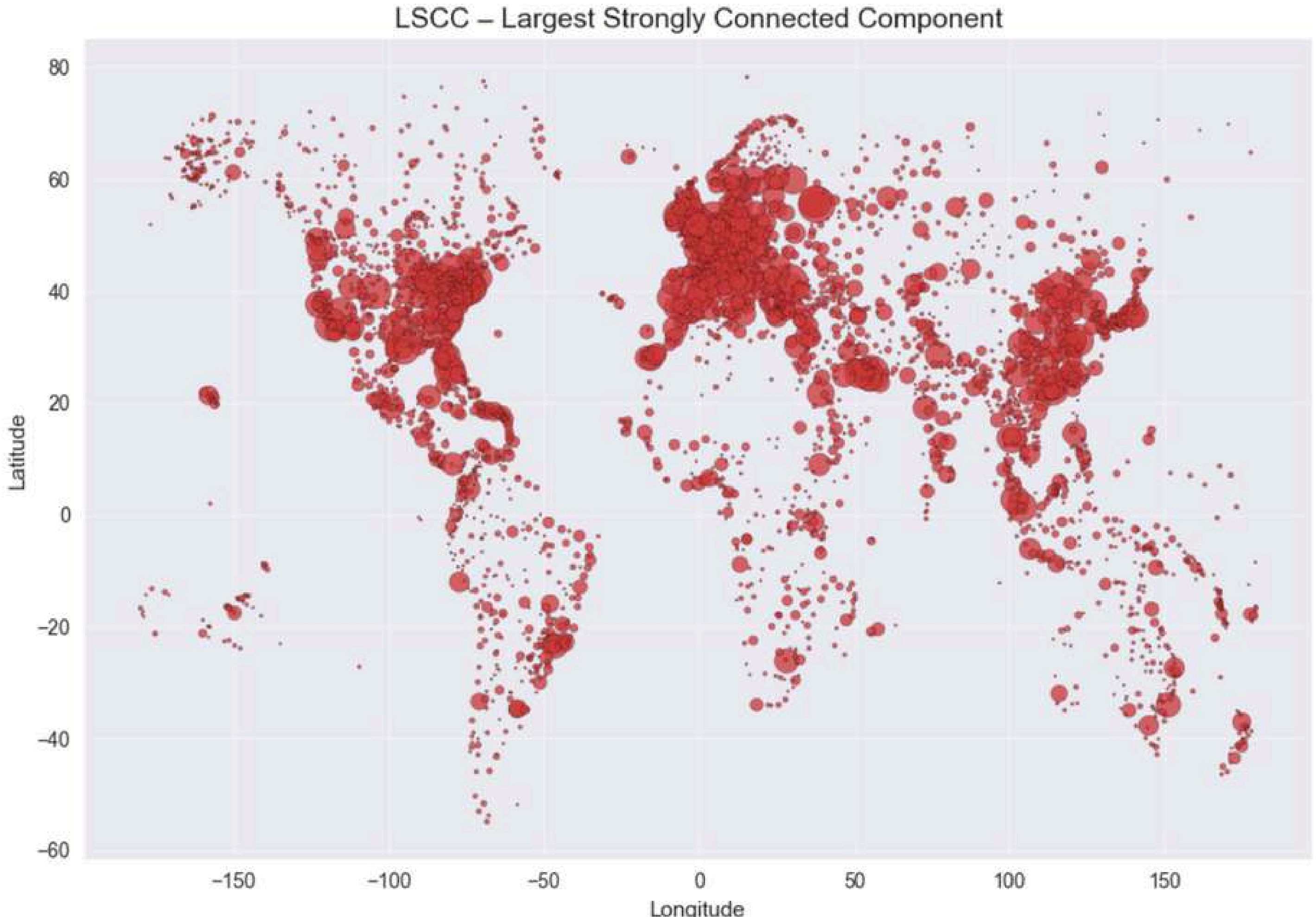
This reveals a *core-periphery structure*.

- Just over half the airports form a dense, interconnected backbone that supports virtually all global air travel,
- While the other half are peripheral.

Key Findings:

The Largest Strongly Connected Component (LSCC) has

- Nodes: 3,190 (52.5% of airports)
- Routes: 99.75% of all routes



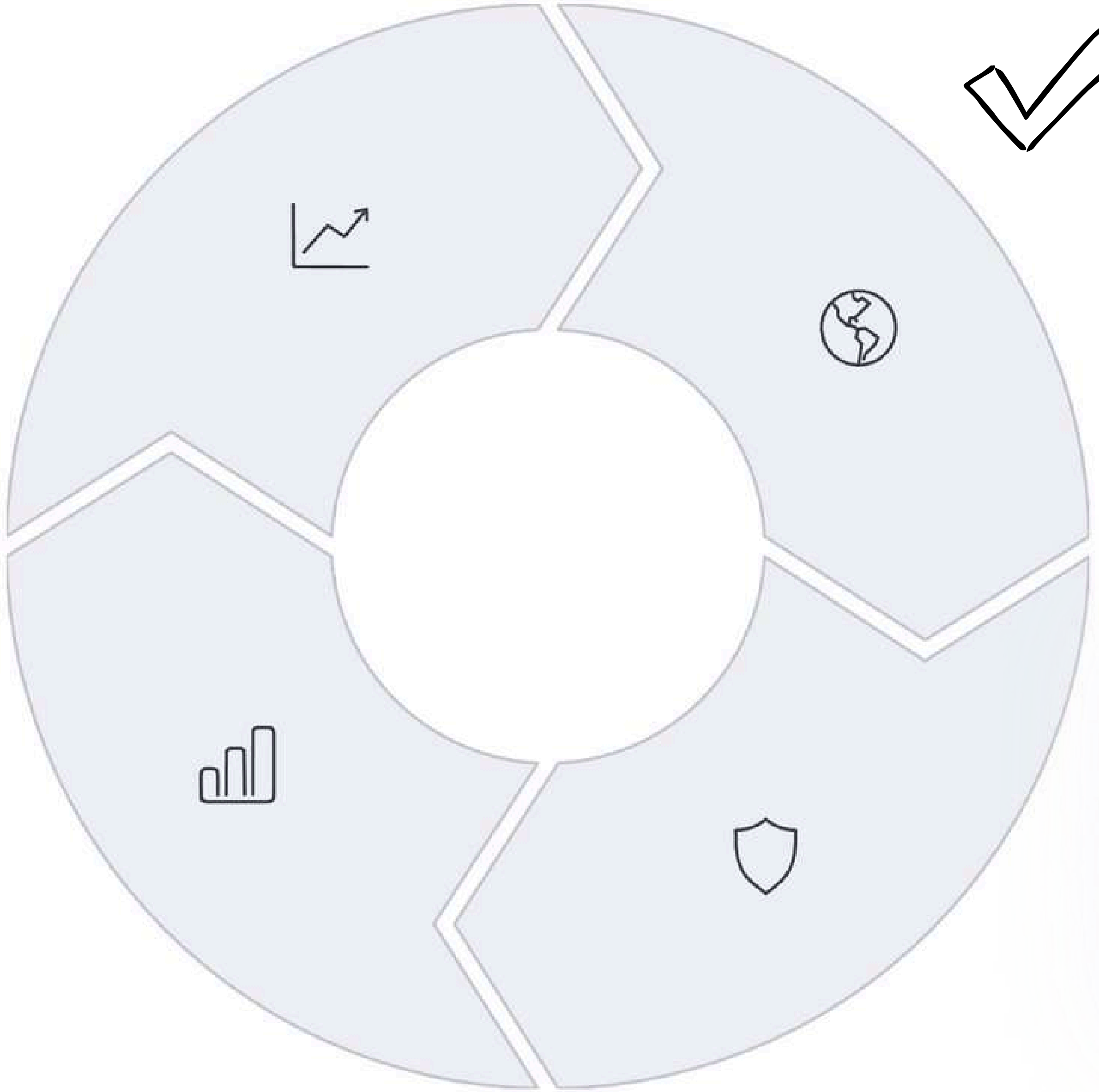
Algorithms for Analysis: Uncovering Insights

Centrality Measures

Betweenness, PageRank, and closeness centrality to identify influential airports.

Community Detection

Louvain method to identify functional clusters or modules within the global network.



Connectivity Analysis

Metrics like the **Largest Strongly Connected component (LSCC)** to understand network structure.

Robustness Simulation

Simulating node removal (random vs. targeted) and observing the impact on network connectivity.

Community Detection

Algorithm Used

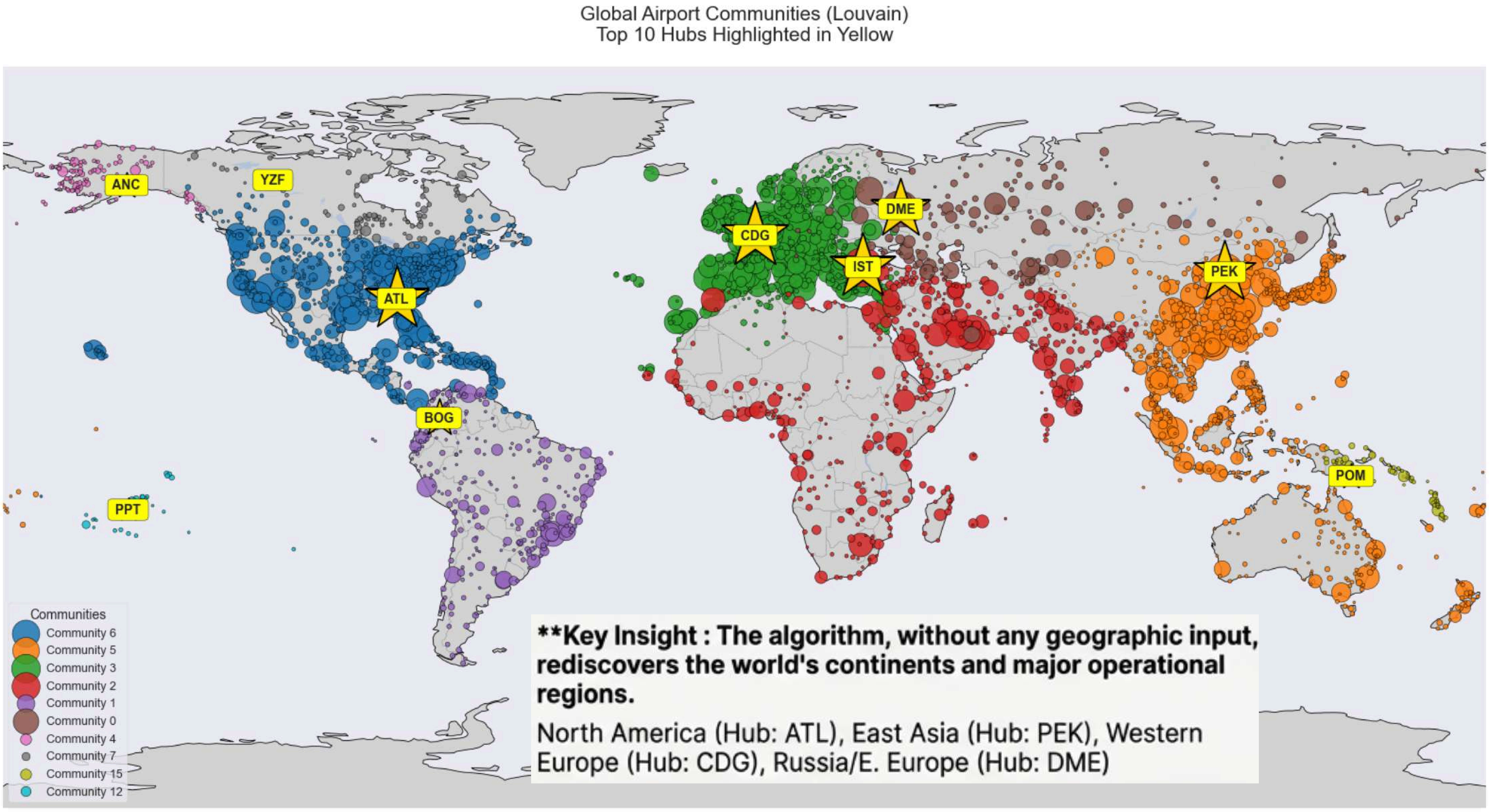
Louvain (PLM) community detection for modularity maximisation.

Why Use Louvain?

- Identifies natural clusters of airports with higher internal connectivity.
- **It calculates modularity score (Q)** indicates a highly meaningful partition.
- Louvain wins on Speed + Accuracy + Scalability + Reproducibility

KEY FINDINGS:

- **Communities Found: ~20**
- Communities align with geographic region.
- Identifies regional hubs (e.g., ATL, PEK, CDG, IST, BOG, DME, ANC) as centers of these communities
- **Modularity: 0.65,** which is > 0.5 indicates a strong and meaningful community structure.


Global Airport Communities (Louvain)
Top 10 Hubs Highlighted in Yellow
ANC
YZF
ATL
BOG
PPT
CDG
DME
IST
PEK
POM
Communities
Community 6
Community 5
Community 3
Community 2
Community 1
Community 0
Community 4
Community 7
Community 15
Community 12
**Key Insight : The algorithm, without any geographic input, rediscovers the world's continents and major operational regions.
North America (Hub: ATL), East Asia (Hub: PEK), Western Europe (Hub: CDG), Russia/E. Europe (Hub: DME)

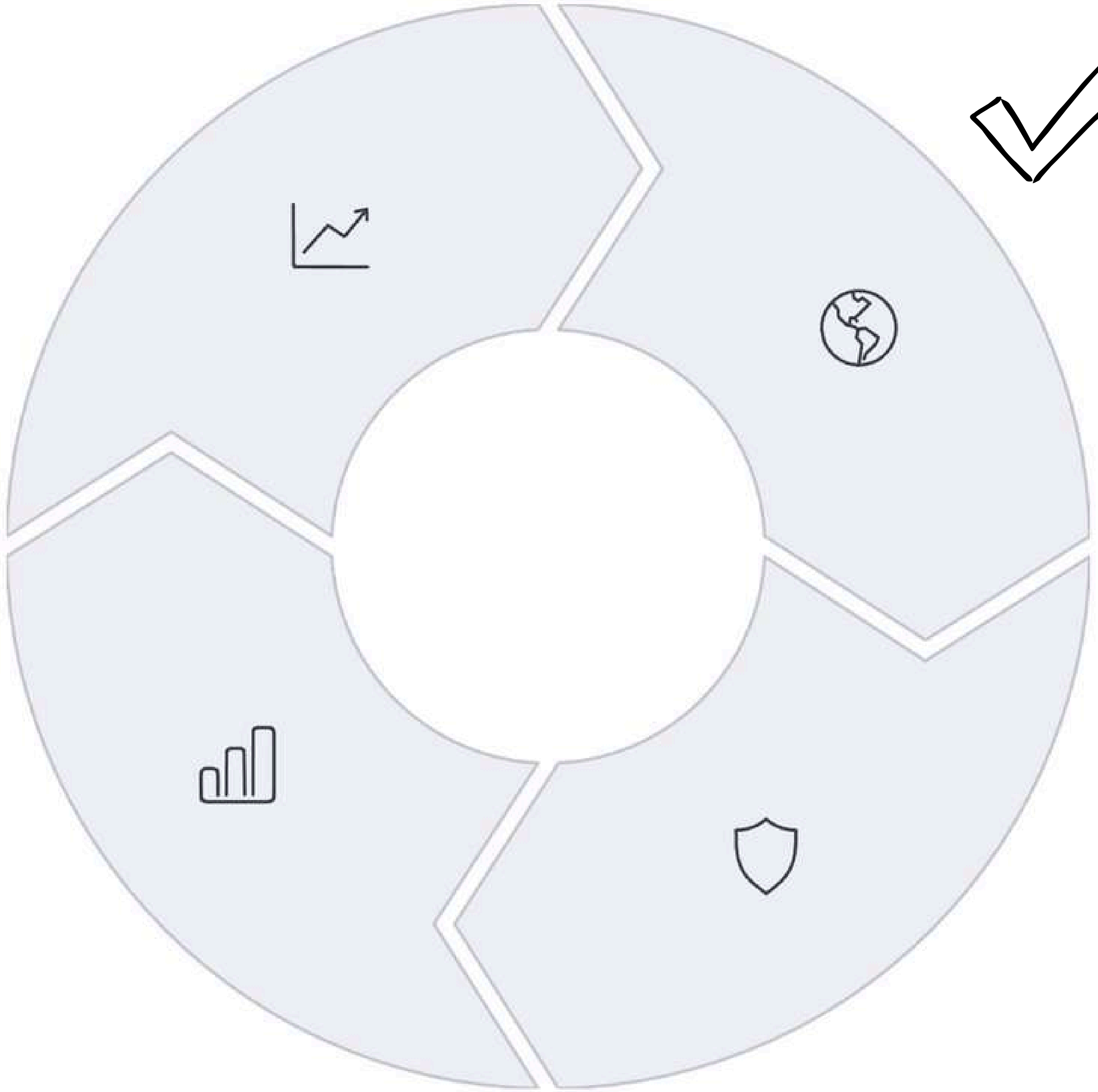
Algorithms for Analysis: Uncovering Insights

Centrality Measures

Betweenness, PageRank, and closeness centrality to identify influential airports.

Community Detection

Louvain method to identify functional clusters or modules within the global network.



Connectivity Analysis

Metrics like the **Largest Strongly Connected component (LSCC)** to understand network structure.

Robustness Simulation

Simulating node removal (random vs. targeted) and observing the impact on network connectivity.

Resilience Analysis

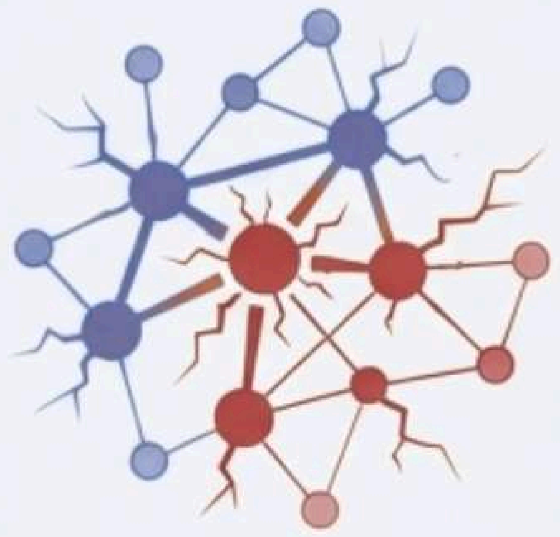
The Stress Test: How Does the Network Behave When It Breaks?

1. Random Failures

Simulating random airport closures or minor, distributed disruptions.

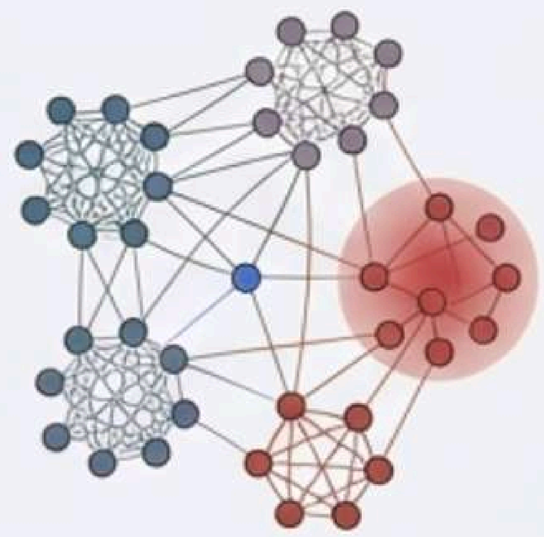
2. Targeted Attacks

Simulating the failure of the most critical hubs (by Degree and Betweenness).



3. Regional Collapse

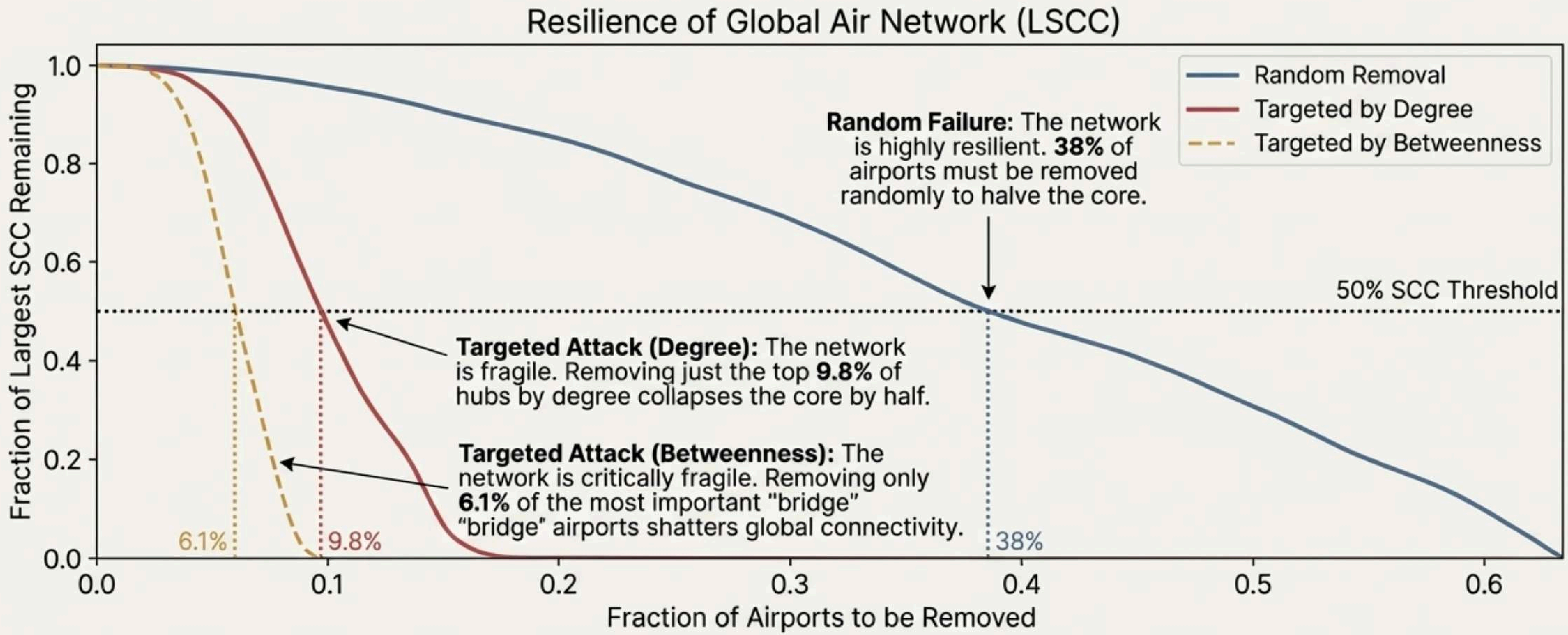
Simulating the failure of entire communities, a novel analysis for events like pandemics or regional conflicts.



Metric for Failure: The size of the largest strongly connected component (LSCC). We track the point where this core shrinks by 50%.

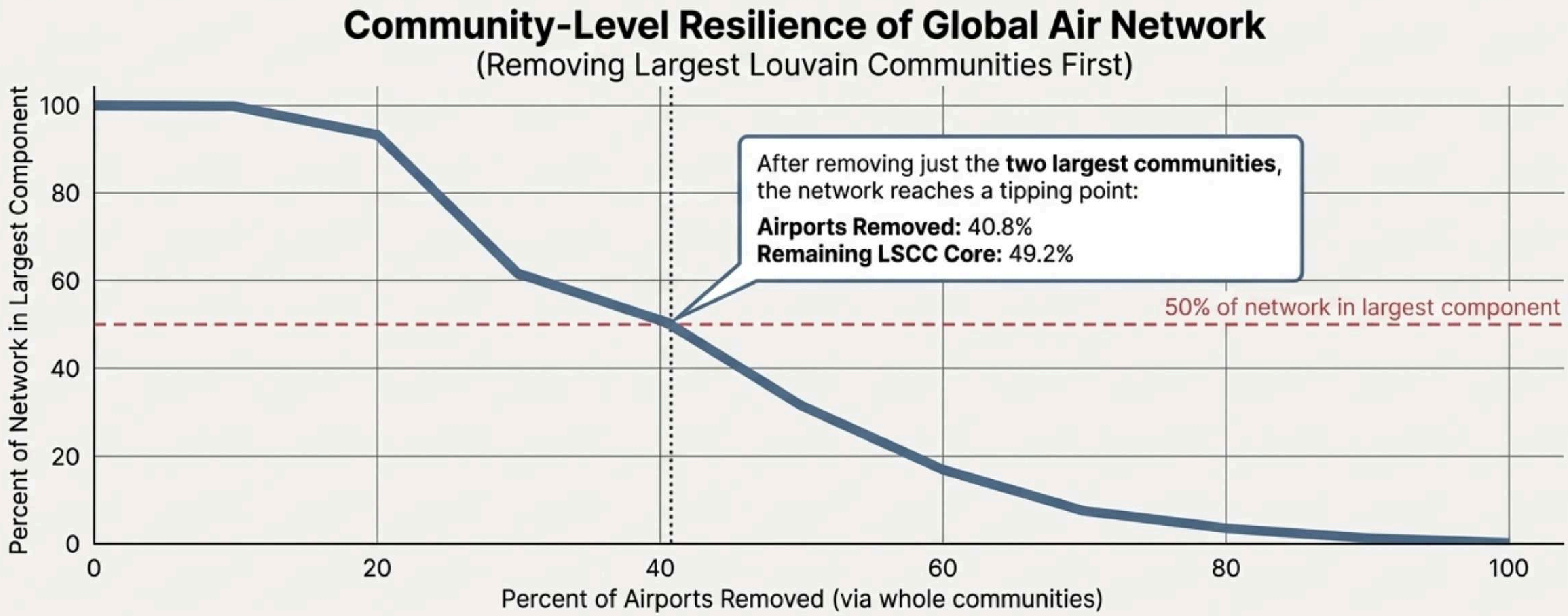
Random vs Targeted Removal

The Network is Robust to Error, but Fragile Under Attack



Communities Removal

The Failure of a Single Region Can Cripple the Global System



Findings of the Stress Test

Finding 1: Robust to Random Failure

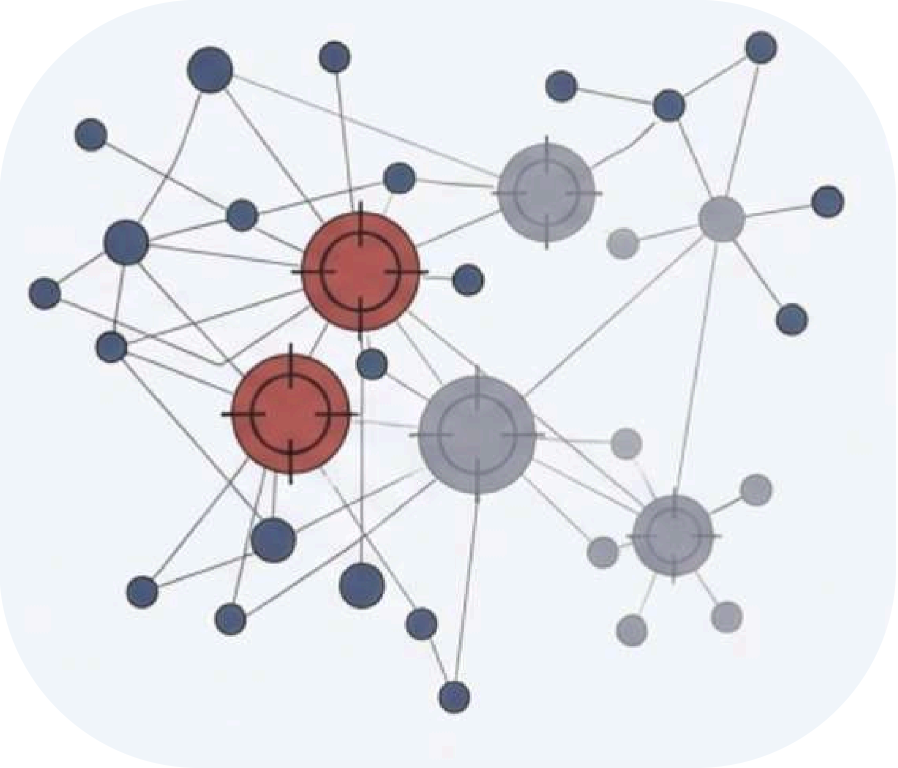
- Random removal barely affects the network; it takes removing 38% of airports randomly to break half of the LSCC.

Finding 2: Fragile to Targeted Attack

- Removing just ~10% of the highest-degree airports fragments the network.
- Removing ~7% of the highest-betweenness airports breaks it even faster.

Finding 3: Extremely Fragile to Communitiy Attack

- The largest component drops below 50% of the network at:
  - On Step: 2 (after removing community ID 5)
  - Airports removed: 1303.0 (40.8% of all airports)
  - Largest component size: 1569.0 airports (49.2% of network)



The Paradox of the Global Air Network

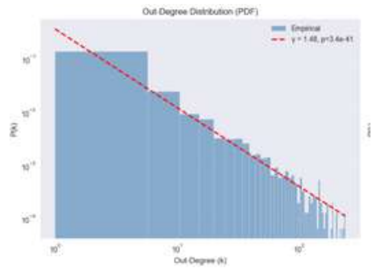
Efficiency Creates Fragility.

Hubs enable short paths but are single points of failure.

Communities create efficient regional flows but introduce risk of large-scale, cascading collapse.

The very architectural properties that make the global air network so remarkably efficient are the same properties that create its critical vulnerabilities. This duality is the fundamental truth of the system.

The Five Core Questions: Now Answered



1. **Scale-Free Structure: (Yes.)** The heavy-tailed degree distribution confirms a hub-dominated architecture.

2. **Small-World Effect: (Yes.)** The combination of short paths (~4 hops) and high clustering (0.26) creates an efficient, locally-structured network.

3. **Key Hubs: (A small, critical set.)** Centrality metrics identify distinct roles: Mega-Hubs, Bridges, and Global Centres.

4. **Communities: (20 distinct regions.)** Louvain detection with high modularity (0.655) reveals geographically coherent communities that form the network's building blocks.

5. **Resilience: (A story of dualities.)** The network is robust to random failure (38% threshold) but fragile against targeted (6-10%) and regional (2 communities) disruption.

The Global Air Network: Efficient but Fragile

This infographic summarizes a complex network analysis of the global air transportation system, using a dataset of 6,072 airports and over 37,000 flight routes. The analysis reveals a system that is remarkably efficient for global travel but paradoxically vulnerable to targeted disruptions.

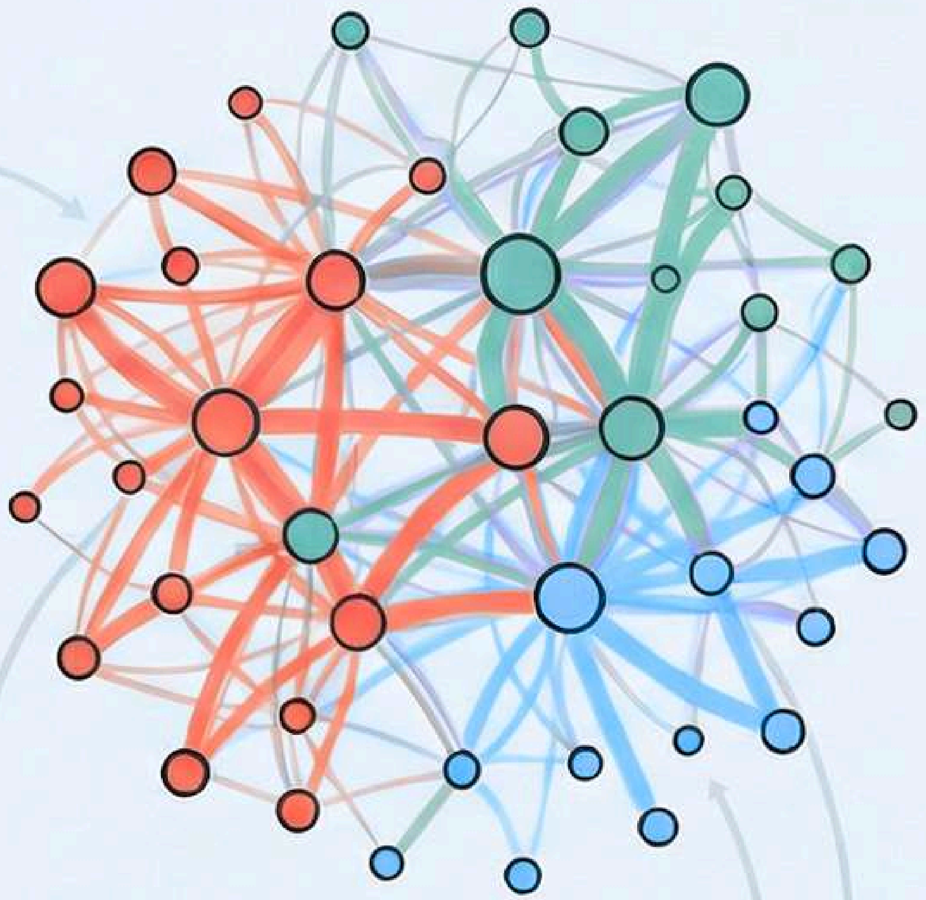
An Efficient, Small-World System

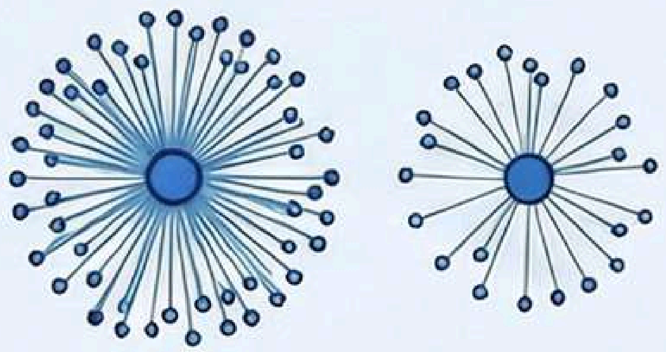
It's a "Small World"

On average, any two sirports in the network's core are only about 4 flights spart.

Dominated by Hubs

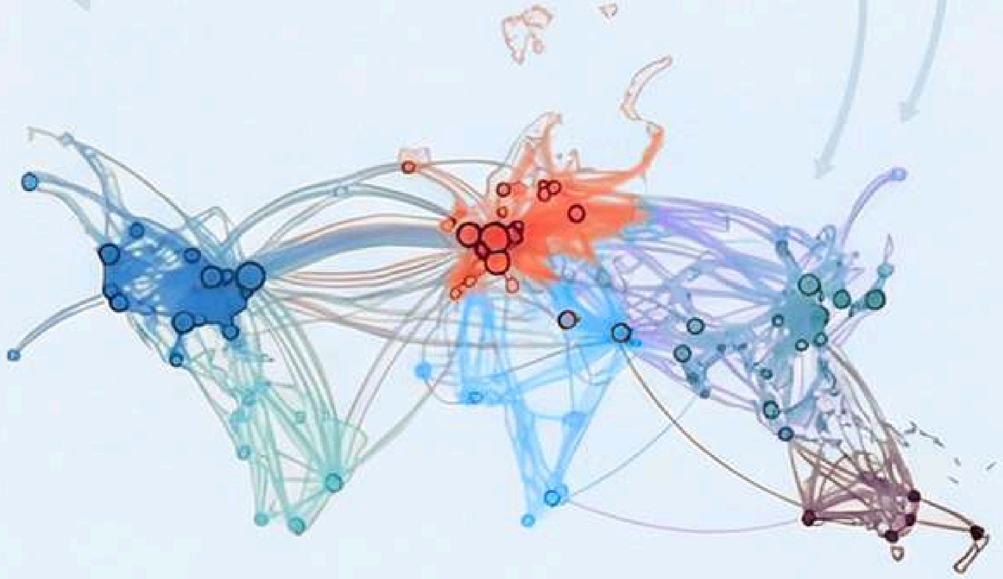
The network is "scale-free": most airports have few connections, while a few megahubs have hundreds.





Organized into Natural Regions

The network clusters into ~20 distinct geographic communities, like North America or East Asia.



A Fragile System Under Stress

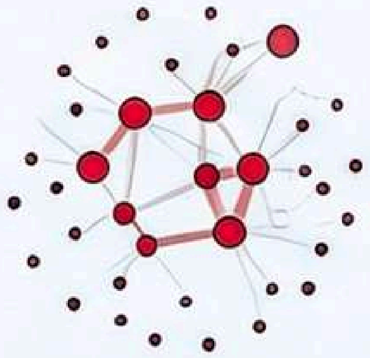
Robust to Random Failures, Fragile to Targeted Attacks

Random Failures (38% of airports)

It takes removing 38% of airports randomly to break the network.

Targeted Attacks (6.1% of "bridge" airports)

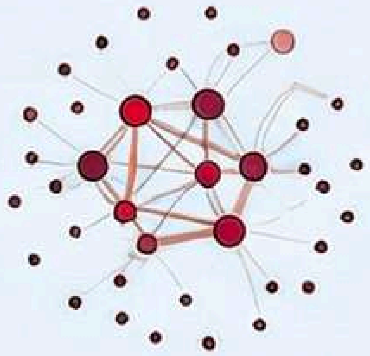
...but only 6.1% of key "bridge" airports.



9.8%

Critical Tipping Point

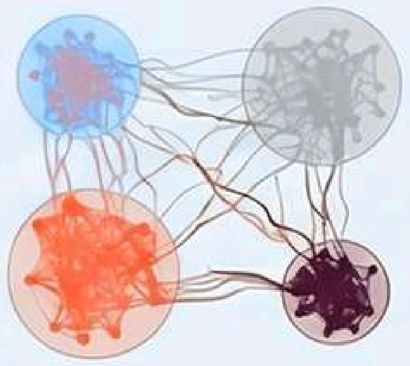
Removing just the top 9.8% of hubs by number of connections collapses the network core by half.



2 Communities

Community Collapse

The fallure of just the two largest regional communities can cripple the entire global system.



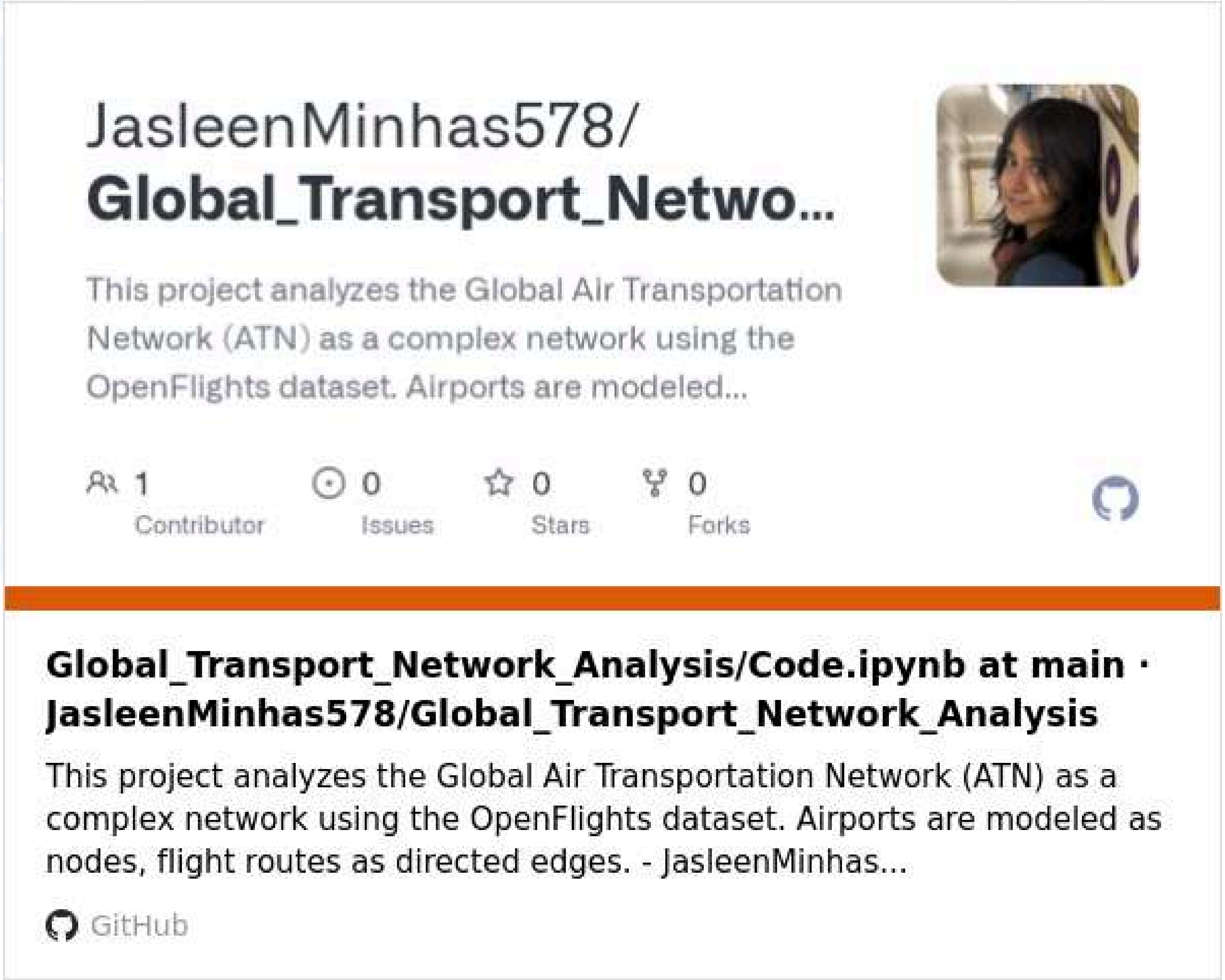
Airport Archetypes & Their Roles

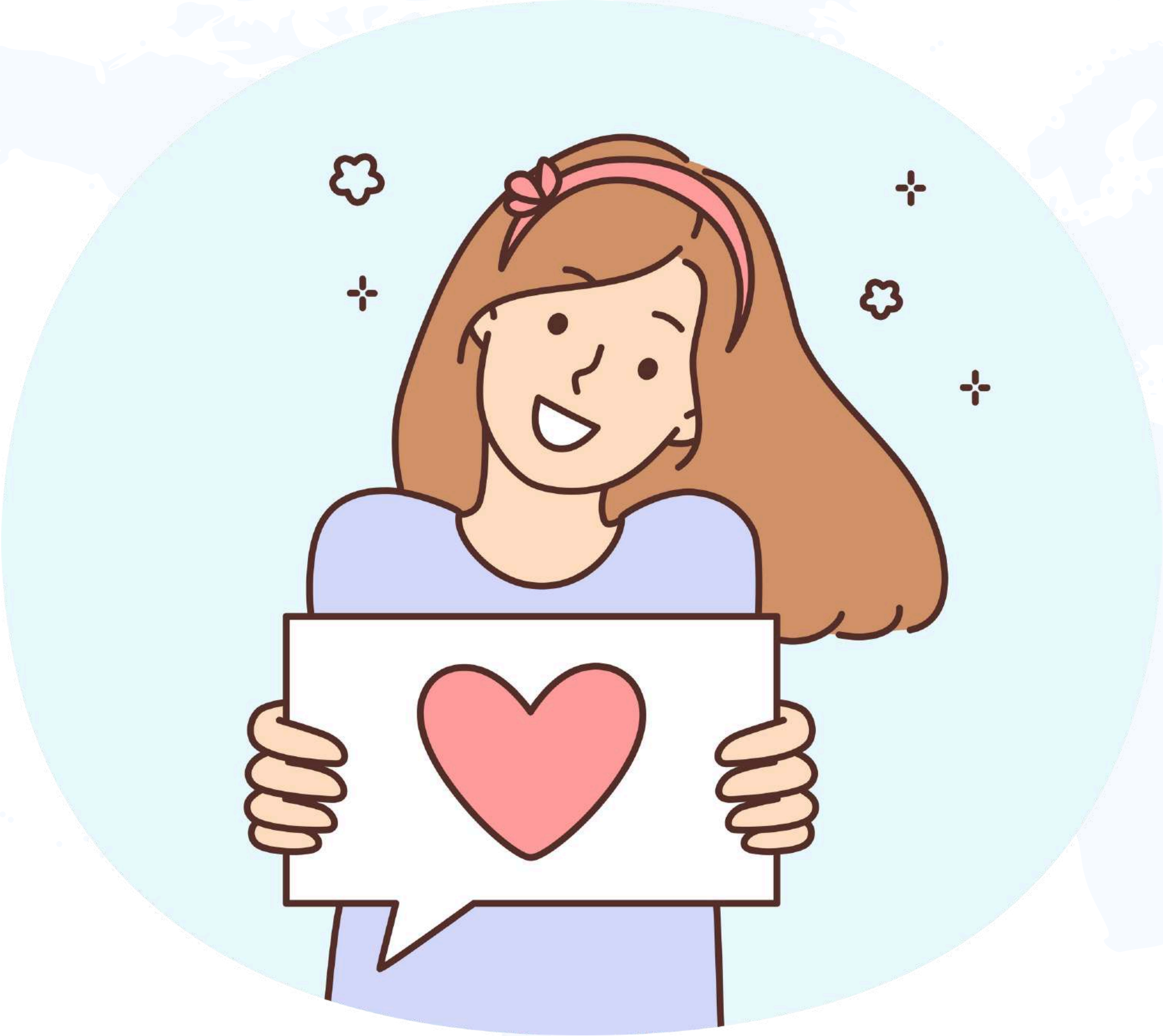
	Megahub (Mest Flights)	Influencer (Well-Connected)	Destination (Gravity Center)	Kaystone (Global Bridge)	Navigator (Efficient Reach)
FRA (Frankfurt)	●	●			●
CDG (Paris)	●	●		●	●
ATL (Atianta)	●		●		
DXB (Dubai)				●	
ANC (Anchorage)				●	

Want to Explore the Investigation?

For More Technical Details

- The complete analysis is documented in a Jupyter Notebook on GitHub
- **You'll find:**
  - Data cleaning and graph construction
  - Centrality, community detection, and resilience code
  - Comments explaining each technical choice
  - In-depth interpretation of all results





Thank You For Your Attention