U.S. Airport Network

By Alan Wang,
Brenden Cheung,
Kun Xu

Introduction

Literature Review

- "A study of the U.S. domestic air transportation network: temporal evolution of network topology and robustness from 2001 to 2016", by Siozos-Rousoulis, L., Robert, D. & Verbeke, W.

Main Goal

- Measure the airport development in terms of:
  - Flight Route Map
  - Relative importance among airports
  - Resilience of air transportation network
- With a span of 10 years

Table of Contents

- Flight Route Map
- Centrality
- Assortativity
- Small-World Effect
- Scale Free Property
- Targeted Attack Tolerance

Dataset, Cleaning, EDA

- Data Set: all domestic flights from 2009 to 2018
- Source: “Airline Delay and Cancellation Data, 2009 - 2018”

Data Cleaning

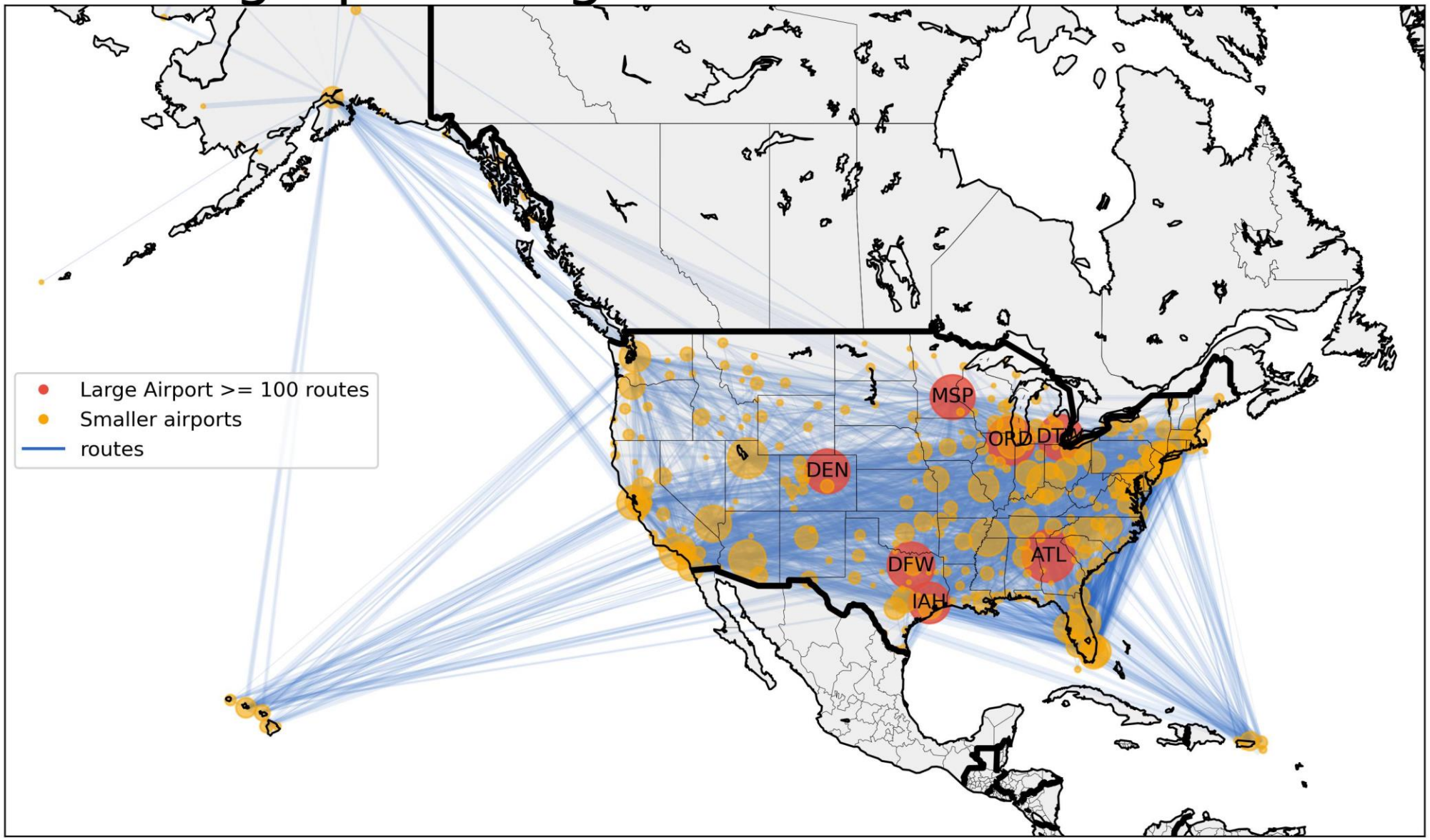
Original dataset: ~ 800 MB / Year; we have 10 years; that’s 8GB

We only want the `ORIGIN`, `DESTINATION`, `AIRLINE`, `Date`, and `DELAY`.

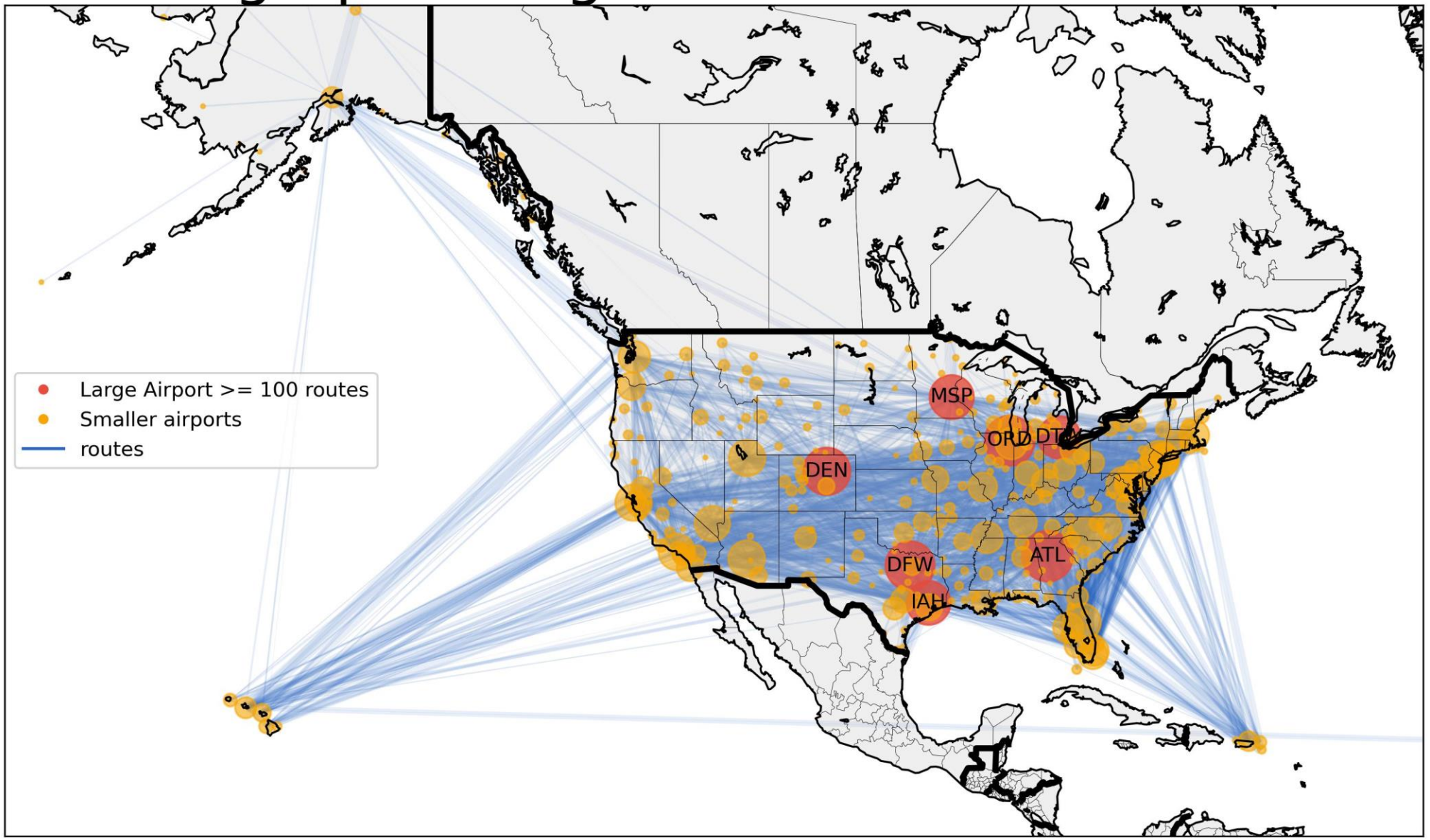
We group everything together by `AIRLINE`, `ORIGIN`, and `DESTINATION` and compute the total number of flights and average delay.

Through SQL => Reduced the dataset to ~10MB / Year

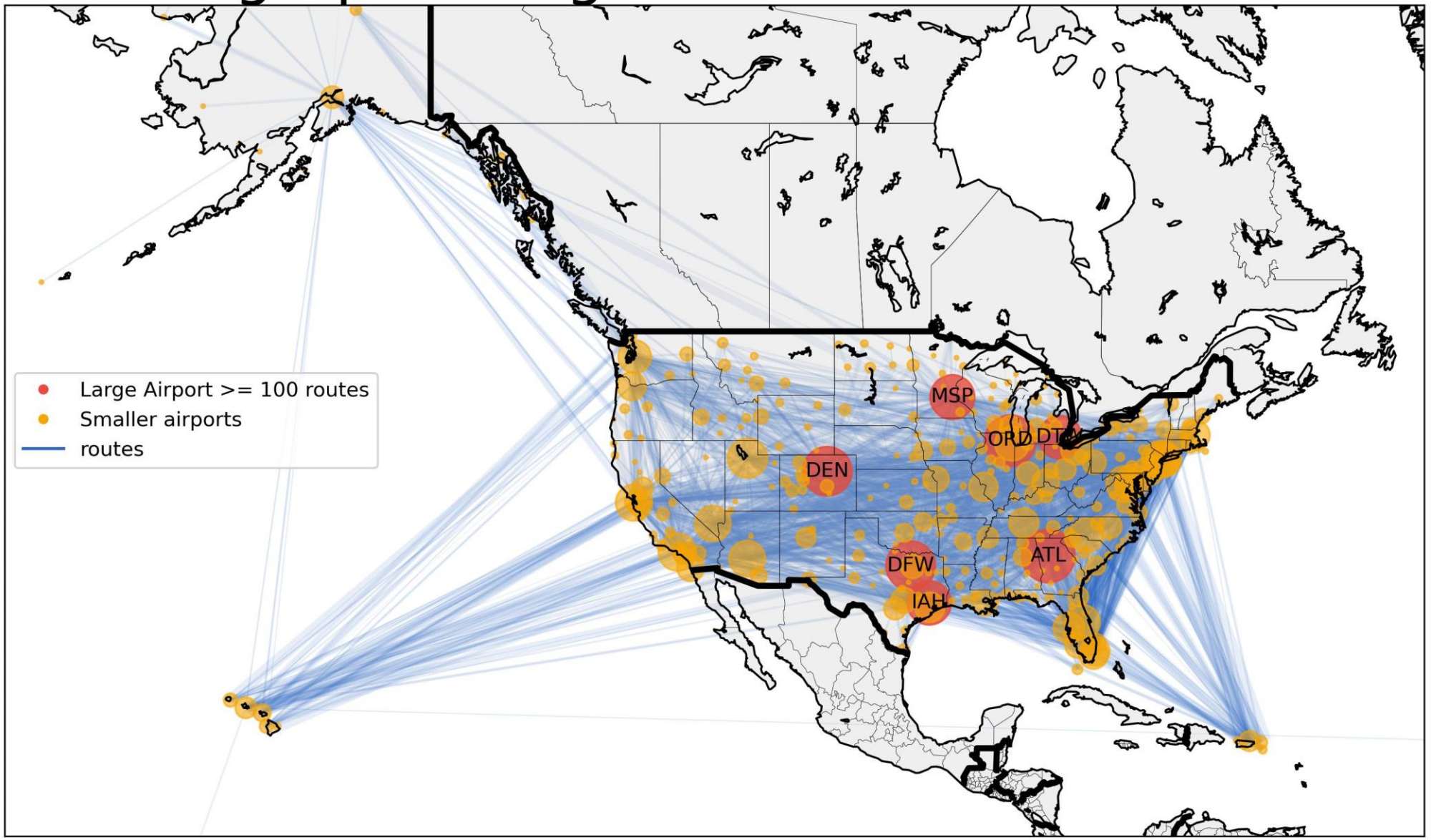
Network graph of flight routes in the USA in 2009



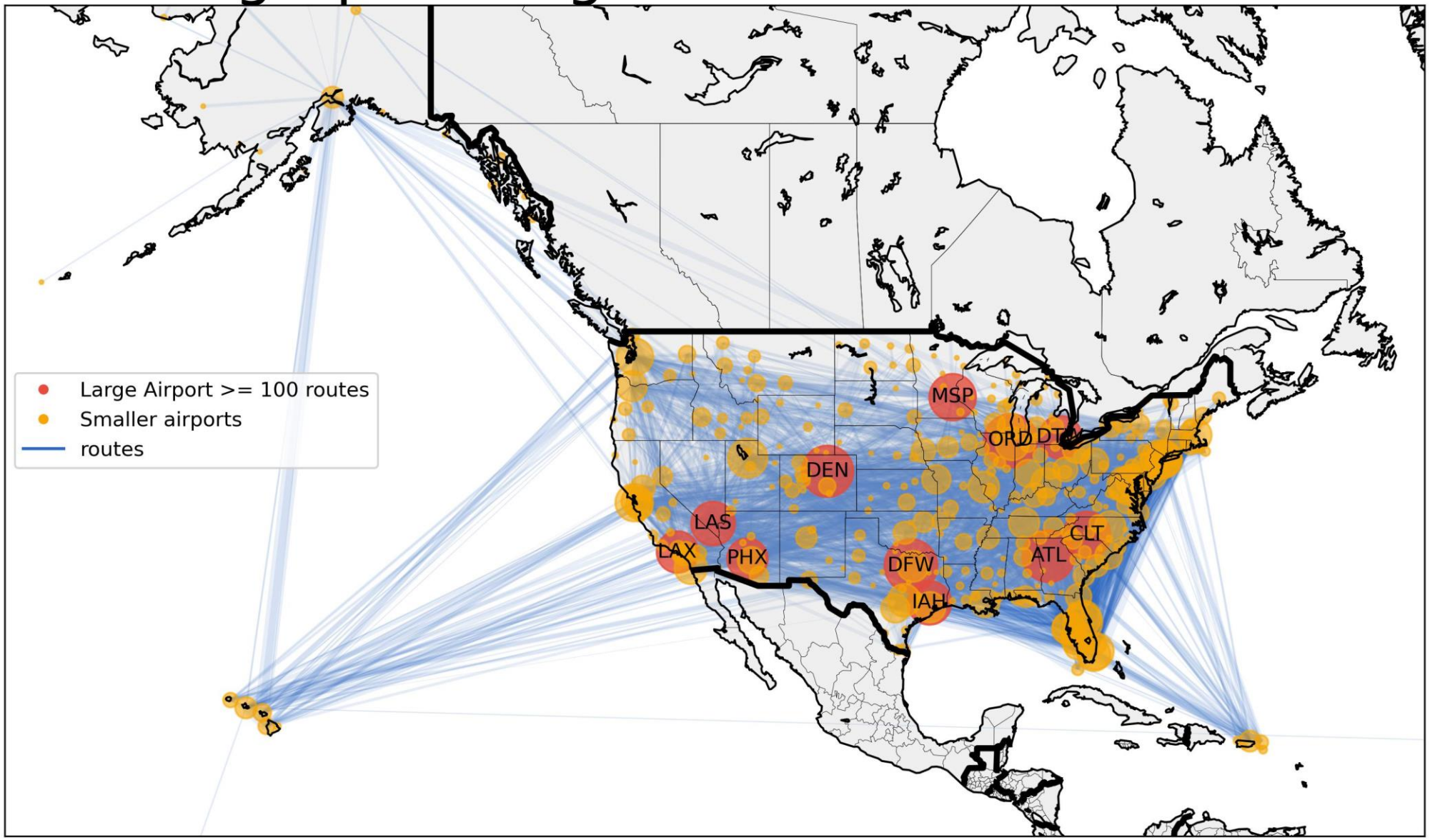
Network graph of flight routes in the USA in 2012



Network graph of flight routes in the USA in 2014



Network graph of flight routes in the USA in 2018



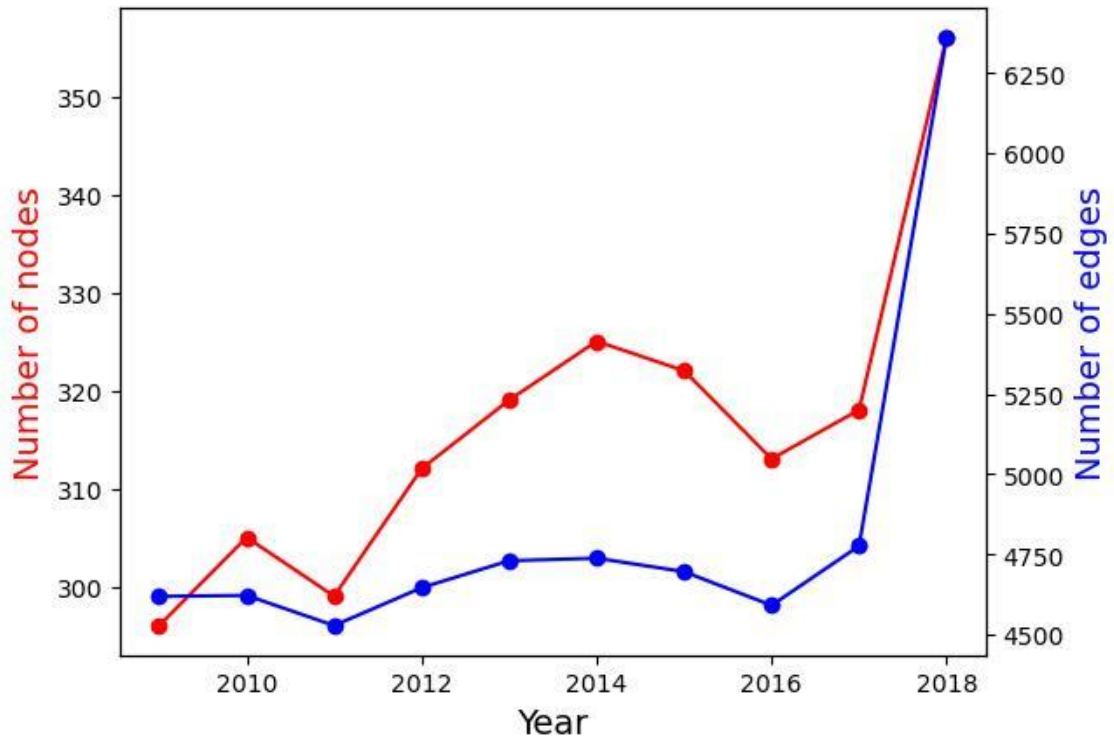
Centrality

Node Degree (k)

- the number of nodes that a node is pointing to
- $k_i = \sum_{j=1}^{N} A_{ij}$
  - k_i = the degree of node i
  - A_{ij} = the adjacency matrix of the network
- As the node degree increases, the importance of a node increases as well

Betweenness Centrality (B)

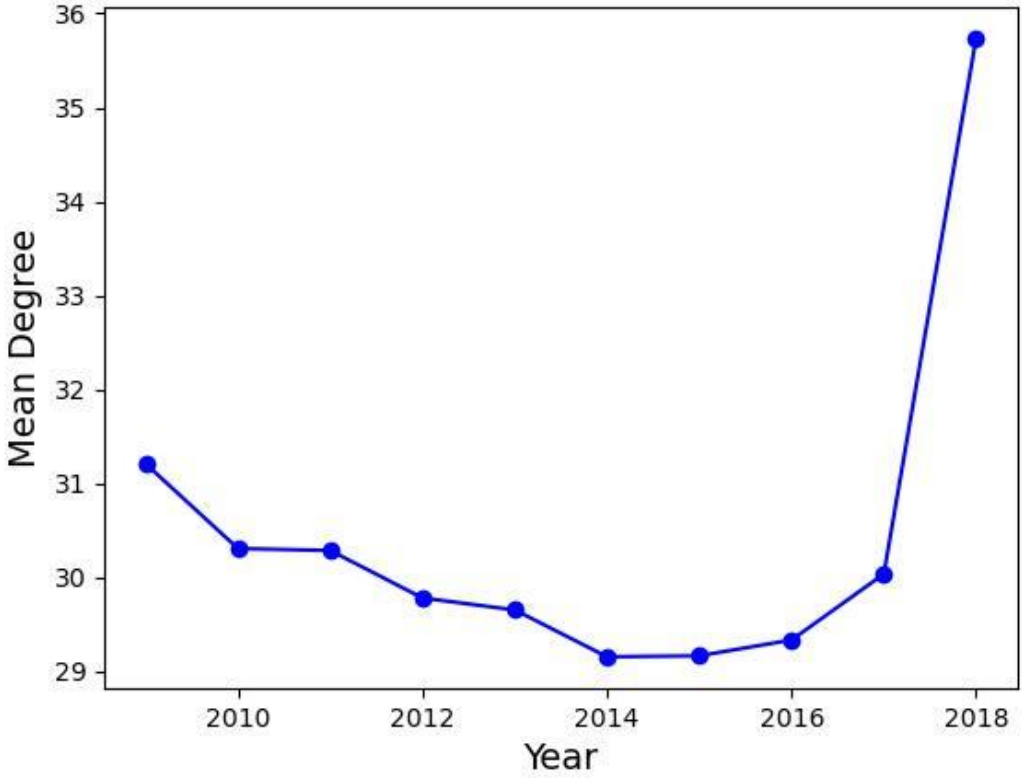
- $B_i = \frac{1}{(N-1)(N-2)} \sum_{i \neq j \neq k} \frac{g_{jk}(i)}{g_{jk}}$
  - $g_{jk}(i)$ = the number of binary shortest paths passing through node i
  - g_{jk} = the number of binary shortest paths between two nodes
- As the number of shortest paths through node i increases, B_i will increase accordingly
- measures the relative importance of an airport to the aviation system



Number of Nodes & Edges

Mean Degree over Years

Smaller airports are constructed

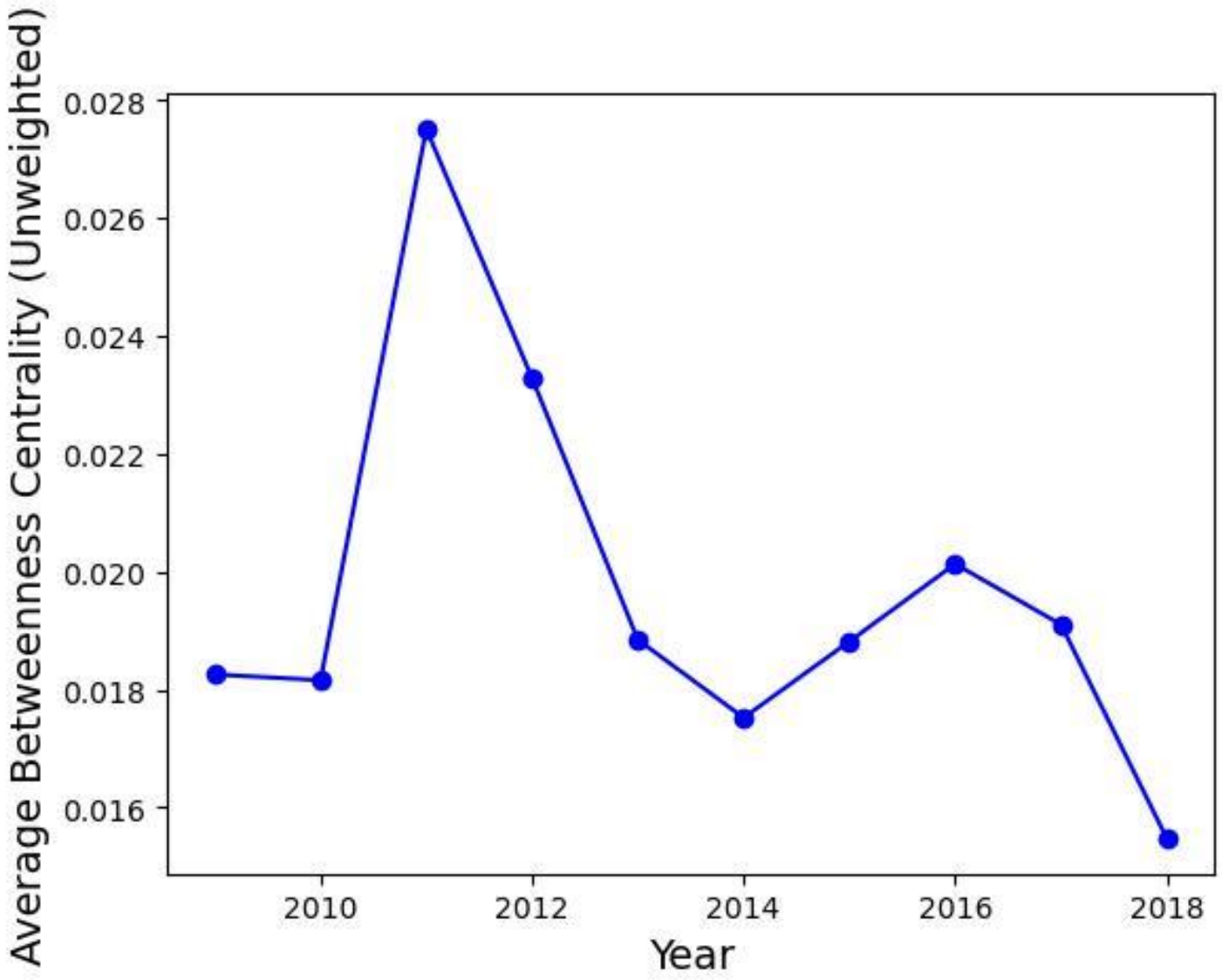


Topeka Regional Airport

Topeka Regional

Airport

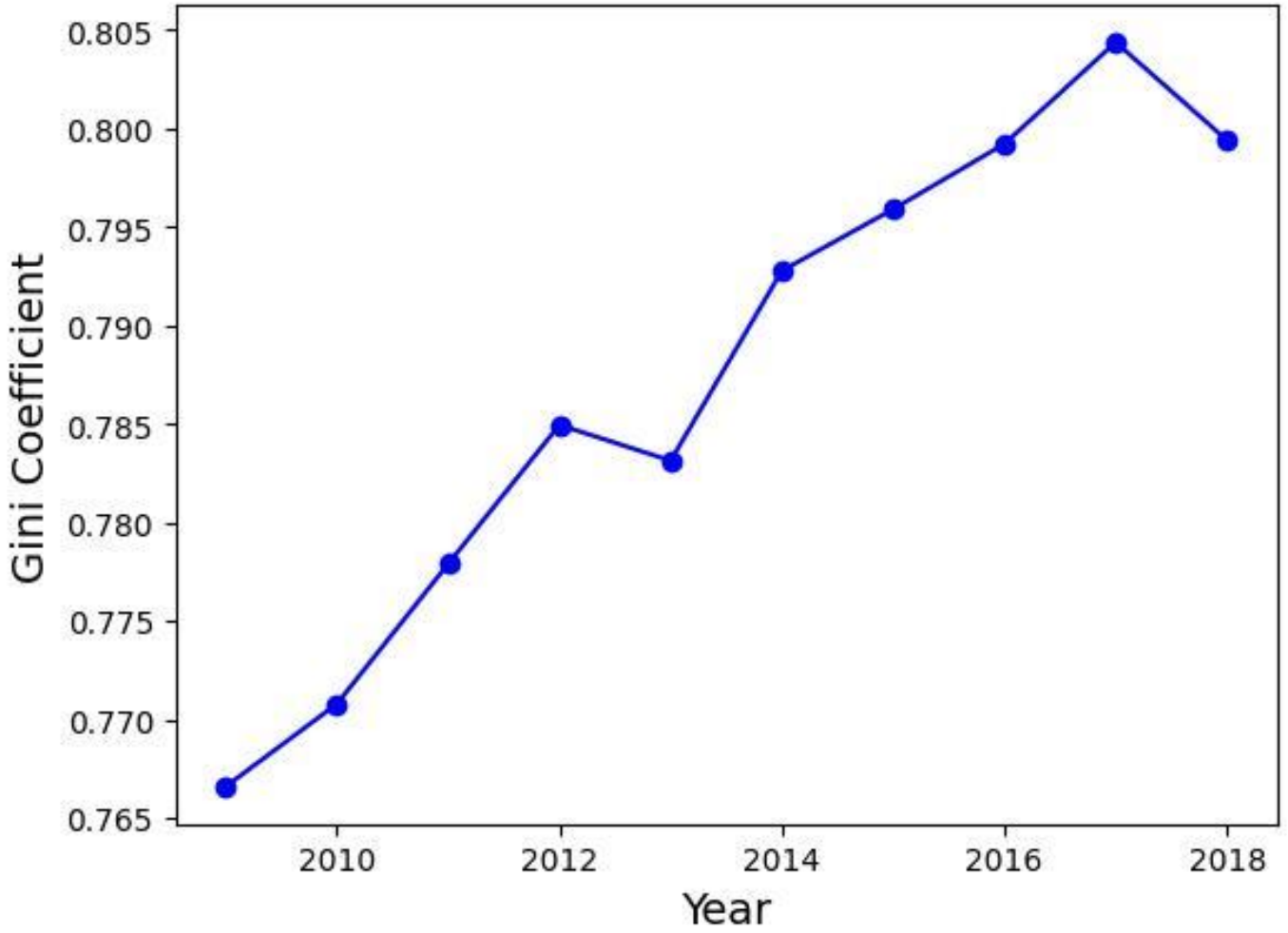
Airline


Average Betweenness Centrality (Unweighted)
0.028
0.026
0.024
0.022
0.020
0.018
0.016
2010
2012
2014
2016
2018
Year

Assortativity

Gini Coefficient (G)

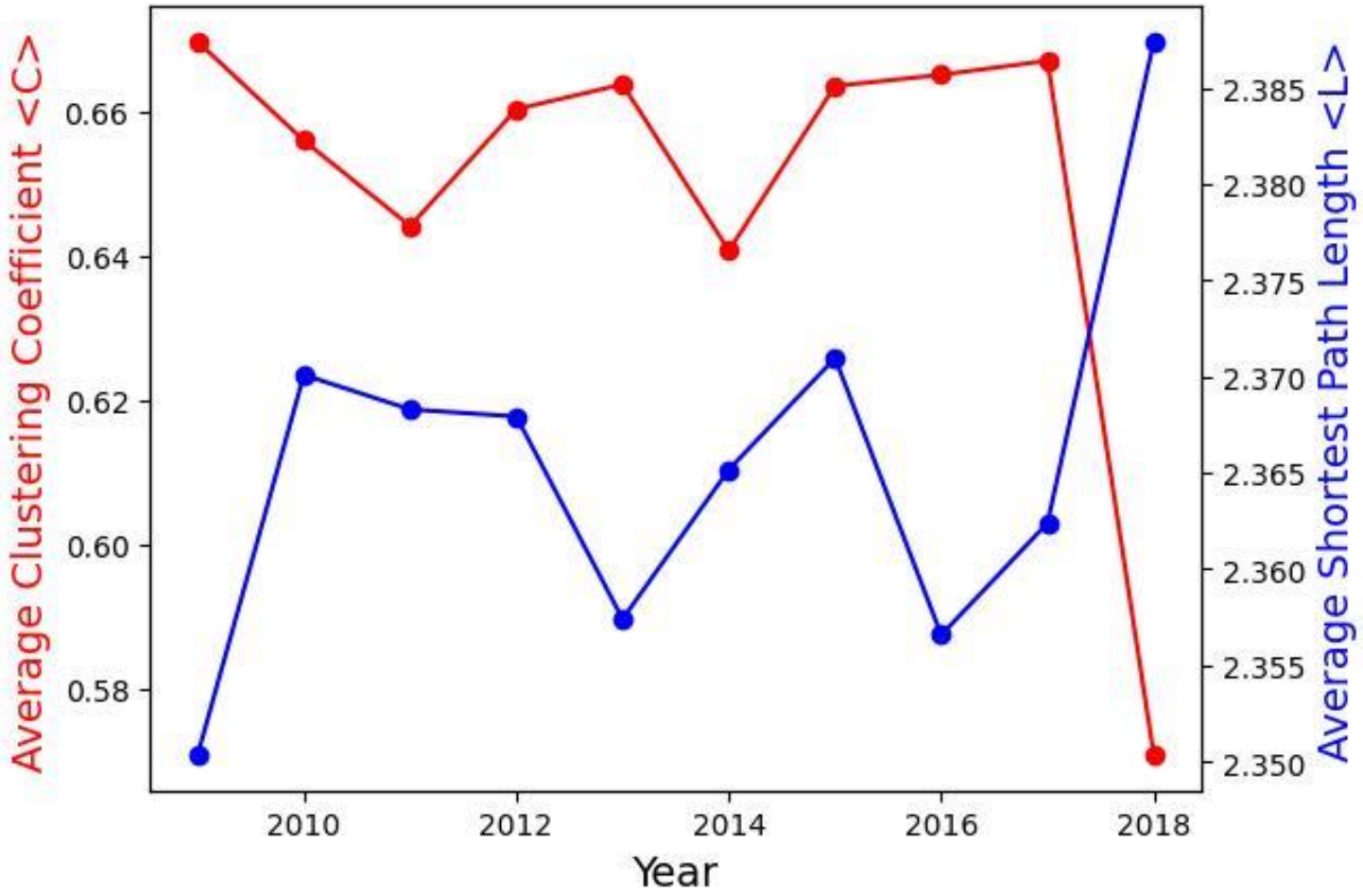
- $G = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|k_i - k_j|}{2MN}$,
  - k_i = weighted degree of node i
  - M = total weights
  - N = total number of nodes
- Commonly used to quantify the level of inequality in economics
- Assortativity: the tendency for nodes to connect to other nodes with similar properties within a network



Small-World Effect

Clustering Coefficient (C)

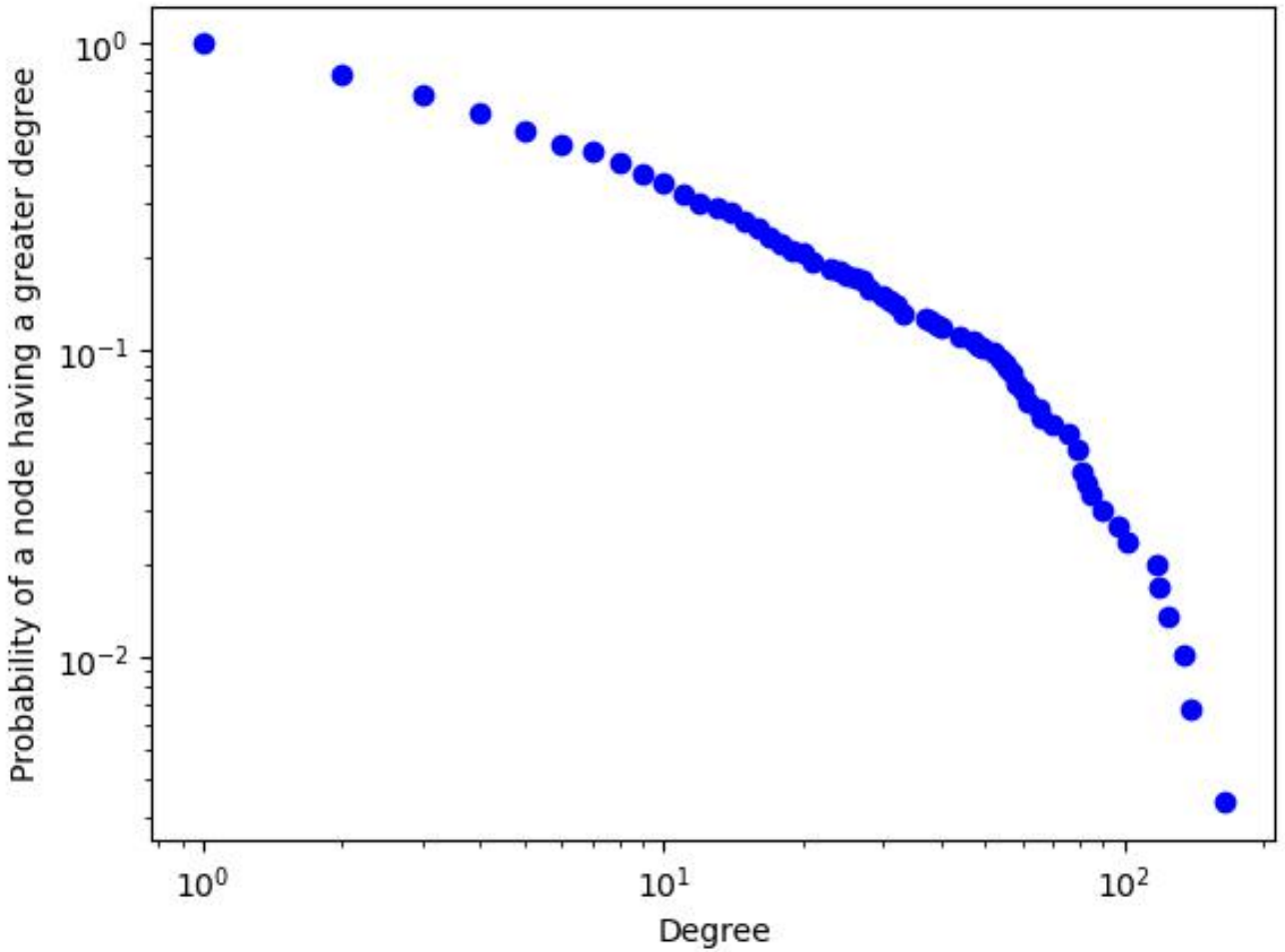
- measures the possibility that two neighbors of a node are connected and is thus is computed to determine the small-world property of a given network
- $C_i = \frac{2m_i}{k_i(k_i-1)}$
  - m_i = the number of edges that connect the neighbors of node i
- Large C_i → small-world property exists → nodes in the network can be reached by a small number of connections



Scale-Free

Power Law with Exponential Cutoff

$$f(x) \propto x^{-\alpha} e^{-\beta x}$$



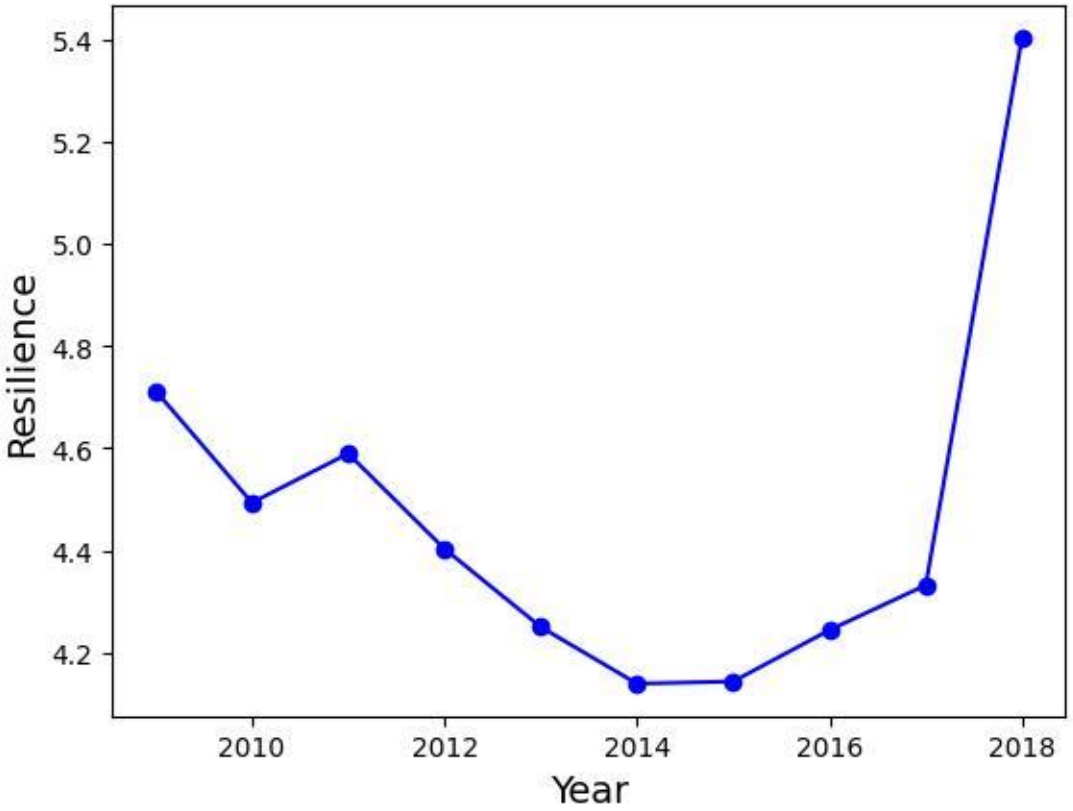
Target Attack Tolerance

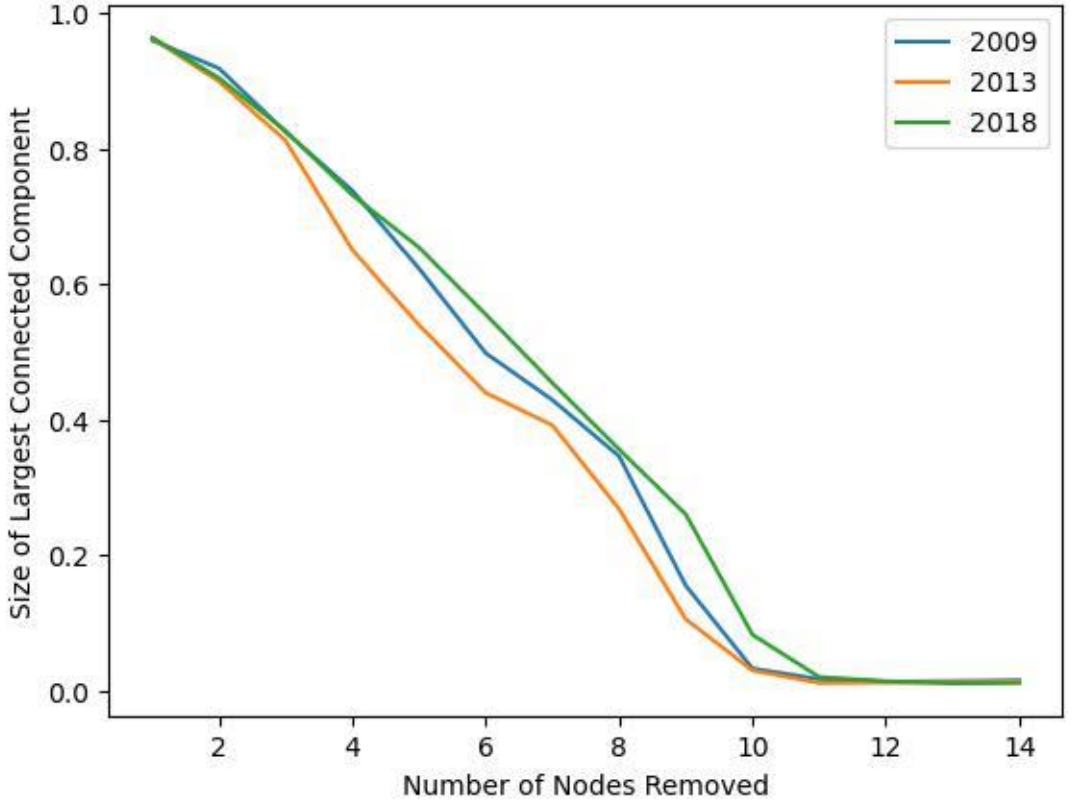
Network Resilience (R)

- $\mathrm{R} = \frac{1}{\mathrm{N}^2} \sum_{i \neq j} m(i,j)$
  - $m(i,j)$ = minimum number of nodes needed to be removed from the network such that j is unreachable from i.
  - N = total number of nodes
- Use Resilience to quantify the robustness of the network

Proportion of Nodes in GCC (S)

- $S = \frac{size(GCC)}{N}$
- Removing highly centralized nodes will split the network up → smaller S


Resilience
5.4
5.2
5.0
4.8
4.6
4.4
4.2
2010
2012
2014
2016
2018
Year


Size of Largest Connected Component
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
12
14
Number of Nodes Removed
2009
2013
2018

Conclusions

- Trend of Small Airports:
  - 2009 – 2014: Following the end of recession, an expansion of smaller regional airports
  - 2014 – 2017: Smaller airports were shutdown due to the saturation of market
  - 2017 - : A reinforcement of the Hub-and-Spoke system; network more interconnected

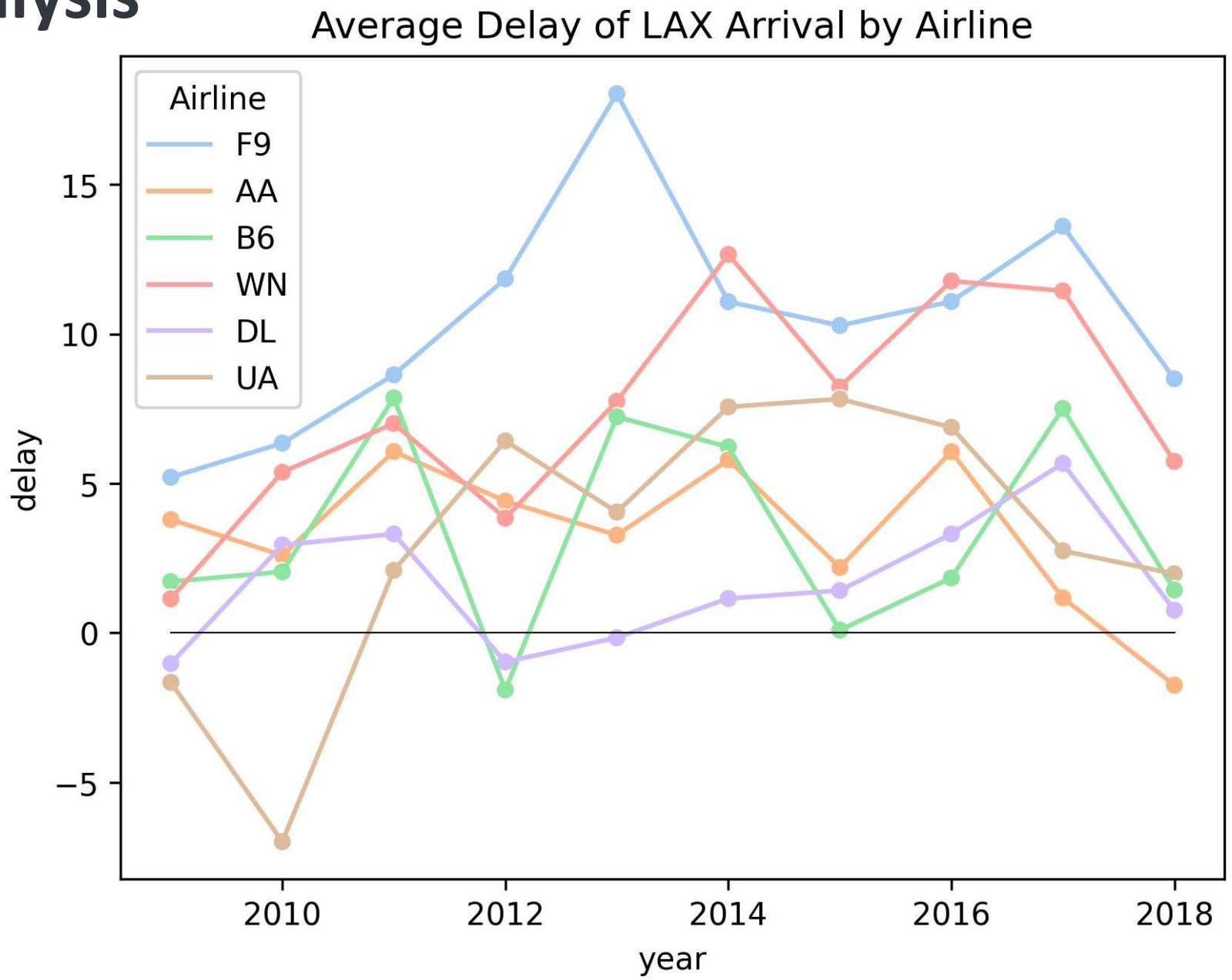
Limitations / Improvements

- Shortest air distance between airports were not provided; could talk about efficiency otherwise
- Could do an in-depth research into different airlines to see the expansion of major carriers
- Observe seasonal/monthly/holiday/major events changes
- When studying targeted attacks, cluster airports into large and small first, as terrorists are more likely attack on major hubs

Roles

Alan Wang	Brenden Cheung	Kun Xu
• literature review • data cleaning • coding simulations	• literature review • data cleaning • coding simulations	• literature review • make slides • write paper

Delay Analysis



Reference

- *Measuring resilience*. Packt. (n.d.). https://subscription.packtpub.com/book/cloud-and-networking/9781789955316/6/ch06lvl1sec41/measuring-resilience. Accessed December 8, 2022.
- Siozos-Rousoulis, L., Robert, D. & Verbeke, W. A study of the U.S. domestic air transportation network: temporal evolution of network topology and robustness from 2001 to 2016. *J Transp Secur* **14**, 55–78 (2021). https://doi.org/10.1007/s12198-020-00227-x. Accessed October 27, 2022.