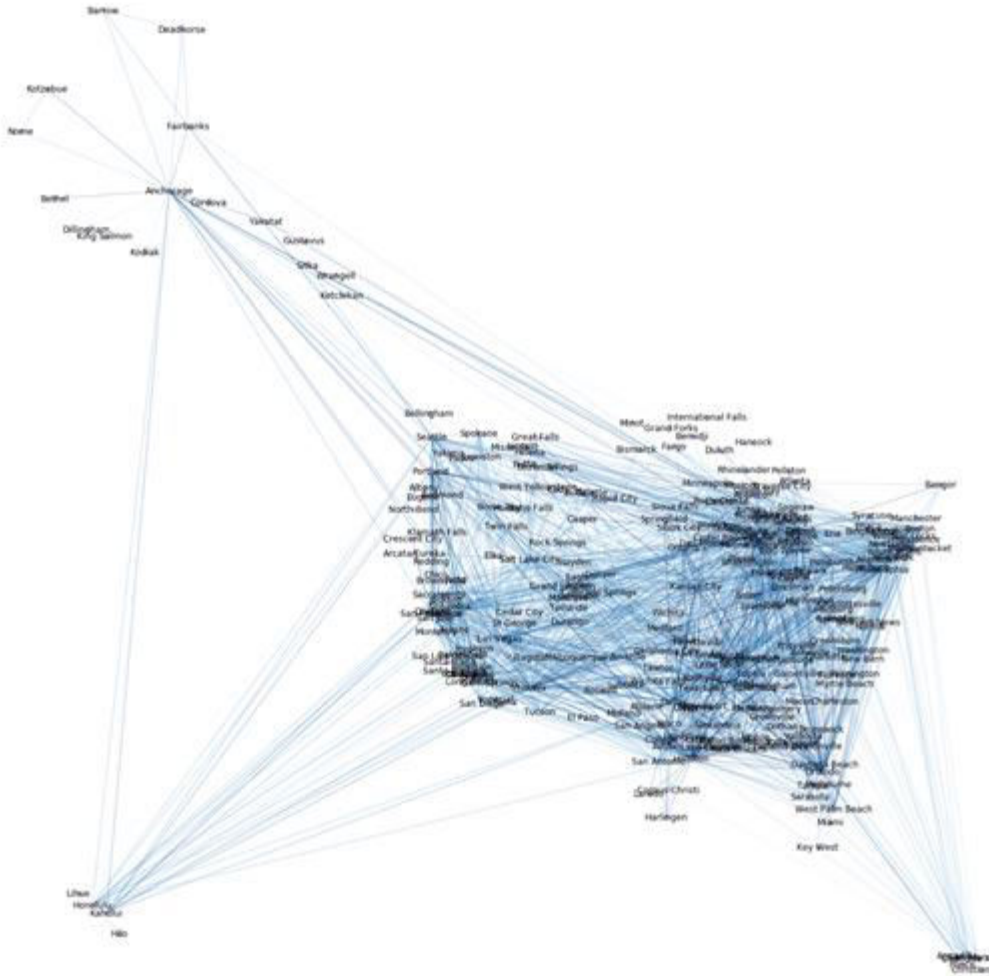
A Case Study of Airport Delay Network

Kaiyuan Chen, Sihui He, Zeyang Xu, and Changyu Yan

1. Introduction



2. Dataset Preparation

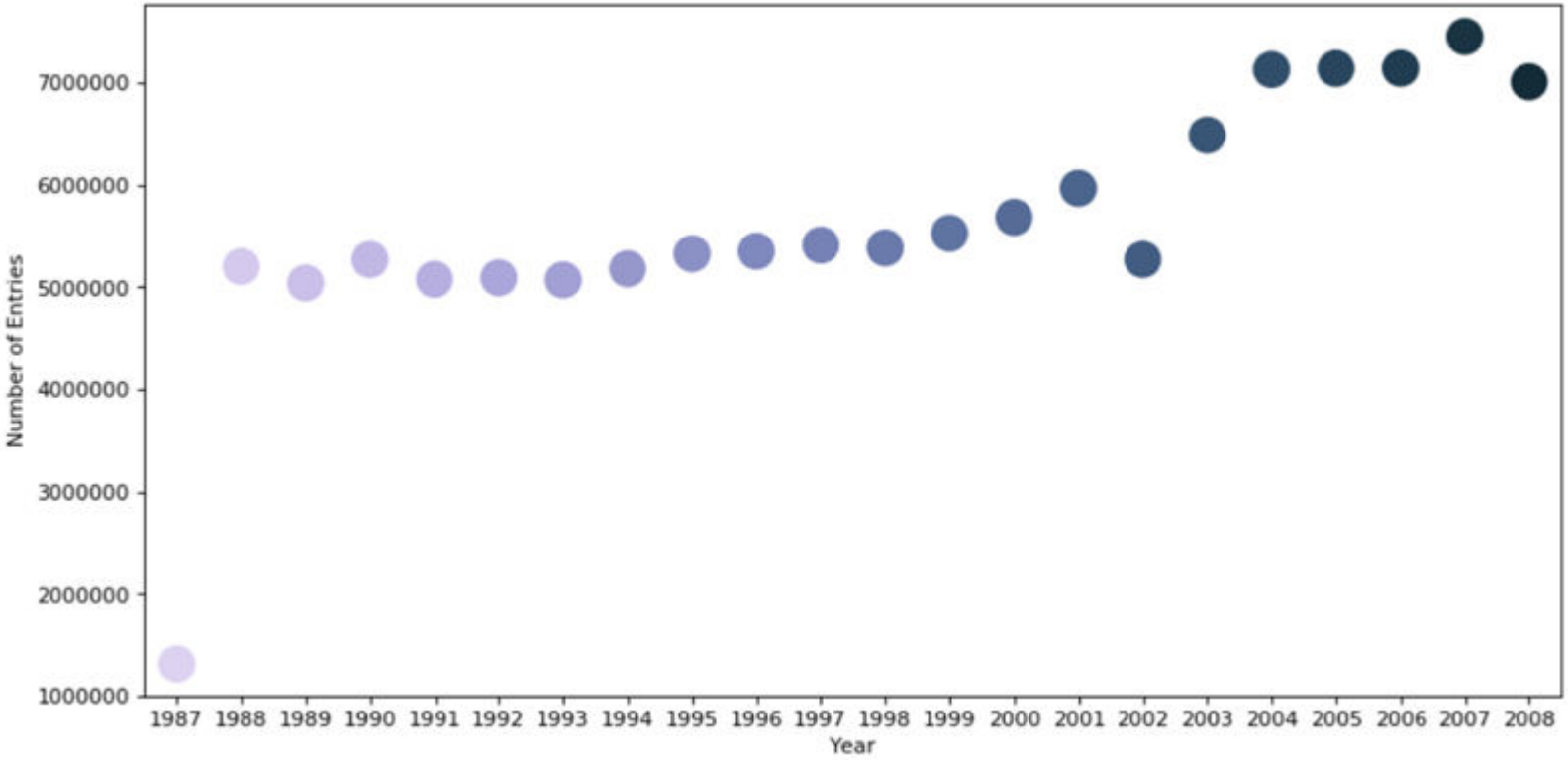
- Stat-Computing, a sub branch of American Statistical Association (ASA)
- Originally collected by Bureau of Transportation Statistics

Variables

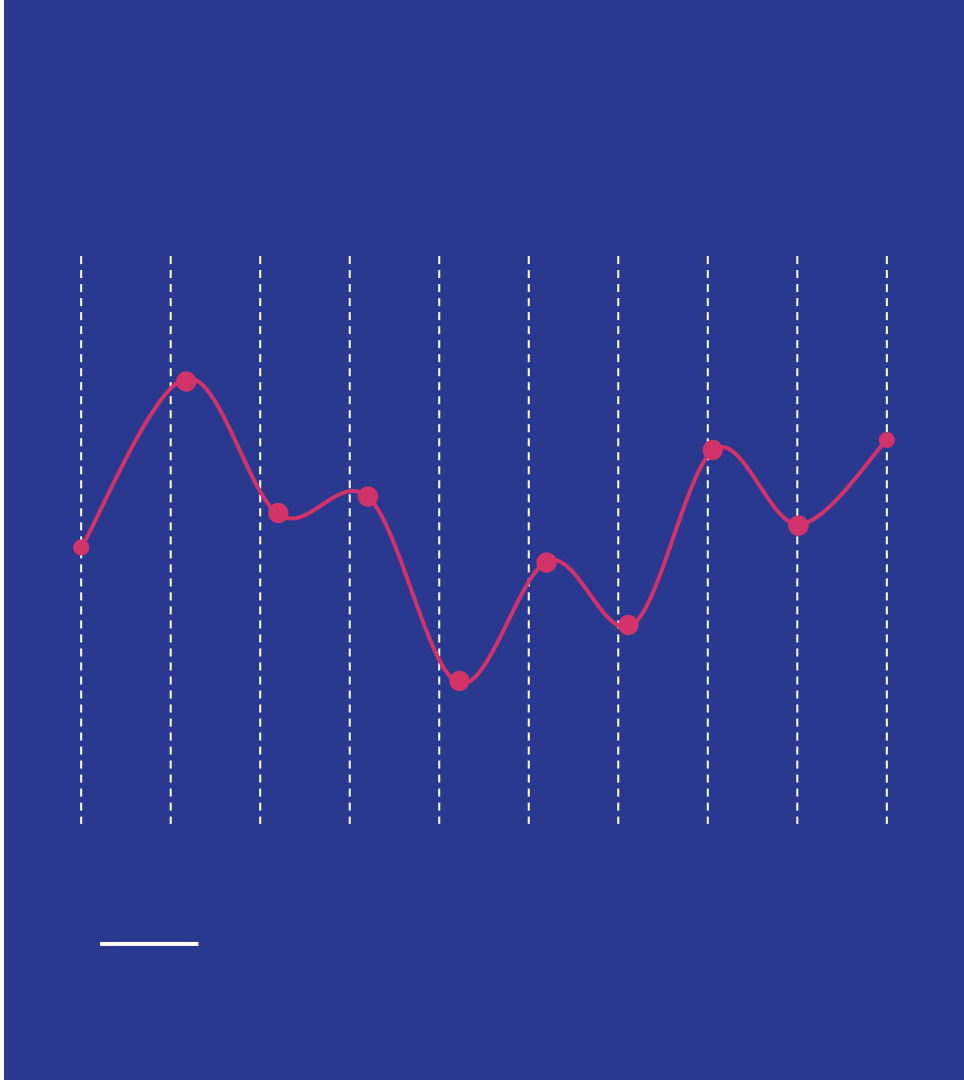
	Year	UniqueCarrier	ActualElapsedTime	CRSElapsedTime	ArrDelay	DepDelay	Origin	Dest	Distance	Cancelled
0	2008	WN	128.0	150.0	-14.0	8.0	IAD	TPA	810	0
1	2008	WN	128.0	145.0	2.0	19.0	IAD	TPA	810	0
2	2008	WN	96.0	90.0	14.0	8.0	IND	BWI	515	0
3	2008	WN	88.0	90.0	-6.0	-4.0	IND	BWI	515	0
4	2008	WN	90.0	90.0	34.0	34.0	IND	BWI	515	0

We also used airport information dataset, which maps airport codes to names and coordinates

Number of Entries Per Year



3.Results and Calculations



I. Relationship between Centrality and Delay

- Centrality Score we used:
  - Degree Centrality
    - Hypothesis: More flights going to/from other places may lead to delay
  - Closeness Centrality
    - Being closer to most of airports
  - Betweenness Centrality
    - Being in between other nodes may lead to more connecting flights
- To measure the performance, we use metric
  - Pearson Correlation
  - Spearman’s coefficient

Top 10 For each Centrality

Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	Atlanta	Atlanta	Atlanta
2	Chicago	Salt Lake City	Chicago
3	Dallas	Dallas	Dallas
4	Denver	Minneapolis	Denver
5	Minneapolis	Anchorage	Minneapolis
6	Detroit	Chicago	Salt Lake City
7	Salt Lake City	Denver	Detroit
8	Houston	Houston	Houston
9	Cincinnati	Detroit	Cincinnati
10	New York	San Francisco	Las Vegas

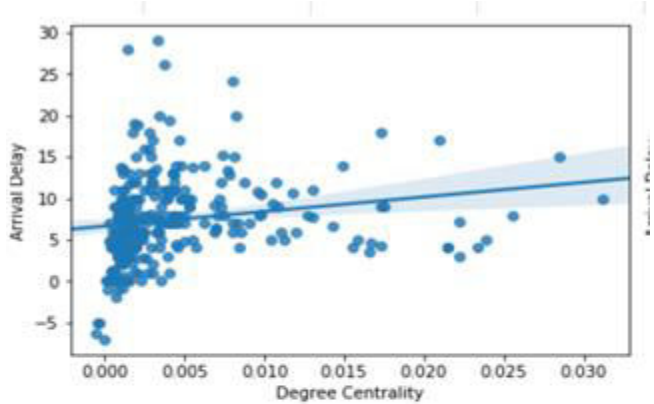
I. Relationship between Centrality and Delay

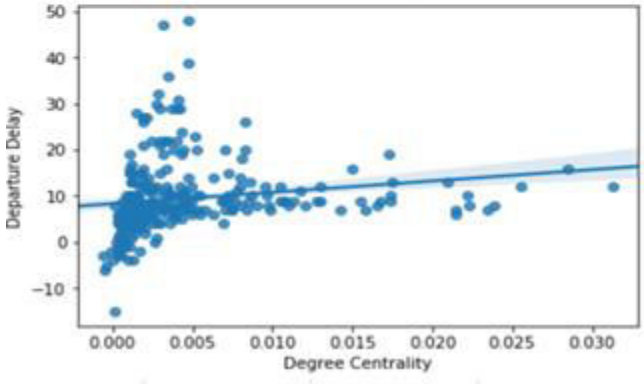
- Hypothesis 1:

  By Degree Centrality,

  More flights going to/from other places may lead to delay.

I.Relationship between Centrality and Delay - Degree Centrality





	Pearson Correlation	Spearman Correlation
Arrival Delay	0.179	0.404
Departure Delay	0.298	0.518

Interesting Pattern in Degree Centrality

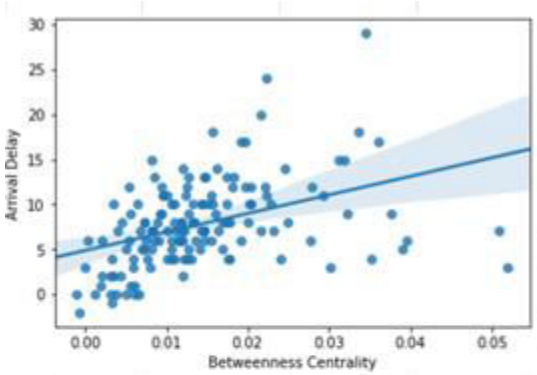
- Departure delay is much more correlated to degree centrality than arrival delay
  - Follow our intuition!
  - Larger airport is more congested, so it has higher delay.
  - Arrival delay is less relevant to size of airport, since planes have to plan ahead when taking off

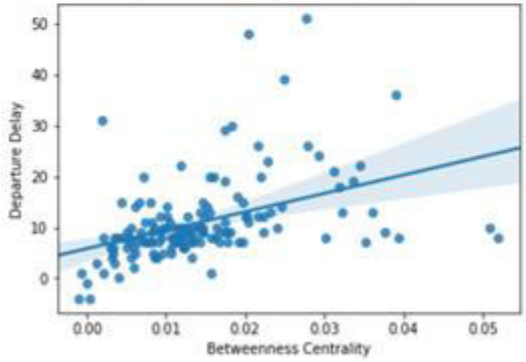
I. Relationship between Centrality and Delay

- Hypothesis 2:

  By betweenness centrality, being in between other nodes may lead to more connecting flights

I. Relationship between Centrality and Delay - Betweenness Centrality





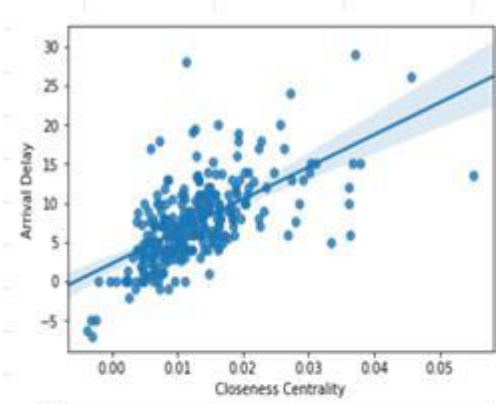
	Pearson Correlation	Spearman Correlation
Arrival Delay	0.413	0.488
Departure Delay	0.430	0.535

I. Relationship between Centrality and Delay

- Hypothesis 3:

  By closeness centrality, being closer to most of nodes may lead to more connecting flights

I. Relationship between Centrality and Delay - Closeness Centrality



	Pearson Correlation	Spearman Correlation
Arrival Delay	0.634	0.643
Departure Delay	0.676	0.634

Interesting Pattern on Closeness and Betweenness Centrality

- Those with many connecting flights and those are closest to all other airports are the busiest airports
- Busiest airports are vulnerable to long arrival and departure delay
- But, how do we measure vulnerability?

II. Vulnerability Index

$$V_i = \frac{D_i}{C_i},$$

Di is the average delay of the airport and Ci is the (closeness) centrality of airport i.
Invulnerability index measures how susceptible an airport in terms of the importance of the airport. An airport with a more significant centrality score is more important.

Fig. 5. Top Ten Vulnerable Airports for Delay

Cities	Vulerablility Index
1. Atlanta	0.160344
2. Salt Lake City	0.128662
3. Minneapolis	0.127824
4. Dallas	0.126783
5. Detroit	0.101268
6. Denver	0.076429
7. Chicago	0.062064
8. Houston	0.059642
9. Anchorage	0.057610
10. Phoenix	0.055963

Table 1. Airport Ranking Based on Different Centrality Measurements

Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality
1	Atlanta	Atlanta	Atlanta
2	Chicago	Salt Lake City	Chicago
3	Dallas	Dallas	Dallas
4	Denver	Minneapolis	Denver
5	Minneapolis	Anchorage	Minneapolis
6	Detroit	Chicago	Salt Lake City
7	Salt Lake City	Denver	Detroit
8	Houston	Houston	Houston
9	Cincinnati	Detroit	Cincinnati
10	New York	San Francisco	Las Vegas

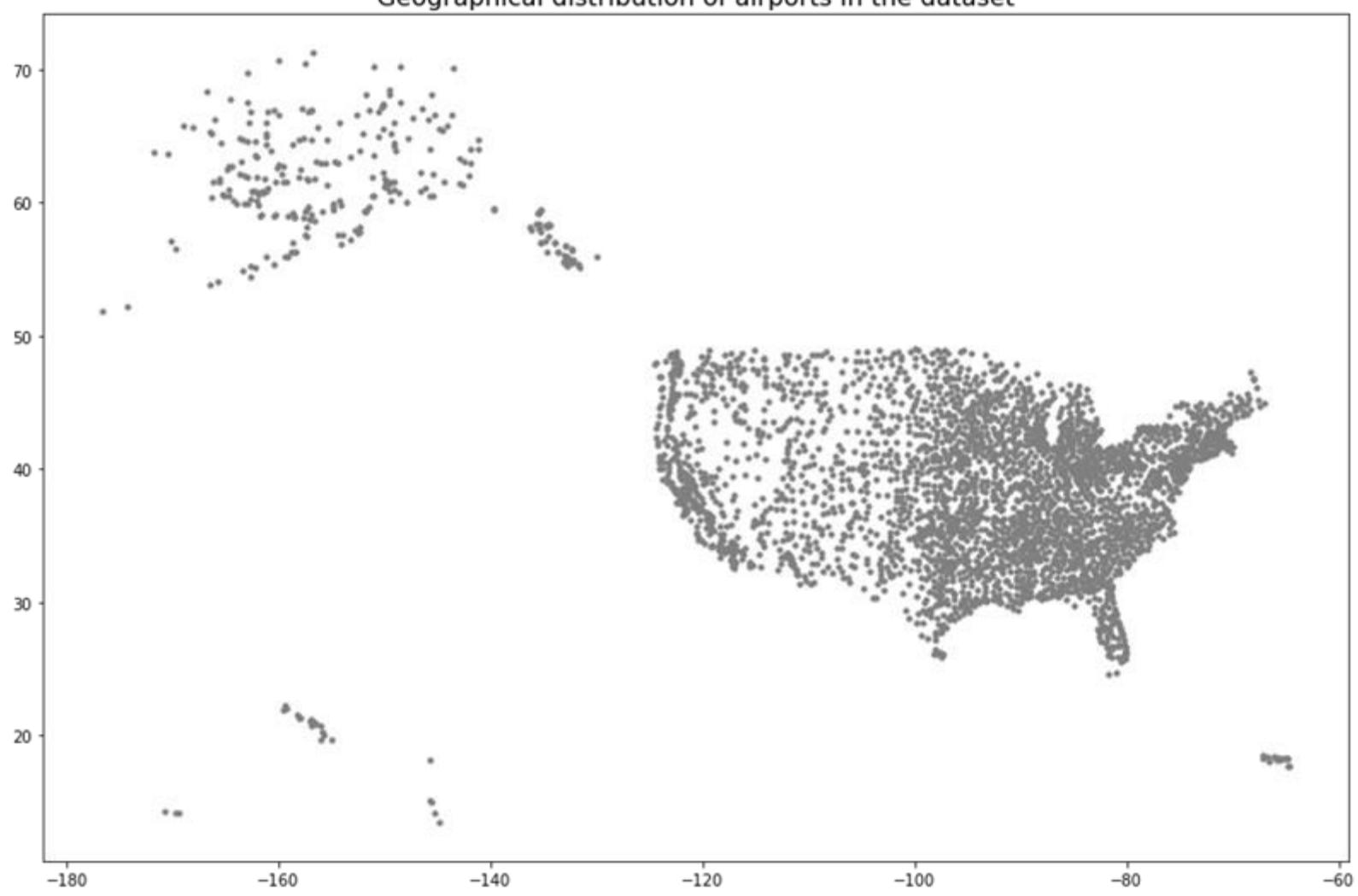
Observation:The airports with high centrality scores tend to have large vulnerability index values.

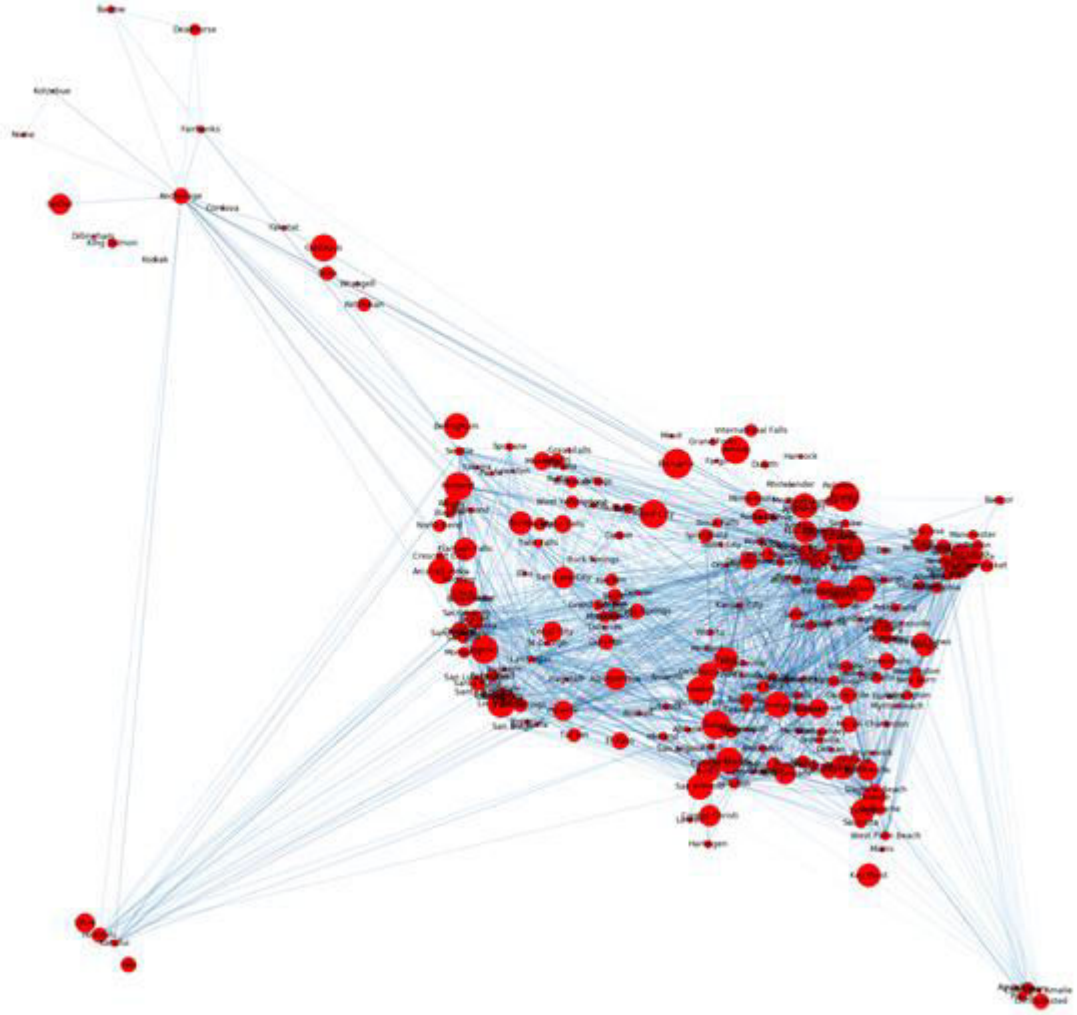
4. Analysis

- Atlanta
- Degree centrality

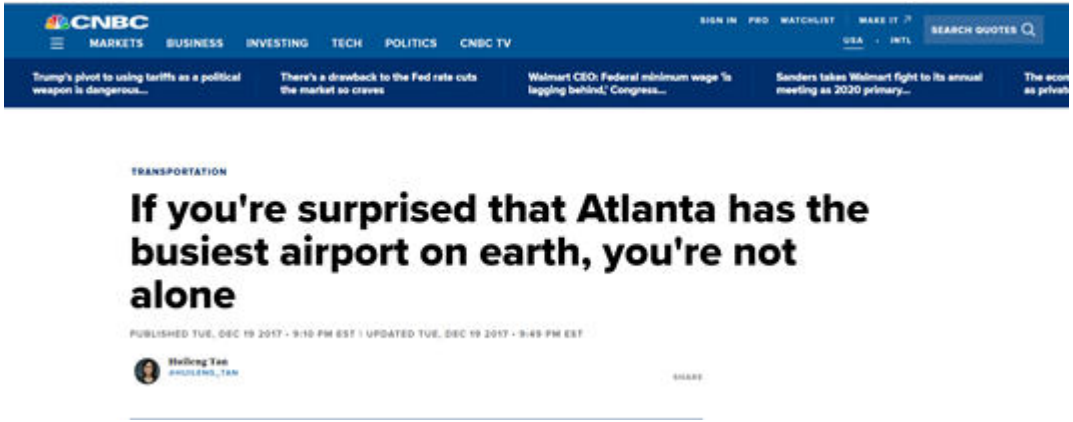
I. Atlanta


British Columbia
Alberta
Saskatchewan
Manitoba
Nunavut
Newfoundland an
Québec
Ontario
Washington
Montana
Minnesota
Michigan
Maine
Oregon
Idaho
Wyoming
New York
Nebraska
Iowa
Ohio
New Jersey
Nevada
Utah
Colorado
Kansas
Missouri
Kentucky
Virginia
California
Oklahoma
North Carolina
New Mexico
Atlanta
Texas
Louisiana
Florida
Mexico


Geographical distribution of airports in the dataset
70
60
50
40
30
20
−180
−160
−140
−120
−100
−80
−60



According to a fact sheet released by Hartsfield-Jackson, Atlanta is within a two-hour flight for 80 percent of the U.S. population.
The airport is a major connection for flights, serving 150 U.S. destinations and more than 75 destinations in 50 countries.
It handles some 2,500 flights and 275,000 passengers a day.



Rank (2017)	Airports (large hubs)	IATA Code	Major city served	2009[11]
1	Hartsfield–Jackson Atlanta International Airport	ATL	Atlanta	42,280,868
3	O'Hare International Airport	ORD	Chicago	31,135,732
2	Los Angeles International Airport	LAX	Los Angeles	27,439,897
4	Dallas/Fort Worth International Airport	DFW	Dallas/Fort Worth	26,663,984
5	Denver International Airport	DEN	Denver	24,013,669
6	John F. Kennedy International Airport	JFK	New York	22,710,272
8	McCarran International Airport	LAS	Las Vegas	19,445,952
15	George Bush Intercontinental Airport	IAH	Houston	19,290,239
13	Phoenix Sky Harbor International Airport	PHX	Phoenix	18,559,647
7	San Francisco International Airport	SFO	San Francisco	18,467,908
10	Charlotte Douglas International Airport	CLT	Charlotte	18,165,476

Fig. 5. Top Ten Vulnerable Airports for Delay

Cities	Vulerablility Index
1. Atlanta	0.160344
2. Salt Lake City	0.128662
3. Minneapolis	0.127824
4. Dallas	0.126783
5. Detroit	0.101268
6. Denver	0.076429
7. Chicago	0.062064
8. Houston	0.059642
9. Anchorage	0.057610
10. Phoenix	0.055963

Hartsfield-Jackson airport in Atlanta is home to one of the world’s largest carriers, Delta Air Lines, which was founded in Georgia

Conclusion

- We explored flight delay dataset
- We measured the relationship between different centrality measures and delay
- In order to analyze our observations, we come up with a metric vulnerability index
- We found related news reports and statistics to verify our observations

Questions?

We would like to thank Professor and TA for this quarter!