# Making Connections: An Analysis of US Commercial Aviation Networks

Alex Cohen, Patrick Maus, Caroline Sklaver

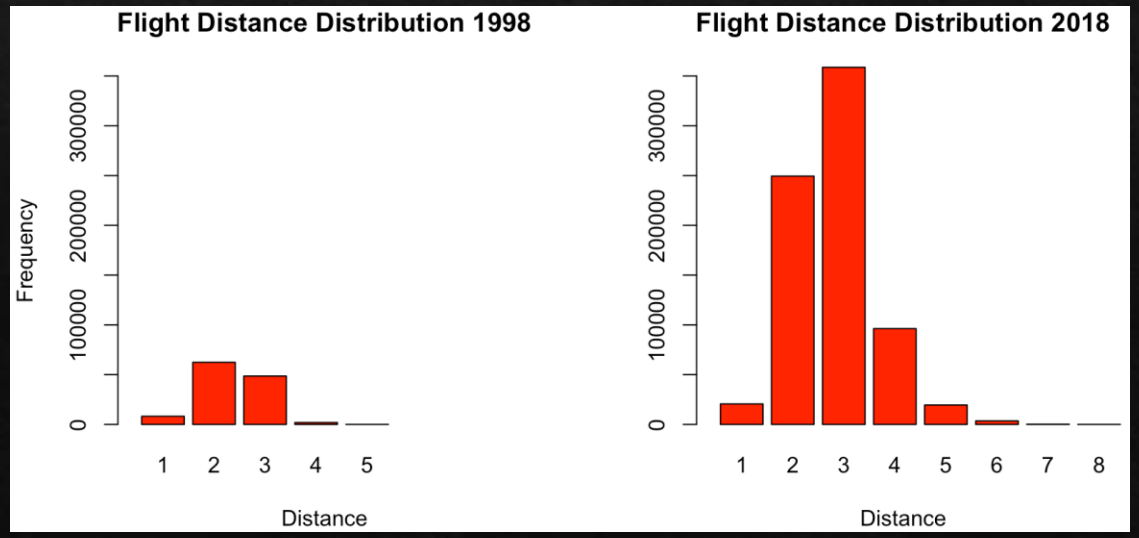
Continental US Aviation in 1998

Continental US Aviation in 2018

Note: Node size is the scaled degree. Edge width and color hue are scaled to number of flights between edges, with thicker lines and darker hues for higher weights.

# *Overview of Total Network*

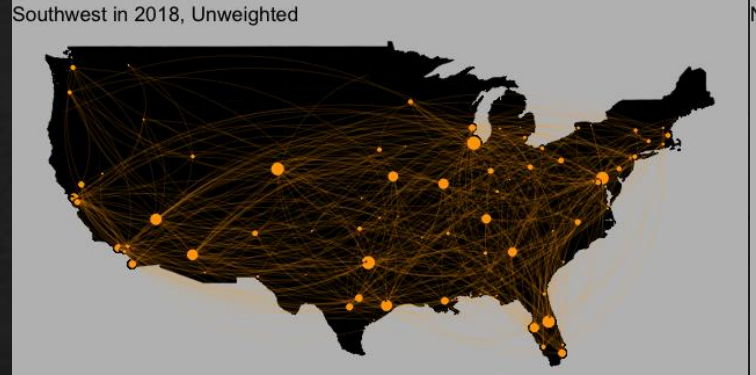
| Metric | 1998 | 2018 |
|---|---|---|
| Unweighted Average Betweenness | 619.4 | 2,700.0 |
| Unweighted Average Degree | 74.7 | 76.6 |
| Unweighted Diameter | 7 | 9 |
| Weighted Average Betweenness | 969.7 | 3,416.3 |
| Weighted Average Degree (Strength) | 35,261.7 | 13,667.0 |
| Number of Airports (Vertices) | 410 | 1,246 |
| Number of Legs (Edges) | 15,323 | 47,716 |
| Number of Flights (Weight) | 7,228,654 | 8,514,517 |
| Number of Passengers | 575,354,256 | 787,782,842 |

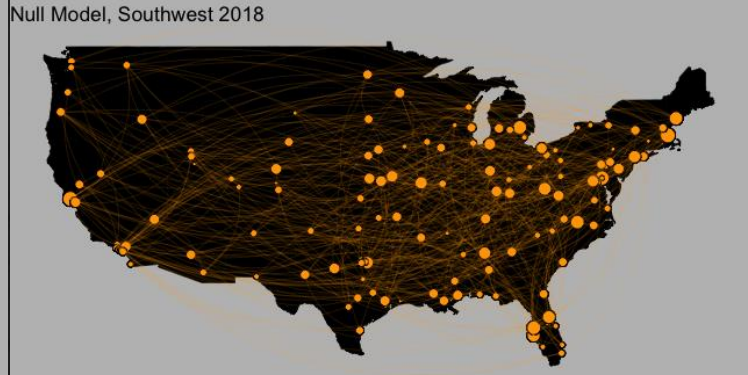


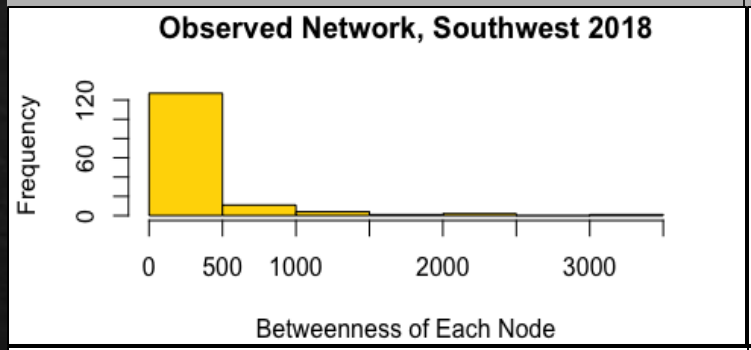
# A Tale of Two Airlines

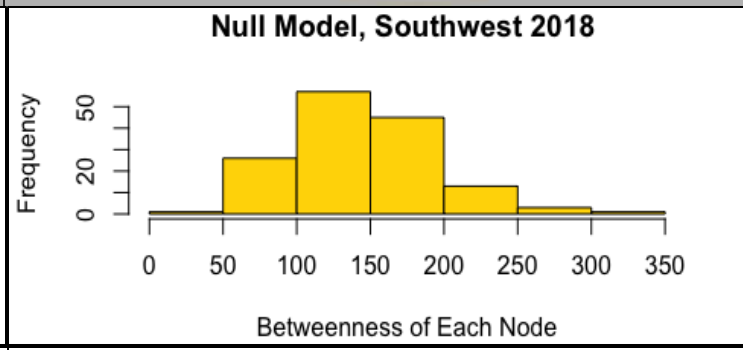
- From 1998 to 2018, Southwest's network nearly tripled the number of nodes, maintained the same average degree, and increased the average betweenness.
- In the same time, Delta's number of nodes and average betweenness both increased by ~44% while the degree increased by 25%.
- When we compared 2018 networks for Delta and Southwest to null models, its clear the observed networks are more centralized than the null model, with a long tail containing a high number of central nodes.
- Delta clearly exhibits this pattern more extremely than Southwest, suggesting Southwest is a flatter network while Delta relies on fewer, larger hubs. An example is Atlanta, a major Delta hub with a betweenness score of 16,812.

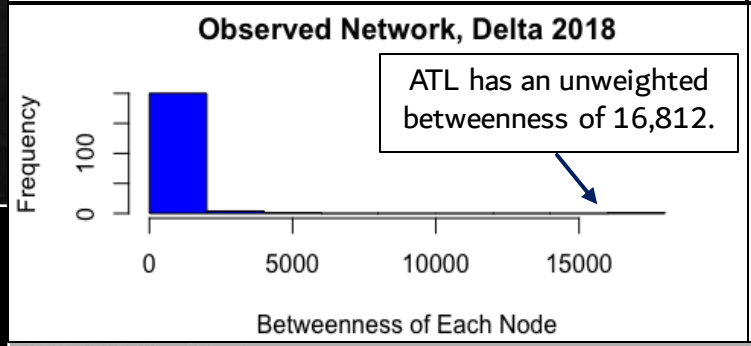
| (measurements are from an unweighted graph) | Number of Nodes | Observed Average Degree | Observed Average Between-ness | Null Model Average Degree | Null Model Average Between-ness |
|---|---|---|---|---|---|
| **Southwest 1998** | 52 | 31.8 | 41.7 | 32.6 | 35.9 |
| **Southwest 2018** | 138 | 35.34 | 201.9 | 34.8 | 144 |
| **Delta 1998** | 131 | 24.9 | 159.1 | 24.7 | 172.3 |
| **Delta 2018** | 187 | 31.3 | 231.1 | 30.9 | 249 |

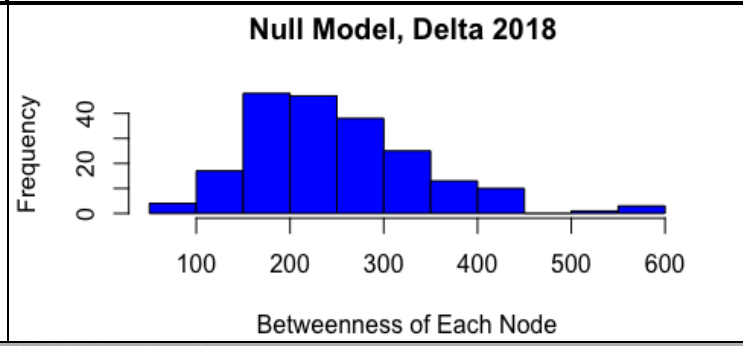


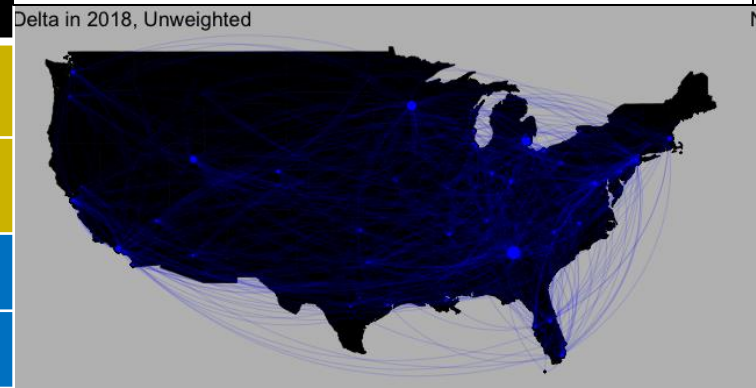



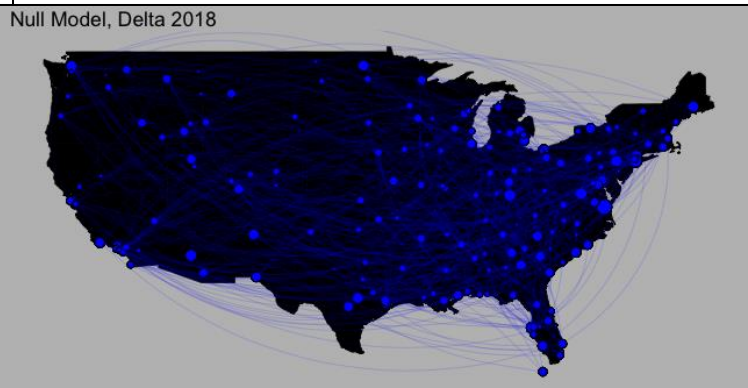



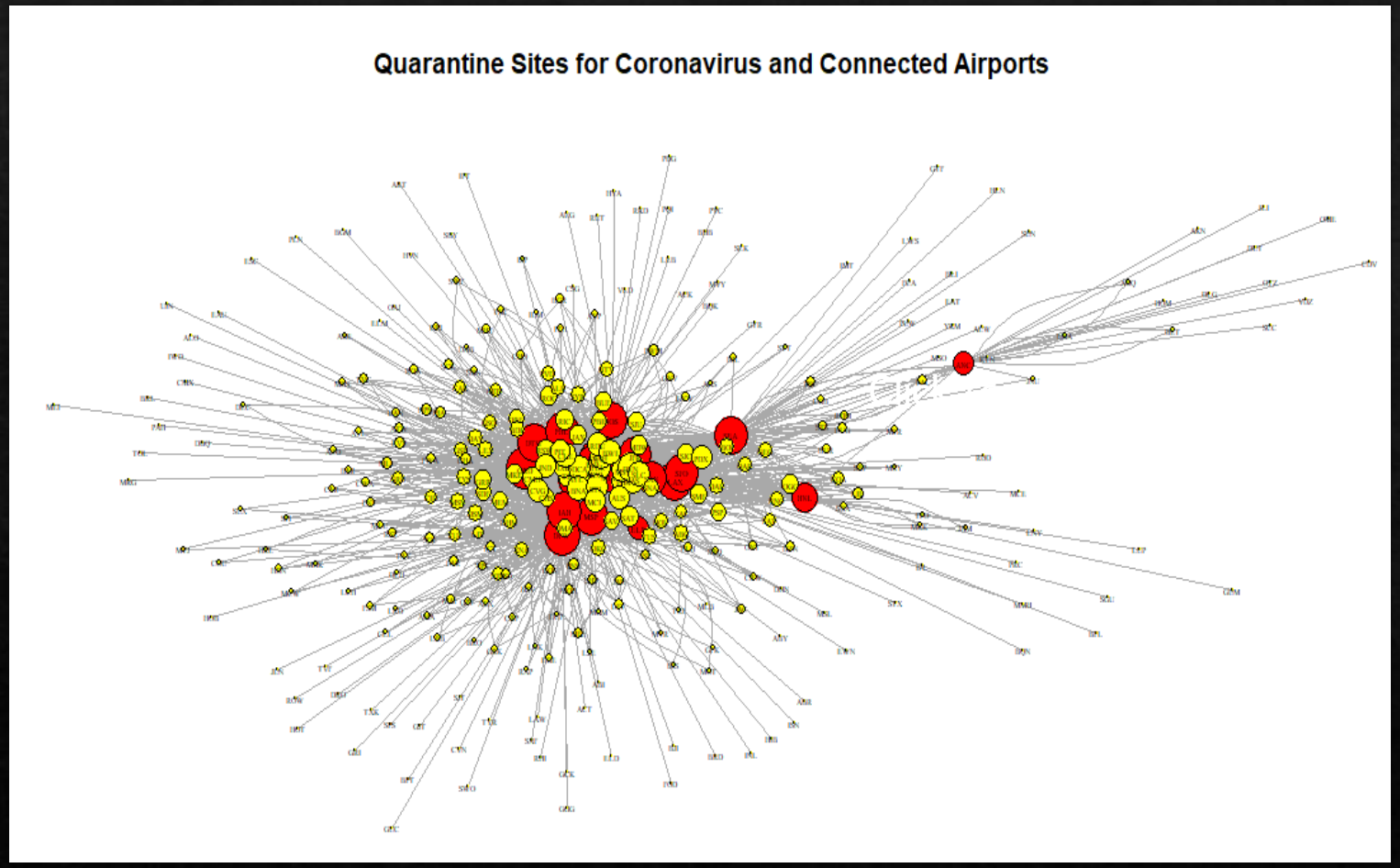



# *Coronavirus Impact on Air Networks*




- 19 airports were designated as quarantine stations by the CDC
- From these 19 airports (excluding Puerto Rico), **298** new airports can be reached in one connection at least once per day
- These 19 airports constitute 7 of the top 10 (13/20) busiest airports in terms of daily departures and 5 of the top 10 (13/20) busiest airports in terms of departing passenger volume
- 68% of passengers traveling domestically in the US will start or end a flight at one of these 19 airports daily
- 65% of all daily domestic flights originated or ended at one of these 19 airports
- This represented 761,000 departing passengers and 7,500 departing flights per day in 02/2018

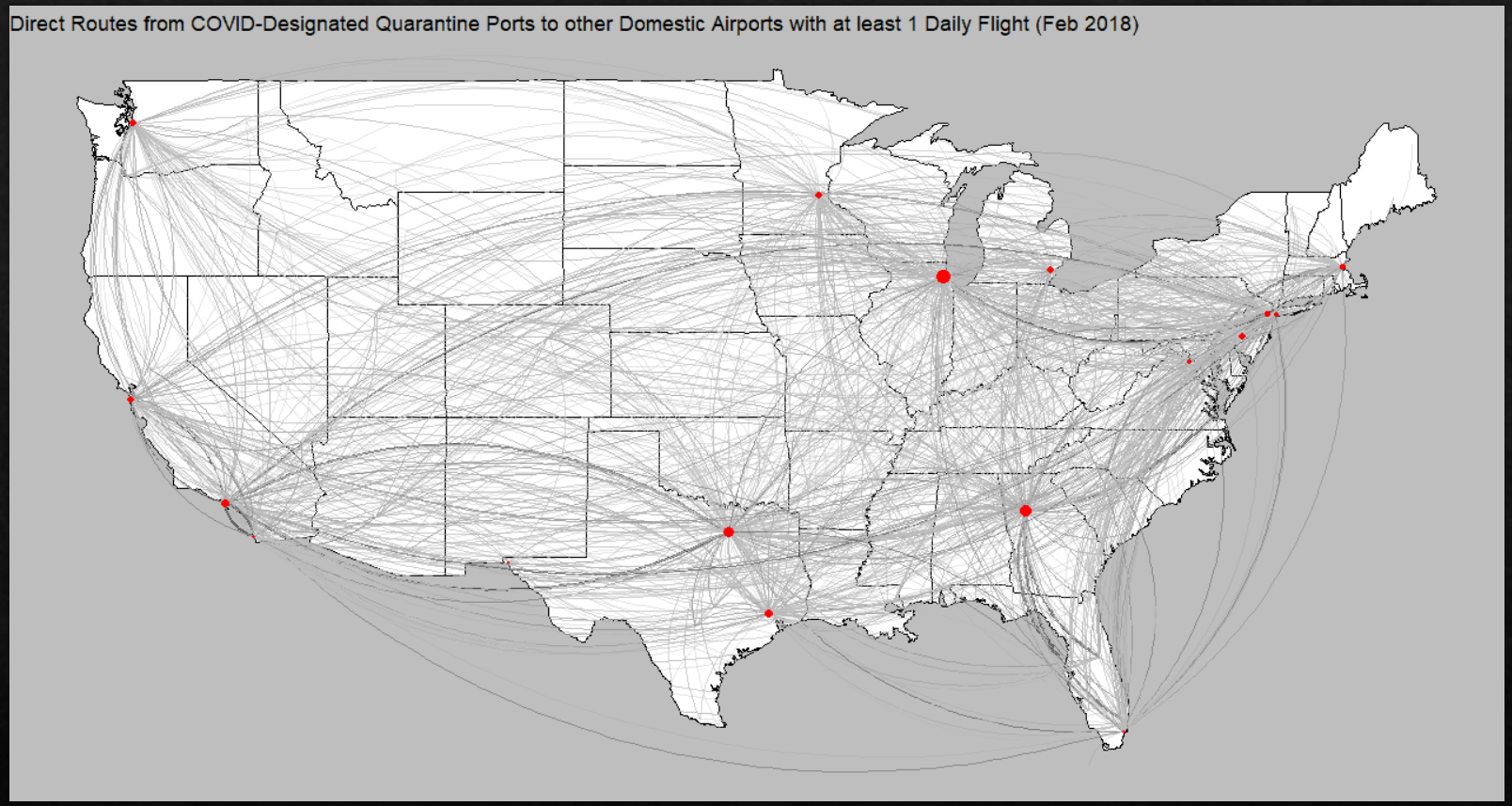
Node size is the log of the degree.
Red Nodes are Screening Airports
Yellow Nodes are within one connection of screening airports
Layout = LGL

*All figures are specific to 02/2018

# Coronavirus Screening Network Metrics

| | All 2018 Flights | All 2018 Flights WITHOUT Screening Airports | To/From Covid-Screening Airports (All 2018) |
|---|---|---|---|
| Degree Centrality (Strength) | 17,495.84 | 6,272.25 | 19,849.05 |
| Diameter | 8 | 8 | 4 |
| Mean Distance | 2.81 | 2.90 | 2.35 |
| Betwteeness | 1776.49 | 1743.66 | 701.22 |
| # Nodes | 901 | 874 | 518 |
| # Edges | 41,541 | 24,918 | 16,623 |
| Total # Flights | 7,881,877 | 2,740,974 | 5,140,903 |

*Numbers are for weighted network with US Territories removed




January 17: Screenings at JFK, LAX, and SFO for passengers from Wuhan, China

January 21: Add screenings at ATL and ORD for passengers from Wuhan.

January 28: Expand screenings to 20 airports; include all passengers arriving from China

Late February: First TSA screeners diagnosed with Coronavirus

March 14: Implement new screening for European travel; announce European travel ban.

April 1: Over 60 TSA screeners have screened positive

# *Appendix*

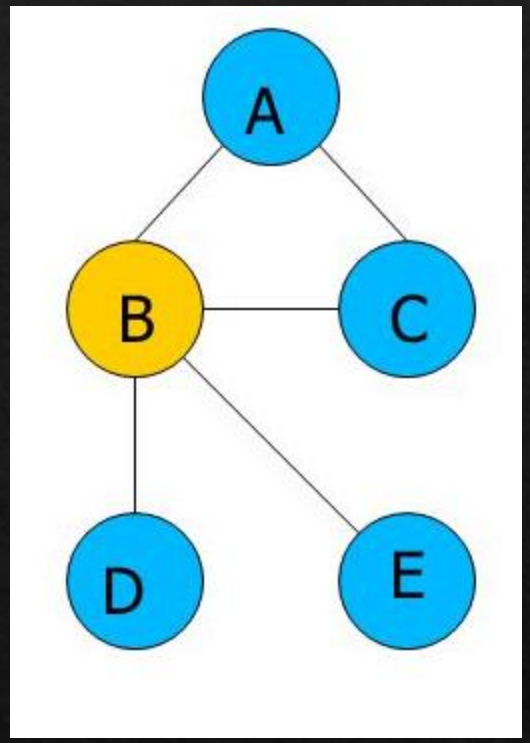
# *Data Source*

- All flight data is from the Bureau of Transportation Statistics, T-100 Domestic Segment (All Carries) table, located here:
  - https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=311
- Data is provided monthly, and annual data from 2018 and 1998 was collected with the following fields:
  - `DepScheduled, DepPerformed, Seats, Passengers, Distance, AirTime, UniqueCarrier, UniqueCarrierName, Carrier, OriginAirportID, Origin, OriginCityName, OriginState, DestAirportID, Dest, DestCityName, DestState, DestAirportID, AircratGroup, AircraftType, Month`
- Data was grouped by `Origin, OriginCityName, Dest, DestCityName, UniqueCarrier`, filtered for `SEATS > 0`, and `Passengers` set to `SUM(Passengers)` and `Flights` set to `SUM(DepPerformed)`

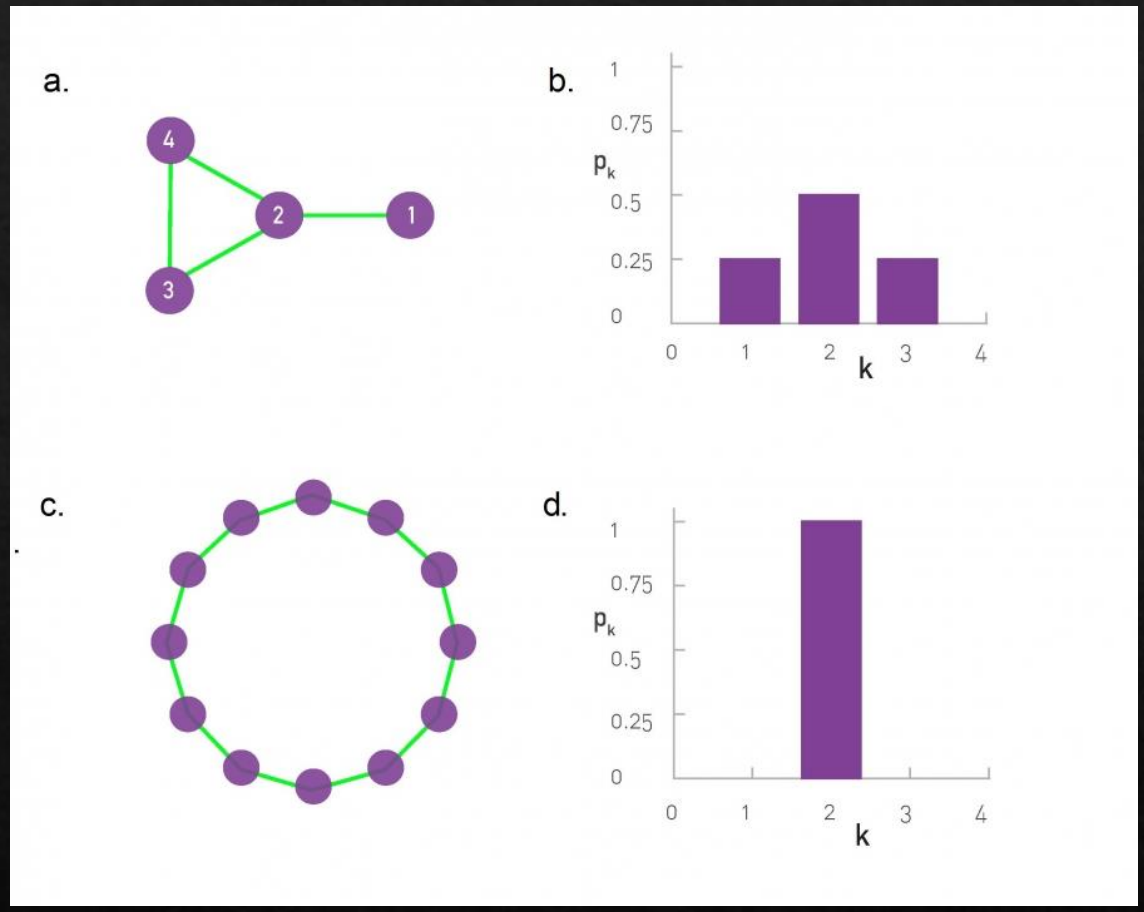
# *igraph Network Measures*

- Betweenness:
  - `betweenness()` - defined by the number of geodesics (shortest paths) going through a vertex
- Mean Distance:
  - `mean_distance()` - calculates the average path length in a graph, by calculating the shortest paths between all pairs of vertices (both ways for directed graphs)
  - This function does not consider edge weights
- Distance histogram:
  - `distance_table()` – calculates a histogram, by calculating the shortest path length between each pair of vertices
  - For directed graphs both directions are considered, so every pair of vertices appears twice in the histogram.
- Diameter:
  - `diameter()` - the length of the longest geodesic (shortest possible line between two nodes)
  - Optional positive weight vector for calculating weighted distances. If the graph has a `weight` edge attribute, then this is used by default
- Degree centrality:
  - `centr_degree()` - centralize a graph according to the degrees of vertices
  - `centr_degree(g)$res` – the node-level centrality scores
  - `Strength()` - Summing the edge weights of the adjacent edges for each vertex. This is the weighted degree
- Random Network
  - `Sample_gnp()` - Generate random graphs according to the G(n,p) Erdos-Renyi model

# Betweenness and Degree

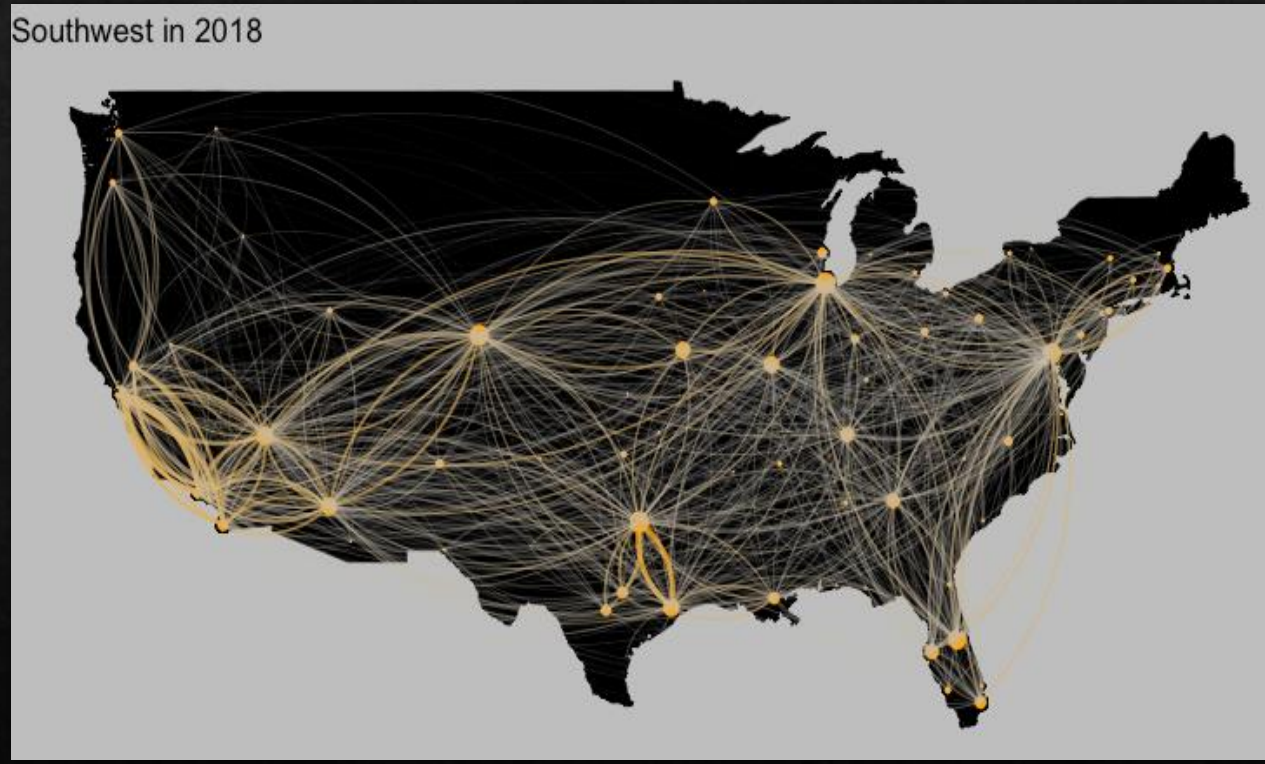



- The shortest paths are AB, AC, ABD, ABE, BC, BD, BE, CBD, CBE, DBE
- Five of those paths (ABD, ABE CBD, CBE, DBE) transit through B.
- B has a betweenness score of 5.




a.For the network in (a) with $N = 4$ the degree distribution is shown in (b).
b.We have $p_1 = 1/4$ (one of the four nodes has degree $k_1 1 = 1$), $p_2 = 1/2$ (two nodes have $k_3 = k_4 = 2$), and $p_3 = 1/4$ (as $k_2 = 3$). As we lack nodes with degree $k \rangle 3$, $p_k = 0$ for any $k \rangle 3$.
c.A one dimensional lattice for which each node has the same degree $k = 2$.
d.The degree distribution of (c) is a Kronecker's delta function, $p_k = \delta(k - 2)$.

# *Cases of Delta and Southwest Growth/Dynamics with Weighted Graphs*

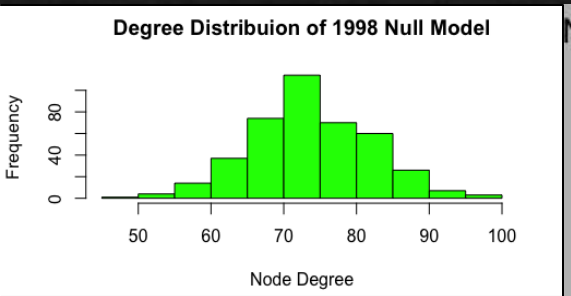
# *Null Model Generation*

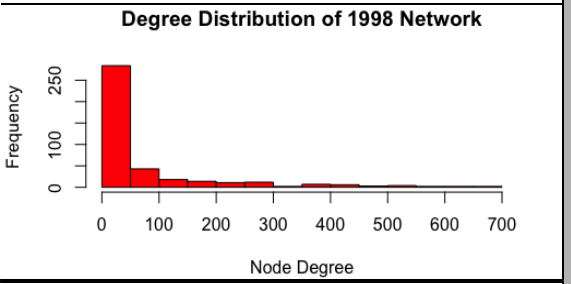
- For a given network, we created a null model by first identifying the number of vertices and the number of edges. From there, we can determine the probability of an edge occurring in a network.
- We used the igraph function sample_gnp() in R to generate our random graph. This function uses the random network model first defined by Gilbert and referred to as *G( N, p).*
- The *G*(*N*, *p*) model fixes the probability *p* that two nodes are connected and the *G*(*N*, *L*)) model fixes the total number of links *L*.
- When needed, we also simulated 1000 random graphs and took their average to build our null model.

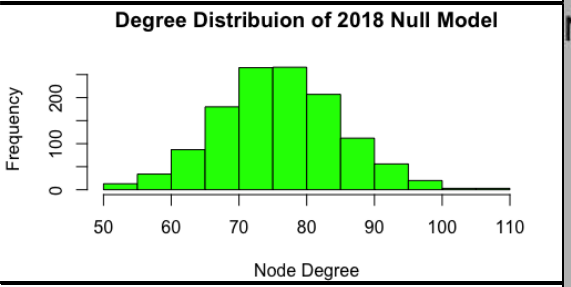
```
# get random network params
 numb_edges <- length(E(g))
 numb_vertices <-length(V(g))
p_edges <- (numb_edges/(numb_vertices*(numb_vertices-1)))


 # make random network
  rand_g <- sample_gnp(n=numb_vertices, p=p_edges, directed=TRUE) %>%
  set_vertex_attr("label", value = V(g)
 V(rand_g)$name <- V(g)$name # name random network vertices same as original graphs
```

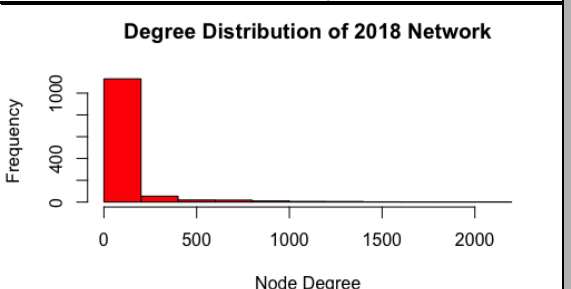
# *Null Models vs. Unweighted Networks*

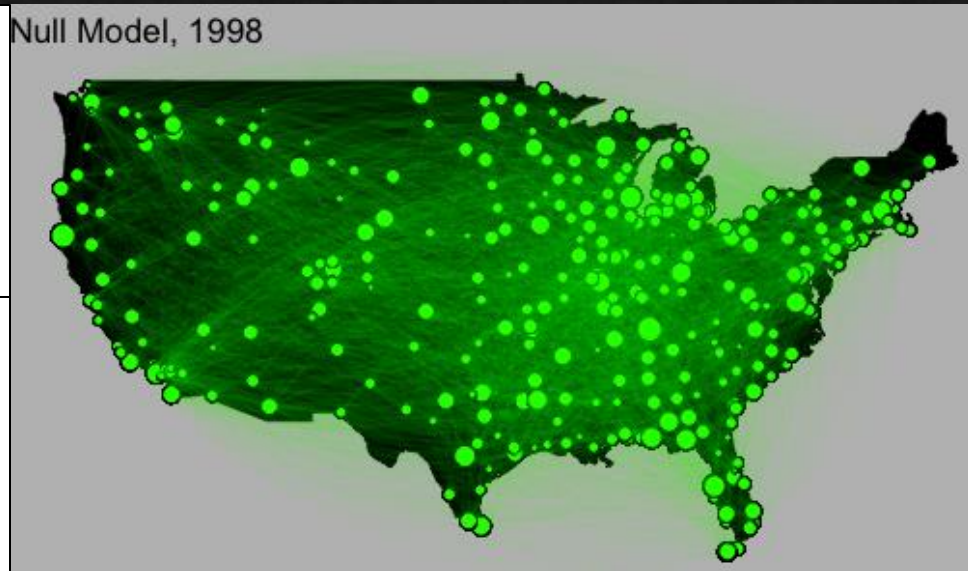
| Metric | 1998 | 2018 | Metric | 1998 | 2018 |
|---|---|---|---|---|---|
| Null Model Average Betweenness | 383.7 | 1576.5 | Observed Average Betweenness | 619.4 | 2,700.0 |
| Null Model Average Degree | 74.6 | 76.8 | Observed Average Degree | 74.7 | 76.6 |
| Null Model Mean Distance | 1.9 | 2.3 | Observed Mean Distance | 2.6 | 3.3 |
| Null Model Diameter | 3 | 3 | Observed Diameter | 7 | 9 |

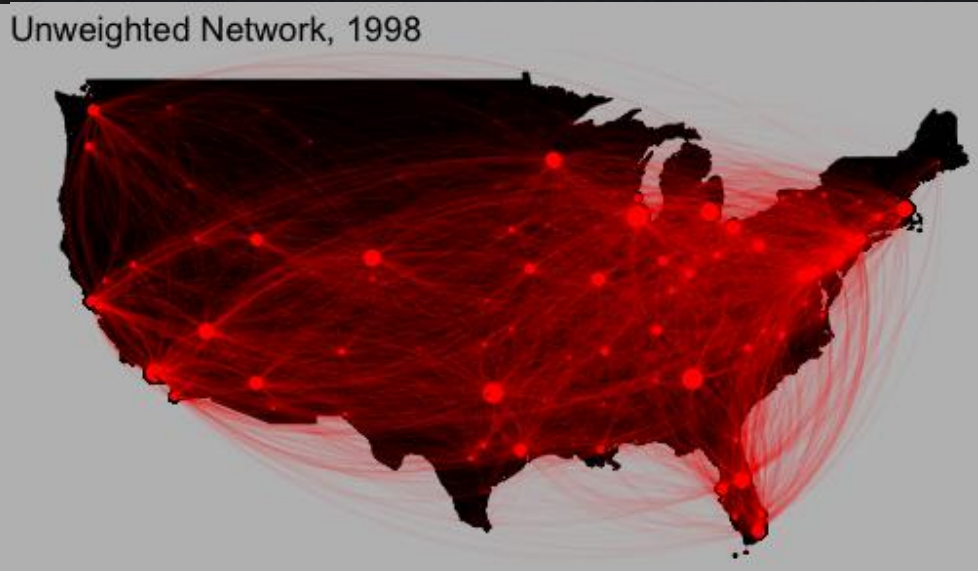


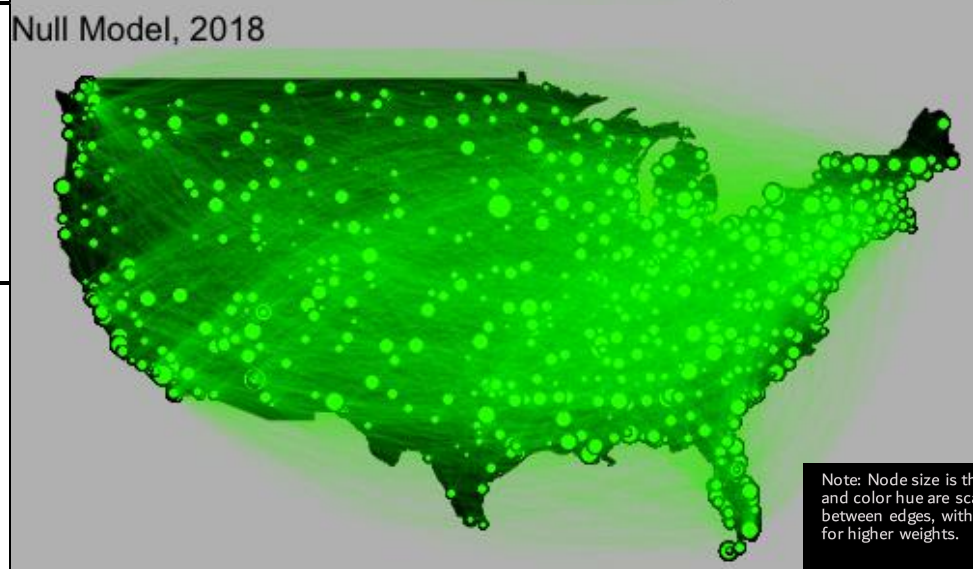



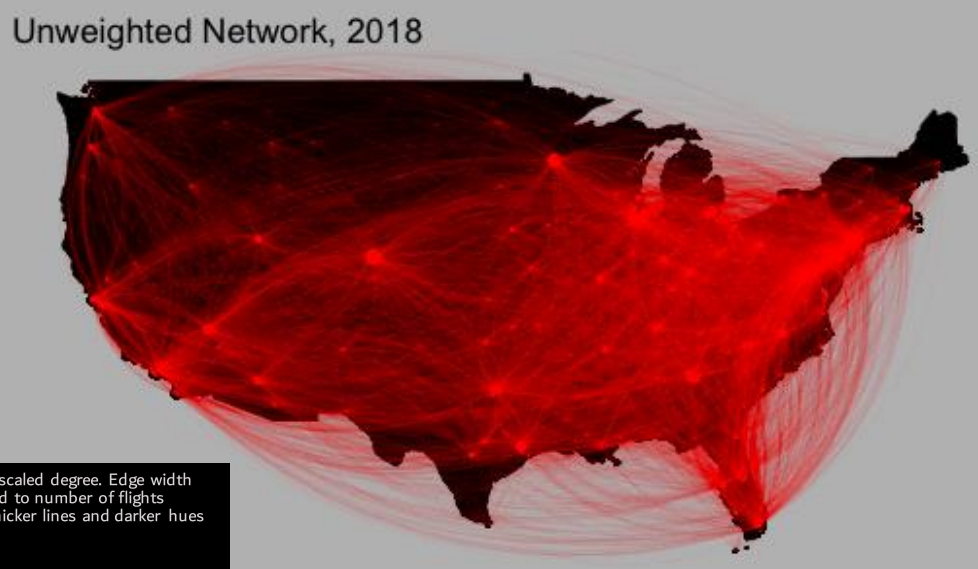



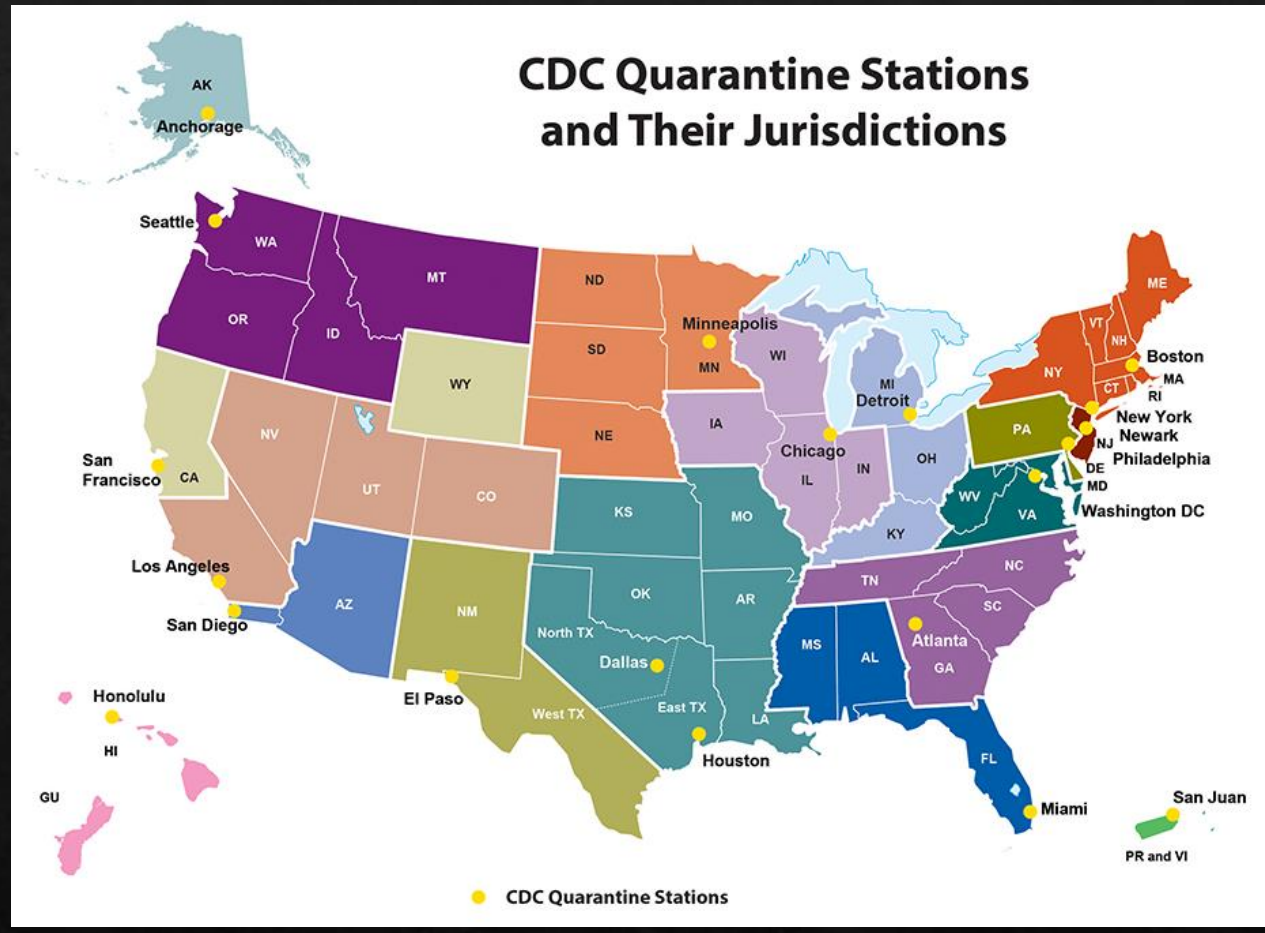




Note: Node size is the scaled degree. Edge width and color hue are scaled to number of flights between edges, with thicker lines and darker hues for higher weights.

# Coronavirus Screening at US Airports

- The administration followed CDC's existing plan to use 20 widely-used ports of entry as quarantine stations, including many of the America's busiest airports.

- However, this plan may not have considered particular traits of COVID-19, including that patients can remain asymptomatic for many days while still spreading the disease. CDC Director assesses as many as 25% of infected are asymptomatic but can still spread disease (1 April).




January 17: Screenings at JFK, LAX, and SFO for passengers from Wuhan, China

January 21: Add screenings at ATL and ORD for passengers from Wuhan.

January 28: Expand screenings to 20 airports; include all passengers arriving from China

Late February: First TSA screeners diagnosed with Coronavirus

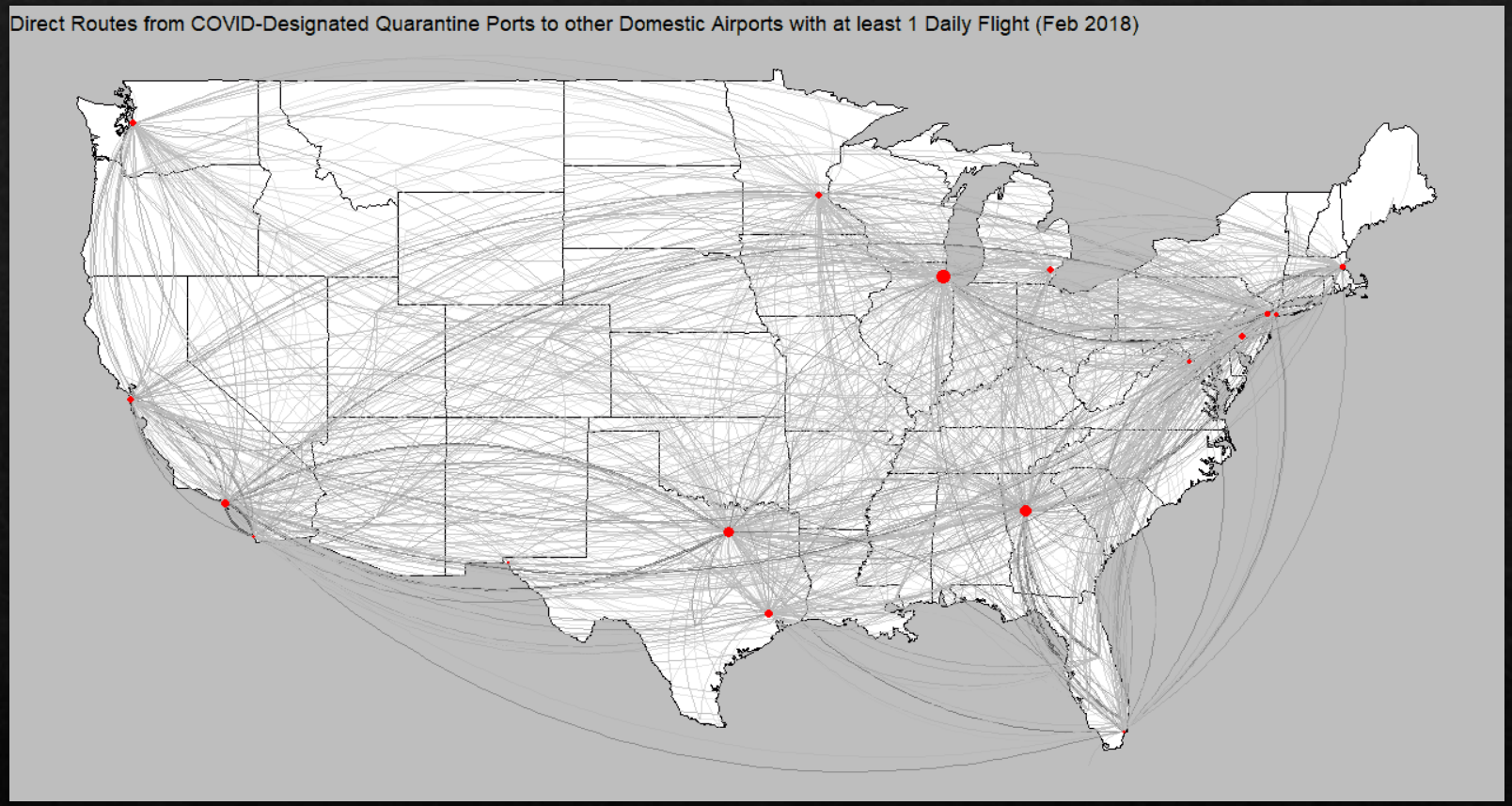
March 14: Implement new screening for European travel; announce European travel ban.

April 1: Over 60 TSA screeners have screened positive

# *Coronavirus Screening Network Metrics*

| | All 2018 Flights | All 2018 Flights WITHOUT Screening Airports | To/From Covid-Screening Airports |
|---|---|---|---|
| **Degree Centrality** | 76.6 | 49.9 | 53.2 |
| **Diameter** | 9 | 13 | 4 |
| **Mean Distance** | 3.3 | 4.2 | 2.6 |
| **Betwteen-ness** | 2700.0 | 3593.3 | 849.6 |
| **# Nodes** | 1264 | 1213 | 657 |
| **# Edges** | 47716 | 30251 | 17465 |
| **Total # Flights** | 8,514, 517 | 3,127,595 | 5,386,922 |

*Numbers are for unweighted network with US Territories included




January 17: Screenings at JFK, LAX, and SFO for passengers from Wuhan, China

January 21: Add screenings at ATL and ORD for passengers from Wuhan.

January 28: Expand screenings to 20 airports; include all passengers arriving from China

Late February: First TSA screeners diagnosed with Coronavirus

March 14: Implement new screening for European travel; announce European travel ban.

April 1: Over 60 TSA screeners have screened positive

Delta in 1998

American in 1998

Delta in 2018

American in 2018

United in 1998

Southwest in 1998

United in 2018

Southwest in 2018

# *Limitations & Next Steps*

- What happens to the network after taking certain airports out of the equation?
  - Which airports have the greatest affect the network efficiency and dynamics?
  - In the case of coronavirus spread, which airports specifically could/should be shut down? Does removing epicenters disconnect the whole network?
- Exploring and building weighted null models in R
- Examining to what extent each network is "scale-free"
- Visualizing this number of nodes and edges with non-geospatial layouts
- Explore dynamics of merging airline networks (e.g. US Airways/AMR/American Airlines, or United Airlines / Continental Airlines)
- Compare to biological or social network (measure assortativity)

# *References*


- Csardi G, Nepusz T (2006). "The igraph software package for complex network research." InterJournal, Complex Systems, 1695. http://igraph.org.
- Network Science by Albert-László Barabási. (n.d.). Retrieved April 13, 2020, from http://networksciencebook.com/
- OST_R | BTS | Transtats. (n.d.). Retrieved April 13, 2020, from https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=311
- The weighted random graph model Related content. (n.d.). https://doi.org/10.1088/1367-2630/11/7/073005
- Verma, T., Araújo, N. A. M., & Herrmann, H. J. (2014). Revealing the structure of the world airline network. Scientific Reports, 4(1), 1–6. https://doi.org/10.1038/srep05638
- Introduction to Network Theory: Modern Concepts, Algorithms, and Applications, Ernest Estrada, University of Strathclyde, Glasgow, Retrieved April 13, 2020, from https://slideplayer.com/slide/3893394/
- E. N. Gilbert. Random graphs. The Annals of Mathematical Statistics, 30:1141-1144, 1959.


- Link to github:
  - https://github.com/redhairedcelt/network_science_flights