Airline Travel Networks and COVID-19

Ryan Edelstein

December 12, 2021

CS 3891: Network Analysis in Healthcare

**Abstract**

This study focuses on airline travel during the COVID-19 pandemic and investigates the correlation between network structure and COVID-19 case rates. By creating a network of airports, where each airport is a node and directed edges are weighted for the number of passengers flying between the two airports, network analysis can be performed to understand air travel as a network. Airports are then associated with a COVID-19 case rate corresponding to the county in which they are located. After calculating connectivity, betweenness, and centrality, the question was how do these network statistics that reveal network structure correlate with COVID-19 case rates. It was discovered that high betweenness and centrality have a significant correlation with high case rates, and high connectivity between airports correlates with slightly more similarity in COVID-19 case rates. This analysis reveals the impact that airline travel has on the spread of infectious disease and the importance of managing and limiting travel in order to control an active pandemic.

**Introduction**

The spread of disease is directly impacted by the movement of the people or carriers of the disease. Globalization presents the possibility of rapid spread of viruses across continents and within countries. With the beginning of COVID-19 in China, the virus quickly reached every country in the world and reached the status of a global pandemic within three months of the recording of the first case[[1]](#footnote-0). Once the virus reached this elevated status, precautions taken to combat the spread largely revolved around limiting social contact between people.

One major policy change was drastic limitations to traveling. Both foreign and domestic flights were stopped for some time, and many countries, including the United States, imposed travel bans on all travel into and out of the country[[2]](#footnote-1). Over the first six months of the pandemic, travel industry companies like Southwest Airlines and Amtrak saw their passenger volumes drop to 10% of typical rates[[3]](#footnote-2). The decrease in travel allowed, accompanied by other social distancing and masking requirements, slowed the spread in the first months of the pandemic.

In the last year, travel has slowly reopened, with travel bans being removed, allowing people to travel to other countries. In addition, airline and train travel has slowly gone back to its normal rates. However, as cases continue to rise at times, and as new variants lead some countries to reimplement early pandemic protocols, it is important to understand the impact that travel is currently having on the spread of COVID-19. The aim of this study is to understand how the return to normal airline travel impacted the spread of COVID-19 in the United States.

This study focuses on airline travel data between airports in the United States during May of 2021. This month was chosen because travel was open back up in general, but there had been spikes in cases due to increased travel and gatherings during the summer months. In addition, vaccines were starting to be widely distributed in the United States, so there is justification for the return to travel. By May, all adults qualified for the COVID-19 vaccine, and 38.5% of the population was fully vaccinated by May[[4]](#footnote-3).

This study leverages network analysis in order to study the correlation between airline travel and COVID-19 rates. A set of airports together with the passengers traveling between each pair of them provides a simple motivation for the creation of a network. In this study, a network was created with airports as nodes and the number of passengers traveling between two airports as the edge weight in a directed graph. The aim was to create the network and then compare network analysis data with COVID-19 data from the regions surrounding the airports to see how the structure of the airline network correlates with the spread of disease.

Clearly, there are going to be other factors that correlate both with airports and COVID-19 case rates. Simply calculating the number of cases, for example, would be uninteresting since large cities, with high populations and case numbers, also have high volume airports. In this study, the aim was to focus not on the specific airports but on the structure of the network in total. In this way, it can be understood exactly how the increase in air travel impacted the spread, and in which regions of the United States.

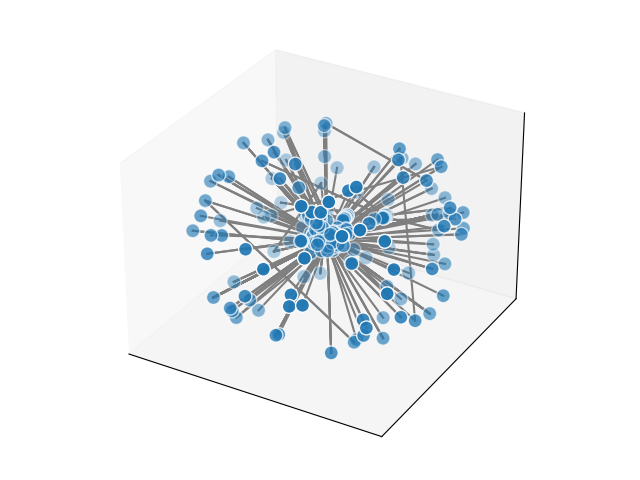
The goal of this study was to understand the correlation between network structure statistics, particularly connectedness and betweenness, and COVID-19 rates. The network was constructed based on airline data and then paired with COVID-19 data relevant to the areas of the airport to perform final analysis.

**Methods**

Network Construction

The first step in this study is to construct the network. This first requires a compilation of airline data. As noted previously, the data was selected to be airline travel during the month of May 2021. This data was taken from the public databases of the United States Department of Transportation[[5]](#footnote-4). The database contained 937 airports and the number of passengers going between each pair of airports during the month. This data was compiled into a summary table, containing each pair of airports for which there was travel, and the number of passengers going in each direction between them.

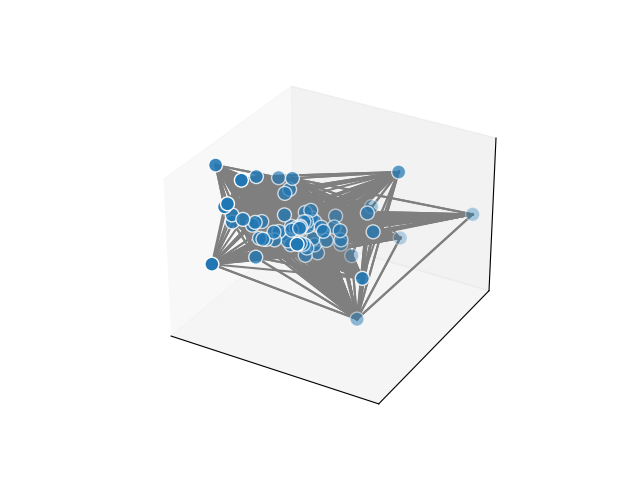
The network was constructed using the Python Networkx package[[6]](#footnote-5). Each airport was constructed as a node in the network. For each set of passengers between two airports, a directed edge was constructed between the associated two nodes. For example, if there were 1,000 passengers between airport A and airport B, then there is an edge of weight 1,000 from airport A to airport B. The following is a graph of the network, as produced by the Networkx package and drawing functionality in Python.



**Figure 1: 3-D graph of airline network with 937 airports**

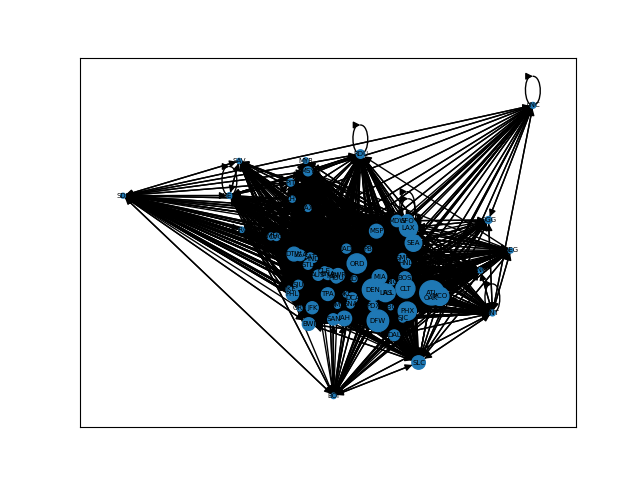
The first interesting thing to note is that there is a clear center to the network, with nodes that are directly connected to many other nodes. On the outside of the network are nodes with only one or two connections to something in the center. This reflects that there are major differences in node structure between large airports and small airports. Local airports typically have fewer flights and fly primarily into larger airports, while the largest airports may have flights to hundreds of different airports.

Having constructed this network, it became apparent that the size of the network needed to be shrunk. Too many of the airports had far less data than the top 100 airports, as the range of passengers was 500 to 3,000,000, and it would be too difficult to accurately measure the network statistics with such discrepancies. It was determined that using the top seventy airports would allow for the most accurate and non-skewed data analysis. The list of airports was trimmed to include only the seventy airports with the most inbound passengers. After trimming the airport list, the range of inbound passengers becomes 123,000 to 3,000,000. Although this is still a large range, it allows for data analysis to be done without statistics being skewed as much by size differences. From this, a very different structure of network appears:



**Figure 2: 3-D graph of airline network with 70 airports**

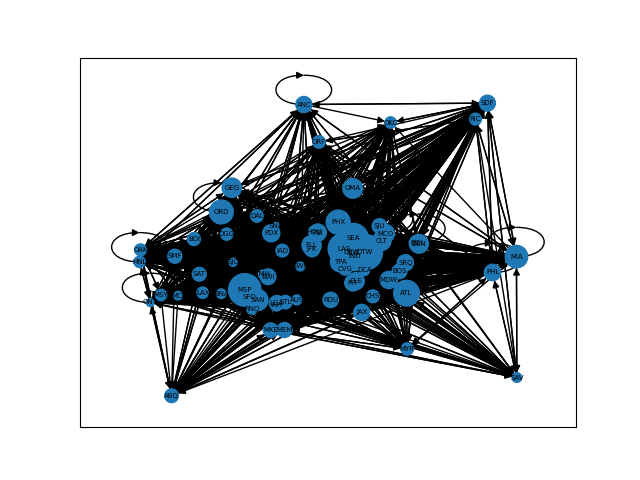
Clearly, this trimming reduced the network down to the most central nodes. Obviously, this drastically changes the network, and this change will be discussed in the limitations section of the paper. However, the more similar passenger volumes are essential for the data analysis in the later parts. The following is the same graph with nodes weighted based on their inbound passenger volume.



**Figure 3: Airline network with 70 airports, node weights for inbound passenger volume**

COVID-19 Data

Based on this network construction, the goal was to then associate each airport with COVID-19 data for its area. In this study, each airport was identified with the county in which it is contained. The airport then was paired with the COVID-19 case rate for that county in May 2021 as published by the New York Times[[7]](#footnote-6). For example, the ATL airport was associated with the May 2021 COVID-19 case rate for Fulton County, Georgia. Of course, this simple pairing method with the county containing the airport is not perfect, and this will be discussed in the limitations section of the paper. After assigning each airport its COVID-19 case rate, the following network emerges.



**Figure 4: Airline network with 70 airports, node weights for COVID-19 case rate**

Network Analysis

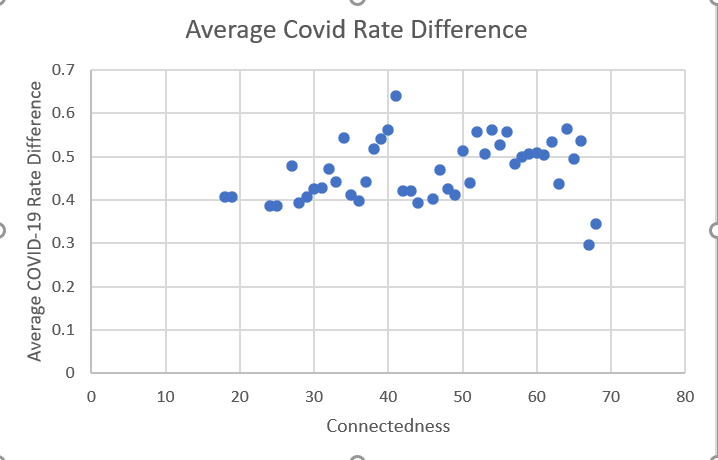
With the network and associated COVID-19 case rate data constructed, the network analysis can be performed. Three main statistics were calculated using the Networkx package in order to understand the local structure of the network. The first is connectivity, which is a value associated with each distinct pair of nodes to indicate their degree of connection in the network. For example, the ATL airport has 69 data points for connectivity, one for its connectivity with each other airport in the network. From this, a table was constructed containing each pair of airports, their associated connectivity value, and the percent difference between the COVID-19 case rates of them. The summary data table can be seen in **Table 1** in the appendix.

The second network statistic that was calculated is the betweenness of each node in the network. The betweenness of a node is based on the number of other nodes that have paths that pass through that node. High betweenness of an airport’s node indicates that the airport has flights going into it from other highly connected airports. From this data, a table was constructed containing each airport, its associated betweenness value as calculated, and its COVID-19 case rate. The data table can be seen in **Table 2** in the appendix.

The final network statistic that was calculated is the centrality of each node in the network. The centrality of a node is the fraction of other nodes that are directly connected by an edge to that node. Note that in this network, the maximum centrality value is two since the graph is directed. High centrality indicates that the node is a center, or hub, of the network and thus the airport is receiving travelers from many other airports. From this data, a table was constructed containing each airport, its associated centrality value, and its COVID-19 case rate. The data table can be seen in **Table 3** in the appendix.

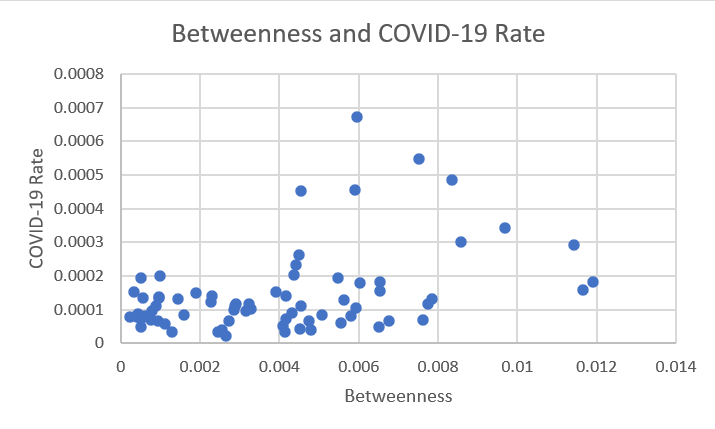
**Results**

The results are based on the correlation between the network structure and the COVID-19 case rates, so the three data tables presented above form the basis for the final analysis. Starting with connectivity, there is a slightly significant (.112 correlation coefficient) positive correlation between connectivity and percent difference in COVID-19 rates. The correlation can be seen in the following graph of the data.



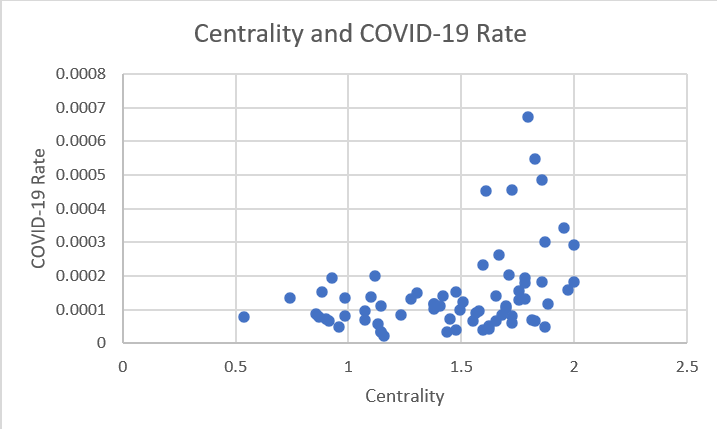
**Figure 5: Graph of connectivity and difference in COVID-19 case rate**

For the betweenness data, there is a moderately significant (.407 correlation coefficient) positive correlation between betweenness and COVID-19 rates. The correlation can be seen in the following graph of the data.



**Figure 6: Graph of betweenness and COVID-19 case rates**

Finally, for the centrality data, there is a moderately significant (.376 correlation coefficient) positive correlation between centrality and COVID-19 rates. The correlation can be seen in the following graph of the data.



**Figure 7: Graph of centrality and COVID-19 case rates**

**Conclusions**

Clearly, there is a correlation with highly central airports and higher COVID-19 case rates in the surrounding region. Cities with major airports pay the price of their large and diverse travel via higher rates of COVID-19 transmission. This is expected, as more people are coming into these counties and thus the passing of disease is more common. The high betweenness correlating with high case rates also reflects that the airports play a major role in the spread of COVID-19. It is clear that managing air travel is key in the limiting of the spread of contagious viruses.

There was a slight correlation between connectivity and percent difference in COVID-19 rates, reflecting that counties that have more flights between them are likely to have more similar COVID-19 case rates. However, this statistic is based on much less data, as each data point is the flights in one direction between two airports. It is possible that with more data there would be a larger observed correlation between these statistics.

Limitations

The first source of limitations is recognizing the plethora of other factors that impact COVID-19 case rates and are related to airports. As mentioned earlier, airports with higher passenger volumes, which also correlates with higher betweenness and centrality, also are in counties with higher populations. Dense population centers are likely to have more case rates, and it is certainly a limitation of this study that data controlled for population size and density.

The trimming of the data based on passenger volume was intended to correct for this. However, this presented a limitation in itself. As was noted earlier, this trimming from 937 airports to the 70 most traveled drastically impacted the network structure. It would be a good source of further research to redo the study with all of the airports. This would lead to more of a drastic difference in values of betweenness and centrality, as **Figure 1** clearly indicates that the less traveled airports are far less connected. However, this would require performing the study over multiple months in order to obtain enough data to supplement the smaller airports. This would result in more factors being introduced, as the pandemic evolved significantly from month to month.

Another limitation is the method by which COVID-19 case rates were associated with airports. County data is the most recorded and thus was the most practical for compiling for this study. However, in some cases the county containing an airport is not the most impacted by passengers traveling to and from the airport. For example, the Dulles Airport (IAD) is in Loudoun County, Virginia. However, the airport is a hub for travel to and from Washington, DC, which is about a half hour drive from the airport. Associating the airports more carefully with the county for which most passengers travel to from the airport would perhaps lead to more interesting and accurate results.

**Appendix**

**Table 1: Summary data table of connectivity values and percent difference COVID-19 case rate**

| Connectedness | Average Covid Rate Difference |
| --- | --- |
| 53 | 0.506060664 |
| 44 | 0.393335056 |
| 31 | 0.426794235 |
| 59 | 0.505929511 |
| 60 | 0.509025932 |
| 63 | 0.436941706 |
| 57 | 0.482609537 |
| 42 | 0.42048375 |
| 61 | 0.503138936 |
| 58 | 0.498454537 |
| 32 | 0.471014975 |
| 62 | 0.534952189 |
| 40 | 0.562221454 |
| 55 | 0.526313332 |
| 38 | 0.518080122 |
| 50 | 0.513377464 |
| 56 | 0.556577639 |
| 35 | 0.410957739 |
| 51 | 0.440222437 |
| 30 | 0.425167614 |
| 46 | 0.401578587 |
| 39 | 0.541093265 |
| 48 | 0.424985227 |
| 52 | 0.556655068 |
| 24 | 0.386183903 |
| 41 | 0.641131542 |
| 27 | 0.4792755 |
| 37 | 0.441466981 |
| 18 | 0.407759475 |
| 47 | 0.469296814 |
| 54 | 0.561639848 |
| 49 | 0.411859264 |
| 65 | 0.495213306 |
| 66 | 0.535450629 |
| 67 | 0.296063321 |
| 64 | 0.564599822 |
| 43 | 0.42048375 |
| 34 | 0.543162227 |
| 68 | 0.343915279 |
| 25 | 0.386183903 |
| 33 | 0.440679722 |
| 36 | 0.397368336 |
| 19 | 0.407759475 |
| 28 | 0.393704084 |
| 29 | 0.407172601 |

**Table 2: Airport betweenness value and COVID-19 case rate**

| Airport | Betweenness | Covid Rate |
| --- | --- | --- |
| MEM | 0.007736851 | 0.000116308 |
| SAT | 0.002864259 | 9.83034E-05 |
| JAX | 0.001450066 | 0.000132602 |
| PBI | 0.000330291 | 0.000150969 |
| BWI | 0.005622654 | 0.000128117 |
| IAD | 0.005073678 | 8.46355E-05 |
| DFW | 0.011650506 | 0.000158953 |
| AUS | 0.004083095 | 5.10204E-05 |
| CHS | 0.001601328 | 8.50741E-05 |
| CLT | 0.006037288 | 0.000178441 |
| RIC | 0.00051238 | 7.25474E-05 |
| DCA | 0.005543688 | 6.06338E-05 |
| SAV | 0.00052309 | 4.83709E-05 |
| ATL | 0.009692649 | 0.000342166 |
| SDF | 0.007845481 | 0.000130387 |
| BNA | 0.006519503 | 4.75406E-05 |
| CVG | 0.004320784 | 8.98214E-05 |
| STL | 0.003151685 | 9.42905E-05 |
| DEN | 0.011915172 | 0.000181412 |
| OMA | 0.000978206 | 0.000199343 |
| MIA | 0.004497719 | 0.000263158 |
| SJU | 0.000893881 | 0.000111772 |
| MSP | 0.007520923 | 0.000547197 |
| ONT | 0.004139598 | 3.34862E-05 |
| IAH | 0.006772777 | 6.57755E-05 |
| TPA | 0.004378821 | 0.000203125 |
| SEA | 0.005894395 | 0.000454229 |
| BDL | 0.000966049 | 0.000137936 |
| PHX | 0.008580833 | 0.000300674 |
| PDX | 0.003914406 | 0.000152549 |
| SMF | 0.003295299 | 0.000101804 |
| HNL | 0.000931377 | 6.66966E-05 |
| LAX | 0.00761839 | 6.78094E-05 |
| JFK | 0.003229747 | 0.000115334 |
| SFO | 0.004803524 | 4.00018E-05 |
| CLE | 0.00190866 | 0.000150607 |
| RNO | 0.001124867 | 5.72617E-05 |
| DTW | 0.005947881 | 0.000672751 |
| LGA | 0.002904964 | 0.000115334 |
| SAN | 0.004420479 | 0.000233153 |
| ORD | 0.011423243 | 0.000293358 |
| MDW | 0.006530586 | 0.000154369 |
| LAS | 0.00834802 | 0.000486788 |
| OAK | 0.004155937 | 7.30102E-05 |
| DAL | 0.005814426 | 8.00455E-05 |
| RDU | 0.002264404 | 0.000121403 |
| PHL | 0.004160131 | 0.000139392 |
| SNA | 0.002663368 | 0.000020466 |
| MSY | 0.00273184 | 6.65225E-05 |
| SLC | 0.00593625 | 0.00010431 |
| FLL | 0.005476691 | 0.000192524 |
| BOS | 0.004553767 | 0.000111953 |
| ANC | 0.000569047 | 0.000133963 |
| HOU | 0.004754682 | 6.57755E-05 |
| EWR | 0.004523245 | 4.13029E-05 |
| ORF | 0.000621942 | 8.08429E-05 |
| PIT | 0.00230734 | 0.00013898 |
| OKC | 0.000771126 | 6.77172E-05 |
| GEG | 0.000512566 | 0.000195104 |
| ABQ | 0.00077927 | 9.71844E-05 |
| MCO | 0.006527515 | 0.000181622 |
| SRQ | 0.000969626 | 0.000133911 |
| OGG | 0.000229067 | 7.76504E-05 |
| MKE | 0.002885444 | 0.000111026 |
| IND | 0.004536734 | 0.000451339 |
| CMH | 0.001301473 | 3.40698E-05 |
| SJC | 0.002460056 | 3.37137E-05 |
| MCI | 0.002561858 | 3.83076E-05 |
| BOI | 0.000424631 | 8.72117E-05 |
| MYR | 0.000404036 | 0.000079078 |

**Table 3: Airport centrality value and COVID-19 case rate**

| Airport | Centrality | Covid Rate |
| --- | --- | --- |
| MEM | 1.884057971 | 0.000116308 |
| SAT | 1.492753623 | 9.83034E-05 |
| JAX | 1.275362319 | 0.000132602 |
| PBI | 0.884057971 | 0.000150969 |
| BWI | 1.753623188 | 0.000128117 |
| IAD | 1.68115942 | 8.46355E-05 |
| DFW | 1.971014493 | 0.000158953 |
| AUS | 1.623188406 | 5.10204E-05 |
| CHS | 1.231884058 | 8.50741E-05 |
| CLT | 1.782608696 | 0.000178441 |
| RIC | 0.898550725 | 7.25474E-05 |
| DCA | 1.724637681 | 6.06338E-05 |
| SAV | 0.956521739 | 4.83709E-05 |
| ATL | 1.956521739 | 0.000342166 |
| SDF | 1.782608696 | 0.000130387 |
| BNA | 1.869565217 | 4.75406E-05 |
| CVG | 1.565217391 | 8.98214E-05 |
| STL | 1.579710145 | 9.42905E-05 |
| DEN | 2 | 0.000181412 |
| OMA | 1.115942029 | 0.000199343 |
| MIA | 1.666666667 | 0.000263158 |
| SJU | 1.144927536 | 0.000111772 |
| MSP | 1.826086957 | 0.000547197 |
| ONT | 1.434782609 | 3.34862E-05 |
| IAH | 1.826086957 | 6.57755E-05 |
| TPA | 1.710144928 | 0.000203125 |
| SEA | 1.724637681 | 0.000454229 |
| BDL | 1.101449275 | 0.000137936 |
| PHX | 1.869565217 | 0.000300674 |
| PDX | 1.47826087 | 0.000152549 |
| SMF | 1.376811594 | 0.000101804 |
| HNL | 0.913043478 | 6.66966E-05 |
| LAX | 1.811594203 | 6.78094E-05 |
| JFK | 1.376811594 | 0.000115334 |
| SFO | 1.594202899 | 4.00018E-05 |
| CLE | 1.304347826 | 0.000150607 |
| RNO | 1.130434783 | 5.72617E-05 |
| DTW | 1.797101449 | 0.000672751 |
| LGA | 1.376811594 | 0.000115334 |
| SAN | 1.594202899 | 0.000233153 |
| ORD | 2 | 0.000293358 |
| MDW | 1.753623188 | 0.000154369 |
| LAS | 1.855072464 | 0.000486788 |
| OAK | 1.449275362 | 7.30102E-05 |
| DAL | 1.724637681 | 8.00455E-05 |
| RDU | 1.507246377 | 0.000121403 |
| PHL | 1.652173913 | 0.000139392 |
| SNA | 1.15942029 | 0.000020466 |
| MSY | 1.550724638 | 6.65225E-05 |
| SLC | 1.695652174 | 0.00010431 |
| FLL | 1.782608696 | 0.000192524 |
| BOS | 1.695652174 | 0.000111953 |
| ANC | 0.739130435 | 0.000133963 |
| HOU | 1.652173913 | 6.57755E-05 |
| EWR | 1.623188406 | 4.13029E-05 |
| ORF | 0.985507246 | 8.08429E-05 |
| PIT | 1.420289855 | 0.00013898 |
| OKC | 1.072463768 | 6.77172E-05 |
| GEG | 0.927536232 | 0.000195104 |
| ABQ | 1.072463768 | 9.71844E-05 |
| MCO | 1.855072464 | 0.000181622 |
| SRQ | 0.985507246 | 0.000133911 |
| OGG | 0.536231884 | 7.76504E-05 |
| MKE | 1.405797101 | 0.000111026 |
| IND | 1.608695652 | 0.000451339 |
| CMH | 1.144927536 | 3.40698E-05 |
| SJC | 1.144927536 | 3.37137E-05 |
| MCI | 1.47826087 | 3.83076E-05 |
| BOI | 0.855072464 | 8.72117E-05 |
| MYR | 0.869565217 | 0.000079078 |

1. https://www.cdc.gov/museum/timeline/covid19.html#:~:text=December%2012%2C%202019%20A,of%20breath%20and%20fever. [↑](#footnote-ref-0)
2. https://www.cdc.gov/coronavirus/2019-ncov/travelers/from-other-countries.html [↑](#footnote-ref-1)
3. https://www.brookings.edu/research/the-covid-19-travel-shock-hit-tourism-dependent-economies-hard/ [↑](#footnote-ref-2)
4. https://ourworldindata.org/covid-vaccinations?country=USA [↑](#footnote-ref-3)
5. https://www.transtats.bts.gov/Fields.asp?gnoyr\_VQ=FIL [↑](#footnote-ref-4)
6. https://networkx.org/ [↑](#footnote-ref-5)
7. https://www.nytimes.com/interactive/2021/us/covid-cases.html [↑](#footnote-ref-6)