CE 263N Scalable Spatial Analytics

Final Project Paper

Method for Benchmarking Diffusion of Infectious Disease at an Airport's Gates Using Bipartite Networks

Alvin Zhou
Pietro Achatz Antonelli

December 14th, 2020

Introduction

With the spread of Covid-19 throughout 2020, the aviation network proved pivotal in spreading the infection across the world with unprecedented speed, turning a local outbreak to a global pandemic in the matter of a few months. Unsurprisingly, the aviation industry is one of the ones that has been most scrutinized and has had devastating losses throughout the pandemic. Although planes undoubtedly contributed in carrying disease vectors at far lengths, making our world a little bit more like a small-world network, the degree at which disease transmission occurs within airplanes and airports is debatable. On one side, there are large spaces, strict mask and social distancing policies, prodigious cleaning services, and advanced air filtration systems that seem to reduce the chance of transmission. On the other side, the fact that airplanes and airports place travelers from many different regions in close proximity makes transmission within airports disproportionately consequential towards the diffusion of disease.

Prior work performed by Wells et. al. in studying the spread of Covid-19 from China through the global airport network showed that border control measures had some effectiveness at slowing the spread of the virus, but ultimately failed to contain it. Their methods involved using daily incidence data from mainland China and data describing the global airport network combined with a Monte Carlo simulation approach to estimate the risk of each country importing the virus over time [1].

There are numerous practices that aim to reduce the risk of transmission. A recently applied one consists of boarding procedures where passengers at the end of the plane are boarded first and the ones in front last to avoid random mixing. On a much larger scale, many studies apply concepts of random mixing on the aviation network by using airports as nodes and the flights as links. These models provide general estimates for how diffusion of a disease can occur and flights can perhaps be altered to mitigate the risk while keeping traffic going. The travel bubbles created during Covid-19 in some parts of the world are clusters of travel that do just this.

But what can be done on an airport management level, given that a set of flights would already have been established? Our team wondered if the concept of travel bubbles or reduced random mixing could be applied differently, on gate assignments for flights. A gate servicing hundreds of gates at other airports, will easily allow disease vectors to spread across all the world. The transmission could occur on infected surfaces such as chairs, in restricted air volumes such as jet bridges, or through the infected personnel that run that gate. If instead that gate only serviced a handful of other airports, a highly infectious disease could be kept in its own compromised channel without diffusing to new regions.

Our hypothesis was that a more clustered assignment between gates and airports would reduce the overall spreading potential, isolating the transmission within high risk travel corridors and keeping the low risk travel corridors protected. We did not have a way to compare more or less clustered gate-airport assignments but we decided to develop a method to understand and benchmark the potential for diffusion of disease at airport gates, given a certain gate-assignment network and a measure of disease incidence across the serviced regions.

Methodology

We had access to the Aircraft Operations Database (AODB) from San Francisco International Airport (SFO), which we used to make an exploratory case study. The AODB is automatically generated by the Aerobahn system, and it tracks every aircraft arrival and departure. Of relevance to this study, they display the origin airport and assigned gate for each arrival and the assigned gate and destination airport for each departure, along with a timestamp and other operational parameters.

We selected the first 17 days of January 2018 as a time range. This time range is not appropriate for the COVID-19 pandemic, but it has the correct data structure that will allow to develop a computational method that could then be applied on more appropriate data.

We selected all arrivals and departures and used them to construct a bipartite network. With each flight being a directed edge, the nodes came in 2 families: the set of gates at SFO and the set of airports SFO's flights arrive from or depart to.

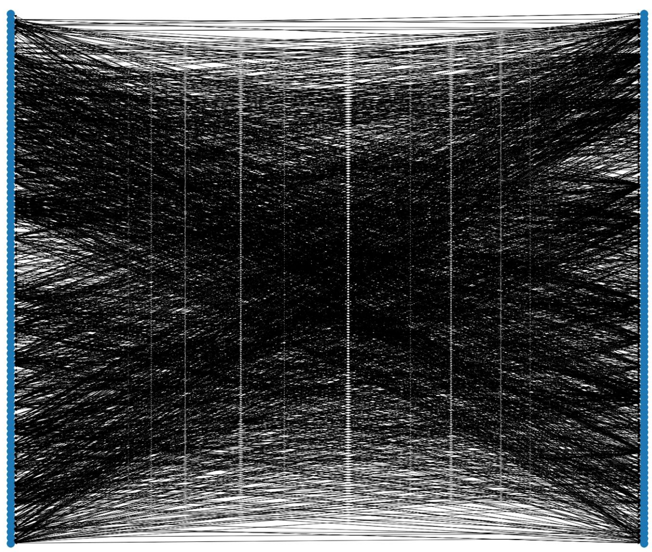

Figure 1. *Directed bipartite network with gates (left) and airports (right)*

We then computed the weighted projection of the bipartite network on the gate nodes. In doing so, the network becomes undirected and unweighted, which will be a simplifying assumption. In the same way, we computed the projection of the bipartite network onto the airport nodes. Both of these projections were graphed on the proper map tesselation to visualize how the SFO gates are connected by other airports and how other airports are connected by SFO gates respectively.

To have an understanding of disease incidence for the flights in our bipartite network, we referred to the Johns Hopkings COVID-19 database. We chose September 25th as the date to understand the incidence of positive cases per 100,000 people in each US state. Then each airport was given an attribute of incidence according to the state it is in.

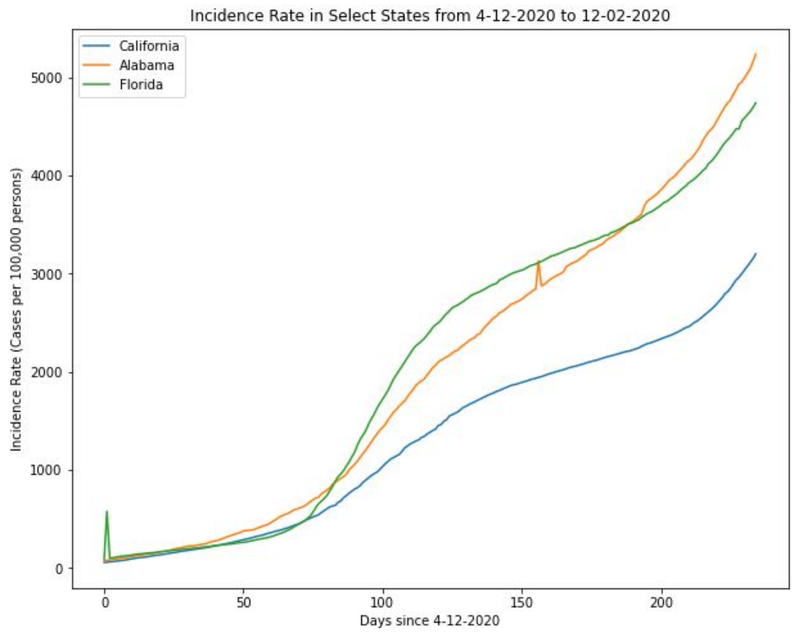

Figure 2. *Covid incidence rates in California Alabama and Florida*

Next, we defined the mixing potential as the abstract measure of how an infectious disease could diffuse from a high incidence airport to a low incidence airport. As a simple initial model, we assumed that diffusion to be a factor of the ratio in incidence between two regions. For each link within the netw ork projected on airports A and B, we calculated:

$$MixingPotential_{AB} = Max\left[\frac{Inc(A)}{Inc(B)}, \frac{Inc(B)}{Inc(A)}\right]$$

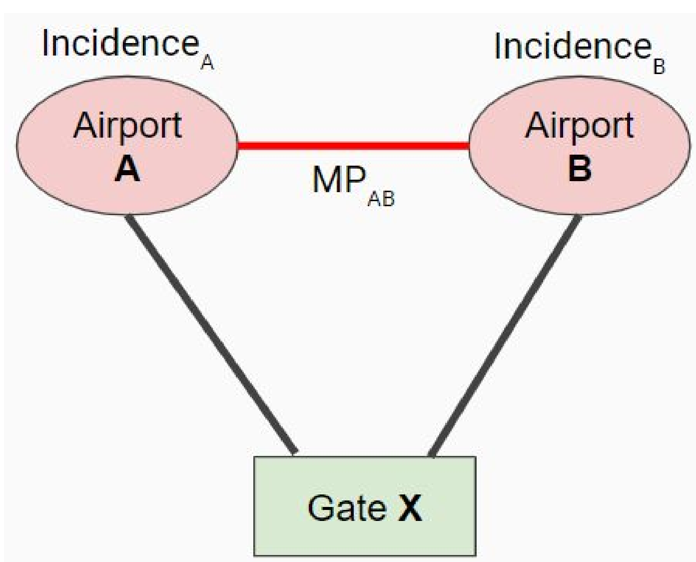

Figure 3. *Model for diffusion between two airports at a gate*

Then, we observed where this mixing potential would actually occur in the airport gates. For each gate X, we defined the Diffusion Risk as the sum of the mixing potential for all airport pairs that that gate connects.

$$MixingRisk(X) = \sum_{i=1}^{len(links(X))} MixingPotential(links(X)_i))$$

For each gate node within the projection on gate nodes, we calculated the number of degrees and the clustering coefficient. Scatterplots were made to observe the correlation between the degree, clustering coefficient and the mixing risk. Where it seemed appropriate, a linear regression was fit.

Results

The projected bipartite networks are shown in Figures 4 and 5. The projection onto gate nodes illustrates how gates are associated with their connected airports. The projection onto airport nodes, even more relevant to understanding diffusion, shows the role of SFO in connecting airports around the United States. Because of the large volume of flights handled by a large airport such as SFO, the resulting networks can be seen to have a large degree of interconnectedness.This indicates the possibility that some of these gates may be inadvertently mixing flights from many healthier regions with more infected regions.

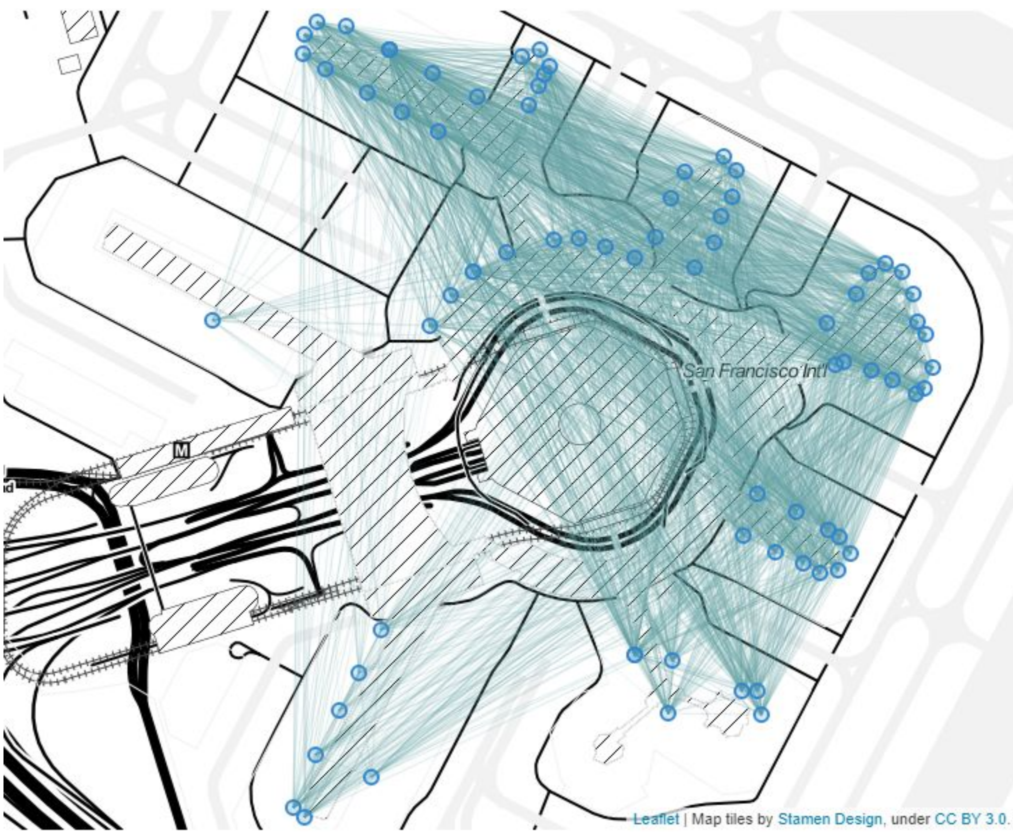

Figure 4. *Projected bipartite network onto the gate nodes*

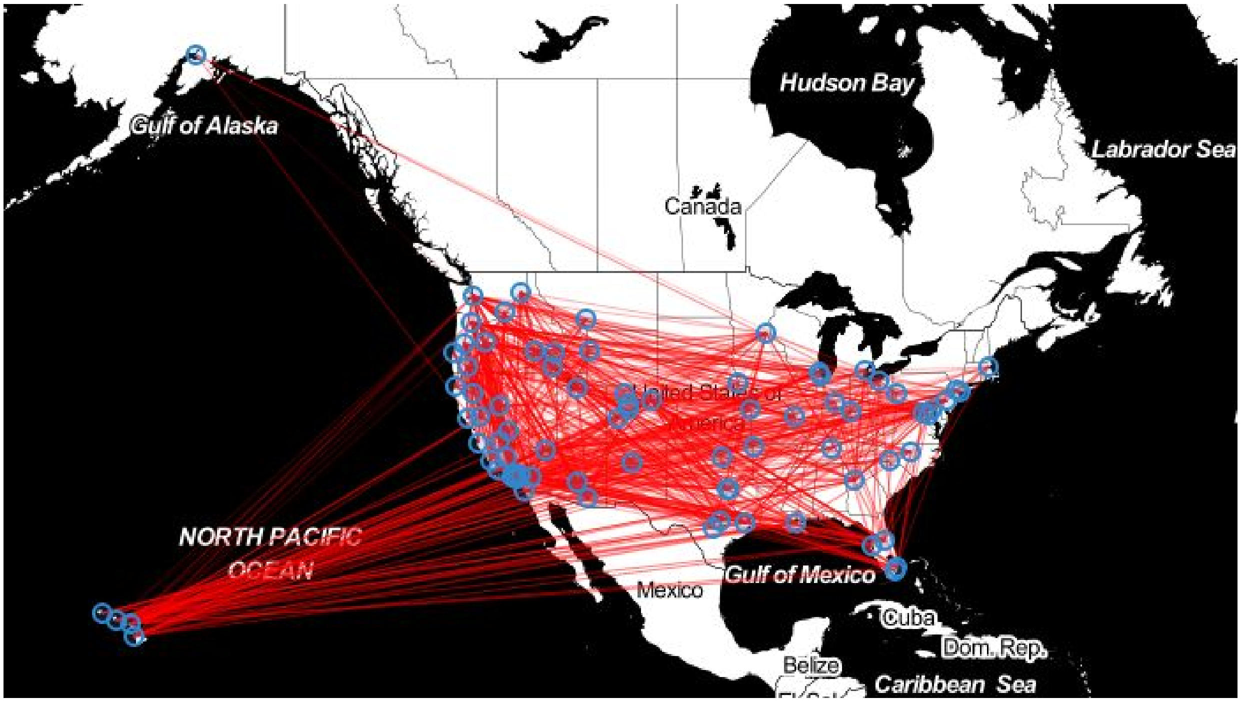

Figure 5. *Projected bipartite network onto the airport nodes.*

Figure 6, the plot displaying the gates with the greatest mixing risk, confirms this possibility. The highest risk value, for Gate 77B, was calculated to be above 139 using the above equations. The lowest risk was 0, likely because of the lack of flights to a certain gate.

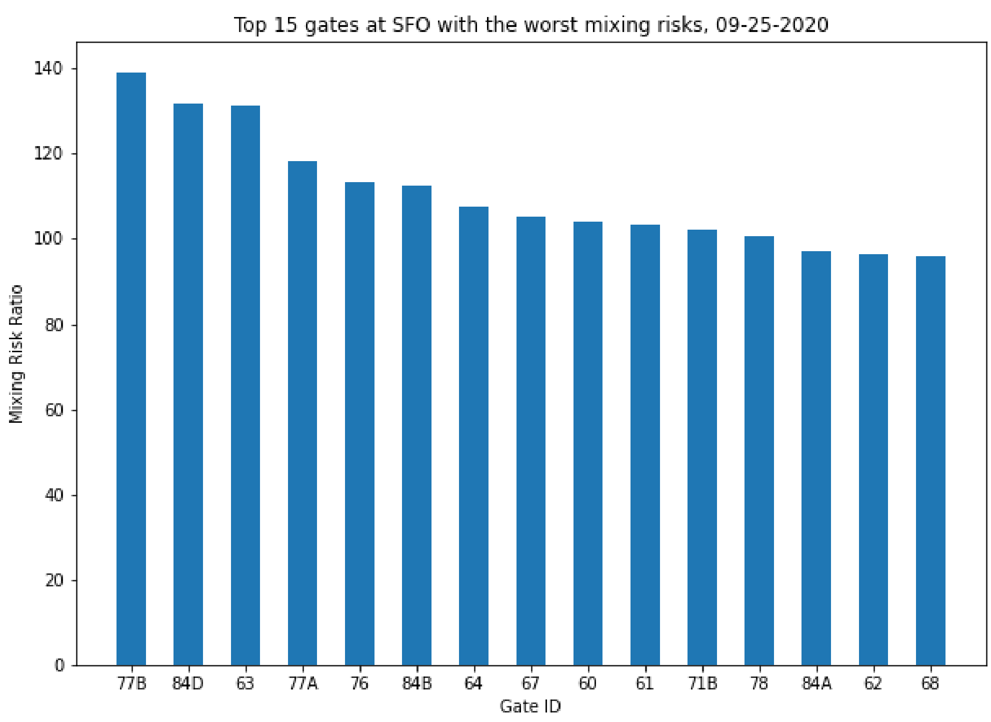

Figure 6. *Gates with the highest mixing risks for September 25th, 2020.*

Figure 7 shows a scatterplot of the clustering coefficient of each gate note within the gate projection versus the calculated mixing risk The result is inconclusive as there does not seem to be a significant correlation. Perhaps there is a correlation between the average correlation coefficient and average mixing risk of a gate assignment system but that relationship is not evident by looking at individual gates. Further data exploration would be necessary.

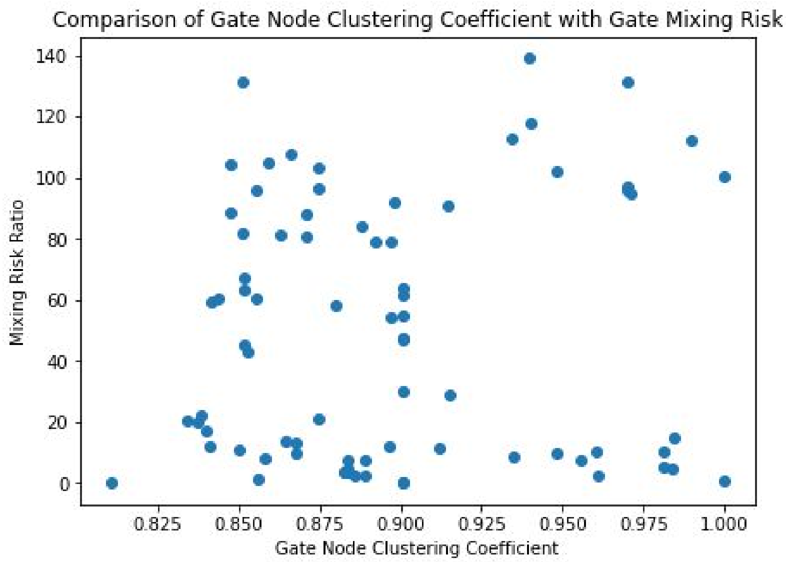

Figure 7. *Clustering Coefficient Comparison*

Figure 8 Shows the same scatterplot of the degree of each gate node within the gate projection versus the mixing risk, the difference being the graph on the right is in log scale. Due to a visible positive correlation, a linear regression was fit to the plot. The correlation is probable, given that the mixing risk is calculated as a sum over every link of a node.

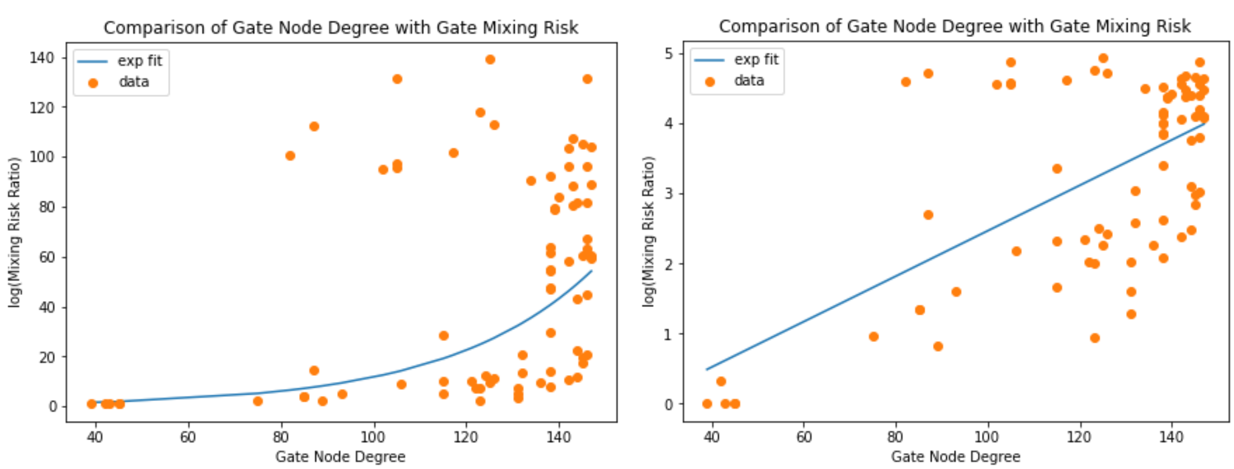

Figure 8. *Clustering Coefficient Comparison*

Conclusion

The Mixing Risk metric proposed provides a succinct summary of the contribution to the spread of infectious disease from each airport gate. This benchmark can then be used to reduce the spread of a specific disease, such as COVID-19, by isolating flights from heavily infected regions from the rest of the traveling population to minimize the risk value. However, this report proposes only a preliminary model for the Mixing Risk benchmark metric.

There is a significant variation between the maximum and minimum diffusion risk, indicating that the metric highlights a problem that could be mitigated. Thorough application of this metric could expose the diffusion risk present at an airport. If they applied it prior to setting gate assignments, they could optimize the system to reduce diffusion risk. Even simply reducing the number of degrees within the network by essentially having gates be specialized in certain airports could help reduce the risk of random diffusion.

This is a simple idea that our team did not find in current airport COVID-19 mitigation strategies. It is an interesting and opaque phenomenon, because failures to address this problem do not necessarily have consequences on the analysed airport, but it has considerable consequence on affecting the low risk regions an airport is connected to. However, we are confident that if transmission is a problem in airports and airplanes, this method of minimizing diffusion could have slowed the unprecedented spread of the pandemic.

Future Work

Topics for improvement in this model include using a more sophisticated epidemic model or dataset to estimate the incidence rate at each airport in the network. A more granular dataset would allow for the mapping of incidence rates at the county or even city level. Such estimations provide an increase in accuracy when predicting the true numbers of infected passengers from a single airport. Additional variables beyond the incidence rate can also be utilized to refine the benchmark further.

For a better study, we would need to collect more appropriate gate assignment data, where the AODB dates match those of the COVID incidence rates. If worth it, this analysis could run in real time by using APIs on our data, which could theoretically be collected from public open sources.

We can refine our diffusion model with the plentiful network data we already have. Passenger counts, flight counts and the direction of passenger flow can all affect the real diffusion risk and they could be integrated in our currently simplified model.

We could also take a step into understanding diffusion in the rest of the airport terminals by looking at gate proximity and understanding passenger flow networks within the terminal building. However, this would be a new and challenging scope to add to the research.

Sources

[1] Chad R.Wells, Pratha Sah, et. al. “Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak” . *Proceedings of the National Academy of Sciences Mar 2020,* 117 (13) 7504-7509. DOI: 10.1073/pnas.2002616117

[2] Ming Liu, Yihong Xiao, "Modeling and Analysis of Epidemic Diffusion within Small-World Network", *Journal of Applied Mathematics*, vol. 2012, Article ID 841531, 14 pages, 2012. https://doi.org/10.1155/2012/841531

[3] Networkx Source Code
https://networkx.org/documentation/stable/_modules/networkx/algorithms/bipartite/projection.html