Tempering the Spread of Epidemics on Aerial Networks

Amaury Combes, Alexandre Daeniker, Manuel Faysse & Alexis Montavon

January 2019

Abstract

In this report, we present an analysis of epidemic spread through flight routes modeled with SIS (Susceptible-Infected-Susceptible). The purpose of this study is to analyse network properties in order to find the best way to slow down a very infectious disease, focusing only on flight transportation and the contamination of people on large distances.

1 Introduction

With the increasing amount of people travelling with planes[1], it is only logical that the number of large distance contamination increases and the chances of large scale epidemic along with it. In this study, we try to use network properties to find the appropriate way to mitigate the development of a widespread disease, without breaking down the network completely. With the aerial route network we use, two main ideas will be developed: removing important edges and taking down central nodes. We will define a node or edge as "important" based on different centrality criteria explained in section 2.3. Choosing such a project allowed us to use crucial concepts we learnt during the class while trying to work on a real world scenario. The idea behind this study was inspired by [1].
We will start by introducing the scenario we have simulated, explain the model we chose and give the techniques we used to select the edges and nodes we removed from the network. We will then describe the dataset we worked on, as well as the way we constructed our graph and discuss the results we obtained. Finally, we will conclude on the improvements that can be done to this project.

2 Methodology

2.1 Pandemic scenario

January 1, 2019: A new type of infectious disease breaks out in Lausanne, and contaminates a small part of the population. With every passing day, the epidemic spreads across the globe... This is the morbid scenario our project is meant to model. As we are working with flight routes, we model the infection rate of a country by the proportion of infected airports it has. The main idea is that the disease starts in Geneva Airport, chosen because it has a node degree of 100, which is much higher than the average degree (21.03), but isn't a major hub either like Atlanta (with degree 911). The infection has the ability to spread to $\approx 80\%$ the world in 200 days and stabilise near 81% of infected persons in the population. [TODO: add simulation graph] The purpose of this study is to find a

[1] https://www.nationalgeographic.com/environment/urban-expeditions/transportation/air-travel-fuel-emissions-environment/

way to reduce the spreading as much as possible while having the smallest possible impact on the network.

2.2 SIS Model

With a small study of epidemic models, as in [2] and the course, we reached the conclusion that the model that would make the most sense for our scenario is the SIS model[2]. As we see it, an airport can either be susceptible (healthy), or it can be infected, having contaminated people inside the airport and ready to travel. Airports become infected if a traveller passes by, at rate $\beta = 0.01$ but can also recover to a healthy (and susceptible once more) state if there are no remaining infected passengers, with rate $\lambda = 0.005$. The idea is then quite simple, at every iteration (seen as days in our scenario) every node (airport) could receive an infected traveller from a neighbouring node and become a source of infection, or it could recover from its infection and become healthy again (e.g. quarantine of infected population).

2.3 Edges & Nodes selection

In order to mitigate the expansion of the disease we will have to remove key edges or nodes. We remove nodes based on 3 different criteria: degree centrality, betweenness centrality and closeness centrality. The degree centrality for each node is the fraction of nodes it is connected to. The betweenness centrality corresponds to the amount of shortest paths between two of the graph's node that pass through the edge/node. Finally, the closeness centrality of a node refers to the reciprocal of the average shortest path between the node and all of the others.

First we removed nodes based on the degree and observed that it broke the graph's connectivity. Consequently, for each criterion, we removed nodes in an iterative way such that the connectivity was not impacted. Finally, we opted for a finer grain strategy by removing edges using the betweenness value.

3 Dataset & Graph construction

3.1 Dataset

We used two different dataset from the OpenFlights project that were available online and suggested to us in class. The first one[3], representing the airports, contains information such as the name, the ID and the location (city, country, longitude, latitude, altitude and timezone) of 12057 airports. The second one[4], represents all commercial flight routes going from one airport to another as of June 2014. It contains information such as the airline, the source and destination airport, the number of stops and the equipment for 67663 different flights. Some flight are operated by different companies but link the same airports.

3.2 Graph construction

We chose to build a directed and weighted graph. Every line of the adjacency matrix represents the source airport and every column represents the destination airport. After a bit of necessary cleaning (flights with no source/destination, IDs that were not matching in both datasets, etc...) we were left

[2] https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology
[3] https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat
[4] https://raw.githubusercontent.com/jpatokal/openflights/master/data/routes.dat

with 7543 airports and 66103 routes. When building the adjacency matrix we incremented the value of each link for every route that existed between each pair of airport, to create the edges weight. In order for our scenario and study to make sense we decided to drop every isolated node or group of nodes and work only with the largest connected component, leaving us with 3143 nodes and 18586 edges. This graph is the base of our analysis and has an average degree of 21.03, a diameter of 12 and an average shortest path of 3.998.

4 Analysis & Results

We ran the simulation 5 times, once as the control, and 4 with different removals of network objects. The control experiment, with the original network led to 83% of the population infected after 800 days and an average shortest path of 3.9.

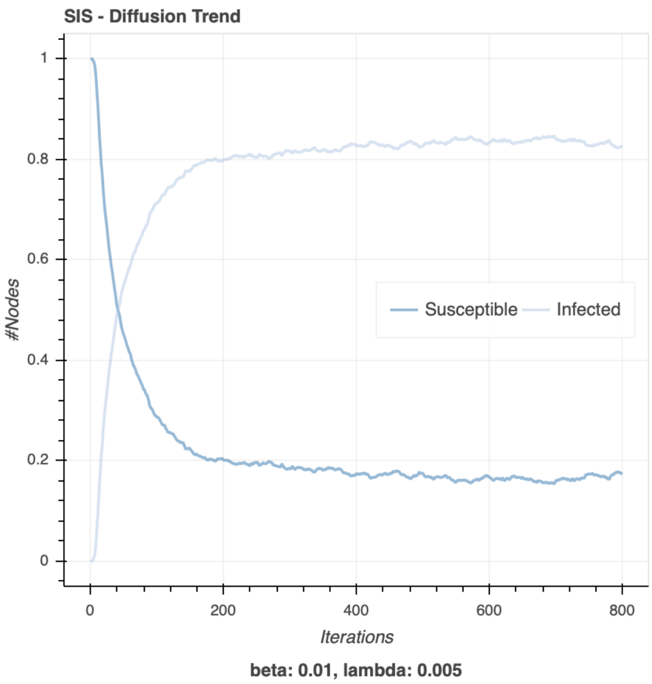

Figure 1: Evolution of the epidemy on the original network (Control experiment)

Removing the nodes with the highest degree led to a rate of 77 % of infected while the average shortest path did not increase significantly. We then removed nodes with regards to their closeness centrality, which also led to improved results, 78% of infected while maintaining a stable average shortest path length. Finally, we removed nodes while looking at their betweenness coefficient and got again a rate of 77%, still maintaining the shortest path.
However, the best results were yielded with the least drastic of the policies, edge (or flight) removal, in regards to betweenness centrality. The graph's nodes were not affected, and this allowed the network to remain fully connected. However, the epidemy had a harder time spreading, considering the 'highly used' traffic routes were taken down. We thus managed to reduce the proportion of the population which would theoretically be infected (76%) while allowing for all airports to remain

accessible to all passengers. Obviously, this policy's major backlash was to significantly augment the average shortest path in the network, which jumped from 3.9 to 15.5 (397% increase).

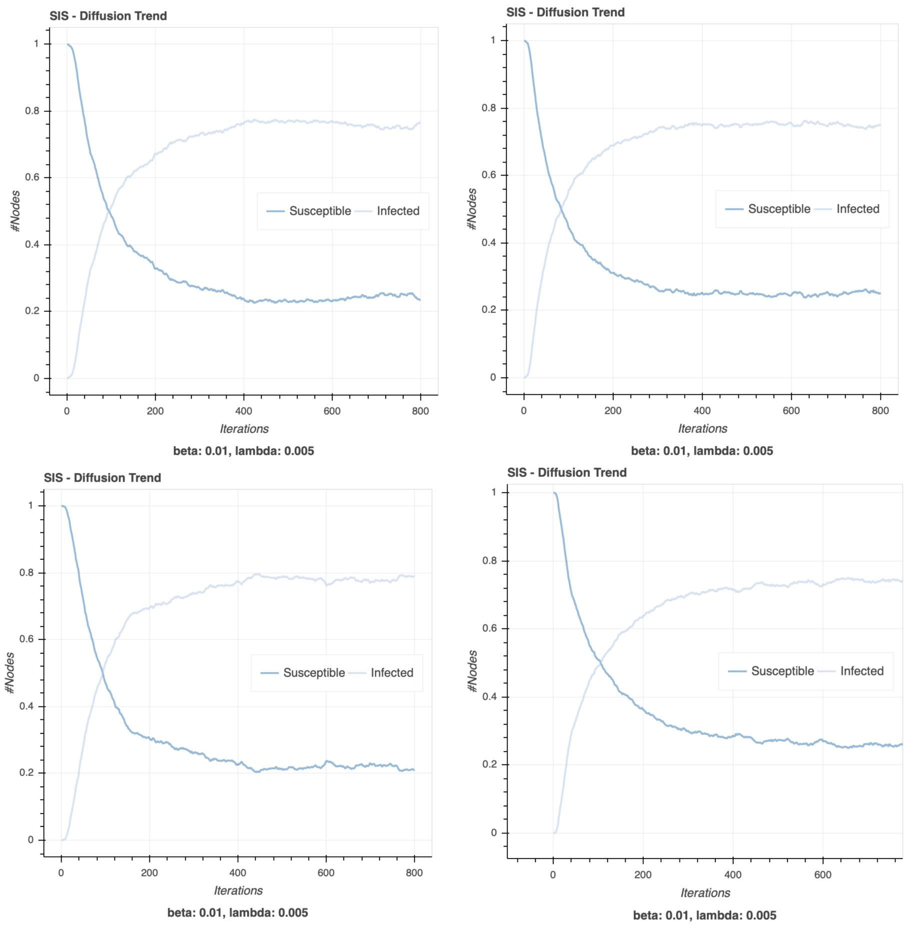

Figure 2: Evolution of the pandemic spread with (from left to right, top to bottom) removal based on node degree, betweenness centrality, closeness centrality, and edge betweenness centrality

5 Conclusion

To conclude, we successfully applied nodes and edges reduction in order to mitigate the propagation of a very infectious disease over an airline network. Completely stopping the epidemic can be done easily but doing without breaking down the network properties was not a trivial task. In the end the best method was to remove edges based on the betweenness coefficient.

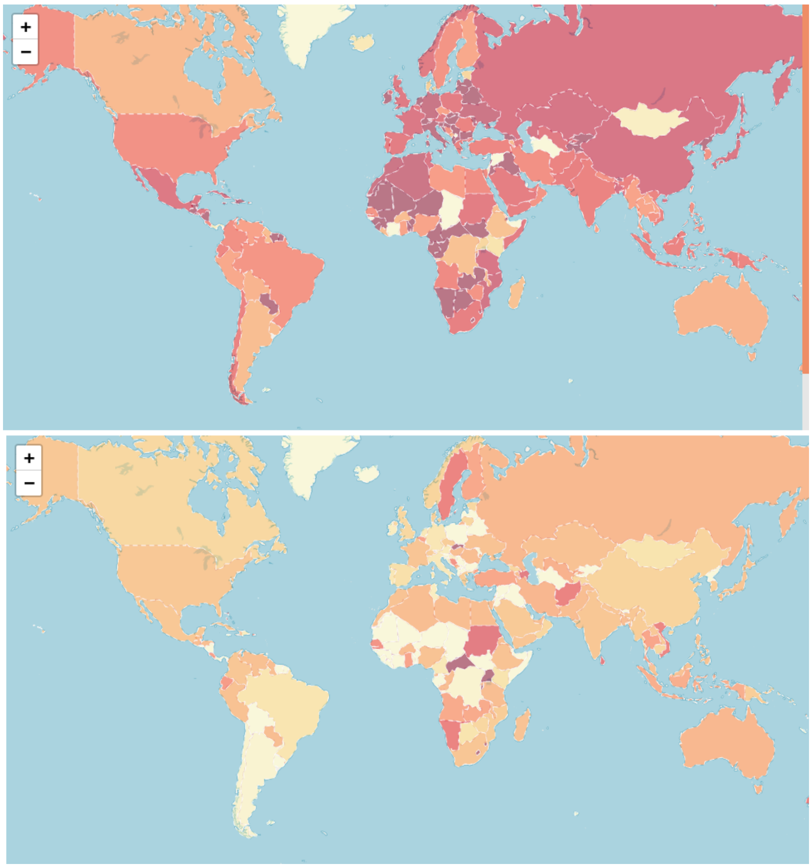

Figure 3 : Contamination map after 100 days of the original network (top) and the altered network with node removal policy by degree (bottom) (animation available as HTML file).

Further work on the subject could include an increase in the network edges, in order to simulate roads. We could measure the distance between two airports, using the latitude and longitude coordinates and add roads under a certain threshold. Having already thought about this, but lacking the time for a significant application, we would compute the distance in this way:

- a = latitude of the source airport
- b = latitude of the destination airport
- c = longitude of the source airport
- d = longitude of the destination airport
- R = equatorial radius $6378,1370\ km$
- D = distance between the two airports

$$D = R * \arccos(sin(a) * sin(b) + cos(a) * cos(b) * cos(c - d))^{5} \tag{1}$$

An interesting study would be the impact of the roads on the global spread of the disease, the effect on the network when we remove certain flight routes and the differences with flight propagation.

[5] $https : //geodesie.ign.fr/contenu/fichiers/Distance_longitude_latitude.pdf$

6 Addendum

This project can be found at: https://github.com/montalex/NTDS_2018_Final_Project

6.1 Libraries used

Python: https://www.python.org
Numpy: http://www.numpy.org
Pandas: https://pandas.pydata.org
Networkx : https://networkx.github.io
Mathplotlib : https://matplotlib.org
NDlib: https://ndlib.readthedocs.io/en/latest/

References

[1] L. Hufnagel, D. Brockmann, and T. Geisel, *Forecast and control of epidemics in a globalized world*, PNAS October 19, 2004 101 (42), (2004), pp. 15124–15129.

[2] T. Tassier, *The Economics of Epidemiology*, Springer, 2013.