

Course Project: It's Airborne

[Complex Networks CS-E4770]

João de Oliveira Gigante Pinheiro
102577172

December 2024

Problem Solutions

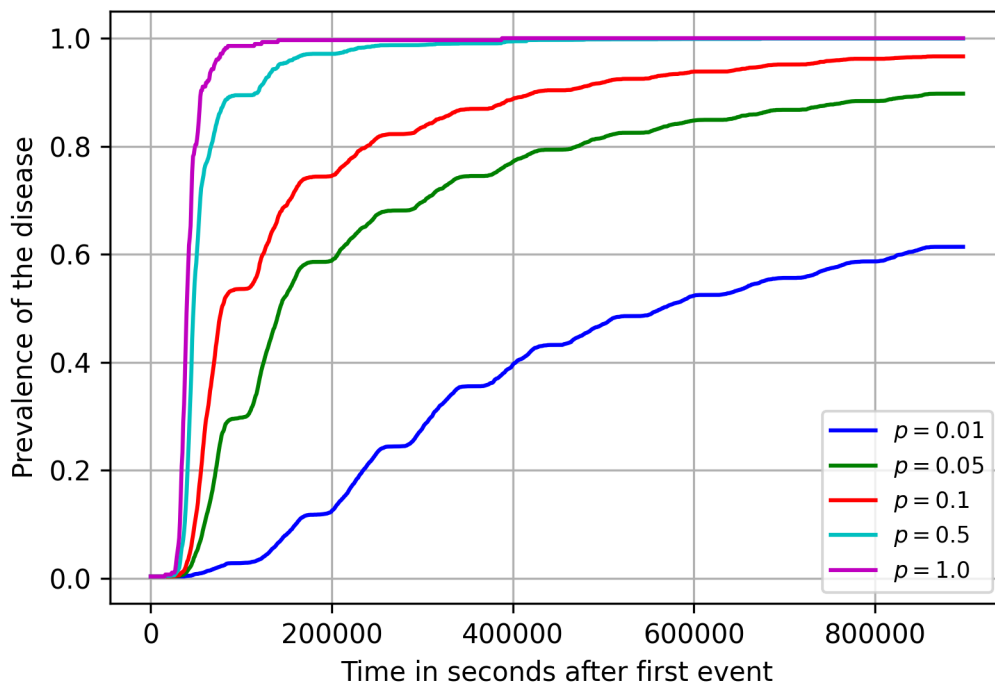
Problem 1: Basic Implementation

(a)

With $p = 1$, the disease is always transmitted and node SLC infected at beginning, node **ANC** becomes infected at **1229283600**, which falls within the expected range of 1229283000-1229284000.

Problem 2: Effect of infection probability p on spreading speed

(a)



(b)

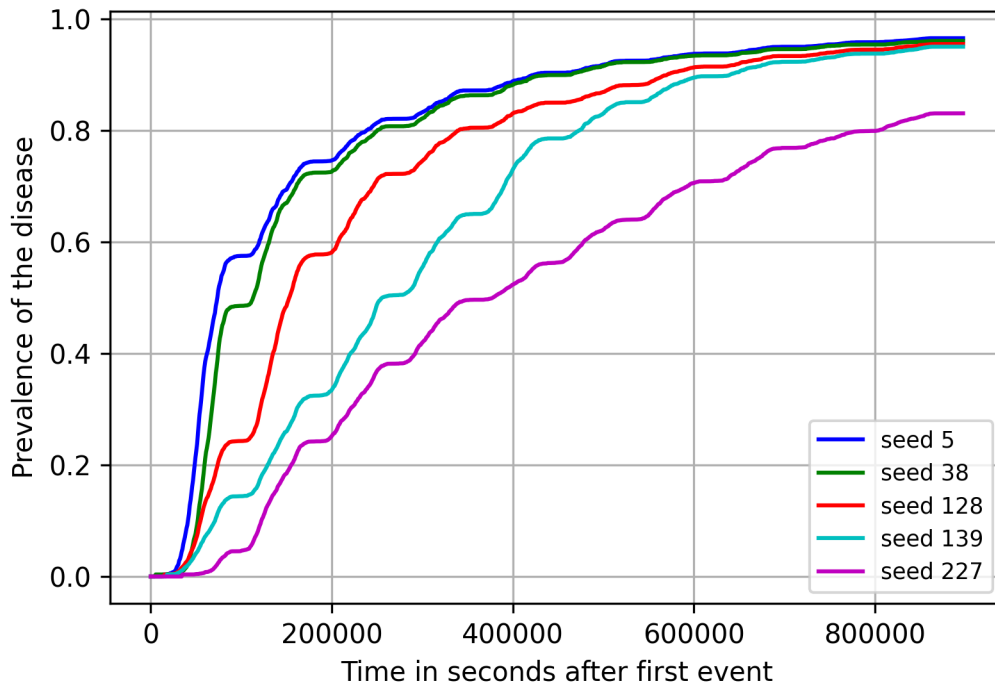
The network only becomes fully infected for probabilities 0.5 and 1.0.

As we are using the same starting node for every probability, the curves are similar. We just get a "delayed" curve with lower probabilities, so the periodic "steps" are reproduced in all of the curves. These "steps" can be described by reaching a new hub, a node with very high degree or that belong to top 1% in other centrality measures, which will be a bridge point to spread the disease and have a fast

evolution. Then when almost all of this community, where the hub belongs, gets infected the number of new nodes infected flattens until a new hub is reached. These steps correspond to the process of disease transmission through clusters connected by high-degree nodes, reflecting the network's structure.

Problem 3: Effect of seed node selection on spreading speed

(a)



(b)

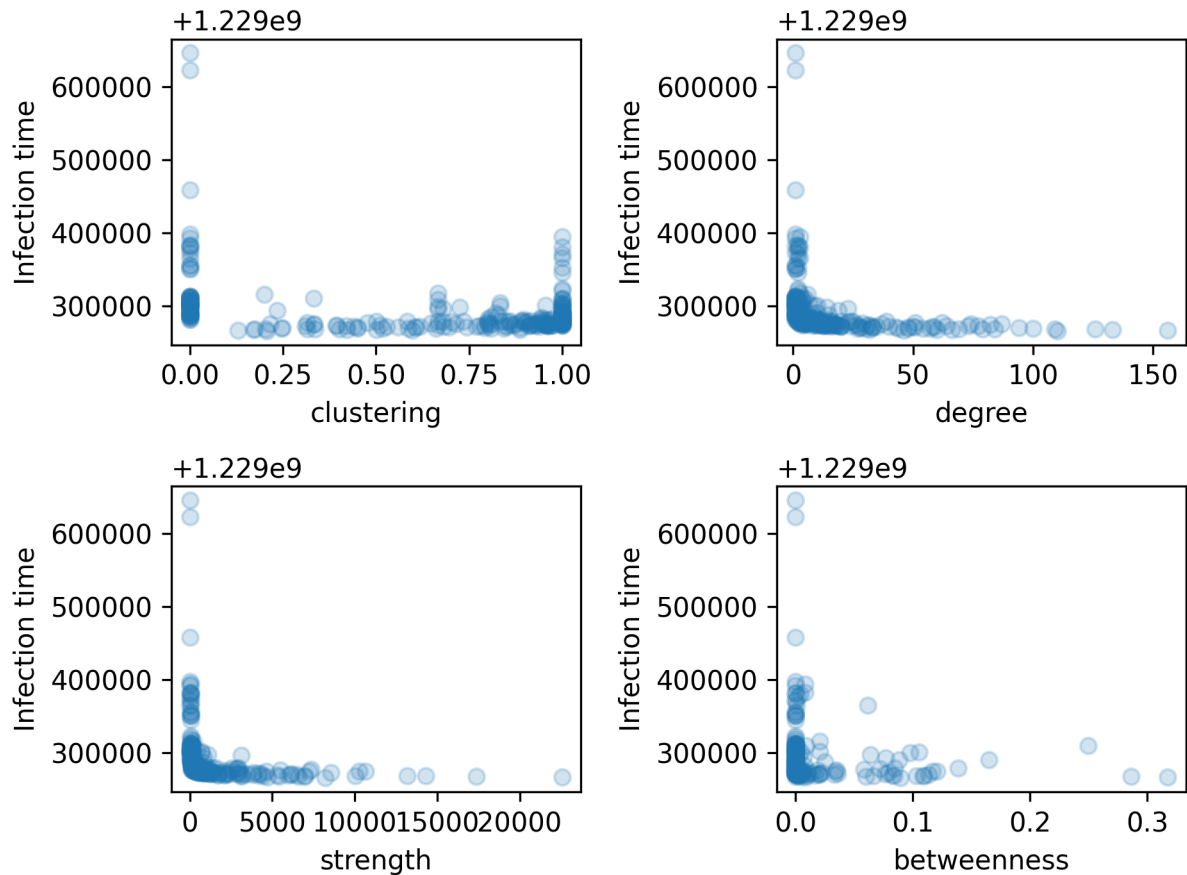
The differences in the spreading speed in the beginning are related to the number of neighbors the initial node has. Node 5 has 110 neighbors, which will be a factor to a rapid spreading of the disease in the start. This is because each infected node can potentially infect its neighbors in the next time step, and a node with more neighbors has more opportunities for infection to propagate rapidly. On the other hand node 227, which only has 5 neighbors, will spread the disease much more slowly. As there is no return from Infected state, the network will eventually reach a stable state. Even with the differences in the start, the disease starting on node with less neighbors will reach the same number of nodes. However we can see the impact it has on how much more time it takes to reach the same stable state.

(c)

Using a single seed to study disease spread can lead to misleading conclusions because the outcome might depend on the characteristics of the starting node, and not on the ones of the network. Using the average results over different seeds helps to reduce bias, improves robustness and helps understand network behavior.

Problem 4: Where to hide?

(a)



(b)

Spearman r between median infection time and clustering: -0.103

Spearman r between median infection time and degree: -0.81

Spearman r between median infection time and strength: -0.891

Spearman r between median infection time and betweenness: -0.397

(c)

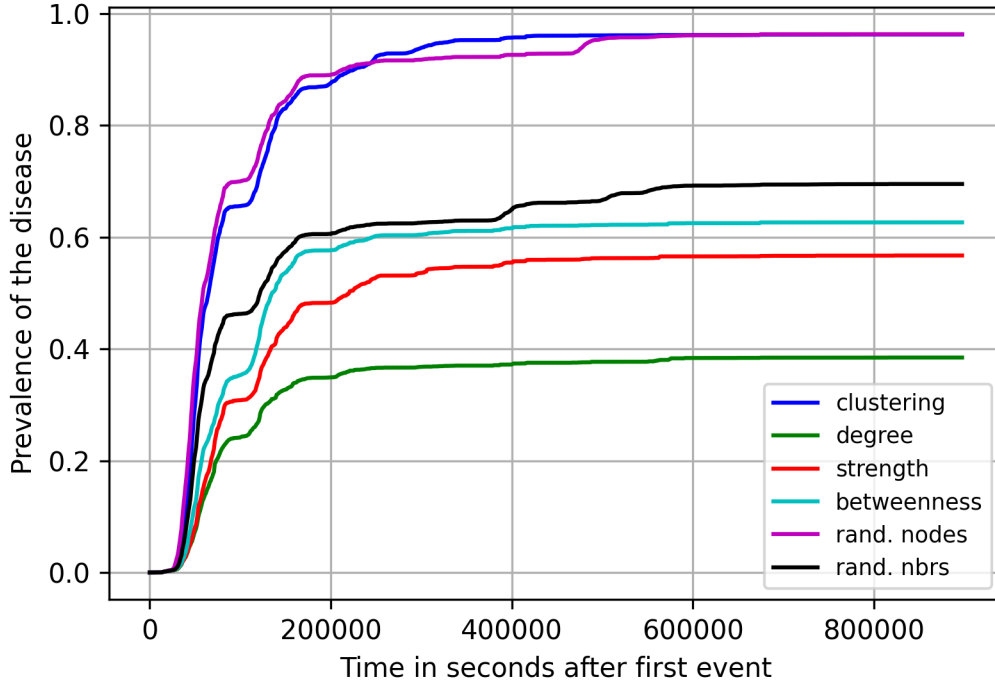
I would base my choice on the degree and strength metrics as this have stronger, even though negative, correlation with the median infection time. That means that the lower this measures, the bigger is the infection time.

Nodes with high betweenness centrality don't mean they are highly connected, means they are important for connecting communities or they have an higher number of connections with other communities than the average. They are strong in spreading the disease, betweenness centrality reflects the undirect importance of a node, while strength and degree are directly connected to the probability of being exposed.

Clustering coefficient, on the other hand, measure how connected it's a node to it's neighbors, this does not have relevance when predicting if a node will take more time to get infected. It's however important in reflecting how fast a disease can spread inside that cluster, but not how fast it got there or to get out.

Problem 5: Shutting down airports

(a)



(b)

The degree immunization strategy performed the best, as expected, between the 4 network measures, we have seen before degree had a strong correlation with the infection times. Degree and strength directly quantify a node exposure, by removing this nodes the network becomes fragmented, as the strength measure was the 2nd best strategy for immunization. Following by betweenness centrality and random neighbors as we will explore this ones more below. Random nodes and clustering performed the worst. Random nodes often leaves critical nodes (high-degree) untouched, allowing the disease to spread, and clustering coefficient only measures how tightly a node it's connected to it's neighbors, so immunizing these nodes does not affect significantly the spreading of the disease. As said before betweenness centrality measures how a node act as a bridge between "communities", removing them is critical in cutting off the transmission of the disease, which is more crucial when looking for an immunization strategy, not on when predicting a safe hiding place as measures like degree or strength are better predictors.

(c)

The network consists of N nodes, where N_k represents the number of nodes with degree k . The probability of choosing a random node with degree k is given by:

$$P(k) = \frac{N_k}{N},$$

When picking a random neighbor of a random node, it is more likely to have a higher degree than the random node itself. This is because nodes with higher degrees are more likely to be chosen as neighbors.

There are $N \cdot P(k)$ nodes with degree k , which means there are $k \cdot N \cdot P(k)$ links leading to nodes of degree k . Thus, the probability of a random neighbor having degree k is:

$$P_{\text{neighbor}}(k) = \frac{k \cdot N \cdot P(k)}{\sum_{k'} k' \cdot N \cdot P(k')} = \frac{k \cdot P(k)}{\langle k \rangle},$$

where $\langle k \rangle = \sum_{k'} k' \cdot P(k')$ is the average degree of the network.

The random neighbors strategy should be more effective, through the friendship paradox as if you pick a random neighbor of a node, that node has on average higher degree than the one picked, and as we can see, level of degree is a determinant factor on slowing down/stopping the disease.

(d)

In the context of social networks, the neighbor immunization strategy, even though worse than other strategies, it targets more influential nodes without needing a full view of the structure. It also more computational efficient as only needs local information, nodes neighbors as contrary to clustering and betweenness centrality.

Problem 6: Disease transmitting links

(a)

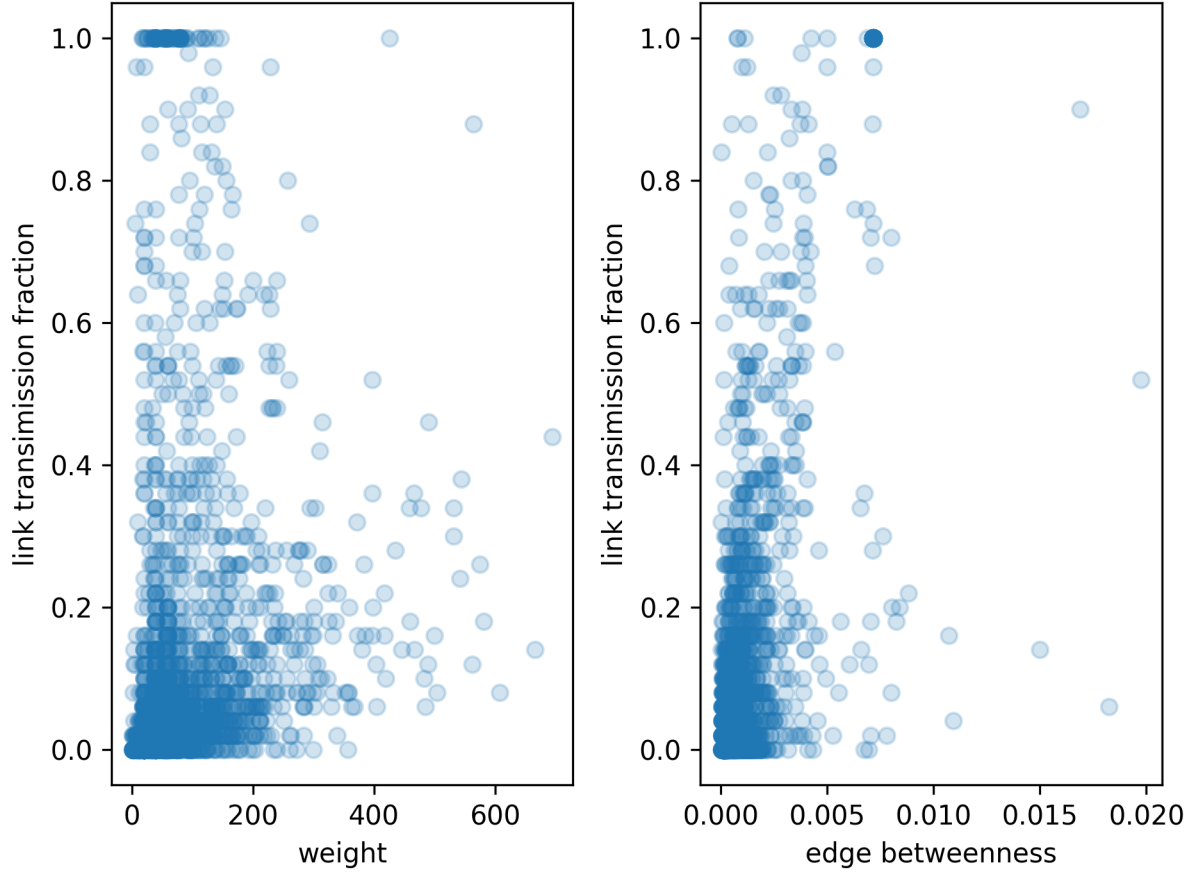


Figure 1: Number of times a link transmitted disease and Maximal Spanning Tree

(b)

Both visualizations are similar due to both looking for the most important pathways for spreading of the disease. They focus on the most frequent used links, in the disease transmission visualization, links with higher number of times they transmitted the disease, highlight critical routes in the network and the MST is constructed by selecting links with the highest weights, prioritizing the strongest connections.

(c)



Spearman r between link transmission fraction and weight: 0.454

Spearman r between link transmission fraction and edge betweenness: 0.535

(d)

The values for the spearmanr correlation indicate somewhat a good indicator in predicting link transmission factor by looking at weights and link betweenness centrality, even though not the strongest. This can be explained by the definitions of both measures. Starting by link weight, measures the number of flights between two nodes, reflecting the frequency of the interaction between two nodes, which is relevant for how many times a link is used to transmit the disease as if they are more connected, the probability for the link to be used to infect is more. However, the limitation of looking only to strength is that it doesn't measure the importance of the link, such as if it acts as a bridge between communities therefore ignoring this. Link betweenness centrality measures how often a link appears on the shortest paths between pairs of nodes, study of bridges in a network reflecting the structure of the same. It has a slightly higher correlation, however once again it does overlook the probability of a node being infected by the number of connections he has.

Problem 7: Discussion

Our model is a simplistic simulation of the real-world. In these section I suggest 4 ways on how to improve our model so it became more realistic.

- **More states** After a node becomes Infected in our node, it reaches his end state. However, as we seen with the COVID-19, there are more states, as Exposed and Recovered which would make the model more realistic. The Exposed state could introduce a delay in spreading, and the Recovered offers immunity during a period of time;
- **Dynamic network** Our SI model accounts for a static network, however every day new pathways are created, as well as erased, or others just may be inactive for some period. Modeling dynamic

networks better represent the real-world behavior. Examples are holiday peaks, weather storms, etc;

- **Fraction of Full Flights** In our SI model we take in account number of flights between nodes, however these can be bigger airplanes (more persons to spread the disease), be in average half of the capacity or always full flights. A feature we could add to turn our model more realistic could be to include the average persons per flight in the weights between nodes. This suggestion can help to simulate where disease how spread depends on travel volume, to help predicting the effect of travel restrictions;
- **Airport different transmission rates** In real epidemics, each infrastructure is prepared differently for these crisis events. More advanced airports can have higher levels of hygiene, as an easier framework for adaption to catastrophic events which should help reduce transmission of the disease.

Ideas were taken and fundament with the help of the resource [1].

References

- [1] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “*Epidemic processes in complex networks,*” arXiv preprint arXiv:1408.2701, 2014.