

Complex Networks

Course Project

Elsa Ollikainen, 709372

t. 1 Basic Implementation

- a.) If Salt Lake City is infected at the beginning of the data set, Anchorage becomes infected at time 1229283600.

t. 2 Effect on infection probability p on spreading speed

- a.)

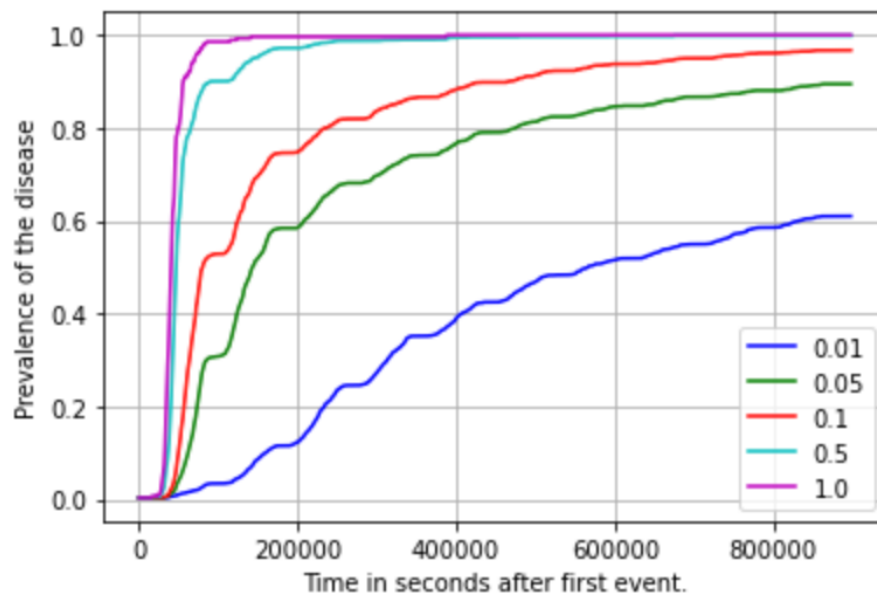


Figure 1. Averaged prevalence of the disease as a function of time for five different infection probabilities. Simulation is run 10 times for each probability.

- b.) We can see from the Figure 1. that when the infection probability is $p = 1.0$ or $p = 0.5$, the whole network becomes fully infected.

From the animated visualizations of the epidemic, we notice that there is remarkably less flights during the nighttime than daytime. That is probably the reason for the periodic “steps” in the curves, as the spreading of the epidemic slows down during nights when there are not that many flights. Also, all the steps are present in the curves at the same time, which indicates that the reason for the steps is the same in every simulation.

t. 3 Effect of seed node selection on spreading speed

a.)

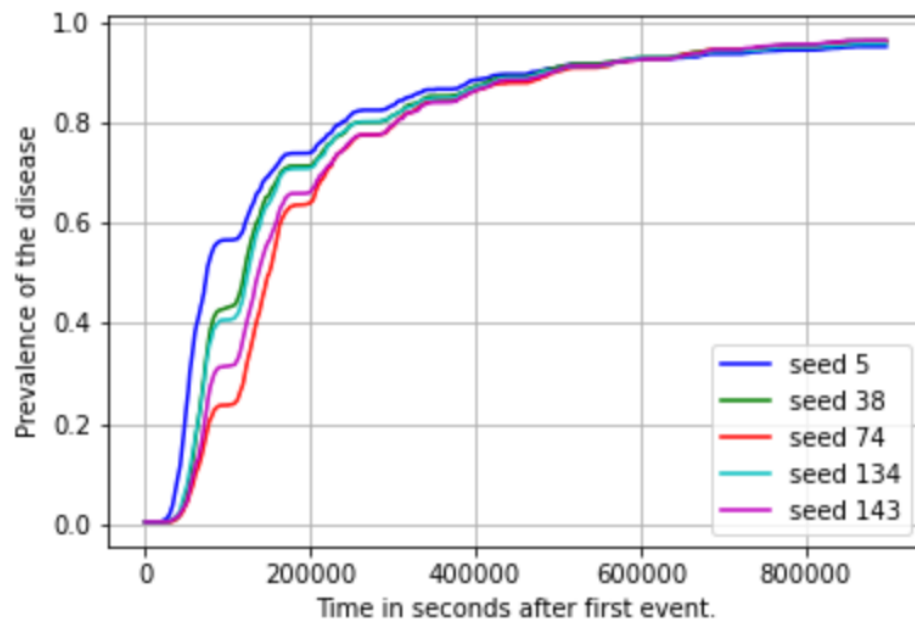


Figure 2. Averaged prevalence of the disease as a function of time for five different seed nodes and $p = 0.1$. Simulation is run 10 times for each probability.

- b.) When the seed node is node number 5, the spreading speed is highest. This is probably due to high value of strength of node 5, which means that there is lot of flights going through node 5 and thus the epidemic spreads fast. When the epidemic propagates, other curves also reach up to the same point, as the epidemic goes through the highly connected nodes.
- c.) As seen in the results of this exercise, the selection of the seed node has a lot of impact on the results of the simulation. The results can vary a lot based on the seed nodes characteristics. For example, from the Figure 2 can be seen that when approximately 100000 seconds are gone, the prevalence with seed node 5 is about 0.57 and with seed node 74 about 0.24. By averaging the results over different seed nodes, the effect of the one particular seed node is diminished, and results are more likely to be correct.

t.4 Where to hide?

a.)

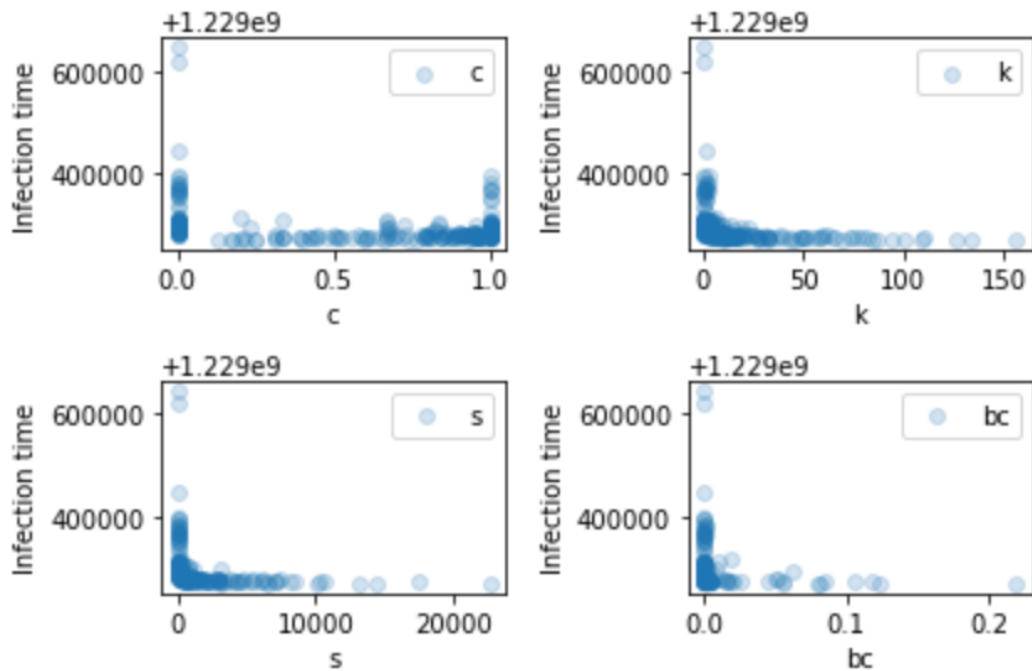


Figure 3. Scatter plots showing the median infection time of a node as a function of four different network measures: unweighted clustering coefficient c , degree k , strength s and unweighted betweenness centrality.

b.)

Network measure	Spearman correlation coefficient
Clustering coefficient	-0.15
Degree	-0.80
Strength	-0.87
Betweenness centrality	-0.61

Table 1. Spearman correlation coefficients between median infection time and four different network measures.

c.) I would primarily pick node strength as a measure to choose the best place to hide. Also, node degree and betweenness centrality are useful when predicting the infection time. Node degree tells that how many routes and other airports are linked to the node, so if the degree of the node is low, there is not that many possible infecting flights that can come to

the node. Node strength is better measure to predict the infection possibility of the node, because it also tells that how many flights are coming to the node. Choosing the best place to hide out of two nodes based on only the degree can go wrong, because the node having the lower degree may still be having more incoming flights than the node with the higher degree.

Betweenness centrality can also be useful in predicting the best node to hide. Betweenness centrality is the measure of the centrality in a network, so the value of it tells where the node is in the network. Nodes with low betweenness centrality are likely to be far away from the center of the network (and thus also likely to be far away from the seed node), so it takes longer time that the epidemic reaches these nodes, as can also be seen from the Figure 3. However, betweenness centrality is not as good measure for the prediction as strength or degree, as the unweighted version of the measure is used in the calculation. Since the weight is not considered, it can't be known how many flights are coming towards a node. Even if the node would be located at the perimeter of the network when there are probably many routes between the node and the seed node, there can still be a lot of flights coming through these routes, which makes the infection time shorter.

Clustering coefficient is not as efficient predictor as the other measures, because even if the node would have a high clustering coefficient meaning that it is strongly clustered to other nodes, it doesn't guarantee a short infection time. From the scatter plot in the Figure 3, it can be seen that the infection times of nodes with high clustering coefficient vary a lot. As we saw in the Figure 5., some of the highly clustered nodes can be island groups with multiple airports connected to each other. These groups are still quite isolated as they don't have many connections to the mainland and thus, the infection time of these nodes can be long.

t. 5 Shutting down airports

a.)

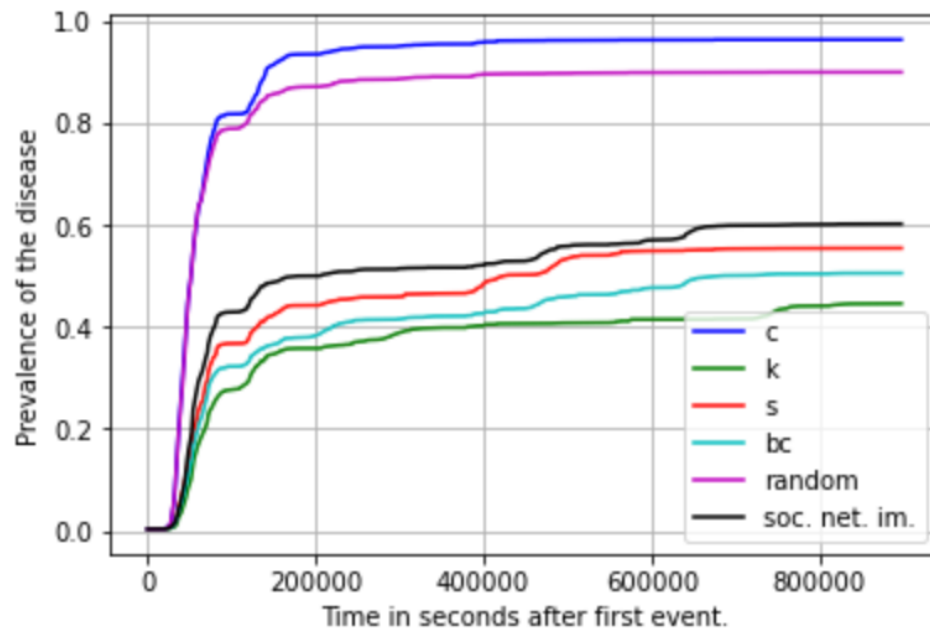


Figure 4. Prevalence of the disease as a function of time for six different immunization strategies with $p = 0.2$. Simulation is run 20 times for each strategy.

b.) Immunization based on the degree of the node seems to perform the best. In this immunization strategy, 10 nodes with the highest degree are removed from the network. These 10 nodes have the most connections to the other nodes and usually this kind of “hubs” in the airport systems have connections to some smaller airports that possibly don’t have any other connections. When the hubs are removed, it can be even impossible for the infection to reach some of the former connection nodes, which makes the prevalence of the disease lower. This is the crucial difference between degree immunization strategy and other strategies. For example, strength immunization is not as efficient technique because it considers the number of flights between nodes. If ten nodes with the highest strength are removed, there can be nodes with low number of connections, so there remains more nodes for infection to connect than with the degree strategy.

Immunization based on betweenness centrality is the second-best strategy. Central hubs that are important to immunize have high betweenness centralities. As stated before, hubs are connected to many smaller airports that doesn’t have many other connections, so many shortest paths to nodes go through the hubs. Hubs are also located at the center of the network, so if there are two nodes completely different sides of the network, the shortest path between these nodes probably also goes through hubs.

- c.) If the degree distribution of the network is $P(k)$, the probability of picking a random node with degree k is the value of $P(k)$. When a random link of a random node is followed, the average degree of the obtained neighbor node is:

$$\langle k_{nn} \rangle = \sum_k k P(k_{nn} = k) = \sum_k k \frac{kP(k)}{\langle k \rangle} = \frac{\langle k^2 \rangle}{\langle k \rangle}$$

where the probability of obtaining a neighbor node of degree k is:

$$P(k_{nn} = k) = \frac{kNP(k)}{N\langle k \rangle} = \frac{kP(k)}{\langle k \rangle}$$

where $NP(k)$ is the number of nodes of degree k , and thus $kNP(k)$ is the number of links leading to nodes of degree k [1]. The average degree of the network is the sum of all the degrees in the network divided by the number of nodes, $\langle k \rangle = 14.9677$. Thus, $\langle k^2 \rangle = 822.8387$ and $\langle k_{nn} \rangle = 54.9743$. The average degree of the network is much smaller than the average degree of the random neighbor. Consequently, when immunizing random neighbors of random nodes, it is more likely that the immunized nodes have high degree, or they are “hubs”, and therefore the immunization strategy is more effective than the random degree immunization.

- d.) Social network immunization strategy is good strategy with social networks, because they are likely to be very large and complex. The network topology needs to be known for the other immunization strategies that are more effective than the social network strategy based on the Figure 4. The social network immunization strategy requires far less information of the network than the other strategies, which is good because the networks are so complex, but it is still effective to immunize the spreading of the epidemic.

t. 6 Disease transmitting links

a.)

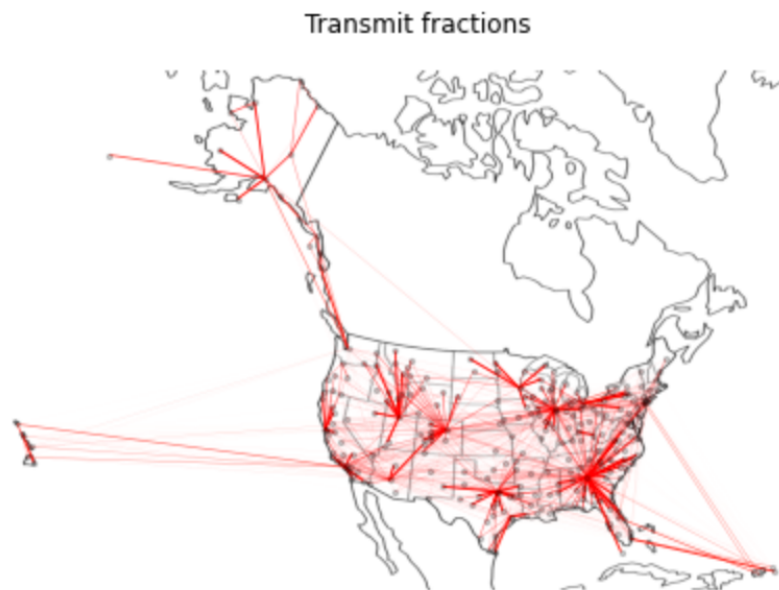


Figure 5. Visualization of the network. Widths of the links are proportional to how likely they are to transmit the disease.



Figure 6. Maximum spanning tree of the network.

In the visualization in figure 5, the fractions of simulations, where each link transmitted the disease, are computed. Simulation was run 20 times using random seed nodes and $p = 0.5$. The widths of the links are adjusted according to their fractions. This was done to see the importance of links in disease transmitting.

The visualized network in the Figure 5 and the maximum spanning tree in the Figure 6 are quite similar, but they have some differences. The maximum spanning tree doesn't have

some of the links with smaller transmission fractions at all, for example there is only one link from the Hawaiian Islands to the mainland. The biggest hubs are more distinct in the visualization in the Figure 5 than in the maximum spanning tree, even though they can be seen in both figures.

b.) The visualization with transmission fractions is quite similar to the maximum spanning tree in the Figure 6. This is because the maximum spanning tree takes into account the links that have the largest weight, i.e., the links that have the most traffic. In the visualization in the Figure 5, the links are the widest links have transmitted the disease the most, so it is likely that there is also the most traffic on those routes. Therefore, the maximum spanning tree and the visualization both show the links with largest weights.

c.)

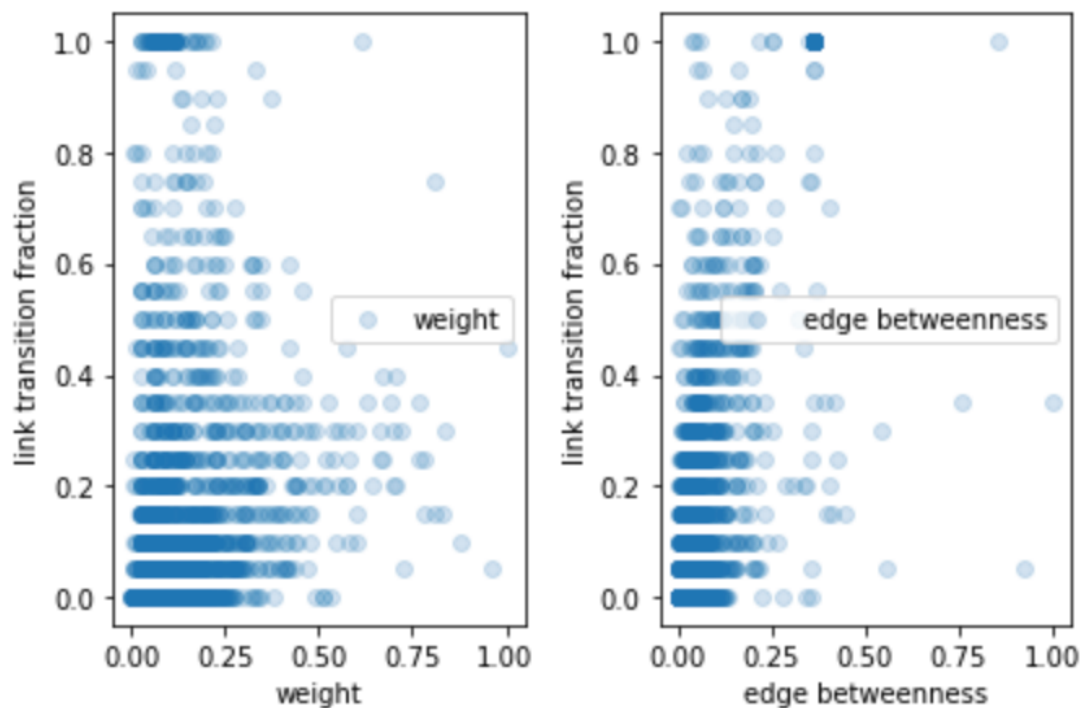


Figure 7. Scatter plots showing transmission fractions as a function of i.) link weight and ii.) unweighted link betweenness centrality.

Link-wise measure	Spearman correlation coefficient
Weight	0.42
Betweenness centrality	0.56

Table 2. Spearman correlation coefficients between transmission fraction and two link-wise measures: weight and betweenness centrality.

d.) Either weight or betweenness centrality are not very good predictors for transmission fraction, as can be seen from the scatter plots in Figure 7 and from the Spearman correlation coefficients in the Table 2. Weights are probably poor predictors because there is lot of links between bigger “hubs” and low degree nodes, that transmit the disease but have a low weight, and therefore are not considered when the transmission fractions are predicted based on weight only. On the other hand, betweenness centrality is not a good predictor, because the unweighted version is used, so it doesn’t consider the weights of the links, only the “centrality”. Centrality correctly predicts the transmission fraction for certain nodes, but not generally.

t.7 Discussion

In this project a simplistic Susceptible-infected (SI) model was generated. Even though the model is far from a realistic one, it can be used to easily get idea and insights of spreading of an epidemic. For example, with the SI model it is possible to estimate strategies on how to prevent the disease from spreading. However, the model needs to be improved if it is wanted to be realistic. There are multiple different ways to make the basic model more advanced. This discussion covers four of this kind of disease spreading models: SIR, SIS, SIRS and SEIR models.

Susceptible-infected-recovered (SIR) model is one of the simplest and classic disease spreading models used. According to its name, in the SIR model an individual can be in three different states: susceptible, infected, or recovered. If susceptible nodes encounter infected ones, they turn infected with some defined probability [2]. Infected nodes turn recovered after they have been infected for some time and recovered from the disease. In this case the individual gets a permanent immunity. Recovering step can also mean that the individual has died to the disease and gets removed from the network.

Susceptible-infected-susceptible (SIS) model is another simple model, that is quite similar to the SIR model. The difference to the SIR model is that when the infected individual recovers from the disease, the individual returns to be susceptible again, i.e., doesn't get immune to the disease at all. In the SIR and SIS models, the last steps where the individual recovers or get susceptible again occur spontaneously after certain defined time [3].

SIS and SIR models are more realistic than SI, but still quite simple. Susceptible-infected-removed-susceptible (SIRS) and susceptible-exposed-infectious-removed (SEIR) models are bit improved versions of SIR, that can be better used to mimic diseases realistic way. SIRS model is otherwise like SIR, but the recovered individuals become susceptible again with a certain probability, so the model assumed that the immunity to the disease doesn't last forever [3]. In the SEIR model, an exposed period is considered, where the individual is exposed to the disease but not infected yet [3]. The SEIR model is usually used to model influenza-like illnesses, and it can for example be used for modelling the current COVID-19 epidemic [4].

References

- [1] Lecture 3
- [2] Lecture 4
- [3] J.M. Carcione, J.E. Santos, C. Bagaini and J. Ba, "A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model", *Frontiers in Public Health*, 2020.
- [4] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *arXiv preprint arXiv:1408.2701*, 2014.