Complex Networks
Mark Ait
779755

Task 1.

a) If Salt Lake City is infected in the beginning of the simulation then based on the results Anchorage gets infected at time <u>1229283600</u>.
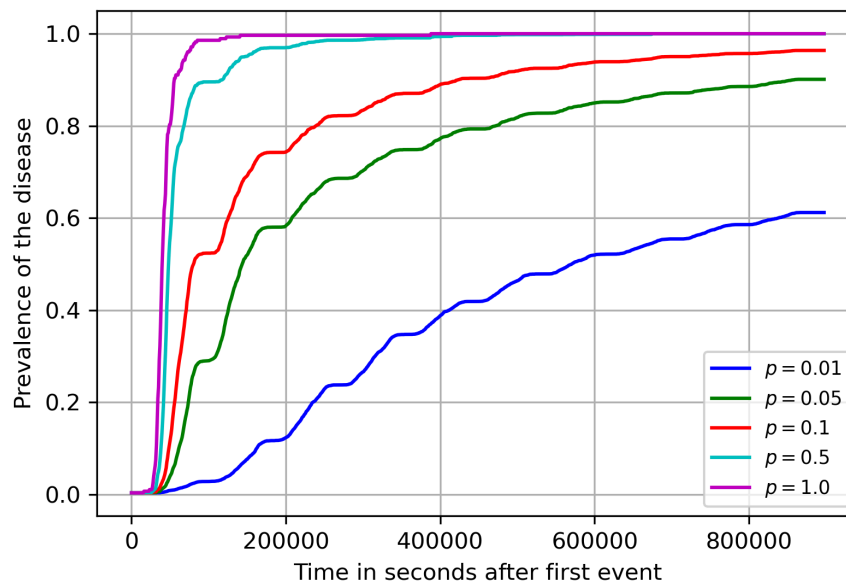
Task 2.

a)

Figure 1. Averaged prevalence ρ(t) of the disease as a function of time for 5 infection probabilities

b) The network gets fully infected with probabilities 1.0 (purple Fig. 1) and 0.5 (cyan in Fig. 1). The periodic steps in the curves can be explained due to hubs. Hubs contain nodes that have a high degree and therefore they have many links going to other nodes. From these nodes the infection can spread rapidly creating these periodic steps.
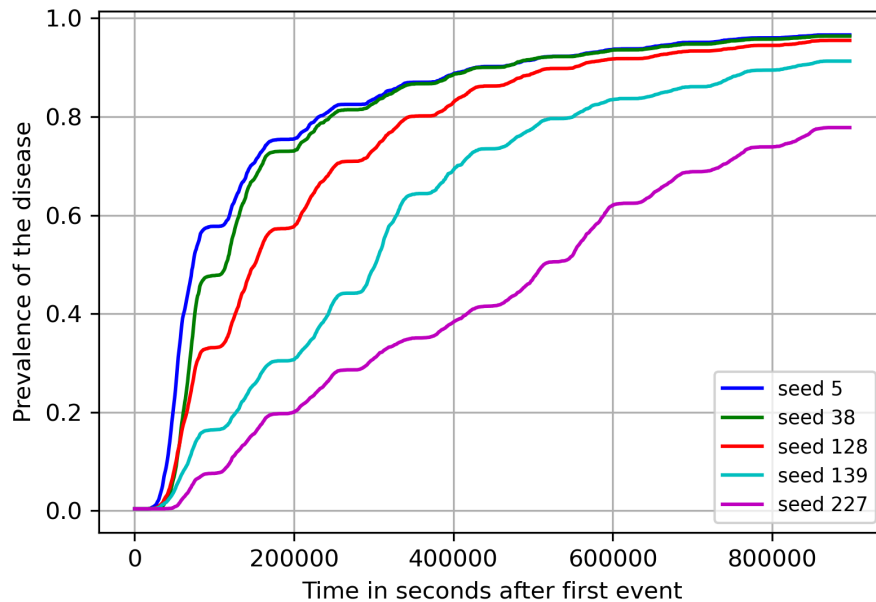
Task 3.

a)

Figure 2. Averaged prevalence $\rho(t)$ of the disease as a function of time for 5 different initial infection nodes

b) The spreading speed in the beginning is different for the given nodes and it is clearly seen from Figure 2. This is due to the connectivity differences in the 5 chosen nodes. A high degree node passes the disease and also gets infected faster. From the graph we can see that nodes 5, 38 and 128 likely spread quicker due to this fact, while nodes 139 and 227 spread slower. Nodes 5, 38, 128 are likely located in hubs while the other nodes aren't as connected.

c) Averaging the results is important due to variability in spreading. When a node is connected to multiple nodes and one of the connected nodes has a high degree then the infection would also spread quicker. When you are averaging the results, different scenarios are taken into consideration and anomalies don't show clearly in the graph. If only one seed is taken into consideration, then it is difficult to see how the infection would spread in different scenarios and this could result in incorrect predictions when looking at the entire network.
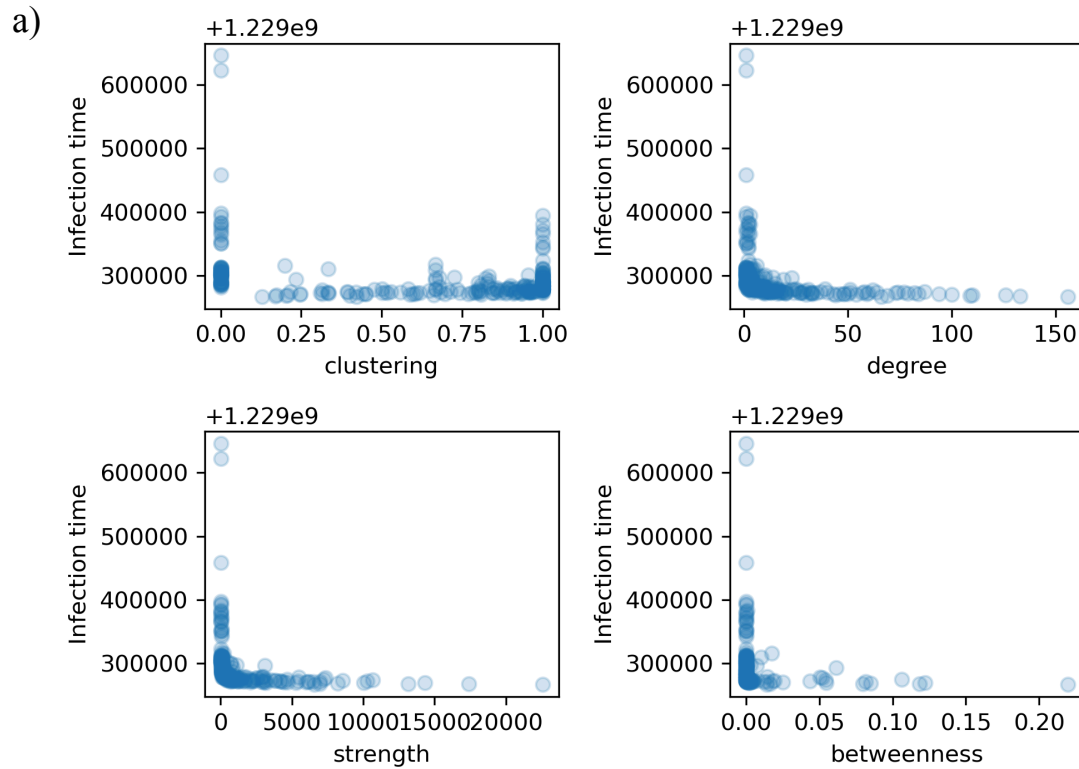
Task 4.

a)

Figure 3. Median infection time of each node as a function of 4 different measures

b)

Spearman r between median infection time and clustering	-0.126
Spearman r between median infection time and degree	-0.813
Spearman r between median infection time and strength	-0.885
Spearman r between median infection time and betweenness	-0.633

Table 1. Spearman rank-correlation coefficient between each of the network measures:

Based on the results, strength seems to be the best predictor for infection times. Node strength measures its connectivity and it is calculated as the sum of its link weights. This implies that nodes with a high strength value either have a lot of links, have links with high weight values or have a

combination of these two mentioned characteristics which makes them vulnerable to infection.

c) Based on the results, degree and strength had the strongest correlation with infection times so I would most consider these two measures when picking a place to hide. Nodes with a high strength value are either important or have a lot of links (or both), and nodes with a high degree have a lot of links which makes them more susceptible to getting infected. If I had to choose a place to hide then I would choose a location that has a small strength and degree value.

Betweenness centrality shows how often a node is traversed through when looking at shortest paths. In our context, hubs and high-strength nodes are more critical because they account for the most traffic and connections. A node with high betweenness centrality might serve as a bridge but has a low degree and due to that it has fewer possibilities to become infected or to spread an infection. On the flipside a node with small betweenness centrality can have a high degree which makes it more susceptible to be infected.

Clustering coefficient evaluates how neighbours of a node are connected with each other. This is a local measurement method that evaluates the nodes neighborhood but does not take the entire network into consideration. For example, a node that has a high clustering coefficient can be very safe from infections and on the flipside a node with a low clustering coefficient can be extremely probable to infection. In our situation the correlation was quite weak (-0.126) which also supports these claims.
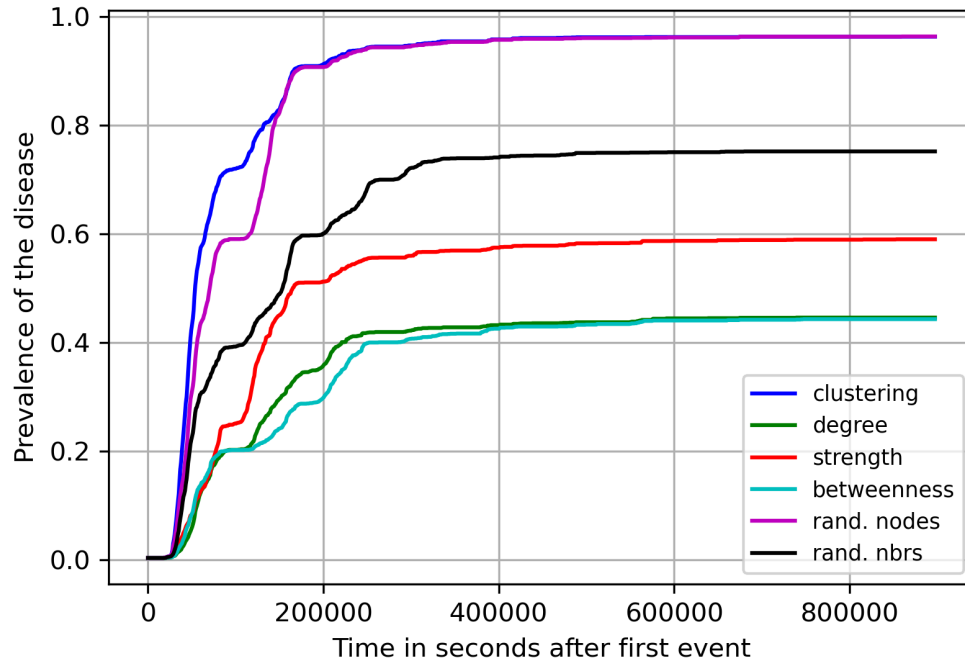
Task 5.

a)

Figure 4. Averaged prevalence ρ(t) of the disease as a function of time for 6 immunization strategies

b) When immunizing nodes based on the six immunization strategies the strategies that worked the best were degree and betweenness. Immunizing nodes that have the highest degree is an effective method due to them not being able to get infected, and them not being able to infect others. These high degree nodes are located in hubs that get a lot of traffic and this means that the disease can not spread as rapidly.

Betweenness was also an effective immunization strategy. It performed better as an immunization strategy because the highest betweenness nodes were selected for the simulation. The nodes with a high betweenness coefficient are often traversed through and when immunizing one of the nodes in the main paths, the disease can not spread as rapidly. In the hiding scenario, all coefficients were used. When using all values for the simulation, there can be nodes that have a high betweenness value but are not traversed through often, and there can also be nodes that have a small betweenness value and are traversed through

often. Therefore it is hard to choose which of the nodes would be safe based on only this measure.

c) If the degree distribution of the network is P(k) then picking a random node of degree k is:

$$P(k) = \frac{N_k}{N}$$ where N_k is the number of nodes with degree k and N is the total amount of nodes.

If we pick a random neighbor of a random node then $P(k)$ can be written as $P(k_{nn} = k)$. $P(k_{nn} = k)$ can be equated to $\frac{k*P(k)}{<k>}$ which implies that the degree of the neighboring node tends to be higher on average compared to the randomly selected node.

$$P(k_{nn} = k) = \frac{k*N*P(k)}{N*<k>} = \frac{k*P(k)}{<k>}.$$

This can be further denoted into the phenomenon called Friendship paradox, where on average the degree of the random nodes neighbor is higher than the random nodes degree. Friendship paradox can be written as: $< k_{nn} > = \frac{<k^2>}{<k>}.$

Due to neighboring nodes having a higher degree on average, immunizing these nodes is more effective than picking random nodes. This is also supported by the conclusion previously that node degree strongly correlates with infection times.

d) In the context of social networks immunizing neighbors makes sense to apply due to it being simple to put into use. When looking at different immunization strategies they all need in depth statistics to track, meanwhile neighbouring immunization can just be implemented easily by immunizing random neighbours.

Task 6.

a)

Figure 5. MST based on infection spreading
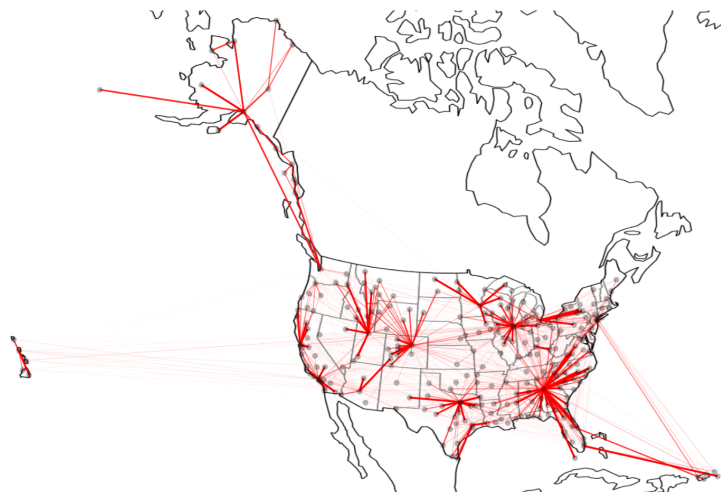
Transmit fractions

Figure 6. Infection spreading visualization based on transmitting fractions

b) The infection spreading visualization is similar to the MST in hub areas
   where the infection spreads rapidly. Overall, in the visualization many of
   the links are of lower opacity whereas the MST has plotted all links with
   maximum opacity and weight. The lower opacity of the visualization
   links come from the link weights that were calculated according to the
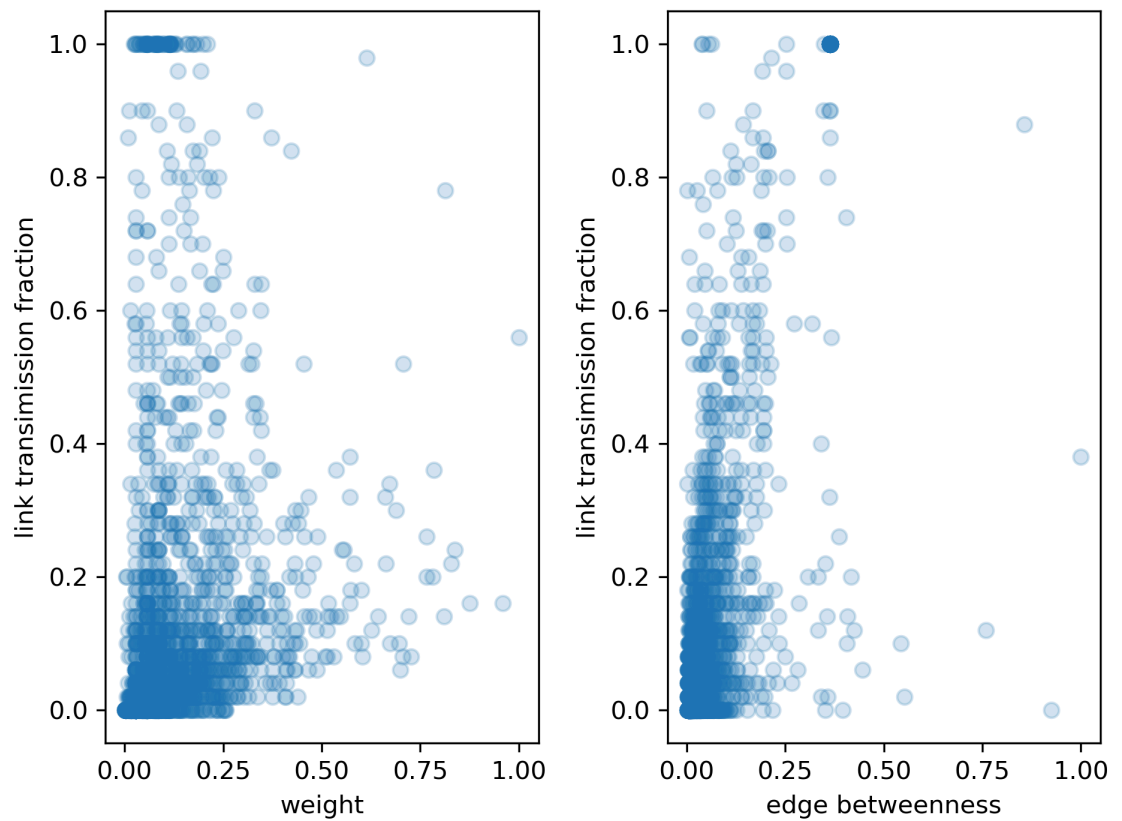   transmission factors.

c)

Figure 7. Scatter plots showing fij as a function of link weight and link betweenness

The Spearman correlation coefficients between f_{ij} and the two link-wise
measures:

Spearman r between link transmission fraction and weight: 0.479
Spearman r between link transmission fraction and edge betweenness: 0.548

d) The two link properties predict transmission factors adequately. When looking at the coefficients, edge betweenness correlated slightly better than link weight (0.548 > 0.479), and therefore it is a better predictor in this situation. Edge betweenness correlates to shortest path distances which is utilized in this network since many routes tend to use centralized or shortest path routes.

Link weight tells how "important" a path is and while these paths are used, this does not neglect that less important paths also play a huge role in transmitting infections especially in the network in question.

Task 7.

Although the current SI-model has produced some interesting results it can be further enhanced with some improvements.

The current model is quite basic in a sense that the nodes only have two states: susceptible and infected. In order to simulate a more probable real life epidemic situation, recovery should also be possible. The model that supports that is called the SIR model. The SIR model is essentially a standard SI model with an additional state that allows for recovery. In the recovery state the node becomes immune and can not be infected in the future.

In addition to the SIR model, various infection prevention methods could be added to the simulation to make it more realistic. These prevention methods could be quarantining or vaccinating susceptible nodes for a certain amount of time and inspecting how it affects the infection spread.

When looking at the epidemic that we ourselves had a couple years ago, information played a huge role in the infection spread. There are rumor spreading models that are variants of the SIR model in which the recovery process does not occur spontaneously, but rather is a consequence of interactions. [1] Our current model could be improved by using certain features of rumor spreading eg. acquiring information about infections from infected nodes and not interacting with these nodes.

When making the model more realistic we must also consider interaction beyond air travel. The model could also support other cross-country travel methods, such as train- and bus travel. This would make the disease spreading more detailed and more information could be extracted from the simulation.

References

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