CS-E5740 Complex Networks, Final Project

Hossein Firooz, Student number: 903929

August 12, 2022

1 Basic implementation

Anchorage airport becomes infected at **1229290800**.

2 Effect of infection probability p on spreading speed

a)

Figure 1 shows the effect of different infection probabilities on the spreading speed.
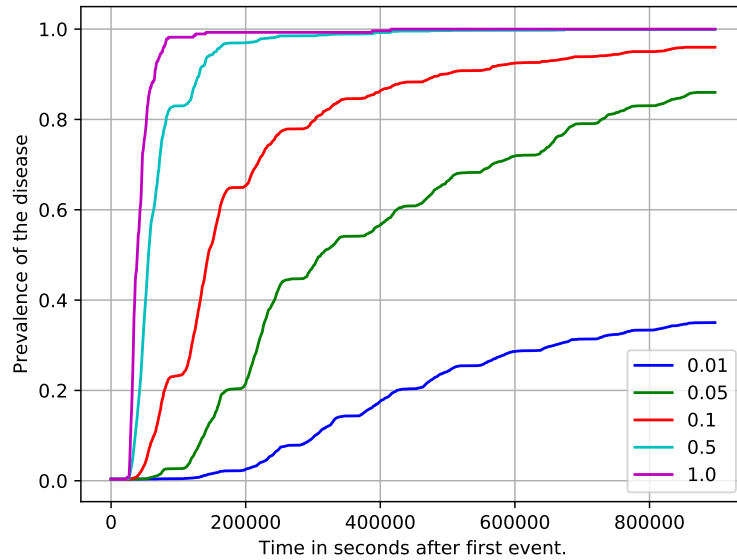

Figure 1: Spread time with different infection probabilities

b)

For both $p = 1.0$ and $p = 0.5$ the whole network gets infected.
Looking at Figure 1 shows that the step behavior of the curves are happening almost at the same time. This give us the intuition that in those periods, there are many flight landing in different cities and this causes the disease to spread faster than usual. Therefore the step behavior of these curves are related to the number of flights.

3 Effect of seed node selection on spreading speed

a)

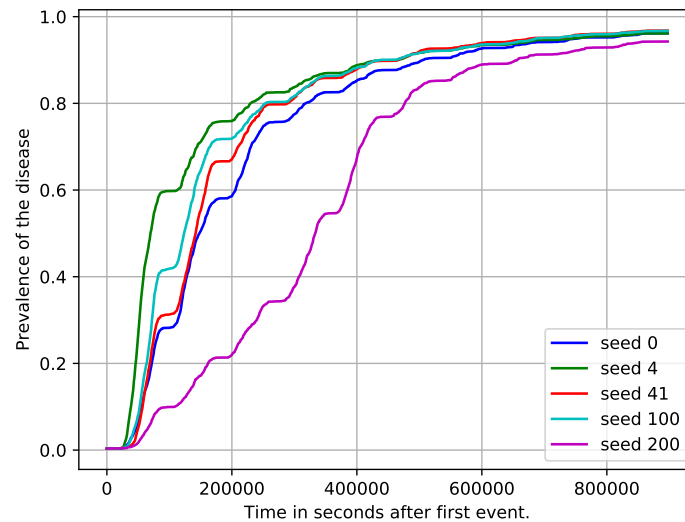Figure 2 shows the effect of different seed node on the spreading speed.

Figure 2: Spread time with different seed infected nodes

b)

This is because the only infected node is the seed node at the beginning. So the spread is mostly based on the seed node and its connections. As time passes, more and more nodes gets infected and infect other nodes. Therefore after while, the effect of the selecting different seed nodes vanishes, and around 97% of the network gets infected for all the seed nodes.

c)

As described in previous section, seed node has a great impact at the beginning on how the infection spreads. Therefore averaging over different seed nodes would reduce the sensitivity of the result to the initial seed node choice and make the results more robust.
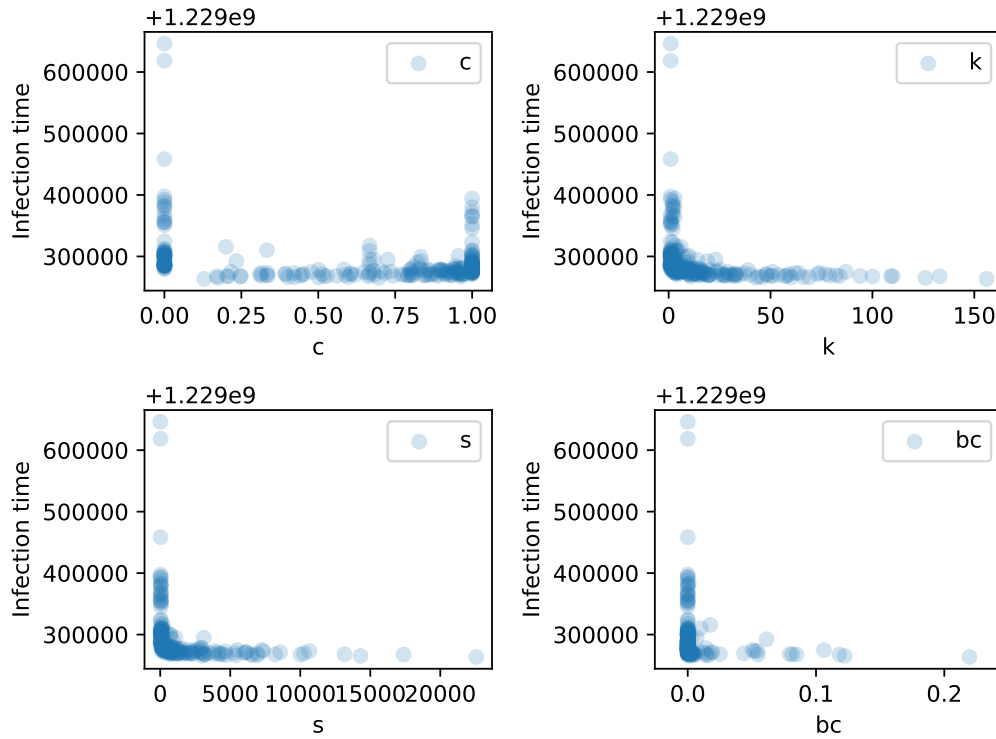
Where to hide?

a)

Figure 3: Median Infection time w.r.t. different network measure

b)

Table 1 shows the Spearman rank-correlation between infection time and network measure.

	Clustering Coefficient	Degree	Strength	Betweenness Centrality
Correlation	-0.132	-0.813	-0.887	-0.629

Table 1: Correlation of Infection Time and Network Measure

c)

By looking at Table 1

1) As the strength of a node has the most anti-correlation with the infection time, it means it could better predict the infection of the nodes better than other measure. Anti-correlation means with a higher strength of a node correlates with a lower infection time which means a faster infection.
High strength for a node means there are high number of flights between this node and other nodes. Therefore it could get infected faster from one of its many connections, as there are many flights in different times. That's why a high strength has anti-correlation with infection time.

2) Betweenness Centrality is a metric for identifying hubs: nodes who behave a a bridge between different communities. Although they have the lowest relative shortest path to other nodes, but the infection only transmits from the direct links to other infected nodes. Therefore as a hub could have low degree to other nodes, it could get infected later than other high degree nodes.

3) Clustering coefficient shows how connected are the neighbours of a node. As the infection only transmits through direct links, it is not important how well connected are the neighbours of a node rather than how well connected the node is other nodes.

Shutting down airports

a)

Figure 4 shows the prevalence of the disease as a function of time for the 6 different immunization strategies.

b)

1) Betweenness Centrality strategy is the best strategy to immune nodes because nodes with high BC are hubs in the network connecting different communities. If these hubs are immune, it would prevent the disease from spreading more from one community to another.

2) BC performs better as an immune strategy rather than a strategy to find a safe place because high BC indicates the hubs. Shutting down a hub would prevent the disease from spreading from one community to another community. However, a node with high BC could still be affected itself as the infection is transmitted through direct links and it doesn't matter if a node has a high or low BC, because BC doesn't indicates

the direct links but the relative shortest path to other links. Therefore choosing a high BC place to hide is not a good strategy as this node could be effected from different communities.
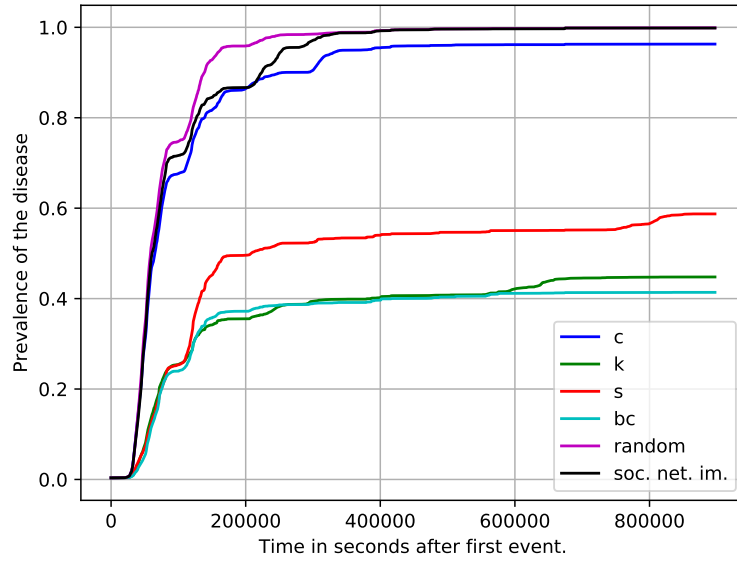

Figure 4: Different Immunization Strategies

c)

1) The probability of picking a node of degree k is the probability of picking one of its connected edges times the probability of the right end (each link has two node): $\frac{1}{2}\frac{k}{e}$ where e indicates the total number of edges. Now as we have n nodes with degree k, the total probability of choosing a node with degree k is: $n\frac{1}{2}\frac{k}{e}$. This probability is also equal to n/N because there are n nodes with degree k among N nodes.

2) The expected outcome of following a random link of a random node is $\frac{\langle k^2 \rangle}{\langle k \rangle}$

3) As the expected degree of choosing a random node in the network is $\langle k \rangle$, the expected degree of following a random link is bigger. Therefore it is a better immunized strategy because it selects nodes with higher expected degrees and it would do a better job in preventing the disease from getting spread. This is also called the friendship paradox which means in a network, a neighbour of a node on average has more neighbours.

d)

In context of social networks, there are obstacles to use other strategies. To calculate other network measure, we need access to full topology of the network. This access is not given all the times with the common APIs. In addition, calculating other measures are usually computationally heavy. Pick-a-neighbour strategy however, doesn't need access to the topology of the network and it is simple to implement and run.

In addition, due to the friendship paradox, which indicated that on average a node's random neighbour has a higher degree than a random node itself. Therefore it is working better than the random node choice as well.

Disease transmitting links

a)

In the 5, both MST and Disease Transmitting links are visualized. They share many links.

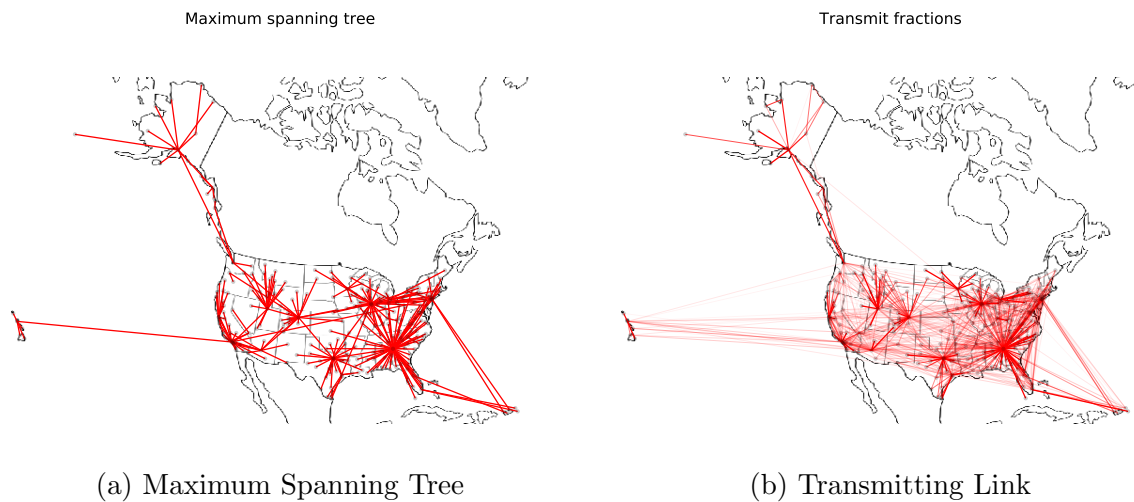(a) Maximum Spanning Tree   (b) Transmitting Link

Figure 5: Comparing MST and Disease Transmitting Links

b)

The similarity between these two graphs are due to the algorithms of how MST is calculated. MST creates a tree with the highest weights in the graph and high weight indicates the frequency of travels between two cities. In the second graph, we used the fraction of times a particular link others as well. A high weights link as shown in 3 has a high correlation with infecting others. Therefore two graph are very similar to each other.

c)

Look at Figure 6 for the plots. Table 2 shows the Spearman correlation coefficients.

	Weight	Edge Betweenness
Correlation	.400	0.513

Table 2: Correlation of f_{ij} and network measures
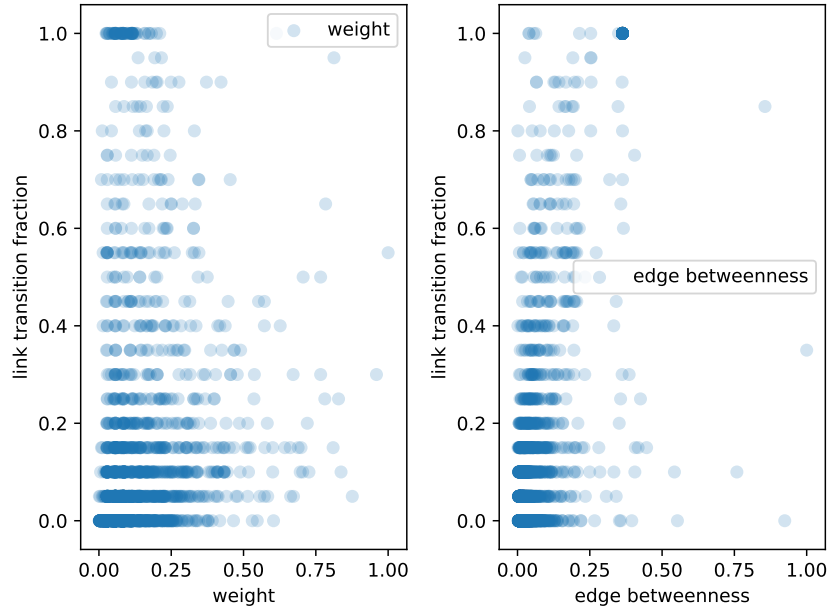

Figure 6: f_{ij} w.r.t Networks measures

d)

Looking at 2, edge betweenness has a higher correlation with f_{ij} and therefore is a better measure. This is because edges with high EB are the bridges between two communities and connecting them together. Therefore these links are more used to spread the disease from one community to another. In contrast, edge weight only shows the number of flights between two nodes and doesn't indicated the position of these nodes to other nodes. Therefore EB is a better predictor for f_{ij}.

Discussion

We can do these improvement to make the model more realistic:

a) After a node gets infected, it could be immune to the disease after a while or even gets back to the susceptible group. For example using the SIR model instead of SI is more realistic.

b) We considered a constant probability for the infection of the disease. However this probability could be dynamic w.r.t other factors like the population of that city, number of flight and even other geo based factors like the behavior of the humans. (For example in Florida due to beaches, the infection rate could be higher than a cold city like Rochester).

c) We treated all the flights similarly. However by considering the flight duration, number of people on the airplane etc we could make the model more realistic.

d) We used a constant flight table but in reality we could change the number of flights from a city when it got infected or it has a high probability of getting infected.