CS-E5740 Complex Networks, Answers to Course Project

Sheetal Borar, Student number: 915263

December 21, 2020

Contents

1 Basic implementation

1.1 a.)

Find the simulation implementation in the code base. Anchorage (ANC, node-id=41) becomes infected at 1229290800.

2 Effect of infection probability p on spreading speed

2.1 a.)

Figure 1 shows the change in Average prevalence over time with different transmission probabilities.

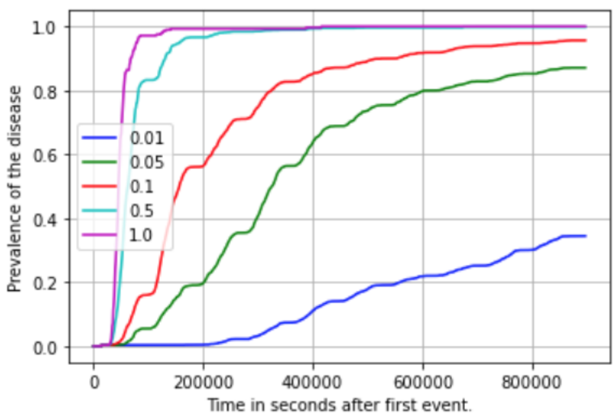

Figure 1: Average prevalence over time with different transmission probabilities

2.2 b.)

With 1 and 0.5 transmission probability, the whole network gets infected. The periodic "steps" are due to the fact that there are periods of time when less number of flights are landing and hence the infection spreading speed is low during those periods. I observed this in the simulation video. Figure 2 also shows this behaviour.

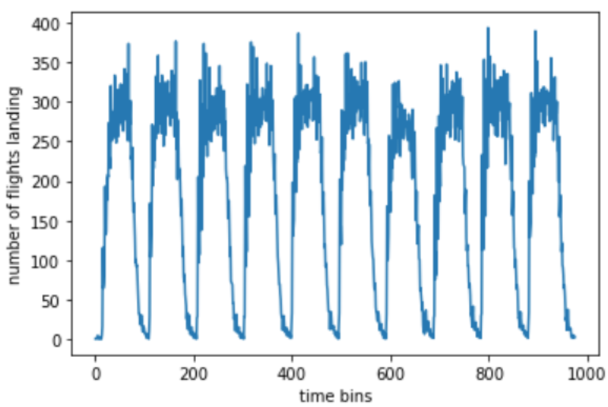

Figure 2: Number of Flights landing over time

3 Effect of seed node selection on spreading speed

3.1 a.)

Figure 3 shows the Average prevalence over time with different seed nodes

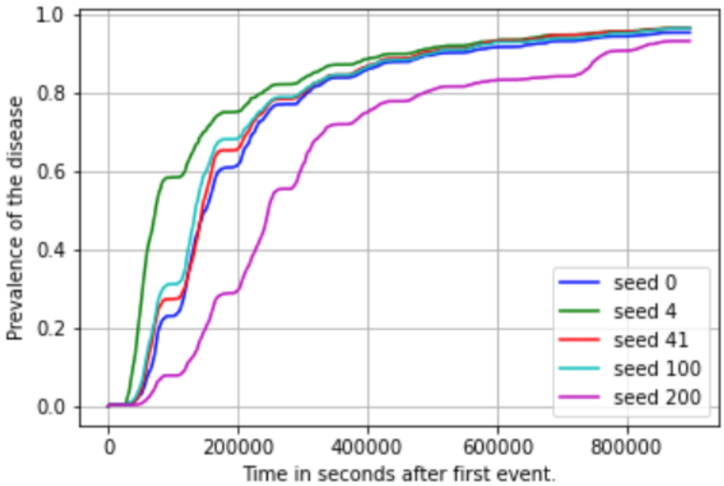

Figure 3: Average prevalence over time with different seed nodes

3.2 b.)

Lets try to understand this behaviour through an example. If one seed has high degree vs another has low degree, in the beginning we can clearly notice which one spreads the infection faster. As the infection spreads to more nodes, the local effect of the seed node becomes less and less important as the low degree node could be connected to a higher degree node and hence might start spreading the infection faster and vice versa. Hence, the differences in spreading speed between seeds should be mostly visible in the beginning of the epidemic.

3.3 c.)

As we noticed in this experiment, different seed nodes can have a significant impact on the speed of infection spreading. We average the spreading result over different seed nodes to account for the seed node bias to ensure that the patterns we observe are robust.

4 Where to hide?

4.1 a.)

Figure 4 shows the scatter plots of median infection times wrt. to different node centrality measures.

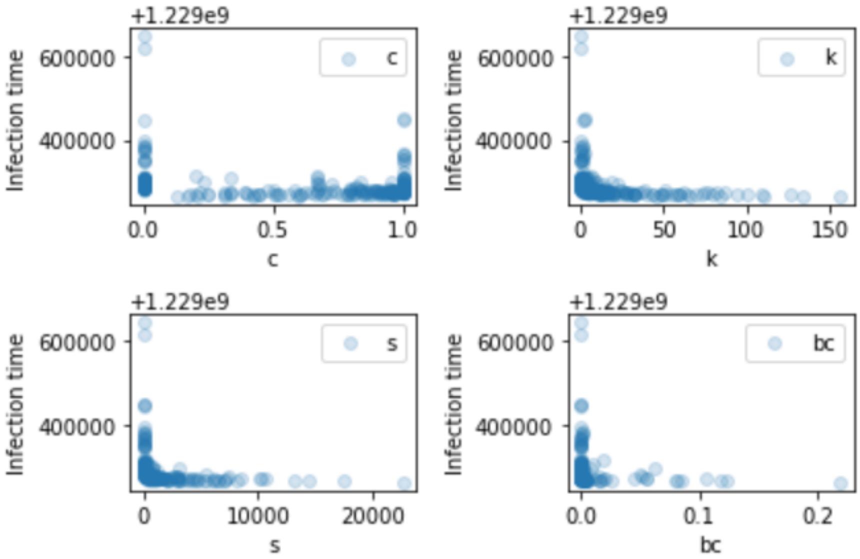

Figure 4: Median infection time as a function of different centrality measures

4.2 b.)

Table 1 shows the Spearman rank-correlation coefficient of median infection time with different centrality measures.

Clustering Coefficient	Degree	Strength	Betweenness
-0.11	-0.78	-0.85	-0.68

Table 1

4.3 c.)

i.) Strength of a node has the highest negative correlation with infection time and hence is the best predictor for infection time. Higher the strength of the node lower the infection time. Weight of the edges corresponds to the number of flights between the nodes. If a node has high weighted degree that means the node has a large number of flights connections to many other nodes. Hence, there is a higher chance for it to get infected from one of its connections faster. To hide, we should find a place with low strength value.

ii.) Nodes with high betweenness serve as bridge between clusters but it is not necessary that they have a high degree or a high weighted degree. If these nodes have a lower degree, then their probability of getting infected is lower as well. Hence the infection behaviour is different for these nodes in comparison to hub nodes.

iii.) Clustering coefficient shows how well connected a node's neighbours are to one another and what is the clustering tendency in the graph. If a node has high clustering coefficient it only means that it is a part of a strong cluster, as this information is not directly linked with the number of flights landing at the node it is a poor predictor of infection time.

5 Shutting down airports

5.1 a.)

Figure 5 shows the change in Average prevalence over time wrt. to different immunization strategies

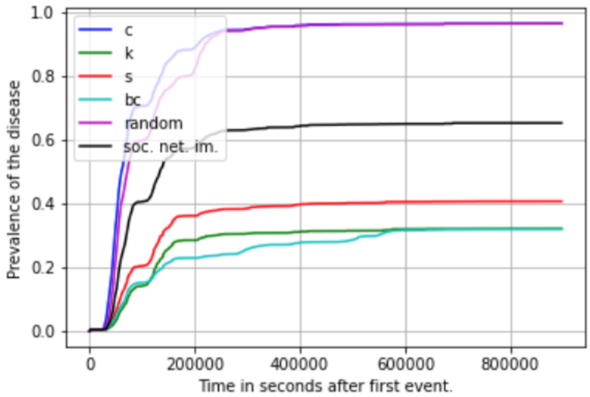

Figure 5: Average prevalence over time wrt. to different immunization strategies

5.2 b.)

i.) Betweenness centrality based strategies performs the best. Nodes with high betweenness are generally the brigde nodes between clusters. By immunizing these nodes we make sure that the infection does not transmit from one cluster to another. This strategy is successful as it can localise the infection and prevent it from spreading to the entire network.

ii.) Betweenness centrality performs better as an immunization strategy than as a predictor for a safe hiding place as the chance of a node with high betweenness getting infected is sufficiently high as its connected to several clusters. So if someone picks places with high betweenness to hide, they are still putting themselves at high risk. But immunizing a node with high betweenness prevents the infection from going from one cluster to another which makes it a great strategy to contain the spread of the infection.

5.3 c.)

In pick-a-neighbour strategy, we are selecting both the nodes(random node and its random neighbour) at random so its the same as selecting a random link from the network. Number of links in the network is l. Probability of selecting a random link is -

$$\frac{1}{l} = \frac{2}{\sum_j k_j}$$

Once we selected a link, there is $\frac{1}{2}$ probability of selecting either of the nodes. A node with degree k will be present in k links and hence the probability of reaching a node i with degree k from a random link is -

$$\frac{2}{\sum_j k_j} * \frac{1}{2} * k = \frac{k}{\sum_j k_j}$$

Probability of reaching a particular node is proportional to its degree, and hence there is higher probability of reaching a node with higher degree. Pick-a-neighbour is a better immunization strategy than picking random nodes as this strategy selects higher degree nodes than random selection on average. If we immunize higher degree nodes we can prevent the spread of the infection better than immunizing random nodes.

5.4 d.)

This strategy exploits an feature of social networks which is called the friendship paradox, which proves that on average a node's random neighbour has a higher degree than a random node itself. This is due to the fact that the hub nodes are connected to a large number of nodes and when selecting a neighbour of a random node, there is higher chance of reaching the hub nodes. This strategy is more efficient in the context of social networks because to implement the other immunization strategies one needs to completely understand the topology of the network and that is a difficult and noisy task for social network. But this strategy can be implemented without knowing about all the social connections.

6 Disease transmitting links

6.1 a.)

Figure 5 shows the network visualized on the map of USA. In the first graph, the edge weights are f_{ij}. The second graph is the maximum spanning tree of the graph.

6.2 b.)

In the first figure, we used the fraction of times a particular node was used to infect as the weight. We can see that edges connected to hub nodes are highlighted in this graph as the hubs are often responsible for spreading the infection due their high degree. Maximum Spanning tree is a spanning tree of a weighted graph having the highest edge weights. The original edge weights represents the number of flight connections between two nodes. As hub nodes have the highest number of flight connections (edge weights), the hubs connections are generally represented in the maximum spanning tree. Hence, the maximal spanning tree is similar to the first figure in figure 5.

6.3 c.)

Figure 6 shows how f_{ij} is correlated to measures like edge betweenness and original edge weights of the network. Table 2 shows the Spearman correlation coefficients between f_{ij} and the two link-wise measures

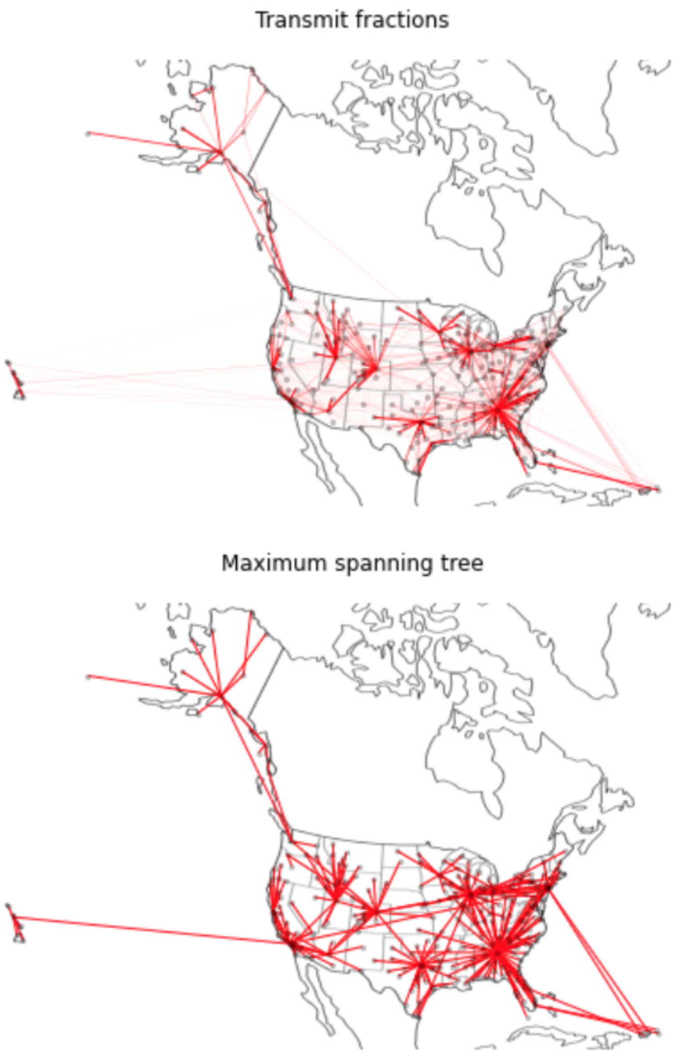

Figure 6: Comparison of fractions f_{ij} as weights to maximum spanning tree

6.4 d.)

Edge betweenness is better at predicting f_{ij} than the edge weight (number of flights between two nodes). Edges with high betweenness will probably be bridge edges between clusters. While edges with high weights are going to be the ones with a lot of flight connections. A lot of flight connections between 2 nodes does not mean that these two nodes are central and will help in spreading the infection faster. But edges with high betweenness are crucial for spreading the infection between clusters and hence predicts $f_{i,j}$ (edges that are spread the virus more often) better.

7 Discussion

a.) We did not account for the population that has already recovered (for example a country got infected and then everyone recovered) from the infection or if infected countries can be infected again or are they immune to the disease.

b.) The probability of transmitting the infection is not static and depends on many

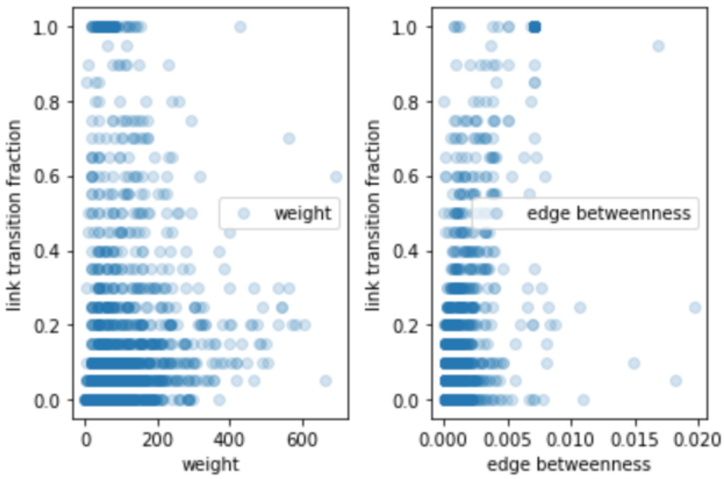

Figure 7: Correlation between fractions f_{ij} to other measures

Weight	Edge Betweenness
0.39	0.50

Table 2

factors like the number of infected individuals who went in the flight from source to destination, the amount of time they spent in the destination etc. So depending on certain factors this probability might need to be adjusted for a more accurate simulation.

c.) The population density of a country and behavioural features of the population could be taken into account to determine whether an infected person could have carried an infection on the flight even if the country is infected.

d.) This model does not take into account changes in the demographics of the countries over time caused by natural process or due to the infection. One possible question to consider is what happens if all the infected people in one country died. In such cases maybe the country can be shifted into susceptible category again.