

# MoCS Assignment3

Kam Bielski  
Nate Gaylinn  
Kevin Motia

November 2023

## 1 Cascade Model

### 1.1 a)

The directed random graph described is acyclic because the edges can only "flow" in a single direction.

In other words, cycles cannot happen on this structure because a node would need to occur twice on any path in the network; this can't happen, because if you pick any node  $i$ , and follow *any* path of edges, you will never reach node  $i$  again because the node numbers  $i'$  that you follow can only decrease, terminating at node 1.

### 1.2 b)

For any node  $i$ , the average in-degree of node  $i$  is:

$$\sum_{i'=k+1}^N kP(k)$$

where  $P(k)$  is the probability of node  $i'$  having  $k$  incoming edges. This is a binomial distribution, because the generative story for a given  $k$  is "for each of the *possible* edges between me and the nodes with higher  $i$  than me, include that edge with probability  $p$ ".

$$\sum_{i'=k+1}^N k \binom{N-i}{k} p^k (1-p)^{N-i-k}$$

### 1.3 c)

We're essentially bisecting the network at node  $i$  where node  $i$  falls on the side of the "lower" indices. There are  $N - i$  nodes with higher indices, and  $i$  nodes with lower indices.

This means there are  $\hat{N} = i * (N - i)$  possible edges between these two halves of the graph. Since we add each edge with probability  $p$ , which is essentially a binomial story, so the average is  $\hat{N}p$  which is

$$i * (N - i) * p = (Ni - i^2)p$$

#### 1.4 d)

The largest value will be  $i = N/2$  because that is the maximum value for the above polynomial.

$$0 = \frac{d}{dx}(Ni - i^2)$$

$$0 = N - 2i$$

$$i = \frac{N}{2}$$

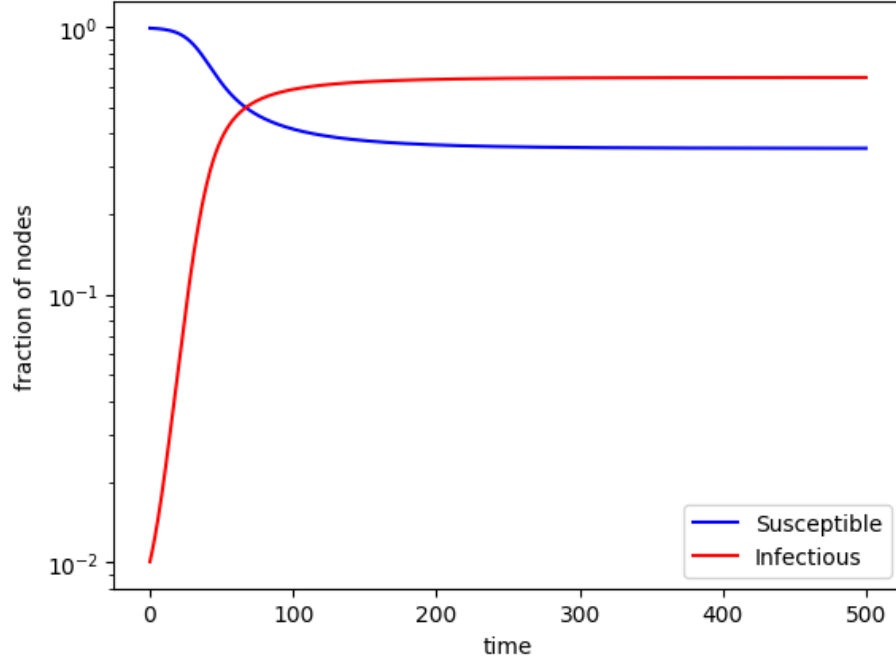
(We know this is a max because the second derivative is negative)

There are two minimum values, where  $i = 1$  and  $i = N$ . These are minimum values because there is only a single node in one of the groups in both cases, and any increase in that factor would increase (or keep the same) the total number of edges crossing the bisection.

## 2 Epidemic Spreading

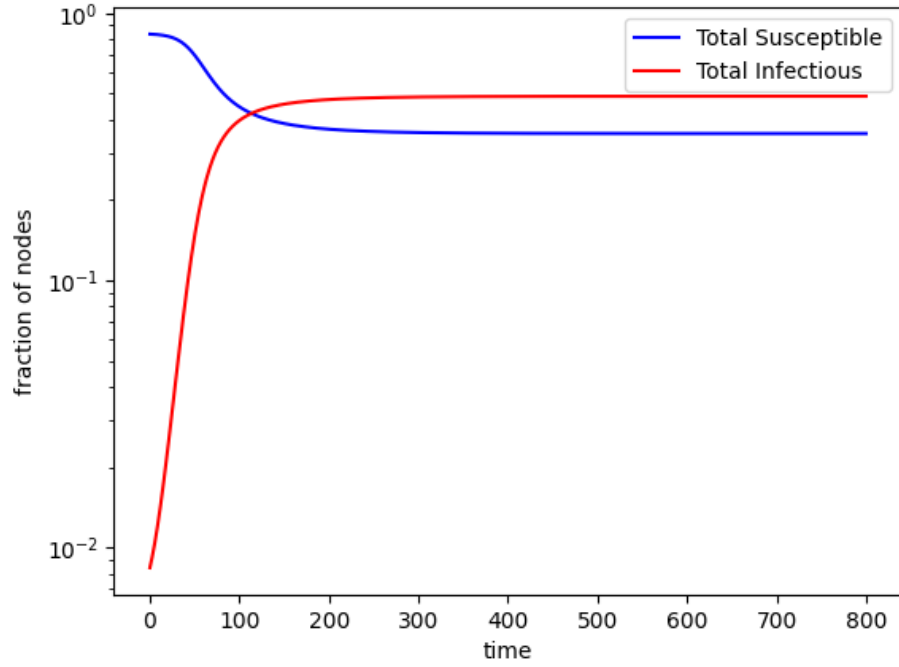
### 2.1 2a)

A power-law degree distribution was chosen such that the configuration model could approximate it. Several degree distributions could have been chosen, but a power-law degree distribution can reasonably occur for an epidemic scenario in the real world, so it was chosen. The existing code was modified to correspond to an SIS model, when it initially corresponded to an SIR model. The resulting plot shows the populations of susceptible individuals and infected individuals both reaching an equilibrium as time increases. Initial conditions were unchanged from the original code, and were the transmission rate, beta, was 0.0025 and the recovery rate, alpha, was 0.01. The initial fraction of individuals was 0.01 of the population.



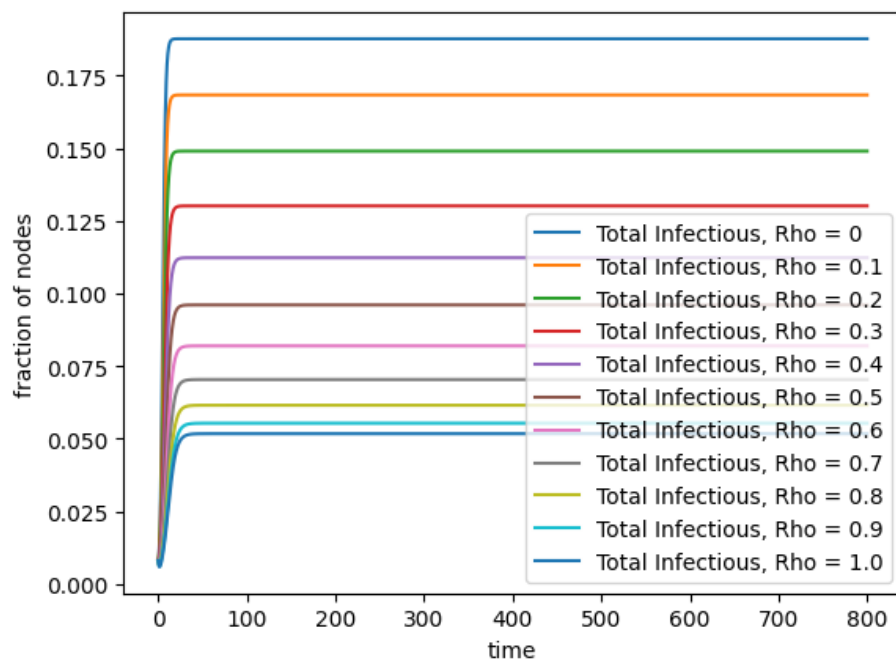
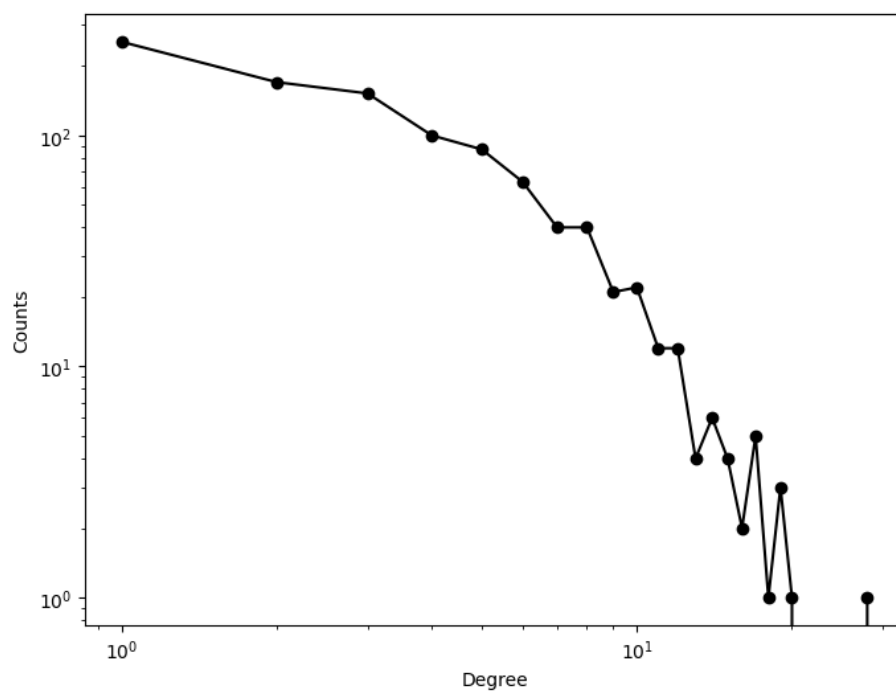
## 2.2 2b)

New compartments were added for vaccinated nodes such that the population of individuals in the system contained S susceptible unvaccinated individuals, I infected unvaccinated individuals, S susceptible vaccinated individuals, I infected vaccinated individuals. The transmission rate is reduced by a multiplicative factor  $(1-\rho)$ . The mean-field,  $\theta$ , was modified to account for vaccinated individuals. The resulting plot shows the populations of total susceptible individuals and total infected individuals both reaching an equilibrium as time increases. The parameters were the same as in 2a, except  $\rho$  was arbitrarily set to 0.35, and the vaccination rate was set to 0.40.



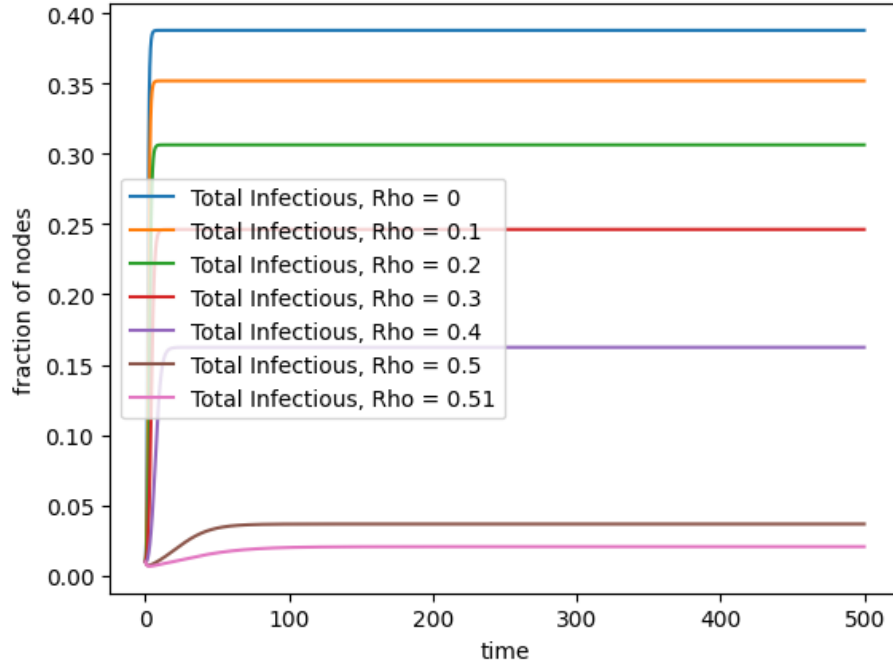
### 2.3 2c)

Specific parameters of  $\beta = 0.30$  and  $\alpha = 1$  were used with a vaccinated population of 40%. A sampled geometric degree distribution was used where  $p_k = p(1-p)^{k-1}$  with  $p = 1/4$ . The resulting plots showed that the fraction of infectious individuals in the steady state decreases with increasing  $\rho$  value (increasing vaccine efficacy). However, since only a random 40% of the individuals are vaccinated, the fraction of infected individuals never goes to 0 regardless of  $\rho$ .



## 2.4 2d)

The same parameters were used as in 2c, however, the 40% of the population that was vaccinated was selected to be the individuals that corresponded to the nodes with the highest degree (most well-connected individuals). The resulting plots showed that the fraction of infectious individuals in the steady state decreases with increasing rho value (increasing vaccine efficacy). The fraction of infectious individuals appears to go to zero after rho reaches 0.5.



## 3 Network Exploration

We chose the Cielnet dataset of Montreal drug runners, with data gathered in 1997, published in 2009 (<https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/cielnet>). In this network, nodes represent individuals in a drug gang and edges represent channels of communication. The edges are directed, and have weight equal to the volume of communication on each channel. Although the nodes in this dataset are unnamed, the background information explains that they represent the "organization leader," that leader's "lieutenant," and subordinates. This can be seen in the graph's structure, where two nodes are like hubs, connected to many other nodes which have relatively few connections. This results in an average degree of 1.8 and a long-tail distribution with more than half of the nodes having two or fewer edges. Just two nodes have a very high degree. These are most likely the "organization leader" and "lieutenant," but since the nodes are unnamed it's unclear which is which.



Figure 1: Visualization of the Cielnet dataset

The modularity of this network is 0.418, which is consistent with two clusters that are well-connected internally, but loosely connected to each other. This intuitively matches the appearance of the network, with two "hub" nodes connected to many "spokes". The two hubs (leader and lieutenant) also have very high betweenness centrality, which makes sense since they're the primary link between two sub-networks that are otherwise mostly separate. This indicates that eliminating the leader and lieutenant from this gang would likely decimate the organization, since most of the remaining nodes are only connected to one of those two and none of the other subordinates. In Figure 1, node color represents betweenness centrality, and edge size / color represents the weight of each edge.

## 4 Voter Model

### 4.1 a)

Running the voter model on the Cielnet graph almost always resulted in consensus (96-98% of the time), and usually in less than 300 iterations. More biased initial conditions tend to settle into consensus more quickly (a 50 / 50 split took about 273 iterations, but an 80 / 20 split took just 157 on average).

Simulating CielNet data

Initial conditions: 50 / 50  
49 of 50 trials had consensus after 273.31 steps on average  
Initial conditions: 60 / 40  
48 of 50 trials had consensus after 268.88 steps on average  
Initial conditions: 80 / 20  
48 of 50 trials had consensus after 156.54 steps on average

## 4.2 b)

Running the same simulations on a randomized configuration model with the same degree structure as the Cielnet data also resulted in consensus most of the time, but not so often (72-84%). It also generally took longer to reach consensus (191-321 steps on average, depending on initial conditions). This is likely because the Cielnet graph has a hub-and-spoke like structure, with two nodes with high in-betweenness centrality connected to many nodes with few neighbors. These two central nodes were much more likely to influence the vote of another node just by chance, so their opinions held more weight and could dominate the whole network. The random configuration models *also* had just two high-degree nodes, but were unlikely to have a structure like that where those two nodes were so central and in control of their own cliques.

Simulating randomized configuration model  
Initial conditions: 50 / 50  
36 of 50 trials had consensus after 321.31 steps on average  
Initial conditions: 60 / 40  
38 of 50 trials had consensus after 307.03 steps on average  
Initial conditions: 80 / 20  
42 of 50 trials had consensus after 190.95 steps on average