PHYS 5116 − Network Science I
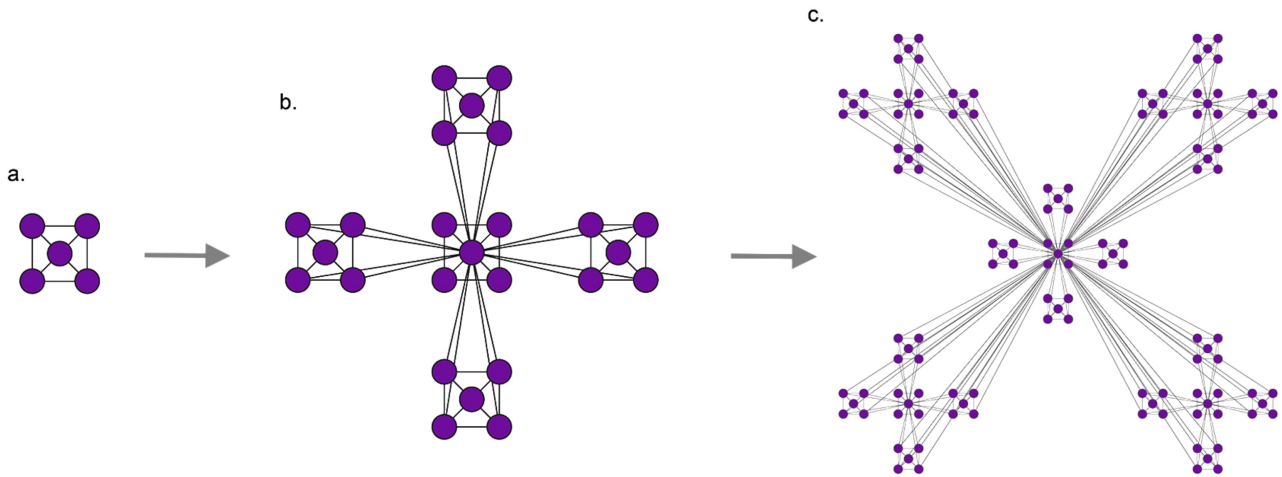# Assignment 3

Giuliano Porciúncula Guedes

November 21, [Fall] 2025

1 Hierarchical Community Network (25 points).

Below is a replica of Figure 9.13 from the textbook, illustrating a hierarchical network generative model. This model constructs networks iteratively, starting from a fully-connected graph of 5 nodes (a; note that two edges are not represented in the drawing), followed by creating 4 replicas of the network at the previous step, where the peripheral (non-center) nodes each connect to the center node of the previous module (b), and so on (c). See Section 9.3 and Advanced Topic 9.B.

Write code to replicate Figure 9.14, creating a 3-panel figure showing:

(a) The degree distribution p_k of your hierarchical network of $N = 15625$ nodes (i.e., cliques of size five for six iterations). Use log-log scaling in your plot.

(b) The relationship between $C(k)$ and k (i.e., the clustering coefficient as a function of degree k vs degree k). Use log-log scaling in your plot.

(c) The relationship between the global clustering of the network and network size, $C(N)$ vs N. Note: this question is referring to the average clustering function in networkx, which is different than some of the previous assignments in this class. Use log-log scaling in your plot.
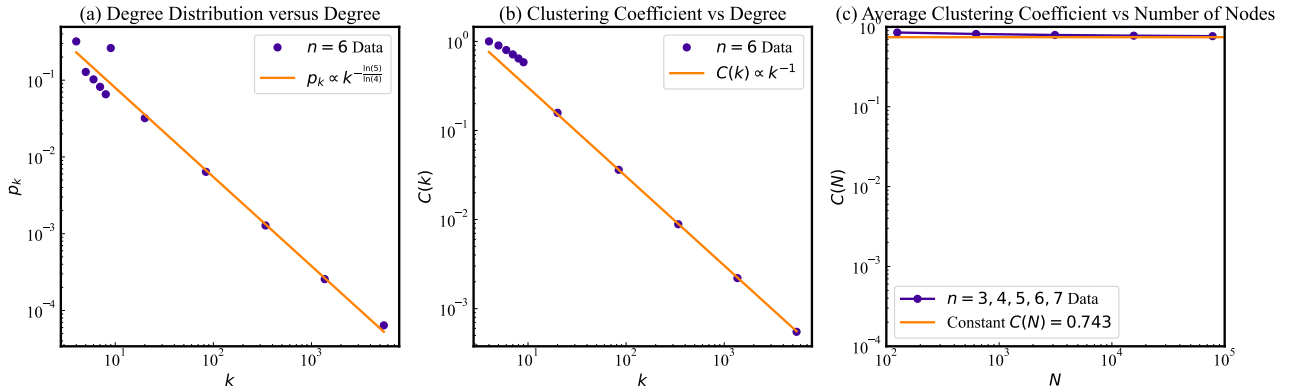
Solution

Figure 1: (a) Degree Distribution versus Degree, (b) Clustering Coefficient versus Degree, and (c) Average Clustering Coefficient versus Number of Nodes. We observe all the tendencies predicted by theory. Both the slopes $-\frac{\ln(5)}{\ln(4)}$ and -1 of, respectively, the Degree distribution and the Clustering Coefficient are present, and the constant $C(N) = 0.743$ behavior of the average clustering coefficient is also present.

∎

2 Conspiracy in Social Networks (25 points).

In a Big Brother society, the Thought Police wants to follow a "divide and conquer" strategy by fragmenting the social network into isolated components. You belong to the resistance and want to foil their plans. There are rumors that the police wants to detain individuals that have many friends and individuals whose friends tend to know each other. The resistance puts you in charge to decide which individuals to protect: those whose friendship circle is highly interconnected or those with many friends. To decide, you simulate two different kinds of attacks on your network by removing (i) the nodes with the highest clustering coefficient or (ii) nodes with the largest degree. Study the size of the giant component as a function of the fraction of removed nodes for both types of attack on the following two networks:

(a) A network with $N = 15625$ nodes generated with the configuration model and power-law degree distribution with $\gamma = 2.5$.

(b) A network with $N = 15625$ nodes generated with the hierarchical model described in Figure 9.13 and Advanced Topic 9.B (same network as 1a above).

Answer the following questions: What is topological information—clustering coefficient or degree—which, if protected, limits the damage best? Would it be better if all individuals' information (clustering coefficient, degree, etc.) could be kept secret? Why? Create figures that show the size of the largest connected component over time (as a fraction of N) for each of the two attack types (degree and clustering) for each of the networks in (a) and (b).

Solution

(a) Power-law degree distribution

$$p(k) = Ck^{-\gamma} \tag{1}$$

Normalization to find C

$$\int_1^\infty p(k)dk = 1 \tag{2}$$

$$C^{-1} = \int_1^\infty k^{-\gamma}dk \tag{3}$$

$$C^{-1} = \left.\frac{k^{-\gamma+1}}{-\gamma+1}\right|_{k=1}^\infty = \frac{1}{(\gamma-1)} \tag{4}$$

So

$$p(k) = (\gamma-1)k^{-\gamma} \tag{5}$$

The cumulative degree distribution is given by

$$f(k) = \int_1^k p(k')dk' \tag{6}$$

Doing the same integral again, now we have

$$f(k) = 1 - k^{-(\gamma-1)} \tag{7}$$

Thus if we generate a random number $u \in [0,1]$, we can find k by inverting $f(k) = u$ to find $k(u)$

$$k(u) = (1-u)^{-\frac{1}{(\gamma-1)}} \tag{8}$$

Remember this is a continuous distribution, but in reality we only have integer ks, thus we can fix it by takin the floor, and making 1 be the minimum. So

$$k(u) = \max\left(\quad 1 \quad , \quad (1-u)^{-\frac{1}{(\gamma-1)}} \quad \right) \tag{9}$$

By sampling $N = 15625$ nodes and creating the network with that degree sequence by the configurational mode, we have the following degree distribution.
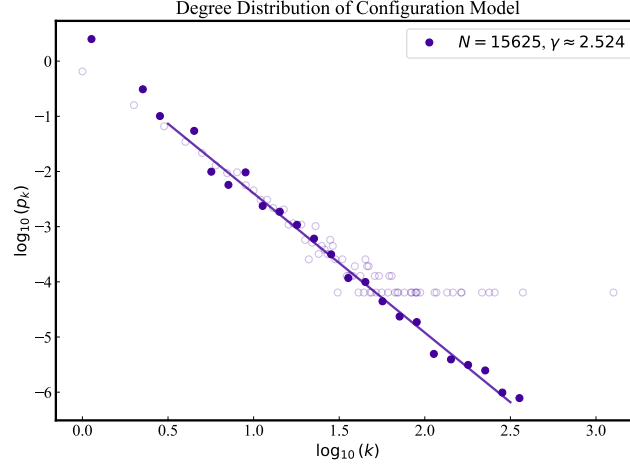
3

Figure 2: (a) Degree Distribution of the configurational model. We successfully created the network with $N = 15625$ and γ "close enough" to 2.5. Hollow circles are the unaltered data, while the filled circles are the log-binned data and the straight line is a fit where we extracted γ.

(a,b) We did three analysis. First we removed nodes uniformly randomly. This was the fastest method. The second method, which was slower than the (uniformly) random one, was by removing the highest degree nodes. If there was more than one node with the highest degree, they would be chosen uniformly randomly between them. The last and slowest method was by focusing the nodes with the highest clustering coefficient. This took a bit to optimize. The straightforward and slow method is to re-calculate the clustering coefficient of all nodes for each step. My optimization was to only re-calculate the clustering coefficient of the nodes that were linked to the removed node. This reduced the computation time by 2 order of magnitudes.
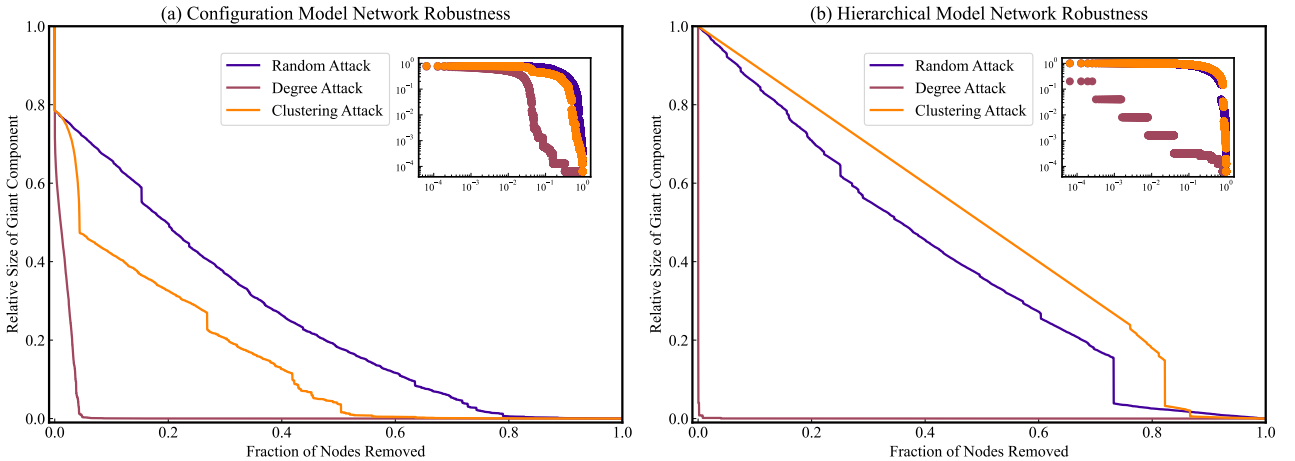

Figure 3: Network Robustness of the (a) Configurational Model and (b) Hierarchical Model. We observe a very devastating effect by the degree attacks. In the configurational model, we observe that the clustering attack was more devastating as the random attack, which is expected. But in the hierarchical model, the random attack was more devastating than the clustering attack, which is not straightforward.

Figure 3 shows the results of the attacks on each network. It shows the Network Robustness of the (a) Configurational Model and (b) Hierarchical Model. (a) The attacks on the

4

Configurational Model behave very intuitively. The degree attacks are the most devastating ones, since they quickly attack the hubs and destroy the connectivity of the network. The next most effective attack was the clustering attack, more even so than the random attack. This is expected by the "friendship paradox". The nodes with the highest clustering not necessarily hold together the connectivity of the network! And the least effective, also as expected, was the random attack. (b) The attacks of the Hierarchical Model were very interesting. The most devastating one was the degree attack, as expected. The higher degree attacks are the hubs, and we have the central node which hold the whole network together, followed by the smaller hubs. Focusing them, it very quickly destroys the connectivity of the network. The very non straightforward result here is that the random attacks are more effective than the clustering attack. If you think thoroughly, it actually makes total sense. The most clustered nodes are not the nodes holding the network together, they are the very small peripherical hubs that are very clustered, but they have little value for the overall connectivity of the network. A random attack, on the other hand, can accidentaly hit one of the more important hubs, destroying the connectivity of the network more quickly. This is a very interesting result.

For the Configurational Model, the resistance want to hid both the degrees and the clustering of their members, while focusing on hiding the degree. For the Hierarchical Model, it's very very important to hide the degree of their members, but it's actually a good strategy to show the clustering of their members. This means that the Big Brother will redirect it's attention to focusing on non-important nodes while thinking it's being smart and being progress, while it would be more productive to just attack the members randomly! Amazing results!

∎

3. Epidemics on networks (25 points)

Use the BA model to obtain a network with 10000 nodes, using $m = 2$.

(a) Calculate the critical spreading rate λ_c for both the SIS and SIR models.

(b) Run some epidemic simulations on your network, using the SIS model. Start by selecting 1000 infected nodes to simulate something endemic like the flu (for this exercise, it will be more interesting if the epidemic can't die out immediately by chance). Then, for 100 timesteps, allow the epidemic to run on the network. Keep in mind that:

- At each timestep, a susceptible node can be infected by any of its infected neighbors with probability β.

- At each timestep, an infected node may recover and become susceptible again with probability μ.

Do this for 3 different values of μ, with $\beta = 0.2$. Plot the fraction of the nodes that are infected at each timestep, to show the evolution of the epidemic. Does the epidemic die out in any of your simulations? Is there a value of μ for which you would expect it to?

(c) Now you'll run some vaccination campaigns, with varying levels of success. Fix $\beta = \mu = 0.2$.

Then randomly immunize 10% of the nodes in your network, and run the simulations again. Repeat for 30%, 50%, and 70% vaccinated. Have you managed to eradicate the epidemic? Would you expect to be able to?

d) Now imagine a world ruled by network scientists. With unilateral power to determine who gets vaccines, you decide to use the "select a random individual, then immunize one of its randomly-selected friends" strategy (which is totally reasonable and everyone would be very happy about). Repeat part (c), but now instead of randomly selecting nodes to immunize, randomly select edges, and then select one of the nodes at the end of the edge to vaccinate (stop until you've reached the target percentage of nodes). Does this strategy allow you to eradicate the epidemic?

Solution

(a) The critical spreadin rate for both SIS and SIR models are given by

$$\lambda_c^{(\text{SIS})} = \frac{\langle k \rangle}{\langle k^2 \rangle} \qquad \lambda_c^{(\text{SIR})} = \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1} \tag{10}$$

We know that $\langle k^2 \rangle$ diverges to infinity as $N \to \infty$. For our generated network with $N = 10000$ and $m = 2$, we calculated $\lambda_c^{(\text{SIS})} \approx 0.07007$ and $\lambda_c^{(\text{SIR})} \approx 0.07535$.

$\square$

(b) Since we've found $\lambda_c^{(\text{SIS})} \approx 0.07007$, we know that $\lambda = \beta/\mu$ and so $\mu = \beta/\lambda$, thus our $m_c = \beta/\lambda_c c$, which with $\beta = 0.2$ gives us $\mu_c \approx 2.85423$. So if we decrease μ from μ_c, we approach a fully infected regime. If we increase μ from μ_c, we approach a zero infection regime.
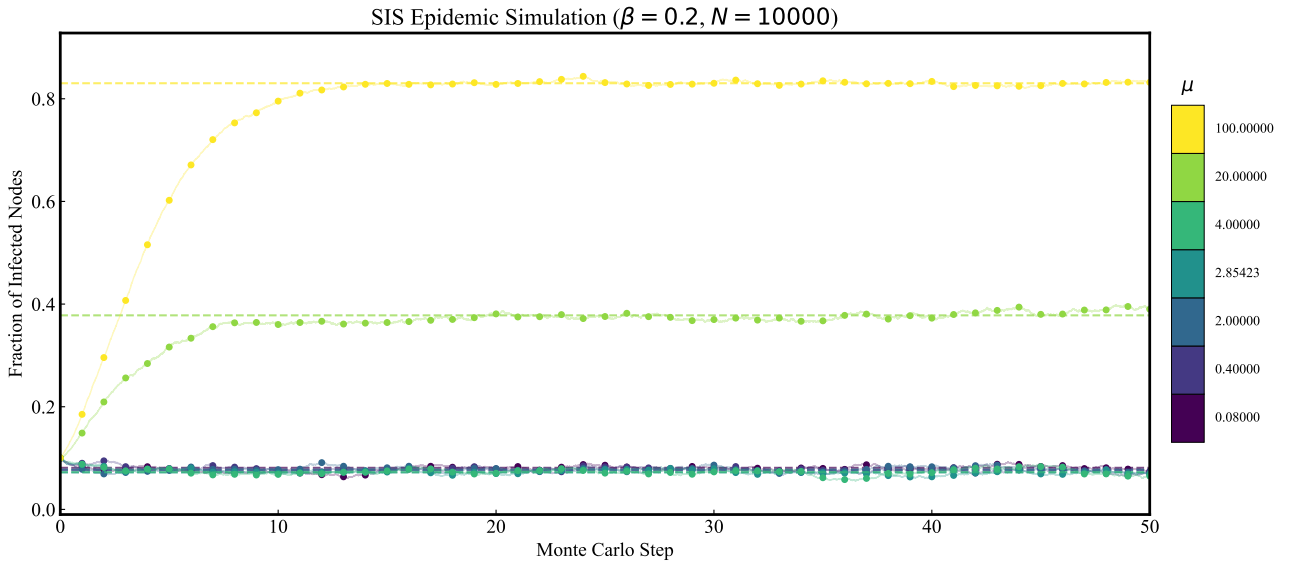

Figure 4: Time evolution of the Fraction of Infected Nodes for different values of μ. We see a mostly-infected / mostly-not-infected regimes separated by μ_c. The dashed lines represent the stationary average.

Figure 4 shows the time evolution of the Fraction of Infected Nodes in our network. 1 Monte Carlo Step consists of N trials, where each trial we randomly select a node. If it's infectious, we check if it becomes susceptible. If it's susceptible, it interacts with all it's infectious neighbors and each interaction it tries to turn infectious with probability β. The dashed lines represent the stationary average, and were taken after 20 MCS to wait for thermalization.

Figure 5 shows the Stationary State versus the recovery rate μ. The middle point is μ_c. We observe the predicted $\mu < \mu_c$ and $\mu > \mu_c$ behaviors discussed above.
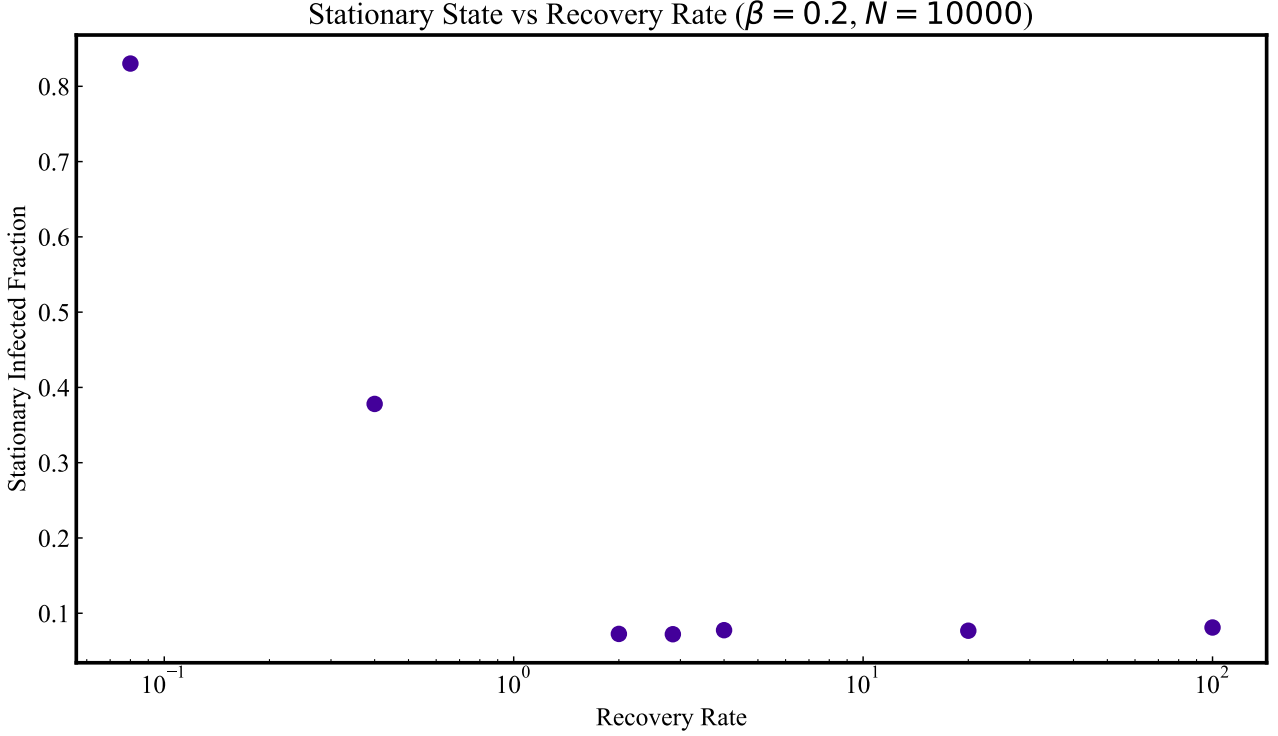

Figure 5: Stationary State versus the recovery rate. The middle point is μ_c. We observe a mostly-infected phase for $\mu < \mu_c$ to a mostly-not-infected phase for $\mu > \mu_c$.

One can go real crazy on the analysis of this system by doing a critical phase transition finite size scaling analysis. It's the exact same toolking and exact same procedure as done by Porciúncula et al. so the system is fully characterized. Here our order parameter would be the fraction of infected nodes and our main changing parameter will be the recovery rate. The main question here for doing all this is in what network exactly are we gonna do it, and why is it a justified pick. That answered, it's a free paper!

□

(c) Figure 6 shows the time evolution of the Fraction of Infected nodes for different vaccination percentile of the population. Here we have equal $\mu = \beta = 0.2$. We followed the same procedure for (b). Vaccination here means that the node cannot turn into infected. What this does for the dynamics is that: that nodes is essentially removed from the network. Since it's always a susceptible node, but cannot get infected, it is effectively removed from the dynamics. This creates an effective subgraph of the network with different $\langle k \rangle$ and $\langle k^2 \rangle$, pushing the λ_c around.

This is what we see in our figure. For high vaccinations, we see that we are in the mostly-non-infected regime, while for a low vaccination we are in the mostly-infected regime. This means that, even tho the epidemic will still be always there, we won't overload our hospitals and all is relatively good.
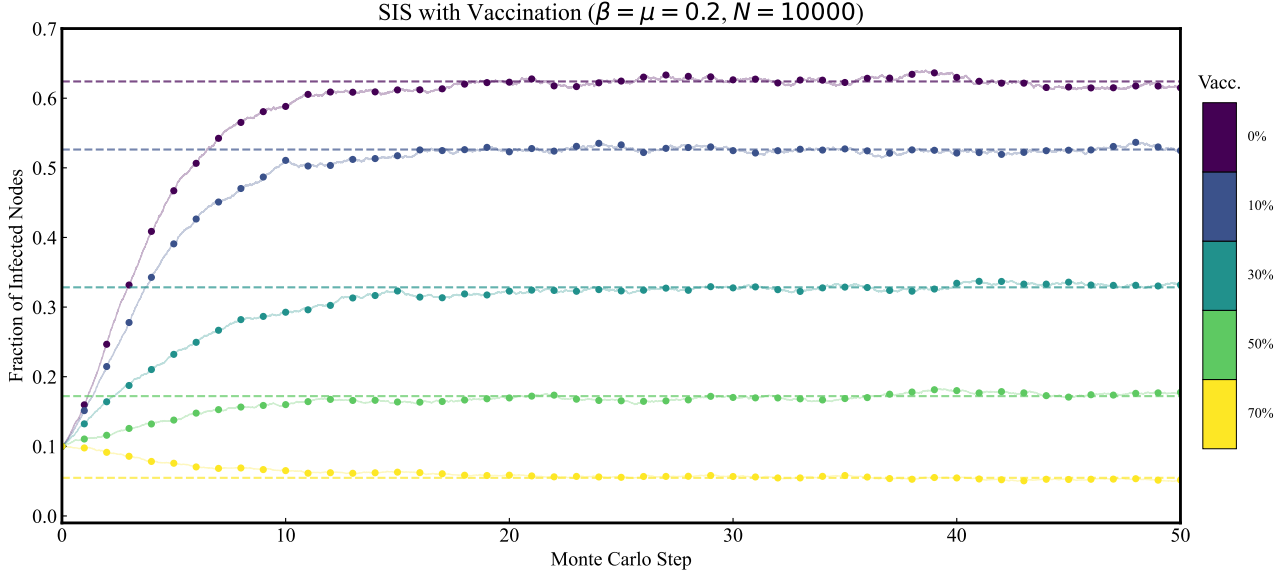

Figure 6: Time evolution of the Fraction of Infected nodes for different vaccination percentile of the population with equal $\mu = \beta = 0.2$. We observe that the epidemic is never fully eradicated.

□

(d) Figure 7 shows the time evolution of the fraction of infected nodes for $\mu = \beta = 0.2$. Again, the infection is never fully eradicated but it works considerably better than the normal random nodes. Here by focused we mean focusing on the edges. It is indeed more effective than the normal vaccination.
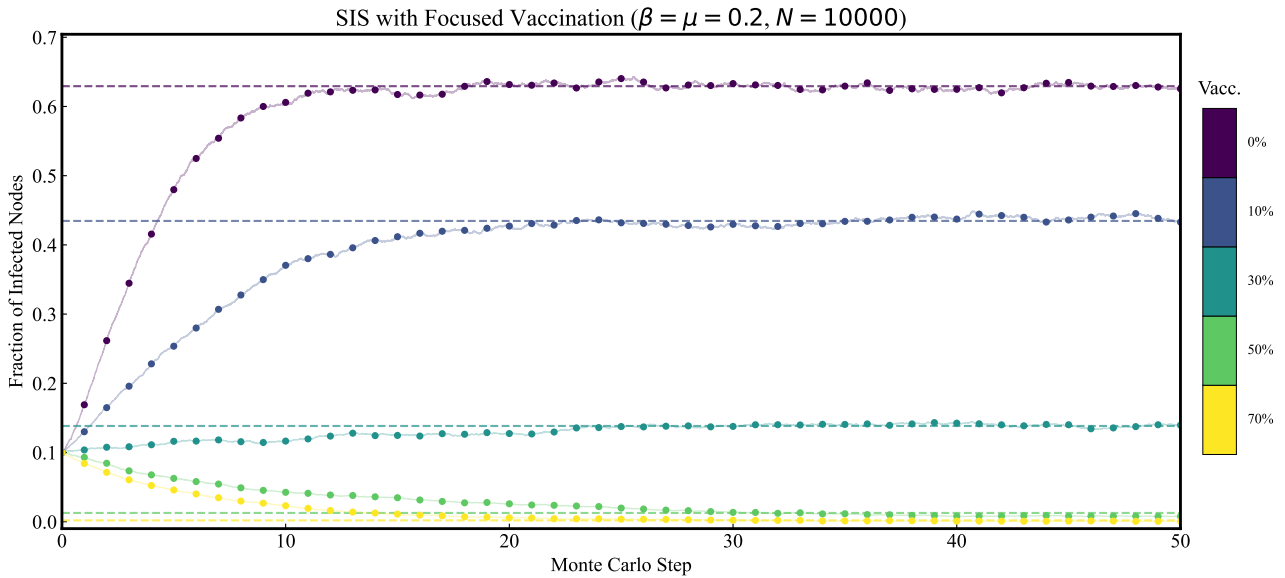

Figure 7: Time evolution of the Fraction of Infected nodes for different focused vaccination percentile of the population with equal $\mu = \beta = 0.2$. We observe that the epidemic is never fully eradicated, but it works better than the normal vaccination.

8

Figure 8 shows the stationary state of both Normal; and Focused Vaccinations. We can clearly see that the focused vaccination is considerably more effective than the normal vaccination. For example, just eyeballing it, 10% focused vaccination has the same effectiveness as 20% normal vaccination, while 30% focused vaccination has the effectiveness of almost 60% normal vaccination. It is not a linear relationship as we can see from the figure, but the focused vaccination is way more effective specially at reducing the infection and to maintain it low. It never is really eradicated, but it gets really close to it. This means that an effective immunization strategy is to focus vaccinate people that interact with a lot of people, i.e., people with jobs that interact with a lot of different people. These jobs might include supermarket workers, office workers, restaurant staff, amusement park workers, hotel staff.
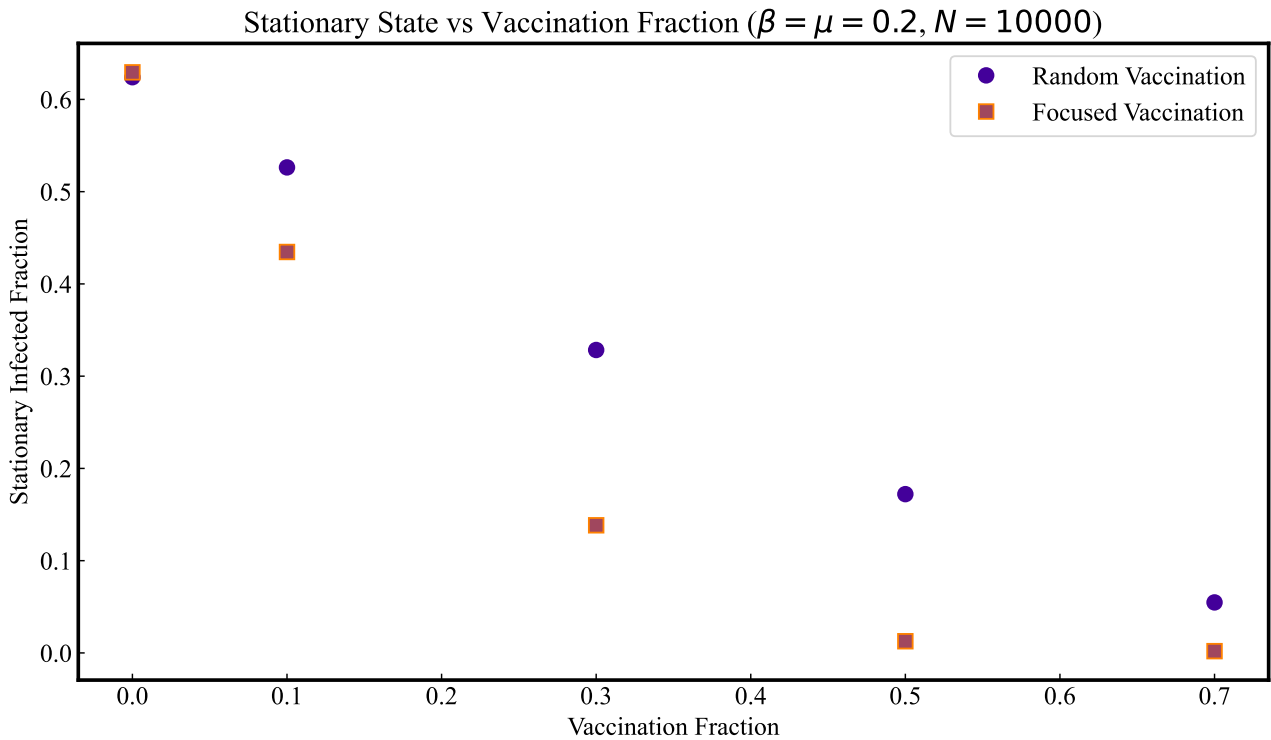

Figure 8: Stationary State versus the recovery rate for both Normal and Focused Vaccinations. We observe that the focused vaccination is always more effective than the normal one.

4. Modularity Resolution Limit: (25 points)

Consider a network consisting of a ring of n cliques of size k, where n is an even number. The neighboring cliques are connected by a single link (each k-clique has an additional 2 links). This network has an obvious intuitive community structure, with each k-clique corresponding to a community.

(a) Determine the modularity M_i of the intuitive partition, and the modularity M_p of the partition in which pairs of neighboring k-cliques are merged into a single community. Express each modularity **only** as a function of k and n.

(b) Relating M_i and M_p, show that only for $n < k^2 - k + 2$ will the modularity maximum

predict the intuitively correct partition.

(c) What are the consequences of increasing n above this limit? Specifically, why does the "pairs" partition look better in terms of modularity when n crosses this threshold? To answer this, derive the expected number of links between two adjacent communities in the intuitive partition according to the null model that is built into the definition of modularity. What does this value have to be less than for the "intuitive" partition to be unlikely by chance and hence a good guess for the community structure? Relate this line of reasoning to the inequality you derived in (b).

Solution

(a) Modularity in general can be expressed by

$$M = \frac{L_c}{L} - \left(\frac{k_c}{2L} \right)^2 \tag{11}$$

Where L_c is the total number of links inside the community, k_c is the sum of all the degrees of the nodes in that community, and L is the total number of links in the network.

Let's start by the intuitive partition. Of course the intuition partition is just the cliques. So it's a clique of size k, this means that each node makes $k - 1$ links to other nodes in the clique, thus we have that $L_i = k(k-1)/2$ (we divide by two to not double count). For k_i we have again that each node has degree $k-1$ with the exception of two nodes that also connect outside of the clique. Thus we have $k_i = k(k-1) + 2$. For the total number of links in the network, we have that each clique has $k(k-1)/2$ links, thus as we have n cliques we have a total of $nk(k-1)/2$ in-clique links. Since our cliques are in a ring, we have an extra n links between adjacent cliques. Thus

$$L = n\frac{k(k-1)}{2} + n \tag{12}$$

So our intuitive partition modularity is

$$M_i = \left(\frac{k(k-1)}{2} \right) \left(n\frac{k(k-1)}{2} + n \right)^{-1} - \left[\frac{1}{2} (k(k-1) + 2) \left(n\frac{k(k-1)}{2} + n \right)^{-1} \right]^2 \tag{13}$$

After some simplification, we have

$$\boxed{M_i = \frac{(n-1)(k^2 - k) - 2}{n^2(k^2 - k + 2)}} \tag{14}$$

For $n \to \infty$, M_i goes to zero. This makes sense since we are going to have infinite communities which are relatively small to the network size, so they just disappear. Now for $k \to \infty$, we have an interesting results. We see that the k^2 terms both in the numerator and denominator dominate over all other terms, and we find a modularity of $M_i = (n-1)/n^2$.

Now for M_p we have that total number of links in the partition are two times the ones from M_i plus one connecting the cliques, thus $L_p = k(k-1) + 1$. For the sum of all the degrees, we have two times the previous one, so $k_p = 2k(k-1) + 4$. To facilitate our life, we can write the general modularity as

$$M = \frac{L_c L - \frac{1}{4} k_c^2}{L^2} \tag{15}$$

So now our modularity for the joint is is

$$M_p = \frac{(k(k-1) + 1)\left(n\frac{k(k-1)}{2} + n\right) - \frac{1}{4}(2k(k-1) + 4)^2}{\left(n\frac{k(k-1)}{2} + n\right)^2} \tag{16}$$

After simplification, we have

$$\boxed{M_p = 2\frac{(n-2)(k^2 - k + 1) - 2}{n^2\left(k^2 - k + 2\right)}} \tag{17}$$

$\square$

(b) Since we want to know which modularity method will predict the correct partitions, we need to calculate the total modularity $M^{(T)}$ of the network. For the intuitive method, we have a total of n communities, thus we multiply the individual modularity by n (The total modularity is the sum of the modularities. Since they are the same we can just multiply).

$$M_i^{(T)} = \frac{(n-1)(k^2 - k) - 2}{n(k^2 - k + 2)} \tag{18}$$

For the other partition method, we multiply by $n/2$ since we have pairs of cliques (remember n is even so we are safe here).

$$M_p^{(T)} = \frac{(n-2)(k^2 - k + 1) - 2}{n\left(k^2 - k + 2\right)} \tag{19}$$

Now we can look at the ratio We can look at the ratio M_i^T / M_p^T

$$\frac{M_i^{(T)}}{M_p^{(T)}} = \frac{(n-1)(k^2 - k) - 2}{(n-2)(k^2 - k + 1) - 2} \tag{20}$$

The turning point is when the ratio is 1. If the ratio is more than 1, then M_i wins. If the ratio is less than one, M_p wins. We can calculate n' by

$$\frac{(n'-1)(k^2 - k) - 2}{(n'-2)(k^2 - k + 1) - 2} = 1 \tag{21}$$

Which by straightforwardly isolating n' we have

$$\boxed{n' = k^2 - k + 2} \tag{22}$$

Since the term summing $(n-2)$ in the denominator is bigger than the term summing $(n-1)$ in the numerator, we know that by increasing n the denominator wins, thus for $n > n'$ M_p wins. Since we want the intuitively partition to win, we want $\boxed{n < n'}$.

$\square$

(c) The degree preserving null model gives us the expected number of edges that two nodes i and j that are connected as

$$e_{i,j} = \frac{k_i k_j}{2L} \tag{23}$$

The expected number of connections $E_{A,B}$ between two sets A and B of nodes is

$$E_{A,B} = \sum_{i \in A} \sum_{j \in B} e_{i,j} \tag{24}$$

This can be rewritten as

$$E_{A,B} = \frac{1}{2L} \left(\sum_{i \in A} k_i \right) \left(\sum_{j \in B} k_j \right) \tag{25}$$

Where each sum is k_A and k_B respectively

$$E_{A,B} = \frac{k_A k_B}{2L} \tag{26}$$

From before we know that each clique has $k_i = k(k-1)+2$, thus after substituting $k_A = k_B = k_i$ and L we have

$$E_{A,B} = \frac{(k^2 - k + 2)}{n} \tag{27}$$

If the denominator is bigger than the numerator $n > (k^2 - k + 2)$, we expect by the null model less than one edge between A and B. If $n < (k^2 - k + 2)$ we expect by the null model more than one edge between them. Since we have exactly one link between adjacent communities A and B, this can be related to the null model.

For $n > (k^2 - k + 2)$, the null model expects less than one edge between two adjacent cliques. Since the real network has one edge, the connectivity between the cliques is not unusually sparse, so modularity tends to join the cliques, favoring the "pairs" partition. For $n < (k^2 - k + 2)$, the null model expects more than one edge between the cliques. Observing only one edge is therefore unusually sparse, so modularity has an incentive to keep the cliques as separate communities, favoring the intuitive partition. These conditions agree with what we found back in (b).

$\blacksquare$

All the code from this assignment can be found: Here.