Final Exam Problem 2

Yuancheng Xu

Dec 17, 2020

Q1

Firstly, we give an expression for the critical transmissibility T_c. As derived in class, we have $T_c = \frac{\sum_{k=1}^{\infty} k p_k}{\sum_{k=1}^{\infty} k(k-1)p_k} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} = \frac{1}{\kappa - 1}$. Note that since the original graph has a giant component, $\kappa > 2$ and thus $T_c < 1$.

Now assume that $T > T_c$ (so there will be an epidemic if we do nothing) and we would like to know if we vaccinate a fraction v of randomly selected nodes (before the spread of disease), can we destroy the epidemic and most importantly, what is the critical value v_c (the minimal fraction of vaccinated node in order to stop the epidemic). Taking the vaccine into account and assuming there is no epidemic, we will first derive the generating function for the transmitting edges, then the generating function for the size of transmitting cluster and finally, the average outbreak size, from which we will extract the critical v_c.

As usual, let $G_0(x) = \sum_{k=0}^{\infty} p_k x^k$ and $G_1(x) = \frac{G_0'(x)}{\langle k \rangle}$ denote the generating function for degree and excess degree distribution. Note that the probability that an edge is transmitting, when taking the vaccine into account, is $(1-v)T$ and is not transmitting with probability $v + (1-v)(1-T)$. Therefore, the generating function for distribution of transmitting edges adjacent to a node is

$$
\begin{aligned}
G_0(x;T) &= \sum_{m=0}^{\infty} \sum_{k=m}^{\infty} p_k \binom{k}{m} ((1-v)T)^m (v + (1-v)(1-T))^{k-m} x^m \\
&= \sum_{k=0}^{\infty} p_k \sum_{m=0}^{k} \binom{k}{m} ((1-v)Tx)^m (v + (1-v)(1-T))^{k-m} \\
&= \sum_{k=0}^{\infty} p_k ((1-v)Tx + v + (1-v)(1-T))^k \\
&= G_0((1-v)Tx + 1 - T + Tv)
\end{aligned}
\tag{1}
$$

Similarly, we can obtain the generating function for distribution of transmitting edges adjacent to a node arrived at by a randomly chosen edge

$$
G_1(x;T) = G_1((1-v)Tx + 1 - T + Tv)
\tag{2}
$$

Then the generating functions for the size of transmitting cluster reached from a random edge and vertex are given by

$$H_1(x;T) = xG_1(H_1(x;T);T) \tag{3}$$

$$H_0(x;T) = xG_0(H_1(x;T);T) \tag{4}$$

By differentiating Equation 1 and Equation 2 and substituting $x = 1$, we have $G_0'(1;T) = T(1-v)G_0'(1)$ and $G_1'(1;T) = T(1-v)G_1'(1)$. By differentiating Equation 3 we have $H_1'(1;T) = 1 + G_1'(1;T)H_1'(1;T) = \frac{1}{1-G_1'(1;T)}$. Finally, using Equation 4, we compute the average outbreak size as

$$
\begin{aligned}
\langle s \rangle &= H_0'(1;T) \\
&= 1 + G_0'(1;T)H_1'(1;T) \\
&= 1 + \frac{G_0'(1;T)}{1 - G_1'(1;T)} \\
&= 1 + \frac{T(1-v)G_0'(1)}{1 - T(1-v)G_1'(1)}
\end{aligned} \tag{5}
$$

Therefore, the phase transition occurs when $1 - T(1-v)G_1'(1) = 0$. Since T is given here (with $T > T_c = \frac{1}{\kappa-1}$), the critical fraction of nodes

$$v_c = 1 - \frac{1}{TG_1'(1)} = 1 - \frac{T_c}{T}$$

Note that $0 \leq v_c \leq 1 - T_c \leq 1$. If $v > v_c$, epidemic cannot occur and if $v < v_c$, one may occur.

Q2

(a)

First we compute the generating function for degree and excess degree distribution. Denote $\text{Li}_\alpha(x) = \sum_{k=1}^{\infty} \frac{x^k}{k^\alpha}$ and therefore $\text{Li}_\alpha(1) = \zeta(\alpha)$. Note that $\text{Li}_\alpha'(x) = \frac{1}{x}\text{Li}_{\alpha-1}(x)$

$$G_0(x) = \frac{1}{\text{Li}_\alpha(1)} \sum_{k=1}^{\infty} k^{-\alpha} x^k = \frac{1}{\text{Li}_\alpha(1)}\text{Li}_\alpha(x)$$

$$G_1(x) = \frac{G_0'(x)}{\langle k \rangle} = \frac{\text{Li}_\alpha'(x)}{\text{Li}_\alpha(1)\frac{\text{Li}_{\alpha-1}(1)}{\text{Li}_\alpha(1)}} = \frac{\text{Li}_{\alpha-1}(x)}{x\text{Li}_{\alpha-1}(1)}$$

Now let u be the solution of $u = G_1(u)$ (we numerically compute u by solving $u = \frac{\text{Li}_{\alpha-1}(u)}{u\text{Li}_{\alpha-1}(1)}$ in Matlab). As we have derived in class, the fraction S of nodes in the giant component can be computed by $S = 1 - G_0(u)$. Therefore,

$$S = 1 - \frac{\text{Li}_\alpha(u)}{\text{Li}_\alpha(1)}$$

2

.

Using $\alpha = 2.2$, we find $u = 0.1963$ and $S = 0.8622$, which is the fraction of nodes in the giant component.

(b)

As derived in the slides, denote $u = H_1(1; T)$ and we have $u = G_1(u; T) = G_1(1 + (u - 1)T)$. Therefore, we numerically solve

$$u = \frac{\text{Li}_{\alpha-1}(1 + (u - 1)T)}{(1 + (u - 1)T)\text{Li}_{\alpha-1}(1)}$$

Then we compute the fraction $S(T)$ of nodes affected by the epidemic (the fraction of the giant component) as $S(T) = 1 - H_0(1; T)$ where $H_0(1; T) = G_0(u; T) = G_0(1 + (u - 1)T) = \frac{\text{Li}_\alpha(1+(u-1)T)}{\text{Li}_\alpha(1)}$. That is,

$$S(T) = 1 - \frac{\text{Li}_\alpha(1 + (u - 1)T)}{\text{Li}_\alpha(1)}$$

Using $\alpha = 2.2, T = 0.4$, we have $u = 0.2883$ and $S(T) = 0.4078$, which is the fraction of nodes affected by the epidemic if it occurs.

(c)

As derived in Question 1,

$$v_c = 1 - \frac{T_c}{T}$$

where $T_c = \frac{1}{\kappa - 1}$. The first moment $\langle k \rangle = \frac{\text{Li}_{\alpha-1}(1)}{\text{Li}_\alpha(1)}$ and the second moment $\langle k^2 \rangle = \frac{1}{\text{Li}_\alpha(1)} \sum_{k=1}^{\infty} k^2 k^{-\alpha}$ does not exist (since $\alpha < 3$), and thus formally $\kappa = \infty$, $T_c = 0$.

Therefore, the critical fraction v_c to vaccinate is $v_c = 1$ (vaccinate everyone), for any $T > 0$.

Q3

In this question, we will run numerical simulations and compare the results with those in Q2.

(a)

We use DFS to find the size of the gaint component. Note that the theoretical value for an infinite graph is $S = 0.8622$.

The value obtained by simulation is 0.8211, averaged over 100 runs.

(b)

To simulate this, we first generate the original graph and then keep each edge with probability T. Then we use DFS to find the size of the giant component if there is one, which corresponds to the fraction of nodes affected by an epidemic. Note that the theoretical value for an infinite graph is $S_T = 0.4078$.

The value obtained by simulation is 0.3691, averaged over 100 runs.

(c)

Since $\alpha < 3$, by theory we have $T_c = 0$ for an infinite graph. Therefore, by theory, as long as $T > 0$, the giant component (epidemic) occurs. To verify if this is true in the simulation, we plot the fraction S versus T in Figure 1. By printing out specific values in this figure, we observe that when $T < 0.03$, we find $S < 0.01$ and S increases slowly with T. Therefore we suspect that $T_c = 0.03$ is the critical value for this finite graph (so below this threshold there is no giant component). When $T > 0.03$, S is roughly linear in T and epidemic obviously occurs.
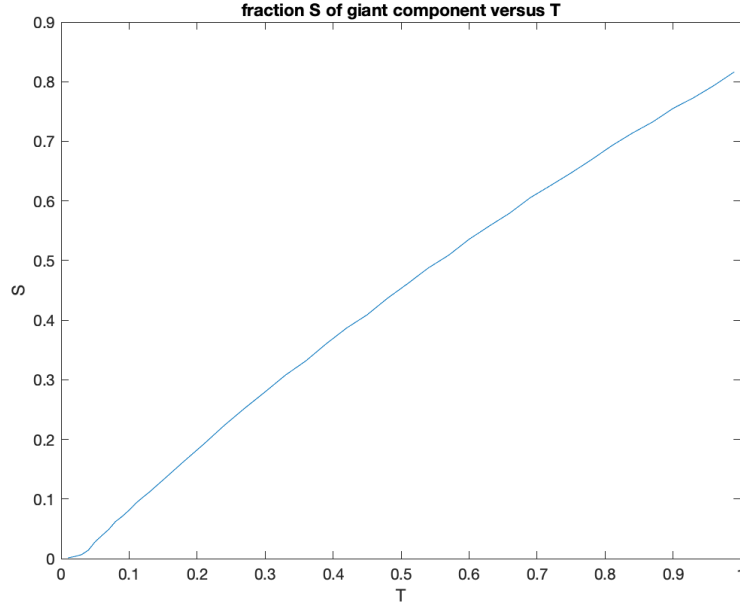

Figure 1: fraction S of the giant component (the fraction affected by an epidemic) versus T. For each T value, S is averaged over 20 runs.

(d)

As we have seen in Q2, the critical value $v_c = 1$ by theory. To simulate this, we first generate the original graph and vaccinate fraction v of nodes. Then

for each edge in the original graph, if the corresponding node is not vaccinated, then with probability T we keep it and otherwise, we remove it. We run DFS on the resulting graph to obtain the fraction of the giant component.

The result is shown in Figure 2. We see that as expected, the fraction S decreases as vaccine v increases. However, in order to fully eliminate the epidemic (say, $S < 0.01$), we need $v > 0.93$ and therefore we conclude that in this finite graph case, $v_c = 0.93$ (this is also observed by printing out the values in Figure 2).
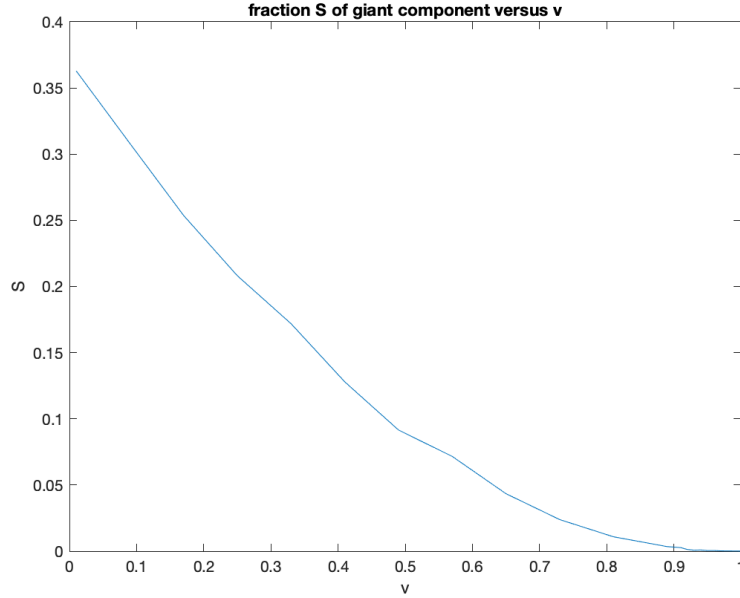

Figure 2: fraction S of the giant component (the fraction affected by an epidemic) versus v, with $T = 0.4$. For each v value, S is averaged over 20 runs.

Q4

Here we run a discrete time SIR model (with $\alpha = 2.2, T = 0.4$ and no vaccine) from a single infecting node. We assume that each infecting node remains infecting for one time step (for example, at time $t - 1$, a node v is uninfected; at time t, it is infected. Then it is infecting at $t + 1$ and not infecting at $t + 2$). The epidemic graph is generated in the same way as in Q3.c. Note that in this question, by construction, the transmitting edges are fixed forever (in the graph generating process) while in practice, it may happen that at time t a node infects a fraction T of its neighbors and at the next time step $t + 1$, it infects potentially a different fraction of its neighbors. However, we only focus on simulating the fixed transmitting edges case.

Using $n = 10000$, we visualize how the fraction of infecting nodes evolves with time in Figure 3. We can observe that when the epidemic does occur (when the starting infecting point is in the giant component), the fraction of infecting nodes first increases to a peak (about 0.27) and then decrease to zero.
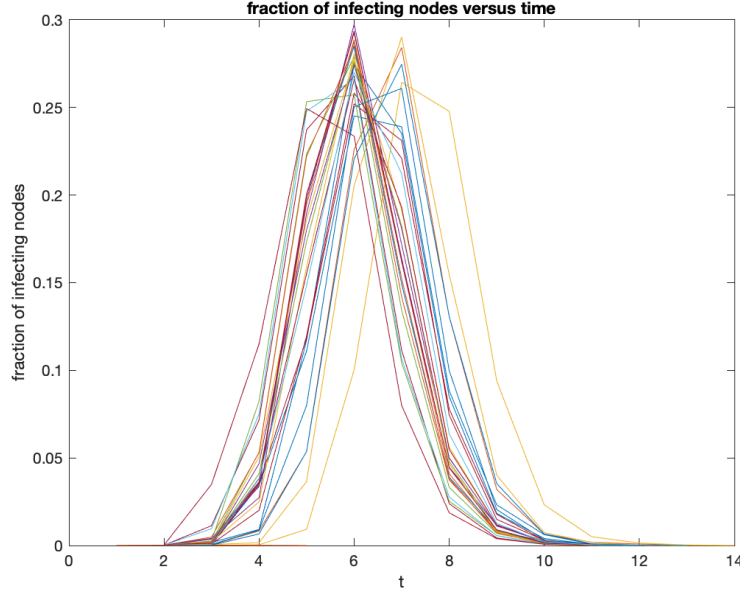

Figure 3: fraction of infecting nodes versus time t. There are 100 trajectories on the graph. 31 out of 100 times, the starting point is in the giant component and therefore there is an epidemic. Other 69 trajectories correspond to the case where the starting point is not in the giant component and therefore there is no epidemic, and these trajectories are not obvious in the figure since the fraction is almost zero and the duration is very short.

How the duration scales with n? Relation with BFS?   This SIR model is very similar to Breadth-first search (BFS). Indeed, SIR can be viewed as a disease exploring the graph (after keeping edges of the original graph with probability T) in a BFS manner, since at each step the disease transmit to neighbors that have never been infected (visited) before. We would like to know how the duration scales with the number of nodes in the graph. Using the intuition of BFS, which explores the graph as if it were a tree, we can predict that the duration should scale with $log(n)$, the height of a tree with n nodes. This is partially verified in Figure 4, where the simulations are run with varying n values. Note that only the case where the starting infecting node is in the giant component (so the epidemic occurs) is recorded.
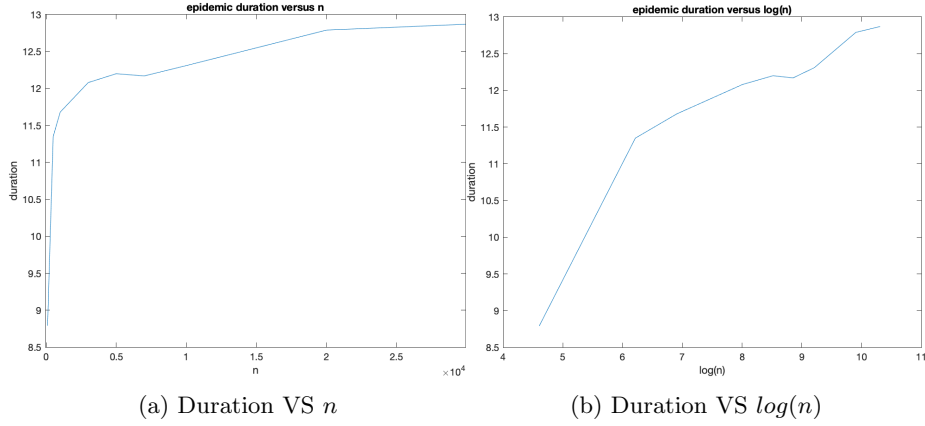
6

(a) Duration VS n (b) Duration VS $log(n)$

Figure 4: Epidemic duration versus n and $log(n)$. For each n value, the duration is averaged over 100 runs. We see that the duration roughly scales with $log(n)$.

Remark   In the problem we only consider static SIR model, where all the infections happen synchronously. However, in real life, the infection may happen asynchronously, with some of paths transmit the disease much faster than the others. In that case, the SIR model is kind of a mixture of BFS and DFS. Also, in this problem we assume that an infecting node will infect all of its neighbors by transmitting edges, while in practice this will not happen. Moreover, the transmitting edges will evolve over time (since population is dynamical), in stead of static. Nevertheless, the simplified model considered in this question conveys some useful information that is potentially useful in practice.