

**14.15**  
**NETWORKS**  
**PROBLEM SET 4**

JOHN WANG

Collaborators: Ryan Liu, Albert Wu

1. PROBLEM 1

**1.1. Problem 1.a. Problem:** Find the degree distribution of this model.

**Solution:** I will consider first the extra edges that are added to the ring lattice. In total, there are on average  $nkp$  shortcuts since each edge has probability  $p$  of obtaining a new edge. This means that there are  $2nkp$  endpoints of shortcuts, so that each vertex has on average  $2kp$  shortcuts connected to it.

Since the number of shortcuts is Poisson distributed, we can write the probability  $p_s$  of having  $s$  shortcut edges as  $p_s = e^{-2kp}(2kp)^s/s!$ . The total degree is given by  $d = s + k$ , since the degree is already  $k$  when starting with a ring lattice. This means that we can substitute  $s = d - k$  in the expression to obtain the degree distribution  $p_d = e^{-2kp}(2kp)^{d-k}/(d-k)!$ .  $\square$

**1.2. Problem 1.b. Problem:** Show that when  $p = 0$  the overall clustering coefficient is given by  $Cl(g) = \frac{3k-3}{4k-2}$ .

**Solution:** To compute the clustering coefficient, we must find the number of triangles and triples in the network. We shall start with the number of triangles. First, we know that when  $p = 0$ , we have a ring lattice and no other shortcut edges. Therefore, the only way to make a triangle for node  $i$  is to move 2 hops in one direction then a single hop in the backwards direction back to node  $i$ .

There are  $k$  possible vertices that one can make two hops in the forwards direction with. Therefore, there are  $\binom{k}{2}$  different ways of making two hops (by choosing two of these vertices for the hops). Once the two hops are chosen, the backwards edge is forced to be from the last hop back to vertex  $i$ . Therefore, for each vertex, there are  $\binom{k}{2}$  different ways of creating a triangle. Thus, there are  $n\binom{k}{2} = nk(k-1)/2$  triangles.

To find the number of triples, we note that for each vertex  $i$ , there are  $2k$  possible vertices that  $i$  can form a triple with. We need to select 2 other vertices to form a triple, and there are  $\binom{2k}{2}$  ways of doing this. Therefore, there are  $\binom{2k}{2}$  triples for each vertex  $i$ , and a total of  $n\binom{2k}{2} = 2nk(2k-1)/2 = nk(2k-1)$  triples. This gives a clustering coefficient of:

$$\begin{aligned}
 (1) \quad Cl(g) &= \frac{3\# \text{ triangles}}{\# \text{ triples}} \\
 (2) \quad &= \frac{3nk(k-1)/2}{nk(2k-1)} \\
 (3) \quad &= \frac{3(k-1)}{2(2k-1)} \\
 (4) \quad &= \frac{3k-3}{4k-2}
 \end{aligned}$$

$\square$

2. PROBLEM 2

**2.1. Problem 2.a. Problem:** Suppose that  $\alpha_i$  is drawn at random from some stationary distribution with a well-defined mean. Show that, in the limit of large  $n$ , the probability that the  $(n+1)$ st vertex added to the network attaches to a previous vertex  $i$  with in-degree  $q_i$  is  $c(q_i + \alpha_i)/n(c + \bar{\alpha})$ , where  $\bar{\alpha}$  is the average value of  $\alpha_i$ .

**Solution:** We know that after  $n$  vertices have been created in the graph, the probability of vertex  $i$  obtaining a new in-degree is proportional to  $q_i + \alpha_i$ . Since we normalize all the probabilities so that the total probability of obtaining any edge is given by 1, we must have  $\sum_j q_j + \alpha_j = 1$ . This means that the

probability of any given  $i$  obtaining a new vertex is  $\frac{q_i + \alpha_i}{\sum_j q_j + \alpha_j}$ . Since there are on average  $c$  new vertices added for the  $n + 1$ st vertex, the probability of attaching to a previous vertex is:

$$(5) \quad \frac{c(q_i + \alpha_i)}{\sum_j q_j + \alpha_j} = \frac{c(q_i + \alpha_i)}{\sum_j q_j + \sum_j \alpha_j}$$

$$(6) \quad = \frac{c(q_i + \alpha_i)}{nc + n\bar{a}}$$

$$(7) \quad = \frac{c(q_i + \alpha_i)}{n(c + \bar{a})}$$

Because  $\sum_j q_j = nc$  is the total degree of the graph and  $\sum_j \alpha_j = n\bar{a}$  is the sum of all the  $\alpha_i$ 's in the graph.  $\square$

**2.2. Problem 2.b. Problem:** Let  $p_q(\alpha)d\alpha$  be the probability that a vertex has degree  $q$  and an  $\alpha$  value in the range  $[\alpha, \alpha + d\alpha]$  in the limit of large network size. Derive master equations and solve them to get an explicit expression for  $p_q(\alpha)$ .

**Solution:** To obtain the master equations, we want to look at the number of vertices of degree  $q$  with  $\alpha$  in the range  $[\alpha, \alpha + d\alpha]$  when  $n + 1$  vertices occur, and related that to when there are  $n$  vertices. We know that there will be  $(n + 1)p_q(\alpha)d\alpha$  vertices with degree  $q$  when there are  $n + 1$  vertices. Now, we want to see the evolution from when there were  $n$  vertices.

First, we start with  $np_q(\alpha)d\alpha$  vertices of degree  $q$  when there are  $n$  vertices. A number of vertices will obtain one extra edge to transition from a degree of  $q - 1$  to  $q$  in the next stage. The number of these vertices is given by  $nc(q - 1 + \alpha)/(n(c + \bar{a}))p_{q-1}(\alpha)d\alpha = c(q - 1 + \alpha)/(c + \bar{a})p_{q-1}(\alpha)d\alpha$  based on the result from the previous problem. Also, a number of vertices will obtain one extra edge to transition from a degree of  $q$  to  $q + 1$ . The number of these vertices is  $nc(q + \alpha)/(n(c + \bar{a}))p_q(\alpha)d\alpha$ . Therefore, we have the following relation as the master equation:

$$(8) \quad (n + 1)p_q(\alpha)d\alpha = np_q(\alpha)d\alpha + \frac{c(q - 1 + \alpha)}{c + \bar{a}}p_{q-1}(\alpha)d\alpha - \frac{c(q + \alpha)}{c + \bar{a}}p_q(\alpha)d\alpha$$

$$(9) \quad p_q(\alpha) = \frac{c(q - 1 + \alpha)}{c + \bar{a}}p_{q-1}(\alpha) - \frac{c(q + \alpha)}{c + \bar{a}}p_q(\alpha)$$

$$(10) \quad = \frac{c(q - 1 + \alpha)}{c + \bar{a} + c(q + \alpha)}p_{q-1}(\alpha)$$

$$(11) \quad = \frac{q - 1 + \alpha}{1 + \bar{a}/c + q + \alpha}p_{q-1}(\alpha)$$

Solving the resulting recurrence (which can be done by computing  $p_q(\alpha)$  iteratively and checking for the pattern), we see that:

$$(12) \quad p_q(\alpha) = \frac{(q + \alpha - 1)(q + \alpha - 2) \dots \alpha}{(q + \alpha + 1 + \bar{a}/c) \dots (\alpha + 2 + \bar{a}/c)} \frac{1 + \bar{a}/c}{\alpha + 1 + \bar{a}/c}$$

We can now use the fact that  $\Gamma(x + 1) = x\Gamma(x)$  to derive the fact that  $\frac{\Gamma(x + n)}{\Gamma(x)} = (x + n - 1)(x + n - 2) \dots x$ . This means we can substitute into our expression using  $\Gamma(\cdot)$  and obtain:

$$(13) \quad p_q(\alpha) = \left(1 + \frac{\bar{a}}{c}\right) \frac{\Gamma(q + \alpha)\Gamma(\alpha + 1 + \bar{a}/c)}{\Gamma(\alpha)\Gamma(q + \alpha + 2 + \bar{a}/c)}$$

Using the fact that  $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}$  we can manipulate our expression to find that

$$(14) \quad p_q(\alpha) = \frac{B(q + \alpha, 2 + \bar{a}/c)}{B(\alpha, 1 + \bar{a}/c)}$$

$\square$

**2.3. Problem 2.c. Problem:** Hence show that the overall degree distribution of the network will have a power-law tail with exponent  $\alpha = 2 + \bar{a}/c$ .

**Solution:** Let us begin with our expression for the degree distribution  $p_q(\alpha) = \frac{B(q + \alpha, 2 + \bar{a}/c)}{B(\alpha, 1 + \bar{a}/c)}$ . We can use Stirling's approximation to note that  $B(x, y) \approx x^{-y}\Gamma(y)$  and see that

$$(15) \quad p_q(\alpha) \approx \frac{(q + \alpha)^{-(2 + \bar{a}/c)}\Gamma(2 + \bar{a}/c)}{\alpha^{-(1 + \bar{a}/c)}\Gamma(1 + \bar{a}/c)}$$

Since the entire denominator is a constant, we see that we have  $p_q(\alpha) \sim (q + \alpha)^{-(2+\bar{a}/c)}$ . The overall degree distribution is thus given by  $p_q \sim q^{-(2+\bar{a}/c)}$ , which has a power law tail with exponent  $2 + \bar{a}/c$ .  $\square$

### 3. PROBLEM 3

**Problem:**

**Solution:**  $\square$

### 4. PROBLEM 4

**4.1. Problem 4.a. Problem:** Find a threshold for the immunity probability (in terms of the moments of the degree distribution) below which the infection spreads to a large portion of the population.

**Solution:** First note that the infected population will form a single, connected component of the graph. This is because the disease begins at a specific vertex and spreads through neighboring edges. There is no way to be diseased without being in contact with another diseased person. Thus, we want to figure out what threshold leads to a component that is a constant portion of the size of the graph. Let us call the infected portion the giant cluster and let  $u$  be the probability that a vertex  $i$  is not connected to the giant cluster via a specific edge.

We see that there are two ways that a vertex can be not connected to the giant cluster. Either the vertex is immune (occurring with probability  $\pi$ ), or the vertex is not immune but the edge doesn't edge in the giant cluster. This occurs with probability  $(1 - \pi)u^k$  if the vertex has  $k$  edges. We also know that  $k$  is distributed according to the excess degree distribution  $q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$ . Thus, if we average over all  $k$ , we have:

$$(16) \quad u = \pi + (1 - \pi) \sum_{k=0}^{\infty} q_k u^k$$

Now a threshold occurs when the change in  $u$  is as large as possible, i.e. when its gradient is 1. We want to see this when  $u = 1$  (see the graph in Newman in figure 16.2 for reasoning). This means that we must have:

$$(17) \quad \frac{d(\pi + (1 - \pi) \sum_{k=0}^{\infty} q_k u^k)}{du} = 1$$

$$(18) \quad (1 - \pi) \frac{d}{du} \sum_{k=0}^{\infty} u^k \frac{(k+1)p_{k+1}}{\langle k \rangle} = 1$$

$$(19) \quad (1 - \pi) \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k u^{k-1} (k+1) p_{k+1} = 1$$

$$(20) \quad (1 - \pi) \frac{1}{\langle k \rangle} \sum_{k=0}^{\infty} k(k-1) p_k = 1$$

$$(21) \quad (1 - \pi) \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} = 1$$

Solving for the threshold value of  $\pi$ , we obtain:

$$(22) \quad \pi = \frac{\langle k^2 \rangle - 2\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

$\square$

**4.2. Problem 4.b. Problem:** What is this threshold for a  $k$ -regular random graph, i.e., a configuration model network in which all nodes have the same degree.

**Solution:** For a  $k$ -regular random graph, we have the degree distribution:

$$(23) \quad p_d = \begin{cases} 1 & \text{if } d = k \\ 0 & \text{else} \end{cases}$$

Therefore, we can compute  $\langle k \rangle = \sum_{d=0}^{\infty} d p_d = k$  and  $\langle k^2 \rangle = \sum_{d=0}^{\infty} d^2 p_d = k^2$  relatively easily. Substituting this into our expression for the threshold in the configuration model, we see that our threshold for a  $k$ -regular graph is:

$$(24) \quad u = \frac{k^2 - 2k}{k^2 - k} = \frac{k - 2}{k - 1}$$

$\square$

**4.3. Problem 4.c. Problem:** What is this threshold for a power law graph with exponents less than 3, i.e.,  $p_k \sim k^{-\alpha}$  with  $\alpha < 3$ ? The Internet graph (representing connections between routers) has a power law distribution with exponent  $\sim 2.1 - 2.7$ . What does this result imply for the Internet graph?

**Solution:** We have the following definition:  $\langle k \rangle = \sum_{k=1}^{\infty} k p_k \sim \sum_{k=1}^{\infty} k k^{-\alpha}$ . Thus, we can use this to compute the threshold value of  $\pi$  from the value we calculated in part a:

$$(25) \quad \pi = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n k^2 k^{-\alpha} - 2k k^{-\alpha}}{\sum_{k=1}^n k^2 k^{-\alpha} - k^{-\alpha}}$$

$$(26) \quad = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n k k^{-\alpha} (k - 2)}{k k^{-\alpha} (k - 1)}$$

$$(27) \quad = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n (k - 2)}{\sum_{k=1}^n (k - 1)}$$

$$(28) \quad = 1$$

Thus, we see that the threshold for a power law graph is simply 1. Thus, all graphs which have  $\pi < 1$  will lead to a giant cluster of contagion which has a number of vertices which is  $\Theta(n)$ . This implies that the Internet graph will always obtain contagions of disease unless everyone in the network is immune.  $\square$

**4.4. Problem 4.d. Problem:** Find the size of the infected population (you can assume that the infection spreads to a large portion of the population).

**Solution:** Let  $S$  denote the proportion of the infected population. We know that for a vertex to not carry that pathogen, it must be the case that none of its edges are pathogen carrying. This occurs with probability  $u^k$  if the vertex has  $k$  edges. Thus, the probability that a vertex does not carry the pathogen can be obtained by using the degree distribution:  $\sum_{k=1}^{\infty} p_k u^k$ . Therefore, the proportion of the infected population is just the inverse of this:  $S = 1 - \sum_{k=1}^{\infty} p_k u^k$ .

This means that the size of the infected population is just  $nS$  or  $n(1 - \sum_{k=1}^{\infty} p_k u^k)$ .  $\square$