JOHN WANG

Collaborators: Ryan Liu, Bonny Jain

1. PROBLEM 1

1.1. **Problem 1.a. Problem:** Let A_1 denote the event that node 1 has at least $l \in \mathbb{Z}^+$ neighbhors. Do we observe a phase transition for this event? If so, finnd the threshold function and explain your reasoning.

Solution: Yes, we do observe a phase transition when $t(n) = \frac{l}{n-1}$. To show that there exists a phase transition, lut us examine the case when $\frac{p(n)}{t(n)} = \frac{p(n)(n-1)}{l} \to 0$. In this case we can use Markov's inequality. We define X as a random variable of the number of neighbors of node 1 and we note that $P(A_1) = P(X \geq l)$. Thus, we want to obtain the expected value of X. However, we know that a node has an expected $(n-1)p(n)$ neighbors, because each neighbor has a $p(n)$ probability of being connected by an edge, and there are $n-1$ neighbors. Thus, we see that $E[A_1] = (n-1)p(n)$. Markov's inequality allows us to obtain the following bound:

$$P(A_1) = P(X \geq l) \leq \frac{E[A_1]}{l} = \frac{p(n)(n-1)}{l} \tag{1}$$

However, since $\frac{p(n)}{t(n)} = \frac{p(n)(n-1)}{l} \to 0$, we see that $P(A_1) \to 0$ as well. Thus, we have shown that the first part of the phase transition, namely that under the threshold, A_1 does not occur. Now we shall show that above the threshold, the event occurs almost surely. We want to see what happens when $\frac{p(n)}{t(n)} = \frac{p(n)(n-1)}{l} \to \infty$. We shall use Chebyshev's inequality:

$$P(|X - E[X]| \geq |E[X] - l|) \quad \leq \quad \frac{\mathrm{Var}(X)}{(E[x] - l)^2} \tag{2}$$

$$= \quad \frac{p(n-1)}{((n-1)p - l)^2} \tag{3}$$

$$= \quad \frac{p(n-1)}{(n-1)^2 p l^2 + l^2 - 2(n-1)pl} \tag{4}$$

$$= \quad \frac{1}{(n-1)l^2 + \frac{l^2}{(n-1)p} - 2l} \tag{5}$$

Where we know that $\mathrm{Var}(X) = E[X] = (n-1)p$ because X can be approximated as a poisson random variable (see Newman p. 402) for large n. Moreover, we know that l is a constant so that as $\frac{p(n)(n-1)}{l} \to \infty$ we have $\frac{l^2}{(n-1)p} \to 0$. Therefore, we also see that $P(|X - E[X]| \geq |E[X] - l|) \leq \frac{1}{(n-1)l^2 - 2l}$ as $\frac{p(n-1)}{l} \to \infty$. Since l is a constant we see that $n \to \infty$ implies that this probability goes to zero. In other words, the probability that X deviates by more than $E[X] - l$ from its expected value goes to zero. This shows that $P(A_1) \to 1$ when $E[X] \geq l$. $\square$

1.2. **Problem 1.b. Problem:** Let B denote the event that a cycle with k edges (for a fixed k) emerges in the graph. Do we observe a phase transition of this event? If so, find the threshold function and explain your reasoning.

Solution: Yes a threshold function does exist for this event. Take $t(n) = \frac{1}{n}$ as the threshold function. We will first show that as $\frac{p}{t(n)} = pn \to 0$, then $P(B) \to 0$. We first want to find the expected number of cycles of length k in the graph. Let X denote the number of cycles of length k. We know that there are $\binom{n}{k}$ ways to select k nodes, and that for each of these subsets of nodes, there are $(k-1)!/2$ ways to create a cycle. This follows because we set the first node of the cycle, then there are $k-1$ ways of picking the second node, $k-2$ ways of picking the third node, etc. We divide by two because we could go either backwards or

forwards in this cycle (clockwise or counterclockwise). Each of these cycles has probability p^k of emerging. Therefore, we have:

$$
\begin{align}
(6) \qquad E[X] &= \binom{n}{k}\frac{(k-1)!}{2}p^k \\
(7) \qquad &= \frac{n!}{k!(n-k)!}\frac{(k-1)!}{2}p^k \\
(8) \qquad &= \frac{n!}{2k(n-k)!}p^k
\end{align}
$$

However, we know that $\frac{n!}{(n-k)!} = n(n-1)\ldots(n-k+1) \leq n^k$. This means that we have $E[X] \leq \frac{(np)^k}{2k}$. Thus, as $np \to 0$, we see that $E[X] \to 0$, which implies that $P(X \geq 1) \leq \frac{E[X]}{1} = E[X] \to 0$. Since $P(X \geq 1) = P(B)$, we see that $P(B) \to 0$ as $\frac{p}{t(n)} \to 0$. Now to show the second half of the phase transition, we need to show that as $\frac{p}{t(n)} \to \infty$ we have $P(B) \to 1$.

To show this we note that $P(X \leq 0) = P(E[X] - X \geq E[X]) \leq \frac{\text{Var}(X)}{E[X]^2}$ by Chebyshev. We can also bound $E[X]$ from below by using the fact that $\frac{n!}{(n-k)!} = n(n-1)\ldots(n-k+1) \geq (n-k)^k$. This shows that $E[X] \geq \frac{(n-k)^k p^k}{2k}$. Since k is a constant we know that $nk - kp \to \infty$ when $np \to \infty$. Thus, we see that $(n-k) \to \infty$, which implies that $E[X] \geq \infty$ when $np \to \infty$.

Now we use the fact that X can be approximated as a poission distribution to note the fact that $\text{Var}(X) = E[X]$. This means that $P(X \leq 0) \leq \frac{1}{E[X]} \to 0$. This implies that $P(X \geq 0) \to \infty$, which means that $P(B) \to \infty$, just as we wanted. $\square$

2. Problem 2

2.1. Problem 2.a. Problem: Show that the mean degree of a vertex in this network is $2c$.

Solution: We know that the expected number of connected trios is the total possible number of triples of nodes times the probability each triple becomes a triangle. This is $\binom{n}{3}\frac{c}{\binom{n-1}{2}} = \frac{n!}{3!(n-3)!}\frac{c}{(n-1)!}(n-1-2)!2! = \frac{2!nc}{3!} = \frac{nc}{3}$. Since we know that the total number of edges is just 3 times the number of triangles we have nc total expected edges. Moreover, the expected total degree is 2 times the total number of expected edges (since each edge has two endpoints). This means we expect $2nc$ total degree in the graph, and since there are n nodes, each node has an expected degree of $\frac{2nc}{n} = 2c$. $\square$

2.2. Problem 2.b. Problem: Show that the degree distribution is

$$
(9) \qquad p_k = \left\{ \begin{array}{ll} e^{-c}c^{k/2}/(k/2)! & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{array} \right\}
$$

Solution: We can assume that the degree distribution is a poisson random variable because of the fact that each triangle is selected randomly with a particular probability. This is a binomial distribution, which converges to a poission distribution as n grows large. Therefore, we only need to find the expected degree of a node. The probability that a node has degree k is really equal to the probability that it is inside of $k/2$ triangles (because each triangle provide two edges). The expected value is given by the number of triangles that a node can connect to times the probability of occuring. This is just $\binom{n-1}{2}\frac{c}{\binom{n-1}{2}}$.

Therefore, we see that $\lambda = c$ in this poisson distribution and that there are $m = k/2$ different triangle possibilities. We therefore have the probability distribution of a possion random variable $\frac{\lambda^m e^{-\lambda}}{m!}$ which we can substitute $\lambda = c$ and $m = k/2$ to obtain $\frac{c^{k/2}e^{-c}}{(k/2)!}$. This is the degree distribution for when k is even, because k cannot be odd. This is because whenever a new triangle is added, 2 more edges are added to each node, and thus, one cannot have an odd degree. Therefore, we have shown that the degree distribution is given by:

$$
(10) \qquad p_k = \left\{ \begin{array}{ll} e^{-c}c^{k/2}/(k/2)! & \text{if } k \text{ is even} \\ 0 & \text{if } k \text{ is odd} \end{array} \right\}
$$

$\square$

2.3. **Problem 2.c. Problem:** Show that the clustering coefficient is $C = \frac{1}{2c+1}$.

Solution: We want to figure out the number of triangles, as well as the number of triples, that we expect to appear in this graph model. We will find the clustering coefficient in the case where multiedges cannot occur, then show that multiedges occur with low enough probability that this approximation is valid. Now consider the case where the graph is simple.

First, we shall find the expected number of triangles. This is just the probability that three vertices i, j, k end up forming a triangle. This could happen in two ways. The first is that the nodes form a trio and this trio becomes a triangle, this happens with probability $p = \frac{c}{\binom{n-1}{2}}$. The other possibility is that the nodes do not form a trio, but that three edges form from other triangles. This happens if $\{i, j\}, \{j, k\}$, and $\{k, i\}$ each are part of a new triangle. This surrounds the area between i, j, k and forms a triangle. Each edge appears with probability $p(n-3)$. This is because for each pair of nodes, there are $n-3$ other nodes that this pair can form a triangle with (the pair cannot form a triangle with the last node in the trio because then we would be in the first case). Forming these triangles is an independent event, so we have $(p(n-3))^3$ probability of having our trio surrounded by triangles. The total probability of the trio being a triangle is thus $p + (1-p)(p(n-3))^3$ because in the second case, we cannot have the trio form a triangle.

Now, let us examine how many triples we obtain. Again consider three vertices i, j, k. When i, j, k form a triangle then 3 triples are formed. Otherwise, a triple is created whenever $\{i, j\}$ and $\{j, k\}$ each are part of a triangle. Each pair of nodes has $(n-3)$ other nodes to form a triangle with, and the event of forming a triangle is independent. So the probability that the $\{i, j\}$ and $\{j, k\}$ edges form is $(p(n-3))^2$. Now the probability that only these two triangle form is $(1-p)(p(n-3))^2$ because otherwise the original i, j, k trio would form a triangle. There are 3 different ways that one can select two edges to create. Therefore, the probability of a triple is just 3 times the probability of forming a triangle plus 3 times the probability of $\{i, j\}, \{j, k\}$ forming a triple. This is $3(p + (1-p)(p(n-3))^3 + (1-p)(p(n-3))^2)$.

The clustering coefficient is therefore:

$$(11) \qquad \frac{3(p + (1-p)(p(n-3))^3)}{3(p + (1-p)(p(n-3))^3 + (1-p)(p(n-3))^2)} = \frac{\frac{p}{(p(n-3))^2} + (1-p)(p(n-3))}{\frac{p}{(p(n-3))^2} + (1-p)(p(n-3)) + (1-p)}$$

We can take the limit as $n \to \infty$ and note that $p = \frac{c}{\binom{n-1}{2}} = \frac{2c(n-3)!}{(n-1)!} = \frac{2c}{(n-1)(n-2)} \approx \frac{2c}{n^2}$ to obtain the following expression:

$$(12) \qquad Cl(G) = \frac{\frac{1}{pn^2} + pn}{\frac{1}{pn^2} + pn + 1 - \frac{2c}{n^2}}$$

$$(13) \qquad = \frac{\frac{1}{2c}}{\frac{1}{2c} + 1}$$

$$(14) \qquad = \frac{1}{2c + 1}$$

Now we want to show that multiedges do not happen with high probability. In particular, I will show that the ratio of multiedges to single edges goes to zero, which means that our above analysis is correct as $n \to \infty$. First, we want to figure out the probability of a k-edge for two nodes i, j. There are $n-2$ total other nodes that these nodes can form triangles with. Thus, to have a k-edge, there must be exactly k triangles which are formed between i, j and k other nodes. Thus, there are $\binom{n-2}{k}$ ways to choose these k other nodes. Each of these choices of a group of nodes has probability $p^k(1-p)^{n-2-k}$ of forming a k-edge. Thus, the total probability of forming a k edge on nodes i, j is $\binom{n-2}{k}p^k(1-p)^{n-2-k}$. The ratio R of multiedges to single edges is:

$$(15) \qquad R = \frac{\binom{n-2}{k}p^k(1-p)^{n-2-k}}{\binom{n-2}{1}p(1-p)^{n-3}}$$

$$(16) \qquad = \frac{\frac{(n-2)!}{(n-2-k)!k!}p^k(1-p)^{n-2-k}}{(n-2)p(1-p)^{n-3}}$$

$$(17) \qquad = \frac{\frac{(n-3)\ldots(n-1-k)}{k!}p^{k-1}}{1-p)^{k+1}}$$

$$(18) \qquad \leq \frac{n^k p^{k-1}}{k!(1-p)^{k+1}}$$

Now, as $n \to \infty$ we know that $p = \frac{2c}{n^2}$ so that we have the following:

$$\text{(19)} \qquad R \ \leq \ \frac{n^k (2c)^{k-1}}{k! n^{2k}}$$

$$\text{(20)} \qquad = \ \frac{(2c)^{k-1}}{n^2 k!}$$

(21)

Which goes to zero. Therefore, we see that the probability of single edges is much higher than the probability of k-edges for $k > 1$, so much so, that the probability goes to zero for all $k > 1$. Thus, we can ignore multiedges in our analysis and our previous analysis of the clustering coefficient is sound. $\square$

2.4. **Problem 2.d. Problem:** Show that when there is a giant component in the network, its expected size S, as a fraction of the network size, satisfies $S = 1 - e^{-cS(2-S)}$.

Solution: Suppose that u_i is the probability that node i does not belong to the giant component. For i to not belong in the giant component we must either have 1) that i is not connected to some triangle j or 2) that i is connected to the triangle j but j itself is not part of the giant component. The first occurs with probability $1 - p$, while the second occurs with probability pu_j^2, since both nodes of the connecting triangle cannot belong to the giant component. This means the probaiblity that i is not connected to the giant component through traingle j is $1 - p + pu_j^2$.

Since each node is independent and triangles are randomly selected, we can say $u_i = u_j = u$. Since there are $\binom{n-1}{2}$ triangles for which we must do this analysis for, we see the following:

$$\text{(22)} \qquad u \ = \ (1 - p + pu^2)^{\binom{n-1}{2}}$$

$$\text{(23)} \qquad = \ \left(1 - \frac{c}{\binom{n-1}{2}} + \frac{c}{\binom{n-1}{2}} u^2 \right)^{\binom{n-1}{2}}$$

Taking logs of both sides we see that we can simplify the expression:

$$\text{(24)} \qquad \ln u \ = \ \binom{n-1}{2} \ln \left(1 - \frac{c}{\binom{n-1}{2}} + \frac{c}{\binom{n-1}{2}} u^2 \right)$$

$$\text{(25)} \qquad = \ \binom{n-1}{2} \frac{cu^2 - u}{\binom{n-1}{2}}$$

$$\text{(26)} \qquad = \ cu^2 - u$$

Where we have used the fact that $\ln(1 + x) = x$ for small x. Exponentiating both sides of the resulting expression gives us that $u = e^{cu^2 - u}$. However, we know that $S = 1 - u$ because u is the probability of any particular node not belong to the giant component, and S is the probability of belonging to the giant component. Therefore, we see that $1 - S = e^{-c + c(1-S)^2}$ which simplifies to $S = 1 - e^{-cS(2-s)}$, which proves our result. $\square$

2.5. **Problem 2.e. Problem:** What is the value of the clustering coefficient when the giant component fills half of the network?

Solution: We choose $S = 1/2$ and we solve for c. We have:

$$\text{(27)} \qquad \frac{1}{2} \ = \ 1 - e^{\frac{1}{2} c \left(\frac{1}{2} - 2\right)}$$

$$\text{(28)} \qquad \frac{1}{2} \ = \ 1 - e^{-\frac{3}{4} c}$$

$$\text{(29)} \qquad e^{-\frac{3}{4} c} \ = \ \frac{1}{2}$$

$$\text{(30)} \qquad -\frac{3}{4} \ = \ -\ln(2)$$

$$\text{(31)} \qquad c = \frac{4}{3} \ln(2)$$

Now, we substitute into our equation for the clustering coefficient of $C = \frac{1}{2c+1}$ so we obtain $C = \frac{1}{8\ln(2)/3 + 1}$.
$\square$

3. Problem 3

Problem: A k-regular graphs (i.e. a graph where all nodes have degree k).
Solution: $\square$
Problem:
Solution: $\square$
Problem:
Solution: $\square$

4. Problem 4

Problem: Consider a graph G with n nodes generated according to the configuration model with a particular degree distribution $P(d)$. Show that the overall clustering coefficient is given by $Cl(G) = \frac{\langle d \rangle}{n} \left[\frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle^2} \right]^2$.

Solution: We note that the clustering coefficient Cl is the average probability that two neighbors of a vertex are neighbor of each other as wel. Thus, we are looking for the probability that if i and j are connected, given that they are connected to v. First, we know that the probability of an edge between nodes i and j is given by $\frac{k_i k_j}{2m}$, were k_i is the degree of node i. This is because there are k_j edges coming into node j, so there is a $k_j/2m$ probability that any edge goes to j. Since there are k_i edges leaving node i, there are k_i times the probability of any given edge going to j. Therefore, we see that the probability of an edge existing between nodes i and j is $\frac{k_i k_j}{2m}$.

Next, we notice that the excess degree distribution for node i can be given by $q_k = \frac{k+1}{k} p_{k+1}$. This can be seen from the derivation in Newman (Equation 13.46). Thus, the probability that any node i, which is already connected to some node v, has degree of k is given by q_k. Thus, we can write the clustering coefficient as the sum over all possible degrees k_i, k_j for nodes i and j of the probability of obtaining k_i and k_j times the probability of an edge between nodes i and j. This is given by:

$$\text{(32)} \qquad Cl(G) \;=\; \sum_{k_i=0}^{\infty} \sum_{k_j=0}^{\infty} q_{k_i} q_{k_j} \frac{k_i k_j}{2m}$$

$$\text{(33)} \qquad =\; \frac{1}{2m} \left(\sum_{k=0}^{\infty} k q_k \right)^2$$

$$\text{(34)} \qquad =\; \frac{1}{2m} \left(\sum_{k=0}^{\infty} (k+1) p_{k+1} \right)^2$$

This was obtaining by noting that $q_k = \frac{k+1}{k} p_{k+1}$. Next, we notice that if we multiply by $\frac{\langle k \rangle^2}{\langle k \rangle^2} = 1$ and make a change of variable for $k = k+1$, we obtain the following expression:

$$\text{(35)} \qquad Cl(G) \;=\; \frac{1}{2m \langle k \rangle^2} \left(\sum_{k=0}^{\infty} k(k+1) p_{k+1} \right)^2$$

$$\text{(36)} \qquad =\; \frac{1}{2m \langle k \rangle^2} \left(\sum_{k=0}^{\infty} (k-1) k p_k \right)^2$$

Since $2m$ is the total number of edges in the graph, we know that $2m = n \langle k \rangle$ so we can substitute in the denominator of the fraction and expand $(k-1)k p_k = k^2 p_k - k p_k$ to obtain:

$$\text{(37)} \qquad Cl(G) \;=\; \frac{1}{n \langle k \rangle^3} \left(\sum_{k=0}^{\infty} k^2 p_k - k p_k \right)^2$$

$$\text{(38)} \qquad =\; \frac{1}{n \langle k \rangle^3} \left(\langle k^2 \rangle - \langle k \rangle \right)^2$$

Now, set $d = k$ and we see that $Cl(G) = \frac{\langle d \rangle}{n} \left[\frac{\langle d^2 \rangle - \langle d \rangle}{\langle d \rangle^2} \right]^2$, which is what we wanted. $\square$