Note: We are including extra bonus questions to let students work on types of prob-lems that interest them more. There is no expectation that you do all the bonus problems.

Problem 1 (Giant Component)**.** Let $p(n) = \lambda/n$ for all n: that is, expected degree is held fixed at λ.

(a) Suppose that as $n \to \infty$ there is a giant component that fills exactly half the network. What is λ?

(b) For the same random graph, what is the probability that a node has degree exactly 5?

(c) Calculate the fraction of nodes in the giant component that have degree exactly 5. [Hint: for any node i, by Bayes' rule, this equals

$$\frac{\Pr(d_i = 5)\Pr(i \text{ in giant component}|d_i = 5)}{\Pr(i \text{ in giant component})}.$$

You should be able to compute all of these terms.]

(d) Give an intuitive explanation for the difference between the answers to parts (b) and (c).

Solution.

(a) Recall from lecture

$$q = 1 - e^{-\lambda q}$$

where q is the fraction of nodes in the giant component. Then

$$0.5 = 1 - e^{0.5\lambda} \iff e^{0.5\lambda} = 0.5 \iff \lambda = -\frac{ln(0.5)}{0.5} \approx 1.3863.$$

(b) The degree distribution of a node i in the ER model converges to a Poisson random variable d_i with parameter λ, which from part a) $\lambda \approx 1.3863$.

$$\mathbb{P}(D = 5) \approx \frac{e^{-1.3863}(1.3863)^5}{5!} = 0.0107.$$

(c) From the problem statement $\mathbb{P}(i \text{ in giant component})$, and from b) we have $\mathbb{P}(d_i = 5) = 0.0107$. Now to think about $\gamma := \mathbb{P}(i \text{in giant component} | d_i = 5)$. Now, $\gamma = 1 - \mathbb{P}(\text{no neighbor of i in giant component} | d_i = 5)$. We can think of i being the last node added. At this time $\frac{n-1}{2}$ nodes are in the giant component. We can think of choosing i's neighbors as picking 5 nodes from the $n-1$ nodes without replacement. When n is large we do not need to consider the without replacement issue. So $\mathbb{P}(\text{no neighbor of i in giant component} | d_i = 5) = \left(\frac{1}{2}\right)^5$. So the overall probability is given by

$$\frac{0.0107\left(1 - \left(\frac{1}{2}\right)^5\right)}{1/2} \approx 0.0207$$

(d) In part b) we calculate the proportion of nodes that have degree 5 in the whole graph, whereas in part c) we calculate the proportion of nodes that have degree 5 in the giant component. It is much more likely that a node has degree five in the giant component than it is just any node has degree 5.

Problem 2 (Configuration Model)**.** Consider the configuration model with degree distribution $P(d) = 2^{-(d+1)}$ for all $d \geq 0$.

(a) Show that the degree distribution is correctly normalized, meaning that $\sum_{d=0}^{\infty} P(d) = 1$.

(b) What is the average degree of a node?

(c) What is the average number of distance-2 neighbors of a node?

(d) Does the network have a giant component? Why or why not?

Solution.

(a) We'll prove by induction that

$$\sum_{d=0}^{T} 2^{-(d+1)} = 1 - 2^{-(T+1)}. \tag{1}$$

For $T = 0$ we have $2^{-1} = 1 - 2^{-1}$ as needed. Assuming (1) holds for $T - 1$, we have that

$$\begin{aligned} \sum_{d=0}^{T} 2^{-(d+1)} &= \sum_{d=0}^{T-1} 2^{-(d+1)} + 2^{-(T+1)} \\ &= 1 - 2^{-T} + 2^{-(T+1)} \\ &= 1 - 2^{-(T+1)}. \end{aligned}$$

Thus

$$\sum_{d=0}^{\infty} 2^{-(d+1)} = \lim_{T\to\infty} \left\{1 - 2^{-(T+1)}\right\} = 1.$$

(b) We can compute that the average degree $\langle d \rangle$ satisfies

$$\begin{aligned} \langle d \rangle = \sum_{d=0}^{\infty} d2^{-(d+1)} &= \sum_{d=1}^{\infty} d2^{-(d+1)} \\ &= \sum_{d'=0}^{\infty} (d'+1)2^{-(d'+2)} \qquad (d' = d-1) \\ &= \frac{1}{2}\left[\sum_{d'=0}^{\infty} d'2^{-(d'+1)} + \sum_{d'=0}^{\infty} 2^{-(d'+1)}\right] \\ &= \frac{1}{2}\left(\langle d \rangle + 1\right) \end{aligned}$$

using our result from part a. in the last step. We can then solve to obtain $\langle d \rangle = 1$.

(c) We know from lecture (via the branching approximation) that the expected number of distance-two neighbors is given by $\langle d^2 \rangle - \langle d \rangle$. We can then similarly compute

$$\begin{aligned}\langle d^2 \rangle = \sum_{d=0}^{\infty} d^2 2^{-(d+1)} &= \sum_{d=1}^{\infty} d^2 2^{-(d+1)} \\ &= \sum_{d'=0}^{\infty} (d'+1)^2 2^{-(d'+2)} \\ &= \frac{1}{2}\left[\sum_{d'=0}^{\infty} (d'^2 + 2d' + 1) 2^{-(d'+1)}\right] \\ &= \frac{1}{2}\Big(\langle d^2 \rangle + 2\langle d \rangle + 1\Big)\end{aligned}$$

Plugging in our value for $\langle d \rangle$ from part b. and solving gives $\langle d^2 \rangle = 3$.

(d) Since $3 = \langle d^2 \rangle / \langle d \rangle > 2$, we will have a giant component.

Problem 3 (Small World Model)**.** Consider a ring network with n nodes in which each node is connected to its neighbors k steps or less away. There are two popular variants of the "small world" model:

Edge-adding For each pair of nodes that are not linked in this network, add a new edge between them with probability p/n, independently across pairs.

Edge-rewiring For each edge (i, j), with independent probability p, replace this edge with an edge chosen uniformly at random from the set of edges not present in the graph.

(a) Find the degree distribution of the edge adding model. (It suffices to find the asymptotic degree distribution for a given node.)

(b) Show that when $p = 0$, the overall clustering coefficient in both models is given by

$$\mathrm{Cl}(g) = \frac{3k-3}{4k-2}.$$

(c) (*Bonus*-3 points) Show that when $p > 0$, the overall clustering coefficient in the *edge rewiring* model satisfies

$$\frac{3k-3}{4k}(1-p)^3 \leq \mathrm{Cl}(g) \leq \frac{3k-3}{4k-2}(1-p)^3.$$

(d) (*Bonus*-3 points) Write a program to generate small world networks according to the edge adding model with $n = 100$, $k = 5$, and $p = 0.1$. Compute the realized overall clustering coefficient and see if it obeys the bounds for the edge rewiring model from part (c).

Solution.

(a) We will first fix a vertex v and compute the limit

$$p^*(d) = \lim_{n\to\infty} \mathbb{P}_{p/n}(d_v = d).$$

Then, to illustrate how this corresponds to a statement about the (random) degree distribution the graph, we will prove that as $n \to \infty$, the *proportion* of vertices with degree d converges in probability to $p^*(d)$.

For the first statement, we know that each vertex v has degree $2k$ before any adges are added, and to this we add $n - 2k - 1$ edges independently with probability p/n. As $n \to \infty$ the probability that the number of additional edges is ℓ is given by

$$\begin{aligned}
\mathbb{P}_{p/n}(d_v - 2k = \ell) &= \binom{n-2k-1}{\ell}\left(\frac{p}{n}\right)^{\ell}\left(1-\frac{p}{n}\right)^{n-2k-1-\ell} \\
&= \underbrace{\left(\frac{(n-2k-1)(n-2k-2)\cdots(n-2k-1-\ell)}{n^{\ell}}\right)}_{\text{converges to } 1}\left(\frac{p^{\ell}}{\ell!}\right)\underbrace{\left(1-\frac{p}{n}\right)^{n-2k-1-\ell}}_{\text{converges to } e^{-p}} \\
&\to \frac{p^{\ell}e^{-p}}{\ell!} = \text{Poisson}(p,\ell)
\end{aligned}$$

This is called the "Poisson limit theorem." Thus, the degree of each vertex has asymptotic distribution $p^*(d) = \text{Poisson}(p, d-2k)$.

To show the asymptotic proportion of vertices with degree d is exactly $p^*(d)$, let N_d denote the number of vertices with degree d. We know that $\mathbb{E}_{p/n}[N_d]/n \to p^*(d)$ by the above. Next, we have that

$$\begin{aligned}
\mathbb{P}_{p/n}\left(|N_d - \mathbb{E}_{p/n}[N_d]| \geq \epsilon n\right) &= \mathbb{P}_{p/n}\left(|N_d - \mathbb{E}_{p/n}[N_d]|^2 \geq \epsilon^2 n^2\right) \\
&\leq \mathbb{E}_{p/n}\left[|N_d - \mathbb{E}_{p/n}[N_d]|^2/(\epsilon^2 n^2)\right] \quad \text{(Markov's inequality)} \\
&= \frac{\mathbb{E}_{p/n}[N_d^2]}{\epsilon^2 n^2} - \frac{\mathbb{E}_{p/n}[N_d]^2}{\epsilon^2 n^2} \to \frac{1}{\epsilon^2}\left(\frac{\mathbb{E}_{p/n}[N_d^2]}{n^2} - p^*(d)^2\right).
\end{aligned}$$

This is called "Chebyshev's inequality," and it tells us that the probability of N_d/n differing from $p^*(n)$ by more than ϵ will go to zero if $\mathbb{E}_{p/n}[N_d^2] = p^*(d)^2 n^2 + o(n^2)$. To verify this, let D_i be 1 when node i has degree d, and note that

$$\mathbb{E}_{p/n}[N_d^2] = \mathbb{E}\left[\sum_{i,j=1}^{n} D_i D_j\right] = \sum_{i,j=1}^{n} \mathbb{E}[D_i D_j].$$

If we can show $\mathbb{E}[D_i D_j] = p^*(d)^2 + o(1)$ then the sum will be equal to $n^2 p^*(d)^2 + o(n^2)$ as needed. The only dependence between D_i and D_j comes from whether there is an (i,j) edge or not. Conditional on there *being* an (i,j) edge, the degrees of both nodes are independent with distribution $\text{Poisson}(p, d-2k-1) = p^*(d-1)$. Conditional on there *not being* an (i,j) edge, the degrees are independent with distribution $p^*(d)$. Thus, if E_{ij} is the event that there is an

(i, j) edge, we have

$$\begin{aligned}\mathbb{E}[D_iD_j] &= \mathbb{E}[D_iD_j|E_{ij}]\mathbb{P}(E_{ij}) + \mathbb{E}[D_iD_j|\bar{E}_{ij}]\mathbb{P}(\bar{E}_{ij}) \\ &= \Big(p^*(d)+1\Big)^2 \frac{p}{n} + \Big(1-\frac{p}{n}\Big) p^*(d)^2 = p^*(d)^2 + \frac{p(2p^*(d)+1)}{n} \\ &= p^*(d)^2 + o(1).\end{aligned}$$

We conclude that for each d, the proportion of nodes with degree d converges in probability to $p^*(d)$ as $n \to \infty$.

(b) Fix a node v. Let's start by counting the number of open triplets centered at v, i.e. triples of distinct nodes (a, v, b) with edges (a, v) and (v, b).

There are $2k$ nodes a that are connected to v. For each choice of a there are $2k-1$ nodes other than a that are connected to v. So the total number of open triplets centered at v is $2k(2k-1)$.

To go from this number to the total number of open triplets, note that (1) by symmetry of the graph there are $2k(2k-1)$ open triplets centered at each node v, and (2) each open triplet is centered at exactly one node. So the total number of open triplets is exactly $n \cdot 2k(2k-1)$.

Now let's count the number of closed triplets centered at v, i.e. triples of nodes (u, v, w) where every pair has an edge.

In order to do this, note that for each $1 \le \ell \le k$ there are exactly 2 nodes at distance ℓ from k. If u has distance ℓ from v, the number of nodes that are connected to both u and v is $2k-\ell-1$: $k-1$ nodes on the same side of v as u and $k-\ell$ nodes on the oppisite side of v. Putting this together, we have

$$\begin{aligned}\sum_{\ell=1}^{k} 2(2k-\ell-1) &= 4k^2 - 2k - \sum_{\ell=1}^{k} \ell \\ &= 4k^2 - 2k - k(k+1) = k(3k-3)\end{aligned}$$

Again, since each closed triple is centered at exactly one node and there are $k(3k-3)$ closed triples centered at each node, the total number is $n \cdot k(3k-3)$.

Dividing the two numbers gives the desired ${3k-3}/{4k-2}$.

(c) Note that there are exactly nk edges to begin with. Let's compute the expected number of times that a new triangle forms. There are $O(n)$ open triangles, and the probability that a given edge gets rewired to a given open triangle is

$\binom{n}{2}^{-1} = O(1/n^2)$, so the probability of getting rerouted to *any* open triangle is $O(1/n)$ by a union bound. Since there are $O(n)$ edges, the total expected number of new triangles is $O(1)$.

On the other hand, the expected number of triangles that remain triangles is exactly $(1-p)^3$ times the previous number of triangles.

Similarly, we can count that there are initially $2(2k-1)$ open triplets using a given edge. Once we reroute it, if we do, there are $2(2k)+2E$ new open triplets formed, where $2E$ is the number of additional edges rewired to one of the same vertices. For each edge e, there are $kn-1$ other edges, each rewired with probability p. The probability each gets rerouted to share a vertex with e is $O(1/n^2)$. Thus $\mathbb{E}[E] = O(1/n) = o(1)$. Thus the number of new potential triangles is *at most*

$$2nk(4k-2) \leq \left(\frac{4k+o(1)}{4k-2}\right) 2nk(4k-2) \leq 2nk4k.$$

As $n \to \infty$ we get

$$\frac{2nk(3k-3)(1-p)^3 + O(1)}{2nk(4k)} \leq \mathrm{Cl}(g) \leq \frac{2nk(3k-3)(1-p)^3 + O(1)}{2nk(4k-2)}$$

The left hand side converges to $(3k-3)(1-p)^3/4k$ and the right hand side converges to $(3k-3)(1-p)^3/(4k-2)$.