CS294-179 Network Structure and Epidemics
Fall 2020
Homework 2 on the Structure of $G(n,p)$

1 Cycles in $G(n,p)$ [6 Points]

(a) Let T be the number of triangles in $G(n,p)$. Calculate $\mathbb{E}(T)$ as a function of n and p.

(b) Show that there are $\frac{(k-1)!}{2}$ ways to have a cycle on k vertices. Then calculate the expected number of cycles of length k in $G(n,p)$. For $k=3$ this should agree with part (a).

For the next part, consider $p = \frac{c}{n}$ *in* $G(n,p)$.

(c) What is the expected number of cycles of length k that pass through a given vertex? Show that for $k = l \log n$, where $l > 0$ is a constant, this goes to zero if l is small enough. What constant do you get?

2 Locally Tree-like [6 Points]

Let $p = c/n$ for some constant $c > 0$. The next three parts shows that $G(n,p)$ is locally tree like: For a random vertex v, and any fixed d, the subgraph you see when restricting yourself to vertices of distance at most d from v is whp a tree.

(a) Take a path of length k_1 attached to a random vertex in $G(n,p)$, and an attached cycle of length k_2. Show that the probability that such an object exists in $G(n,c/n)$ is bounded by $c^{k_1+k_2}/n$.

(b) Consider a connected graph H, and a vertex v in H. Assume that all vertices in H have distance at most d from v, and the H is not a tree. By considering first a spanning tree, and then adding one edge back in, show that H must contain a path with a cycle attached to it, with the length of the line plus that of the cycle bounded by a constant times d (One can also get $2d+1$, but you get the full credit if you prove it for any constant).

(c) Let l be the constant you found in (b). Now use a union bound over all k_1 and k_2 that $k_1 + k_2 \leq ld$ to show there exists a d and c dependent constant $K = K(d,c)$, such that with probability at least $1 - K/n$, the induced ball of radius d around a random vertex is a tree.

3 Phase Transition [6 Points plus 6 Bonus Points]

Solve one of the next two (if you do both, you get credit for the one on which you achieve more points, and bonus points for the other).

(a) Let I denote the number of isolated vertices (vertices of degree 0) in $G(n, p)$. Prove that if for some $\epsilon > 0$, $p \leq \frac{(1-\epsilon)\log(n)}{n}$ then $\lim_{n\to\infty} \mathbb{P}(I > 0) = 1$. Also, show that when $p \geq \frac{(1+\epsilon)\log(n)}{n}$ then $\lim_{n\to\infty} \mathbb{P}(I = 0) = 1$.
Hint: Use the first and second moment method (see below). Please establish both.

(b) Let T be the number of triangles in $G(n, p)$. Find $\lim_{n\to\infty} \mathbb{P}(T = 0)$, when $p = c/n$ for some constant $c > 0$.
Hint: Use Janson's Inequality, see below. You don't need to establish it - just use it and say how/why it applies.

Background (for both parts): Assume $X_1, \ldots, X_N \geq 0$ are independent non-negative random variables. We can calculate the probability that all are zero via $\Pr(\sum_i X_i = 0) = \prod_i \Pr(X_i = 0)$. You can't use this for either part, since the number of triangles, or the number of points of degree 0 is not a sum of independent random variables. But if $\mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j]$ for most of the pairs i and j, we often can still make progress, using either the first and second moment method, or Janson's inequality.

First moment method: If the X_i are integer valued, non-negative random variables, then $\Pr(\sum_i X_i \neq 0) \leq \mathbb{E}(\sum_i X_i)$ (please prove that - it is one line plus a few words). Thus if $\mathbb{E}(\sum_i X_i) \to 0$, we know that whp all the random variables are zero.

Second moment method: If X is a non-negative random variable, then $\Pr(X = 0) \leq \frac{Var(X)}{(\mathbb{E}[X])^2}$ (please prove this, using, e.g., Chebyshev's Inequality - it is again just one line plus a few words). Applying this to $X = \sum_i X_i$, where all $X_i \geq 0$, this allows us to conclude that whp, at least one X_i is non-zero if $\frac{Var(X)}{(\mathbb{E}[X])^2} \to 0$.

Janson's inequality: This gives not just bounds on probabilities of dependent events which go to zero or one, but sometimes allows us to calculate them asymptotically as if they were independent. (In the formulation below, M is the number you would get if the events $\bar{B}_i$ were independent).

Theorem 1 *(Janson's inequality)*
Let Ω be a finite ground set, I be a finite index set, and let A_i, $i \in I$, be subsets of Ω. Choose a random subset R of Ω by including each $\omega \in \Omega$ independently with probability p_ω, and let B_i be the event $A_i \subset R$. Define a "dependency graph" by declaring i and j neighbors, $i \sim j$, when $i \neq j$ and $A_i \cap A_j \neq \emptyset$. Define

$$M = \prod_{i\in I} \Pr[\bar{B}_i] \quad \textit{and} \quad \Delta = \sum_{i\sim j} \Pr[B_i \wedge B_j],$$

where Δ is over all ordered pairs $i \sim j$ (so $\frac{1}{2}\Delta$ is the sum over unordered pairs). If $\Pr[B_i] \leq \epsilon$ for $i \in I$, then

$$M \leq \Pr[\wedge_{i\in I} \bar{B}_i] \leq M \exp\left(\frac{1}{1-\epsilon}\frac{\Delta}{2}\right).$$