Problem Set 4

Electronic submission (email as attachment to psetCS265@gmail.com) due 11am Thursday 10/23. If submitting a hard copy, there is a box on the 1st floor of Gates Building, by the East entrance, labelled CS265/CME309. Hard copies must be submitted by 10am Thursday 10/23.

[You may discuss these problems with classmates. Please do not troll the internet looking for solutions to these problems. All writing must be done independently, and you must fully understand everything you write.]

In this problem set we characterize the extinction probability of the Galton-Watson branching process, and prove the threshold behavior of the size of the largest component of random graphs.

1. The Galton-Watson branching process corresponding to a random variable $X$ that takes non-negative integer values is defined as follows: at time $t = 0$, there is one node. At time $t = 1$, the number of nodes is distributed according to the random variable $X$, and in general, at time $t$, each of the nodes at time $t - 1$ has a number of children distributed according to (independent) copies of $X$. Let $Z_t$ denote the random variable denoting the number of nodes that exist at time $t$. We will prove the following theorem:

**Theorem 1.** Provided $\Pr[X = 1] < 1$ and $\Pr[X = 0] > 0$, then:

- If $E[X] \leq 1$ then $\lim_{t \to \infty} \Pr[Z_t = 0] = 1$.
- If $E[X] > 1$ then $\lim_{t \to \infty} \Pr[Z_t = 0] = p$ for $p \in (0, 1)$ with $p$ being the unique solution in $(0, 1)$ to the equation $p = \sum_{i \geq 0} \Pr[X = i] p^i$.

First, let us understand the relationship between the $Z_i$’s:

(a) Show that $Z_t$ is distributed according to the sum of $Z_1$ independent copies of $Z_{t-1}$.

Second, define $p_t = \Pr[Z_t = 0]$ to be the probability of extinction by time $t$:

(b) Prove that $p_t = \sum_{i \geq 0} \Pr[X = i] \Pr[Z_{t-1} = 0]^i$.

Since $p_1 \leq p_2 \leq \ldots$ is monotonically increasing and bounded by 1, by the Monotone Convergence Theorem, a limit $p = \lim_{t \to \infty} p_t$ exists. Define function $f(x) = \sum_{i \geq 0} \Pr[X = i] x^i$. By part (b) we know that $f(p_t) = p_{t+1}$, and combining with the definition of $p$, we conclude that $p = f(p)$. Let us explore some properties of $f$:

(c) Prove that $f(1) = 1$, $f'(1) = E[X]$, and $f(x)$ is convex on the interval $(0, 1)$.

Finally, to complete our proof:

(d) Show that if $E[X] > 1$, $f(x) = x$ will have a unique solution in $(0, 1)$, and if $E[X] \leq 1$, then there is no solution to $f(x) = x$ for $x \in (0, 1)$. 

For problem 2 and 3, we consider the sizes of the connected components of random graphs. Let \( G_{n,p} \) denote the Erdos-Renyi random graph model, where each edge exists (independently) with probability \( p = c/n \) for some constant \( c \) that does not vary with \( n \).

**Theorem 2.** Let \( G \) be drawn from \( G_{n,p} \), with \( p = c/n \) for some constant \( c \):

- If \( c < 1 \), with probability tending to 1 as \( n \to \infty \), the largest connected component of \( G \) has size \( O(\log n) \).

- If \( c > 1 \), with probability tending to 1 as \( n \to \infty \), the largest connected component of \( G \) has size \( (1 - p)n \pm o(n) \), where \( p \) is the probability of extinction of the Galton-Watson branching process for the Poisson random variable with expectation \( c \), and the second-largest component of \( G \) has size \( O(\log n) \).

2. In this problem we prove the \( c < 1 \) case of the above theorem.

   (a) For a given vertex \( v \), prove that
   \[
   \Pr[v \text{ in connected component of size } \geq k] \leq \Pr[X \geq k - 1],
   \]
   where \( X \) is distributed according to \( \text{Binomial}[k \cdot n, c/n] \). [Hint: consider doing a breadth-first search of the neighborhood of \( v \) in the graph.]

   (b) Assuming the above, using a union bound over Chernoff bounds, prove that
   \[
   \Pr[\text{there is a connected component of size } \geq 10 \log n \cdot (1 - c)^2] \leq 1/n.
   \]
   This completes the proof.

3. In this problem, we prove the \( c > 1 \) case of the above theorem.

   (a) Given a random node \( v \) in the graph, prove that for any \( \frac{100c \log n}{(c-1)^2} \leq k \leq n^{3/4} \), the probability that the connected component of \( v \) has size \( k \) is no more than \( n^{-10} \). [Hint: given that the connected component of \( v \) has size at least \( k \), show that with high probability it will have size at least \( k + 1 \). Be mindful of the way you condition events.]

   (b) Prove that we do not expect any connected components to have size in the interval \( \left[ \frac{100c \log n}{(c-1)^2}, n^{3/4} \right] \).

   (c) Prove that with probability tending to 1 as \( n \to \infty \), there is at most one connected component of size \( \geq n^{3/4} \). [Hint: conditioned on the neighborhood of both \( v \) and \( u \) having size at least \( n^{3/4} \), show that the probability that they are not connected is tiny, then union bound over the at most \( n \) such neighborhoods.]

   (d) Using the theorem proved in problem 1, show that the expected size of the large component is as claimed at the beginning of Theorem 2.

   (e) BONUS: Show that the size of the large component is within \( o(n) \) of its expectation with probability tending to 1 as \( n \to \infty \). [Hint: bound the variance of the number of nodes that are in "small" components of size at most \( \frac{100c \log n}{(c-1)^2} \), then use Chebyshev’s inequality.]

Spend a few minutes thinking about the theorem you have just proved, and the intuition behind why, with very high probability, there are never any medium-sized components.