1 Problem Statement and Initial Ideas

Consider the Erdős-Rényi random graph $G(n,p)$, with $p = p_n$ a function of n. Today we will give some insight and a proof of the following classical result:

Theorem 1. *If $p \ll \frac{\log n}{n}$, then $G(n,p)$ is disconnected with probability tending to 1, and if $p \gg \frac{\log n}{n}$, then $G(n,p)$ is connected with probability tending to 1.*

How can we prove such a result? Let us start with the definitions. A graph is disconnected if there are at least two connected components, or in other words if there is a nonempty (and non-full) set of vertices $S \subset [n]$ such that S has no edges to $[n] \setminus S$. Since S and $[n] \setminus S$ are symmetric here, we need only consider sets with size at most $\frac{n}{2}$. So we have the following bound, which will help us prove that $G(n,p)$ is connected:

$$\mathbb{P}[G(n,p) \text{ is disconnected}] \leq \sum_{k=1}^{n/2} \sum_{\substack{S \subset [n] \\ |S|=k}} \mathbb{P}[\text{there are no edges between } S \text{ and } [n] \setminus S \text{ in } G(n,p)].$$

Since each edge is independent, the probability inside the sum on the right is $(1-p)^{|S| \times (n-|S|)} = (1-p)^{k(n-k)}$. Since $k \leq \frac{n}{2}$ so that $n - k \geq \frac{n}{2}$, and since there are $\binom{n}{k} \leq n^k$ sets S of size k, we obtain

$$\begin{aligned}
\mathbb{P}[G(n,p) \text{ is disconnected}] &\leq \sum_{k=1}^{n/2} \binom{n}{k} (1-p)^{k(n-k)} \\
&\leq \sum_{k=1}^{n/2} n^k (1-p)^{k\frac{n}{2}} \\
&\leq \sum_{k=1}^{\infty} \left(n(1-p)^{n/2}\right)^k \\
&= O\left(n(1-p)^{n/2}\right),
\end{aligned}$$

if we assume that $n(1-p)^{n/2}$ is smaller than 1, since the sum of a geometric series is comparable with its initial term. In fact, the above statement will only be useful for us if $n(1-p)^{n/2} \to 0$ as $n \to \infty$. When does this happen? Let's use the bound $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$ to see that

$$n(1-p)^{n/2} \leq \exp\left(\log n - p\frac{n}{2}\right),$$

and so $n(1-p)^{n/2} \to 0$ if $p \gg \frac{\log n}{n}$. In other words, we have just proved that $G(n,p)$ is connected with probability tending to 1 if $p \gg \frac{\log n}{n}$.

Now let's examine the above proof to see how we might try to prove the other part of the theorem. If $G(n,p)$ *is* actually disconnected, then we know that the above proof can't work. The step where we bound the whole geometric sum by the first term indicates that if the first term is small, then the whole sum will also be small. So for the graph to be disconnected, we must in fact have the first term of the sum be large.

The first term represents the probability that there is an isolated vertex in $G(n,p)$. Thus the strategy suggested is to try and prove that if $p \ll \frac{\log n}{n}$, then there *will* be an isolated vertex with probability tending to 1. This in turn would prove that the graph is disconnected with probability tending to 1 if $p \ll \frac{\log n}{n}$.

2 Moment Methods

How can we prove that there *is* an isolated vertex in $G(n,p)$ with high probability? Recall that the first moment method can tell us when there *does not* exist something with high probability if the expected number of such things is small: If we let X denote the number of things (which is a nonnegative integer-valued random variable), then

$$\mathbb{P}[X > 0] = \sum_{k=1}^{\infty} \mathbb{P}[X = k] \leq \sum_{k=1}^{\infty} k \times \mathbb{P}[X = k] = \mathbb{E}[X].$$

But in order to go the other way we must use the second moment method:

$$\mathbb{P}[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Intuitively, if the spread of the random variable is smaller than the expectation, then there is a high chance that the variable will be closer to its expectation than to zero. This is a special case of the Paley-Zygmund inequality.

Theorem 2 (Paley-Zygmund)**.** *If X is a nonnegative random variable and $0 \leq \theta \leq 1$, then*

$$\mathbb{P}[X > \theta\mathbb{E}[X]] \geq (1-\theta)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Proof. Decompose $\mathbb{E}[X]$ as

$$\mathbb{E}[X] = \mathbb{E}[X\mathbf{1}_{\{X \leq \theta\mathbb{E}[X]\}}] + \mathbb{E}[X\mathbf{1}_{\{X > \theta\mathbb{E}[X]\}}] \leq \theta\mathbb{E}[X] + \sqrt{\mathbb{E}[X^2]\mathbb{P}[X > \theta\mathbb{E}[X]]}$$

(using Cauchy-Schwarz) and solve for $\mathbb{P}[X > \theta\mathbb{E}[X]]$. ∎

3 Isolated Vertices

Now let X_n denote the number of isolated vertices of $G(n,p)$ and suppose that $p \ll \frac{\log n}{n}$. Let's compute the first and second moments of X_n. First, since each vertex has probability $(1-p)^{n-1}$ to be isolated,

$$\mathbb{E}[X_n] = n(1-p)^{n-1}.$$

Now, by writing X_n as a sum of indicators, one for each vertex, we obtain

$$\begin{aligned}
\mathbb{E}[X_n^2] &= \sum_{i,j \in [n]} \mathbb{P}[\text{both } i \text{ and } j \text{ are isolated}] \\
&= \sum_{i \in [n]} \mathbb{P}[i \text{ is isolated}] + \sum_{i \neq j} \mathbb{P}[\text{both } i \text{ and } j \text{ are isolated}] \\
&= n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} \\
&\leq n(1-p)^{n-1} + n^2(1-p)^{2n-3},
\end{aligned}$$

since there are a total of $2(n-2)+1$ edges incident to any two distinct vertices.

So the second moment method tells us that

$$\begin{aligned}
\mathbb{P}[\text{there is an isolated vertex}] &\geq \frac{n^2(1-p)^{2n-2}}{n(1-p)^{n-1} + n^2(1-p)^{2n-3}} \\
&= \left(\frac{1}{n(1-p)^{n-1}} + \frac{1}{1-p}\right)^{-1}.
\end{aligned}$$

Since $p \to 0$, we have $\frac{1}{1-p} \to 1$. To control the other term, notice that we should have something like $(1-p)^{n-1} \approx e^{-p(n-1)}$, which means that if $p \ll \frac{\log n}{n}$ then the other term in the denominator above goes to 0, and we get $\mathbb{P}[\text{there is an isolated vertex}] \geq (1+o(1))^{-1} = 1 - o(1)$.

To make this rigorous, use the inequality $1 - x \geq e^{-x-x^2}$, which holds for $0 \leq x \leq \frac{1}{2}$, to obtain that

$$n(1-p)^{n-1} \geq \exp\left(\log n - p(n-1) - p^2(n-1)\right).$$

The $p^2(n-1)$ term disappears since $p^2(n-1) \ll \frac{\log^2 n}{n}$. This finishes the proof.