Random Graphs

Electrical Engineering 126 (UC Berkeley)

Fall 2018

1 Introduction

In this note, we will briefly introduce the subject of **random graphs**, also known as **Erdös-Rényi random graphs**. Given a positive integer n and a probability value $p \in [0,1]$, the $\mathcal{G}(n,p)$ random graph is an undirected graph on n vertices such that each of the $\binom{n}{2}$ edges is present in the graph independently with probability p. When $p = 0$, $\mathcal{G}(n,0)$ is an empty graph on n vertices, and when $p = 1$, $\mathcal{G}(n,1)$ is the fully connected graph on n vertices (denoted K_n). Often, we think of $p = p(n)$ as depending on n, and we are usually interested in the behavior of the random graph model as $n \to \infty$.

A bit more formally, $\mathcal{G}(n,p)$ defines a distribution over the set of undirected graphs on n vertices. If $G \sim \mathcal{G}(n,p)$, meaning that G is a random graph with the $\mathcal{G}(n,p)$ distribution, then for every fixed graph G_0 on n vertices with m edges, $\mathbb{P}(G = G_0) := p^m(1-p)^{\binom{n}{2}-m}$. In particular, if $p = 1/2$, then the probability space is uniform, or in other words, every undirected graph on n vertices is equally likely.

Here are some warm-up questions.

Question 1. What is the expected number of edges in $\mathcal{G}(n,p)$?

Answer 1. There are $\binom{n}{2}$ possible edges and the probability that any given edge appears in the random graph is p, so by linearity of expectation, the answer is $\binom{n}{2}p$.

Question 2. Pick an arbitrary vertex and let D be its degree. What is the distribution of D? What is the expected degree?

Answer 2. Each of the $n-1$ edges connected to the vertex is present independently with probability p, so $D \sim \text{Binomial}(n-1,p)$. For every $d \in \{0,1,\dots,n-1\}$, $\mathbb{P}(D = d) = \binom{n-1}{d}p^d(1-p)^{n-1-d}$, and $\mathbb{E}[D] = (n-1)p$.

Question 3. Suppose now that $p(n) = \lambda/n$ for a constant $\lambda > 0$. What is the approximate distribution of D when n is large?

Answer 3. By the Poisson approximation to the binomial distribution, D is approximately Poisson(λ). For every $d \in \mathbb{N}$, $\mathbb{P}(D = d) \approx \exp(-\lambda)\lambda^d/d!$.

Question 4. What is the probability that any given vertex is isolated?

Answer 4. All of the $n-1$ edges connected to the vertex must be absent, so the desired probability is $(1-p)^{n-1}$.

2 Sharp Threshold for Connectivity

We will sketch the following result (see [1]):

Theorem 1 (Erdös-Rényi, 1961)**.** *Let*

$$p(n) := \lambda \frac{\ln n}{n}$$

for a constant $\lambda > 0$*.*

- *If* $\lambda < 1$*, then* $\mathbb{P}\{\mathcal{G}(n, p(n)) \text{ is connected}\} \to 0$*.*
- *If* $\lambda > 1$*, then* $\mathbb{P}\{\mathcal{G}(n, p(n)) \text{ is connected}\} \to 1$*.*

In the subject of random graphs, threshold phenomena like the one above are very common. In the above result, nudging the value of λ slightly around the critical value of 1 causes drastically different behavior in the limit, so it is called a *sharp* threshold. In such cases, examining the behavior near the critical value leads to further insights. Here, if we take $p(n) = (\ln n + c)/n$ for a constant $c \in \mathbb{R}$, then it is known that

$$\mathbb{P}\big\{\mathcal{G}\big(n, p(n)\big) \text{ is connected}\big\} \to \exp\{-\exp(-c)\},$$

see [2, Theorem 7.3]. Notice that the probability increases smoothly from 0 to 1 as we vary c from $-\infty$ to ∞.

Why is the threshold $p(n) = (\ln n)/n$? When $p(n) = 1/n$, then the expected degree of a vertex is roughly 1 so many of the vertices will be joined together (a great deal is known about the evolution of the so-called *giant component*, see e.g. [2]), but it is too likely that one of the vertices will have *no* edges connected to it, making it isolated (and thus the graph is disconnected).

Proof of Theorem 1. First, let $\lambda < 1$. If X_n denotes the number of isolated nodes in $\mathcal{G}(n, p(n))$, then it suffices to show that $\mathbb{P}(X_n > 0) \to 1$, i.e., there is an isolated node with high probability (this will then imply that the random graph is disconnected).

- $\mathbb{E}[X_n]$: Define I_i to be the indicator random variable of the event that the ith vertex is isolated. Using linearity of expectation and symmetry, $\mathbb{E}[X_n] = \sum_{i=1}^{n} \mathbb{E}[I_i] = \sum_{i=1}^{n} \mathbb{P}(\text{node } i \text{ is isolated}) = nq(n)$, where we define $q(n) := \mathbb{P}(\text{a node is isolated}) = [1 - p(n)]^{n-1}$.

  Observe that

  $$\ln \mathbb{E}[X_n] = \ln n + (n-1)\ln\{1 - p(n)\} \sim \ln n - \frac{n-1}{n}\lambda \ln n \to \infty,$$

  since $\lambda < 1$. Here, if f and g are two functions on $\mathbb{N}$, then the notation $f(n) \sim g(n)$ means $f(n)/g(n) \to 1$ (asymptotically, f and g have the same behavior). The above line also uses the first-order Taylor expansion $\ln(1 - x) = -x + o(x)$ as $x \to 0$.

  Thus $\mathbb{E}[X_n] \to \infty$ which is reassuring, since we want to prove that $\mathbb{P}(X_n > 0) \to 1$, but in order to prove the probability result we will need to also look at the variance of X_n.

- $\operatorname{var} X_n$: We claim that

  $$\mathbb{P}(X_n = 0) \le \frac{\operatorname{var} X_n}{\mathbb{E}[X_n]^2}.$$

  Here are two ways to see this. First, from the definition of variance,

  $$\begin{aligned}\operatorname{var} X_n &= \mathbb{E}[(X_n - \mathbb{E}[X_n])^2] \\ &= \mathbb{E}[X_n]^2 \mathbb{P}(X_n = 0) + (1 - \mathbb{E}[X_n])^2 \mathbb{P}(X_n = 1) + \cdots \\ &\ge \mathbb{E}[X_n]^2 \mathbb{P}(X_n = 0).\end{aligned}$$

  The second way is to use Chebyshev's Inequality:

  $$\mathbb{P}(X_n = 0) \le \mathbb{P}(|X_n - \mathbb{E}[X_n]| \ge \mathbb{E}[X_n]) \le \frac{\operatorname{var} X_n}{\mathbb{E}[X_n]^2}.$$

  The use of the variance is often called the **Second Moment Method**. We must show that the ratio $(\operatorname{var} X_n)/\mathbb{E}[X_n]^2 \to 0$. Since $I_1, \ldots, I_n$ are not independent, we must use $\operatorname{var} X_n = n \operatorname{var} I_1 + n(n-1)\operatorname{cov}(I_1, I_2)$. Since I_1 is a Bernoulli random variable, $\operatorname{var} I_1 = q(n)[1-q(n)]$, and by definition $\operatorname{cov}(I_1, I_2) = \mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]\,\mathbb{E}[I_2] = \mathbb{E}[I_1 I_2] - q(n)^2$.

In order to find $\mathbb{E}[I_1 I_2]$, we interpret it as a probability:

$$\mathbb{E}[I_1 I_2] = \mathbb{P}(\text{nodes } 1, 2 \text{ are isolated}).$$

In order for this event to happen, $2n-3$ edges must be absent:

$$\mathbb{P}(\text{nodes } 1, 2 \text{ are isolated}) = [1-p(n)]^{2n-3} = \frac{q(n)^2}{1-p(n)}.$$

So, $\operatorname{cov}(I_1, I_2) = q(n)^2/[1-p(n)] - q(n)^2 = p(n)q(n)^2/[1-p(n)]$, and

$$\begin{aligned}\frac{\operatorname{var} X_n}{\mathbb{E}[X_n]^2} &= \frac{nq(n)[1-q(n)] + n(n-1)p(n)q(n)^2/[1-p(n)]}{n^2 q(n)^2}\\ &= \frac{1-q(n)}{nq(n)} + \frac{n-1}{n}\frac{p(n)}{1-p(n)}.\end{aligned}$$

Since $nq(n) = \mathbb{E}[X_n] \to \infty$, the first term tends to 0, and since $p(n) \to 0$, the second term tends to 0 as well.

Next, let $\lambda > 1$. The key idea for the second claim is the following: the graph is disconnected if and only if there exists a set of k nodes, $k \in \{1, \dots, \lfloor n/2 \rfloor\}$, such that there is no edge connecting the k nodes to the other $n-k$ nodes in the graph. We can apply the union bound twice.

$$\begin{aligned}&\mathbb{P}\bigl\{\mathcal{G}\bigl(n, p(n)\bigr) \text{ is disconnected}\bigr\}\\ &\quad= \mathbb{P}\Bigl(\bigcup_{k=1}^{\lfloor n/2 \rfloor} \{\text{some set of } k \text{ nodes is disconnected}\}\Bigr)\\ &\quad\le \sum_{k=1}^{\lfloor n/2 \rfloor} \mathbb{P}(\text{some set of } k \text{ nodes is disconnected})\\ &\quad\le \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} \mathbb{P}(\text{a specific set of } k \text{ nodes is disconnected})\\ &\quad= \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} [1-p(n)]^{k(n-k)}.\end{aligned}$$

The rest of the proof is showing that the above summation tends to 0 via tedious calculations, which will be given in the Appendix. $\square$

Appendix: Tedious Calculations

Here, we will argue that

$$\sum_{k=1}^{\lfloor n/2\rfloor}\binom{n}{k}[1-p(n)]^{k(n-k)} \leq \sum_{k=1}^{\lfloor n/2\rfloor}\binom{n}{k}\exp\{-k(n-k)p(n)\} = \sum_{k=1}^{\lfloor n/2\rfloor}\binom{n}{k}n^{-\lambda k(n-k)/n}$$

tends to 0 as $n\to\infty$. One way to do this is to break up the summation into two parts. Since $\lambda>1$, choose n^* so that $\lambda(n-n^*)/n>1$, which means we can take $n^*=\lfloor n(1-\lambda^{-1})\rfloor$. The first part of the summation is

$$\sum_{k=1}^{n^*}\binom{n}{k}n^{-\lambda k(n-k)/n} \leq \sum_{k=1}^{n^*}n^{-k[\lambda(n-k)/n-1]} \leq \sum_{k=1}^{n^*}n^{-k[\lambda(n-n^*)/n-1]} \leq \frac{n^{-[\lambda(n-n^*)/n-1]}}{1-n^{-[\lambda(n-n^*)/n-1]}} \to 0.$$

For the second part of the summation, we will use the bound

$$\binom{n}{k} \leq \frac{n^k}{k!} = \left(\frac{n}{k}\right)^k\frac{k^k}{k!} \leq \left(\frac{n}{k}\right)^k\sum_{j=0}^{\infty}\frac{k^j}{j!} = \left(\frac{\mathrm{e}n}{k}\right)^k.$$

Using this bound:

$$\sum_{k=n^*+1}^{\lfloor n/2\rfloor}\binom{n}{k}n^{-\lambda k(n-k)/n} \leq \sum_{k=n^*+1}^{\lfloor n/2\rfloor}\left(\frac{\mathrm{e}n^{1-\lambda(n-k)/n}}{k}\right)^k \leq \sum_{k=n^*+1}^{\lfloor n/2\rfloor}\left(\frac{\mathrm{e}n^{1-\lambda(n-k)/n}}{n^*+1}\right)^k \leq \sum_{k=n^*+1}^{\lfloor n/2\rfloor}\left(\frac{\mathrm{e}n^{-\lambda(n-k)/n}}{1-\lambda^{-1}}\right)^k \leq \sum_{k=n^*+1}^{\lfloor n/2\rfloor}\left(\frac{\mathrm{e}n^{-\lambda/2}}{1-\lambda^{-1}}\right)^k$$

For n sufficiently large, $\mathrm{e}n^{-\lambda/2}/(1-\lambda^{-1})<\delta$ for some $\delta<1$.

$$\leq \sum_{k=n^*}^{\infty}\delta^k = \frac{\delta^{n^*}}{1-\delta} \to 0$$

since $n^*\to\infty$.

References

[1] Daron Acemoglu and Asu Ozdaglar. *Erdös-Rényi graphs and branching processes*. 2009. URL: http://economics.mit.edu/files/4621.

[2] Béla Bollobás. *Random graphs*. Second. Vol. 73. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2001, pp. xviii+498. ISBN: 0-521-80920-7; 0-521-79722-5. DOI: 10.1017/CBO9780511814068.