Note 17. Random graphs

Alex Fu

Fall 2022

Weighted, directed transition diagrams describe random one-directional transitions through time, but simply undirected graphs can also describe the relation of *connectivity*. The study of random graphs is largely focused on finding the regularity presented in random connectivity, especially asymptotic behaviors and *thresholds*, for instance in the evolution of random networks.

A good starting point would be a definition of a uniformly random graph. Uniform(Graph(n, m)), the uniform distribution over the set of all undirected graphs with n vertices and m edges, is one such definition, but we will focus on another, more popular "uniform" model.

Definition 1 (Erdős–Rényi random graph).

The **Erdős–Rényi model** of random graphs is the probability distribution $\mathcal{G}(n, p)$ over graphs with n vertices such that every possible edge appears independently with probability $p \in [0, 1]$. Equivalently, the probability of any undirected graph with m edges is

$$p^m(1-p)^{\binom{n}{2}-m}.$$

p is often denoted $p(n)$ for its dependence on n. The trivial cases of $p = 0$ and $p = 1$ are the empty graph and the complete graph K_n respectively, while the case of $p = \frac{1}{2}$ defines a uniform distribution over all graphs with n vertices.

Proposition 1 (Basic properties of ER graphs).

a. The expected number of edges in $G \sim \mathcal{G}(n, p)$ is $\binom{n}{2}p$.

b. The degree of any particular vertex is distributed as Binomial$(n-1, p)$, so its expected number of neighbors is $(n-1)p$.

c. The probability that any given vertex is isolated is $(1-p)^{n-1}$.

d. The distribution of $\deg v$ converges to Poisson(λ) as $n \to \infty$ if $\lambda = \lim_{n\to\infty} np(n)$.

The proofs are left to the reader as exercises in using independence.

A common phenomenon in the study of random graphs is the existence of a *threshold function* for a given property, where the probability that a random graph has that property tends to 0 if it grows asymptotically slower than the threshold, to 1 if it grows asymptotically faster, and undetermined in the critical case. The following result is an example of a *sharp* threshold:

Theorem 1 (Sharp threshold for connectivity).

Suppose that $p(n) = \lambda \ln(n)/n$ for a constant $\lambda > 0$.

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

b. If $\lambda > 1$, then $\mathbb{P}(\mathcal{G}(n, p(n))$ is connected$) \to 1$.

c. If now $p(n) = (\ln(n) + c)/n$ for $c \in \mathbb{R}$, then $\mathbb{P}(\mathcal{G}(n, p(n))$ is connected$) \to \exp(-e^{-c})$.

Proof for $\lambda < 1$. Let X_n be the number of isolated vertices in $G \sim \mathcal{G}(n, p(n))$, so that $\mathbb{P}(X_n > 0) \to 1$ suffices to show that G is almost surely disconnected as $n \to \infty$. Now let us use the second moment method presented in the note on concentration inequalities:

$$\mathbb{P}(X_n = 0) \leq \frac{\mathsf{var}(X_n)}{\mathbb{E}(X_n^2)}.$$

Here, we will actually use a weaker bound that is simpler to work with. By Chebyshev's inequality,

$$\mathbb{P}(X_n = 0) \leq \mathbb{P}(|X_n - \mathbb{E}(X_n)| \geq \mathbb{E}(X_n)) \leq \frac{\mathsf{var}(X_n)}{\mathbb{E}(X_n)^2}.$$

Alternatively, the definition of variance gives the same bound:

$$\mathsf{var}(X_n) = \sum_{x=0}^{\infty} \mathbb{P}(X_n = x) \cdot (x - \mathbb{E}(X_n))^2 \geq \mathbb{P}(X_n = 0) \cdot \mathbb{E}(X_n)^2.$$

Let $q(n) := (1 - p(n))^{n-1}$ be the probability that a vertex is isolated. Then

$$\mathbb{E}(X_n) = \sum_{i=1}^{n} \mathbb{E}(\mathbb{1}\{\text{vertex } i \text{ is isolated}\}) = \sum_{i=1}^{n} \mathbb{P}(\text{vertex } i \text{ is isolated}) = nq(n).$$

Using the first-order Taylor approximation $1 - x \sim e^{-x}$ for $x \to 0$, where $f(n) \sim g(n)$ denotes asymptotic equivalence $f(n)/g(n) \to 1$,

$$\mathbb{E}(X_n) = n\left(1 - \frac{\lambda \ln(n)}{n}\right)^{n-1} \sim n \exp\left(-\frac{\lambda \ln(n)}{n} \cdot (n-1)\right) \sim n^{1-\lambda} \to \infty.$$

We can again write X_n as the sum of i.i.d. indicators $I_i := \mathbb{1}\{\text{vertex } i \text{ is isolated}\}$ to find

$$\begin{aligned}\mathsf{var}(X_n) &= \sum_{i=1}^{n} \mathsf{var}(I_i) + \sum_{i \neq j} \mathsf{cov}(I_i, I_j) \\ &= n\,\mathsf{var}(I_1) + n(n-1) \cdot [\mathbb{E}(I_1 I_2) - \mathbb{E}(I_1)\,\mathbb{E}(I_2)]\end{aligned}$$

$$= n\,\mathsf{var}(\mathsf{Bernoulli}(q(n))) + n(n-1)\cdot[\mathbb{P}(\text{vertices 1 and 2 are isolated}) - \mathbb{E}(I_1)^2]$$
$$= n(q(n)(1-q(n)) + n(n-1)\cdot[(1-p(n))^{2n-3} - q(n)^2].$$

Finally, noting that $(1-p(n))^{2n-3} = q(n)^2/(1-p(n))$, we can simplify

$$\begin{aligned}
\frac{\mathsf{var}(X_n)}{\mathbb{E}(X_n)^2} &= \frac{nq(n)(1-q(n)) + n(n-1)\cdot p(n)q(n)^2/(1-p(n))}{(nq(n))^2} \\
&= \frac{1-q(n)}{nq(n)} + \frac{n-1}{n}\cdot\frac{p(n)}{1-p(n)} \\
&\to 0
\end{aligned}$$

as $nq(n) = \mathbb{E}(X_n) \to \infty$ and $p(n) \to 0$. □

Proof for $\lambda > 1$. The key idea: a graph is disconnected iff it has a cut of size $1 \le k \le \lfloor n/2 \rfloor$, a partition into two subsets of k and $n-k$ vertices not connected by any crossing edge. By the union bound applied twice,

$$\begin{aligned}
\mathbb{P}(\mathcal{G}(n,p(n)) \text{ is disconnected}) &= \mathbb{P}\left(\bigcup_{k=1}^{\lfloor n/2 \rfloor} \{\text{there exists a cut of size } k\}\right) \\
&\le \sum_{k=1}^{\lfloor n/2 \rfloor} \mathbb{P}(\text{there exists a cut of size } k) \\
&\le \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k}\cdot\mathbb{P}(\text{a particular set of } k \text{ vertices is disconnected}) \\
&= \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k}(1-p(n))^{k(n-k)} \\
&\le \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k}\exp(-k(n-k)p(n)) \\
&= \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} n^{-\lambda k(n-k)/n}.
\end{aligned}$$

To show that this sum tends to 0, we break it into two parts: cuts of almost negligible size and other cuts. As $\lambda > 1$, we can choose $n^* := \lfloor n(1-\lambda^{-1}) \rfloor$ so that $\lambda(n-n^*)/n > 1$. Then

$$\begin{aligned}
\sum_{k=1}^{n^*} \binom{n}{k} n^{-\lambda k(n-k)/n} &\le \sum_{k=1}^{n^*} n^{-\lambda k(n-k)/(n-1)}. \\
&\le \sum_{k=1}^{n^*} n^{-\lambda k(n-n^*)/(n-1)}.
\end{aligned}$$

$$\leq \frac{n^{-\lambda(n-n^*)/(n-1)}}{1-n^{-\lambda(n-n^*)/(n-1)}}$$

tends to 0. As for the second part of the sum, we bound the binomial coefficient:

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

Using this bound,

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

Since $n^{-\lambda/2} \sim en^{-\lambda/2}/(1-\lambda^{-1}) < \delta$ for some $\delta < 1$,

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

As $n^* \to \infty$, the second part of the sum tends to 0 as well, and we are finally done. □

Thresholds are also known as **phase transitions**. Beyond Theorem 1, there are many, many other interesting phase transitions to explore, of which Erdős and Rényi's 1959 and 1960 papers provide a good introduction, and some of which we list below.

Theorem 2 (Phase transitions for connected components).

Let $G \sim \mathcal{G}(n,p)$. As $n \to \infty$,

a. If $np < 1$, then G will almost surely have no connected components of size $> O(\log(n))$.

b. If $np = 1$, then G will almost surely have a largest component of size $\sim n^{2/3}$.

c. If $np \to c > 1$, then G will almost surely have a unique *giant component* containing a positive fraction of the vertices, and no other component will be of size $> O(\log(n))$.

■