Lecture 5
Models of Network Formation

1 Models of Network Formation

This lecture introduces random graphs and the basic probabilistic methods that allow us to analyze them. This lets us model networks by a random graphs, and then determine whether they satisfy a given property with high probability. We find that there are often (potentially sharp) thresholds that allow us to determine the properties of a given random graph.

2 Random Graphs

We begin by introducing Erdős-Rényi random graphs. We then discuss the First and Second Moment Methods that will allow us to analyze these models. We then rigorously introduce graph properties, and conclude with a discussion of threshold functions.

2.1 Introduction to Random Graphs

In 1959 Paul Erdős and Alfréd Rényi published a seminal paper in which they introduce the concept of a random graph. There are two common definitions for Erdős-Rényi random graphs. The first and simplest is as follows:

Definition 1. *An Erdős-Rényi Random Graph, $\boldsymbol{G_{n,m}}$, is a labeled graph with n vertices and m edges chosen uniformly at random from all such graphs.*

One of the first questions we consider is what is the probability that a random graph $G_{n,m}$ matches some fixed graph G^*. We can easily see that

$$\mathbb{P}[G_{n,m} = G^*] = \begin{cases} 1/\binom{\binom{n}{2}}{m}, & \text{if } |V(G^*)| = n, \text{ and } |E(G^*)| = m; \\ 0, & \text{otherwise;} \end{cases}$$

since there are $\binom{n}{2}$ possible edges. However, while the definition is simple, the calculations soon become tedious. Thus, we consider an alternative definition.

Definition 2. *An Erdős-Rényi Random Graph, $\boldsymbol{G_{n,p}}$, is a labeled graph of n vertices where each possible edge appears independently with probability p.*

As before, let G^* be a fixed graph, and let us find the probability that a random graph $G_{n,p}$ matches G^*. Then,

$$\mathbb{P}[G_{n,p} = G^*] = \begin{cases} p^{|E(G^*)|}(1-p)^{\binom{n}{2}-|E(G^*)|}, & \text{if } |V(G^*)| = n; \\ 0, & \text{otherwise.} \end{cases}$$

We point out that $G_{n,p}$ is sometimes called the *binomial random graph* because $|E(G_{n,p})| \sim B(\binom{n}{2}, p)$ (i.e. the number of edges follow the binomial distribution on $\binom{n}{2}$ points with probability p). Notice that this implies $\mathbb{E}[|E(G_{n,p})|] = \binom{n}{2}p \approx (n^2/2)p$.

The two definitions of Erdős-Rényi are "the same", in a way which we will describe in Section 2.3. Thus we can often choose to work with whichever definition is most convenient.

2.2 Some Probability Theory

In order to work with random graphs we need to make use of a couple basic facts from probability theory. Specifically, the First and Second Moment Methods.

Theorem 1 (First Moment Method)**.** *If X is a random variable that is integer valued and non-negative, then*

$$\mathbb{P}[X \neq 0] \leq \mathbb{E}[X].$$

Proof. Since X is integer valued and nonnegative,

$$\begin{aligned} \mathbb{P}[X \neq 0] &= \sum_{i=1}^{\infty} \mathbb{P}[X = i] \\ &\leq \sum_{i=1}^{\infty} i\mathbb{P}[X = i] \\ &= \mathbb{E}[X], \end{aligned}$$

as desired. □

Notice that the above proof works as long as every value for X is at least 1, and there are a countable number of such values. While it might be tempting to try to generalize to rational numbers, we would have to take into account the distribution to ensure that we can fold in all necessary cost.

The First Moment Method gives a convenient upper bound for $\mathbb{P}[X \neq 0]$. We now consider a lower bound.

Theorem 2 (Second Moment Method)**.** *If X is a random variable with finite variance, then*

$$\frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]} \leq \mathbb{P}[X \neq 0].$$

Proof. Let $Y = \begin{cases} 0, & \text{if X = 0;} \\ 1, & \text{otherwise.} \end{cases}$

So by definition $\mathbb{E}[Y] = \mathbb{P}[X \neq 0]$. By the definition of Y, we see that $XY = X$, and $Y^2 = Y$. Thus, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}[X]^2 &= \mathbb{E}[XY]^2 \\ &\leq \mathbb{E}[X^2]\mathbb{E}[Y^2] \\ &= \mathbb{E}[X^2]\mathbb{E}[Y] \\ &= \mathbb{E}[X^2]\mathbb{P}[X \neq 0], \end{aligned}$$

so $\mathbb{E}[X]^2/\mathbb{E}[X^2] \leq \mathbb{P}[X \neq 0]$ as desired. □

Both these methods become very useful when proving properties about random graphs.

2.3 Graph Properties

Now that we have some basic tools, we want to describe and analyze random graphs in an objective manner.

Definition 3. *A* ***graph property*** $\mathscr{P}$ *is a set of graphs that is closed under graph isomorphism.*

In other words, if some graph $G \in \mathscr{P}$, then any graph G' that is isomorphic to G will also be in $\mathscr{P}$. For example, G contains a triangle, G is k-colorable, and G is non-planar are all graph properties since they are closed under graph isomorphism.

Given a graph property $\mathscr{P}$, a typical question we might ask is "For what p is $G_{n,p}$ in $\mathscr{P}$ with high probability?". Let $\mathscr{P}_{tri}$ be the graph property "has a triangle", and let us consider the above question.

Theorem 3. *If* $p \in o(1/n)$*, then* $\mathbb{P}[G_{n,p} \in \mathscr{P}_{tri}] \to 0$ *as* $n \to \infty$*.*

Proof. Let $1_{a,b,c}$ be the indicator variable for the event that edges a, b, and c form a triangle. Thus $X = \sum_{a,b,c} 1_{a,b,c}$ is the number of triangles in a graph. Thus, by the First Moment Method,

$$\begin{aligned} \mathbb{P}[G_{n,p} \in \mathscr{P}_{tri}] &= \mathbb{P}[X \neq 0] \\ &\leq \mathbb{E}[X]. \end{aligned}$$

Additionally, by the linearity of expectation,

$$\begin{aligned} \mathbb{P}[G_{n,p} \in \mathscr{P}_{tri}] &\leq \sum \mathbb{E}[1_{a,b,c}] \\ &= \binom{n}{3} p^3 \\ &\leq n^3 p^3. \end{aligned}$$

However, since $p \in o(1/n)$, we know that $n^3p^3 \to 0$ as $n \to \infty$, which gives the desired result. $\square$

Notice that we can refine the idea of a graph property in several ways.

Definition 4. *A graph property* $\mathscr{P}$ *is* ***monotone increasing*** *if* $G \in \mathscr{P}$ *implies* $G + e \in \mathscr{P}$*.*

In other words, adding edges to a graph does not change the fact that it has a certain property. Clearly $\mathscr{P}_{tri}$ is monotone increasing. However, the k-colorability and planar properties are not. Also, note that we can similarly define a **monotone decreasing** property.

We can now use this fact about properties to explain why $G_{n,m}$ and $G_{n,p}$ are almost the same.

Theorem 4. *If a graph property* $\mathscr{P}$ *is monotone increasing, then for* $pn = 2m/n$

$$\mathbb{P}[G_{n,m} \in \mathscr{P}] \leq 3\mathbb{P}[G_{n,p} \in \mathscr{P}].$$

Proof. Let $c_{m'}$ be the probability that a graph $G_{n,p}$ has exactly m' edges. Since $\mathscr{P}$ is monotone increasing, note that $\mathbb{P}[G_{n,m_i} \in \mathscr{P}] \leq \mathbb{P}[G_{n,m_j} \in \mathscr{P}]$ when $m_i < m_j$. Thus,

$$\begin{aligned}
\mathbb{P}[G_{n,p} \in \mathscr{P}] &= \sum_{m'=0}^{\binom{n}{2}} \mathbb{P}[G_{n,p} \in \mathscr{P} : |E(G_{n,p})| = m'] c_{m'} \\
&= \sum_{m'=0}^{\binom{n}{2}} \mathbb{P}[G_{n,m'} \in \mathscr{P}] c_{m'} \\
&\geq \sum_{m'=m}^{\binom{n}{2}} \mathbb{P}[G_{n,m'} \in \mathscr{P}] c_{m'} \\
&\geq \mathbb{P}[G_{n,m} \in \mathscr{P}] \sum_{m'=m}^{\binom{n}{2}} c_{m'}.
\end{aligned}$$

Note that $c_{m'} = \binom{\binom{n}{2}}{m'} p^{m'} (1-p)^{\binom{n}{2}-m'}$, which is the same as $\mathbb{P}[B(\binom{n}{2}, p) = m']$ as discussed in Section 2.1. Additionally, recall that the expected number of edges $\mu = \mathbb{E}[|E(G_{n,p})|] \approx n^2 p/2 = m$, since $pn = 2m/n$ by assumption. Thus, for sufficiently large values of n,

$$\begin{aligned}
\mathbb{P}[G_{n,p} \in \mathscr{P}] &= \mathbb{P}[G_{n,m} \in \mathscr{P}] \sum_{m'=m}^{\binom{n}{2}} \binom{\binom{n}{2}}{m'} p^{m'} (1-p)^{\binom{n}{2}-m'} \\
&= \mathbb{P}[G_{n,m} \in \mathscr{P}] \cdot \mathbb{P}\left[B\left(\binom{n}{2}, p\right) \geq \mu = m\right] \\
&\geq 1/3 \mathbb{P}[G_{n,m} \in \mathscr{P}]
\end{aligned}$$

since $\mathbb{P}[B(\binom{n}{2}, p) \geq \mu] \to 1/2$ as $n \to \infty$. Thus, $\mathbb{P}[G_{n,m} \in \mathscr{P}] \leq 3\mathbb{P}[G_{n,p} \in \mathscr{P}]$ as desired. $\square$

This allows us to compare $G_{n,m}$ and $G_{n,p}$, and we can now prove bounds on one in terms of the other. For example, we can now simply state the result for $\mathbb{P}[G_{n,m} \in \mathscr{P}_{tri}]$ since we have it for $G_{n,p}$.

2.4 Threshold Functions

A phenomenon that is seen when considering monotone properties is that of some sort of threshold that determines when a graph $G_{n,p}$ will be in $\mathscr{P}$ with high probability.

Definition 5. *Given a monotone increasing property $\mathscr{P}$, a **threshold function** is a function $f : \mathbb{N} \to [0,1]$ such that*

$$p \in o(f(n)) \implies \mathbb{P}[G_{n,p} \in \mathscr{P}] \to 0,$$

and

$$p \in \omega(f(n)) \implies \mathbb{P}[G_{n,p} \in \mathscr{P}] \to 1.$$

A fact that we will not prove, is that all monotone increasing properties have a threshold function.[1] Here we prove the special case that $\mathscr{P}_{tri}$ has a threshold function.

Theorem 5. *The function $f(n) = 1/n$ is a threshold function for $\mathscr{P}_{tri}$.*

Proof. Above we proved that $p \in o(f(n))$ implies $\mathbb{P}[G_{n,p} \in \mathscr{P}_{tri}] \to 0$. Thus we need only prove the second statement. Let $p \in \omega(f(n)) = \omega(1/n)$. Let $X = \sum_{a,b,c} 1_{a,b,c}$ be the number of triangles in $G_{n,p}$ as before. By the Second Moment Method, $\mathbb{P}[X \neq 0]$ is lower bounded by $\mathbb{E}[X^2]/\mathbb{E}[X]^2$. Thus it suffices to show that the latter goes to 1 as n approaches infinity.

Recall that $\mathbb{E}[X] = \binom{n}{3}p^3$. Additionally,

$$\mathbb{E}[X^2] = \sum_{a,b,c}\sum_{d,e,f} \mathbb{E}[1_{a,b,c}1_{d,e,f}]$$
$$= \binom{n}{3}p^3\left[\binom{n-3}{3}p^3 + 3\binom{n-3}{2}p^3 + 3\binom{n-3}{1}p^2 + 1\right]$$

since we must consider the cases where a, b, c and d, e, f overlap. Thus,

$$\mathbb{E}[X^2]/\mathbb{E}[X]^2 = \frac{\binom{n-3}{3}}{\binom{n}{3}} + \frac{3\binom{n-3}{2}}{\binom{n}{3}} + \frac{3\binom{n-3}{1}}{p\binom{n}{3}} + \frac{1}{p^3\binom{n}{3}}.$$

Notice that the last three terms in this sum are in $O(1/n)$. Thus we need only consider the first term. As n approaches infinity, $\binom{n-3}{3}/\binom{n}{3}$ approaches 1. Thus, $\mathbb{E}[X^2]/\mathbb{E}[X]^2 \to 1$, which means $\mathbb{P}[X \neq 0] \to 1$ as $n \to \infty$. Hence, $f(n) = 1/n$ is a threshold function for $\mathscr{P}_{tri}$. $\square$

We now refine our definition of a threshold function to something stronger.

Definition 6. *Given a monotone increasing property $\mathscr{P}$, a **sharp threshold** is a function $f : \mathbb{N} \to [0, 1]$ such that for any $\epsilon > 0$*

$$p = (1-\epsilon)f(n) \implies \mathbb{P}[G_{n,p} \in \mathscr{P}] \to 0,$$

and

$$p = (1+\epsilon)f(n) \implies \mathbb{P}[G_{n,p} \in \mathscr{P}] \to 1.$$

Note that we did not prove that $1/n$ is a sharp threshold. In fact, no sharp threshold exists for $\mathscr{P}_{tri}$.

Definition 7. *A threshold that is not sharp is said to be **coarse**. Given a property $\mathscr{P}$, if no sharp threshold exists, we say $\mathscr{P}$ has a **coarse threshold**.*

Thus, $1/n$ is a coarse threshold for $\mathscr{P}_{tri}$.

[1] A survey covering this and other threshold results is given by Ehud Friedgut in *Hunting for sharp thresholds*, and can be found at `http://www.ma.huji.ac.il/~ehudf/docs/survey.ps`.

3 Further Reading

For further reading, take a look at *Random Graphs and Complex Networks* by Remco van der Hofstad, which can be found at `http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf`. Chapter 1 of this set of notes includes an introduction to Erdős-Rényi random graphs and social networks. The set of notes also includes an introduction to various probabilistic methods, and goes into much further detail on specific graph models.

Additionally, *Random Graphs* by B. Bollobás and *Random Graphs* by S. Janson, A. Łuczak, and A. Ruciński are standard references for this material.

Finally, some easy-to-read lecture notes on Random Graphs can be found at `http://www.math.cmu.edu/~af1p/RandomGraphs/`.