Cambridge University Press
978-1-107-11850-8 - Introduction to Random Graphs
Alan Frieze and Michał Karoński
Excerpt
More information

PART I

Basic Models

1

Random Graphs

Graph theory is a vast subject in which the goals are to relate various graph properties i.e. proving that Property A implies Property B for various properties A,B. In some sense, the goals of Random Graph theory are to prove results of the form "Property A almost always implies Property B." In many cases Property A could simply be "Graph G has m edges." A more interesting example would be the following: property A is "G is an r-regular graph, $r \geq 3$" and Property B is "G is r-connected." This is proved in Chapter 10.

Before studying questions such as these, we need to describe the basic models of a random graph.

1.1 Models and Relationships

The study of random graphs in their own right began in earnest with the seminal paper of Erdős and Rényi [276]. This paper was the first to exhibit the threshold phenomena that characterize the subject.

Let $\mathcal{G}_{n,m}$ be the family of all labeled graphs with vertex set $V = [n] = \{1,2,\ldots,n\}$ and exactly m edges, $0 \leq m \leq \binom{n}{2}$. To every graph $G \in \mathcal{G}_{n,m}$, we assign a probability

$$\mathbb{P}(G) = \binom{\binom{n}{2}}{m}^{-1}.$$

Equivalently, we start with an empty graph on the set $[n]$, and insert m edges in such a way that all possible $\binom{\binom{n}{2}}{m}$ choices are equally likely. We denote such a random graph by $\mathbb{G}_{n,m} = ([n], E_{n,m})$ and call it a *uniform random graph*.

We now describe a similar model. Fix $0 \leq p \leq 1$. Then for $0 \leq m \leq \binom{n}{2}$, assign to each graph G with vertex set $[n]$ and m edges a probability

$$\mathbb{P}(G) = p^m (1-p)^{\binom{n}{2}-m},$$

where $0 \leq p \leq 1$. Equivalently, we start with an empty graph with vertex set $[n]$ and perform $\binom{n}{2}$ Bernoulli experiments inserting edges independently with probability p. We call such a random graph, a *Binomial random graph* and denote it by $\mathbb{G}_{n,p} = ([n], E_{n,p})$. This was introduced by Gilbert [367].

As one may expect there is a close relationship between these two models of random graphs. We start with a simple observation.

Lemma 1.1 *A random graph $\mathbb{G}_{n,p}$, given that its number of edges is m, is equally likely to be one of the $\binom{\binom{n}{2}}{m}$ graphs that have m edges.*

Proof Let G_0 be any labeled graph with m edges. Then since

$$\{\mathbb{G}_{n,p} = G_0\} \subseteq \{|E_{n,p}| = m\}$$

we have

$$\begin{aligned}
\mathbb{P}(\mathbb{G}_{n,p} = G_0 \mid |E_{n,p}| = m) &= \frac{\mathbb{P}(\mathbb{G}_{n,p} = G_0, |E_{n,p}| = m)}{\mathbb{P}(|E_{n,p}| = m)} \\
&= \frac{\mathbb{P}(\mathbb{G}_{n,p} = G_0)}{\mathbb{P}(|E_{n,p}| = m)} \\
&= \frac{p^m (1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}} \\
&= \binom{\binom{n}{2}}{m}^{-1}.
\end{aligned}$$

□

Thus $\mathbb{G}_{n,p}$ conditioned on the event $\{\mathbb{G}_{n,p}$ has m edges$\}$ is equal in distribution to $\mathbb{G}_{n,m}$, the graph chosen uniformly at random from all graphs with m edges.

Obviously, the main difference between those two models of random graphs is that in $\mathbb{G}_{n,m}$ we choose its number of edges, while in the case of $\mathbb{G}_{n,p}$ the number of edges is the Binomial random variable with the parameters $\binom{n}{2}$ and p. Intuitively, for large n random graphs $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$ should behave in a similar fashion when the number of edges m in $\mathbb{G}_{n,m}$ equals or is "close" to the expected number of edges of $\mathbb{G}_{n,p}$, i.e. when

$$m = \binom{n}{2} p \approx \frac{n^2 p}{2}, \tag{1.1}$$

or, equivalently, when the edge probability in $\mathbb{G}_{n,p}$

$$p \approx \frac{2m}{n^2}. \tag{1.2}$$

Throughout the book, we use the notation $f \approx g$ to indicate that $f = (1+o(1))g$, where the $o(1)$ term depends on some parameter going to 0 or ∞.

We next introduce a useful "coupling technique" that generates the random graph $\mathbb{G}_{n,p}$ in two independent steps. We then describe a similar idea in relation to $\mathbb{G}_{n,m}$. Suppose that $p_1 < p$ and p_2 is defined by the equation

$$1 - p = (1 - p_1)(1 - p_2), \tag{1.3}$$

or, equivalently,

$$p = p_1 + p_2 - p_1 p_2.$$

Thus an edge is not included in $\mathbb{G}_{n,p}$ if it is not included in either of $\mathbb{G}_{n,p_1}$ or $\mathbb{G}_{n,p_2}$.

It follows that

$$\mathbb{G}_{n,p} = \mathbb{G}_{n,p_1} \cup \mathbb{G}_{n,p_2},$$

where the two graphs $\mathbb{G}_{n,p_1}, \mathbb{G}_{n,p_2}$ are independent. So when we write

$$\mathbb{G}_{n,p_1} \subseteq \mathbb{G}_{n,p},$$

we mean that the two graphs are *coupled* so that $\mathbb{G}_{n,p}$ is obtained from $\mathbb{G}_{n,p_1}$ by superimposing it with $\mathbb{G}_{n,p_2}$ and replacing eventual double edges by a single one.

We can also couple random graphs $\mathbb{G}_{n,m_1}$ and $\mathbb{G}_{n,m_2}$ where $m_2 \geq m_1$ via

$$\mathbb{G}_{n,m_2} = \mathbb{G}_{n,m_1} \cup \mathbb{H}.$$

Here $\mathbb{H}$ is the random graph on vertex set $[n]$ that has $m = m_2 - m_1$ edges chosen uniformly at random from $\binom{[n]}{2} \setminus E_{n,m_1}$.

Consider now a graph property $\mathscr{P}$ defined as a subset of the set of all labeled graphs on vertex set $[n]$, i.e. $\mathscr{P} \subseteq 2^{\binom{n}{2}}$. For example, all connected graphs (on n vertices), graphs with a Hamiltonian cycle, graphs containing a given subgraph, planar graphs, and graphs with a vertex of given degree form a specific "graph property."

We state below two simple observations that show a general relationship between $\mathbb{G}_{n,m}$ and $\mathbb{G}_{n,p}$ in the context of the probabilities of having a given graph property $\mathscr{P}$.

Lemma 1.2 *Let $\mathscr{P}$ be any graph property and $p = m/\binom{n}{2}$ where $m = m(n), \binom{n}{2} - m \to \infty$. Then, for large n,*

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq 10 m^{1/2}\, \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}).$$

Proof By the law of total probability,

$$\begin{aligned}\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) &= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P} \mid |E_{n,p}| = k)\, \mathbb{P}(|E_{n,p}| = k) \\ &= \sum_{k=0}^{\binom{n}{2}} \mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P})\, \mathbb{P}(|E_{n,p}| = k) \\ &\geq \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P})\, \mathbb{P}(|E_{n,p}| = m).\end{aligned}$$

Recall that the number of edges $|E_{n,p}|$ of a random graph $\mathbb{G}_{n,p}$ is a random variable with the Binomial distribution with parameters $\binom{n}{2}$ and p. Applying Stirling's formula:

$$k! = (1 + o(1)) \left(\frac{k}{e}\right)^k \sqrt{2\pi k}, \tag{1.4}$$

and putting $N = \binom{n}{2}$, we get

$$\begin{aligned}\mathbb{P}(|E_{n,p}| = m) &= \binom{N}{m} p^m (1-p)^{\binom{n}{2}-m} \\ &= (1 + o(1)) \frac{N^N \sqrt{2\pi N}\, p^m (1-p)^{N-m}}{m^m (N-m)^{N-m}\, 2\pi \sqrt{m(N-m)}} \qquad (1.5) \\ &= (1 + o(1)) \sqrt{\frac{N}{2\pi m(N-m)}}.\end{aligned}$$

Hence,

$$\mathbb{P}(|E_{n,p}| = m) \geq \frac{1}{10\sqrt{m}},$$

so

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq 10 m^{1/2}\, \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}).$$

□

We call a graph property $\mathscr{P}$ *monotone increasing* if $G \in \mathscr{P}$ implies $G + e \in \mathscr{P}$, i.e. adding an edge e to a graph G does not destroy the property. For example, connectivity and Hamiltonicity are monotone increasing properties. A monotone increasing property is *non-trivial* if the empty graph $\bar{K}_n \notin \mathscr{P}$ and the complete graph $K_n \in \mathscr{P}$.

A graph property is *monotone decreasing* if $G \in \mathscr{P}$ implies $G - e \in \mathscr{P}$, i.e. removing an edge from a graph does not destroy the property. Properties of a graph not being connected or being planar are examples of monotone decreasing graph properties. Obviously, a graph property $\mathscr{P}$ is monotone increasing if and only if its complement is monotone decreasing. Clearly not all graph properties are monotone. For example, having at least half of the vertices having a given fixed degree d is not monotone.

From the coupling argument it follows that if $\mathscr{P}$ is a monotone increasing property then, whenever $p < p'$ or $m < m'$,

$$\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) \leq \mathbb{P}(\mathbb{G}_{n,p'} \in \mathscr{P}), \tag{1.6}$$

and

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq \mathbb{P}(\mathbb{G}_{n,m'} \in \mathscr{P}), \tag{1.7}$$

respectively.

For monotone increasing graph properties we can get a much better upper bound on $\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P})$, in terms of $\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P})$, than that given by Lemma 1.2.

Lemma 1.3 *Let $\mathscr{P}$ be a monotone increasing graph property and $p = \frac{m}{N}$. Then, for large n and p such that $Np, N(1-p)/(Np)^{1/2} \to \infty$,*

$$\mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \leq 3\,\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}).$$

Proof Suppose $\mathscr{P}$ is monotone increasing and $p = \frac{m}{N}$, where $N = \binom{n}{2}$. Then

$$\begin{aligned}\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) &= \sum_{k=0}^{N} \mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P})\,\mathbb{P}(|E_{n,p}| = k)\\ &\geq \sum_{k=m}^{N} \mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P})\,\mathbb{P}(|E_{n,p}| = k).\end{aligned}$$

However, by the coupling property we know that for $k \geq m$,

$$\mathbb{P}(\mathbb{G}_{n,k} \in \mathscr{P}) \geq \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}).$$

The number of edges $|E_{n,p}|$ in $\mathbb{G}_{n,p}$ has the Binomial distribution with parameters N, p. Hence,

$$\begin{aligned}\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) &\geq \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \sum_{k=m}^{N} \mathbb{P}(|E_{n,p}| = k)\\ &= \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) \sum_{k=m}^{N} u_k, \end{aligned} \tag{1.8}$$

where

$$u_k = \binom{N}{k} p^k (1-p)^{N-k}.$$

Now, using Stirling's formula,

$$u_m = (1+o(1)) \frac{N^N p^m (1-p)^{N-m}}{m^m (N-m)^{N-m} (2\pi m)^{1/2}} = \frac{1+o(1)}{(2\pi m)^{1/2}}.$$

Furthermore, if $k = m + t$ where $0 \le t \le m^{1/2}$ then

$$\frac{u_{k+1}}{u_k} = \frac{(N-k)p}{(k+1)(1-p)} = \frac{1 - \frac{t}{N-m}}{1 + \frac{t+1}{m}}$$
$$\ge \exp\left\{ -\frac{t}{N-m-t} - \frac{t+1}{m} \right\} = 1 - o(1),$$

after using Lemma 21.1(a), (b) to obtain the first inequality and our assumptions on N, p to obtain the second.

It follows that

$$\sum_{k=m}^{m+m^{1/2}} u_k \ge \frac{1-o(1)}{(2\pi)^{1/2}}$$

and the lemma follows from (1.8). □

Lemmas 1.2 and 1.3 are surprisingly applicable. In fact, since the $\mathbb{G}_{n,p}$ model is computationally easier to handle than $\mathbb{G}_{n,m}$, we repeatedly use both lemmas to show that $\mathbb{P}(\mathbb{G}_{n,p} \in \mathcal{P}) \to 0$ implies that $\mathbb{P}(\mathbb{G}_{n,m} \in \mathcal{P}) \to 0$ when $n \to \infty$. In other situations we can use a stronger and more widely applicable result. The theorem below, which we state without proof, gives precise conditions for the asymptotic equivalence of random graphs $\mathbb{G}_{n,p}$ and $\mathbb{G}_{n,m}$. It is due to Łuczak [535].

Theorem 1.4 *Let $0 \le p_0 \le 1$, $s(n) = n\sqrt{p(1-p)} \to \infty$, and $\omega(n) \to \infty$ arbitrarily slowly as $n \to \infty$.*

(i) Suppose that $\mathcal{P}$ is a graph property such that $\mathbb{P}(\mathbb{G}_{n,m} \in \mathcal{P}) \to p_0$ for all

$$m \in \left[\binom{n}{2}p - \omega(n)s(n), \binom{n}{2}p + \omega(n)s(n) \right].$$

Then $\mathbb{P}(\mathbb{G}_{n,p} \in \mathcal{P}) \to p_0$ as $n \to \infty$.

(ii) Let $p_- = p - \omega(n)s(n)/n^3$ and $p_+ = p + \omega(n)s(n)/n^3$. Suppose that $\mathcal{P}$ is a monotone graph property such that $\mathbb{P}(\mathbb{G}_{n,p_-} \in \mathcal{P}) \to p_0$ and $\mathbb{P}(\mathbb{G}_{n,p_+} \in \mathcal{P}) \to p_0$. Then $\mathbb{P}(\mathbb{G}_{n,m} \in \mathcal{P}) \to p_0$, as $n \to \infty$, where $m = \lfloor \binom{n}{2} p \rfloor$.

1.2 Thresholds and Sharp Thresholds

One of the most striking observations regarding the asymptotic properties of random graphs is the “abrupt” nature of the appearance and disappearance of certain graph properties. To be more precise in the description of this phenomenon, let us introduce *threshold functions* (or just *thresholds*) for monotone graph properties. We start by giving the formal definition of a threshold for a monotone increasing graph property $\mathscr{P}$.

Definition 1.5 A function $m^* = m^*(n)$ is a *threshold* for a monotone increasing property $\mathscr{P}$ in the random graph $\mathbb{G}_{n,m}$ if

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,m} \in \mathscr{P}) = \begin{cases} 0 & \text{if } m/m^* \to 0, \\ 1 & \text{if } m/m^* \to \infty, \end{cases}$$

as $n \to \infty$.

A similar definition applies to the edge probability $p = p(n)$ in a random graph $\mathbb{G}_{n,p}$.

Definition 1.6 A function $p^* = p^*(n)$ is a *threshold* for a monotone increasing property $\mathscr{P}$ in the random graph $\mathbb{G}_{n,p}$ if

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) = \begin{cases} 0 & \text{if } p/p^* \to 0, \\ 1 & \text{if } p/p^* \to \infty, \end{cases}$$

as $n \to \infty$.

It is easy to see how to define thresholds for monotone decreasing graph properties and therefore we leave this to the reader.

Notice also that the thresholds defined above are not unique, since any function which differs from $m^*(n)$ (resp. $p^*(n)$) by a constant factor is also a threshold for $\mathscr{P}$.

A large body of the theory of random graphs is concerned with the search for thresholds for various properties, such as containing a path or cycle of a given length, or, in general, a copy of a given graph, or being connected or Hamiltonian, to name just a few. Therefore the next result is of special importance. It was proved by Bollobás and Thomason [150].

Theorem 1.7 *Every non-trivial monotone graph property has a threshold.*

Proof Without loss of generality assume that $\mathscr{P}$ is a monotone increasing graph property. Given $0 < \varepsilon < 1$ we define $p(\varepsilon)$ by

$$\mathbb{P}(\mathbb{G}_{n,p(\varepsilon)} \in \mathscr{P}) = \varepsilon.$$

Note that $p(\varepsilon)$ exists because

$$\mathbb{P}(\mathbb{G}_{n,p} \in \mathscr{P}) = \sum_{G\in\mathscr{P}} p^{|E(G)|}(1-p)^{N-|E(G|}$$

is a polynomial in p that increases from 0 to 1. This is not obvious from the expression, but it is obvious from the fact that $\mathcal{P}$ is monotone increasing and that increasing p increases the likelihood that $\mathbb{G}_{n,p} \in \mathcal{P}$.

We will show that $p^* = p(1/2)$ is a threshold for $\mathcal{P}$. Let $G_1, G_2, \ldots, G_k$ be independent copies of $\mathbb{G}_{n,p}$. The graph $G_1 \cup G_2 \cup \ldots \cup G_k$ is distributed as $\mathbb{G}_{n,1-(1-p)^k}$. Now $1-(1-p)^k \leq kp$, and therefore by the coupling argument

$$\mathbb{G}_{n,1-(1-p)^k} \subseteq \mathbb{G}_{n,kp},$$

and so $\mathbb{G}_{n,kp} \notin \mathcal{P}$ implies $G_1, G_2, \ldots, G_k \notin \mathcal{P}$. Hence,

$$\mathbb{P}(\mathbb{G}_{n,kp} \notin \mathcal{P}) \leq [\mathbb{P}(\mathbb{G}_{n,p} \notin \mathcal{P})]^k.$$

Let ω be a function of n such that $\omega \to \infty$ arbitrarily slowly as $n \to \infty$, $\omega \ll \log\log n$. (We say that $f(n) \ll g(n)$ or $f(n) = o(g(n))$ if $f(n)/g(n) \to 0$ as $n \to \infty$. Of course in this case we can also write $g(n) \gg f(n)$.) Suppose also that $p = p^* = p(1/2)$ and $k = \omega$. Then

$$\mathbb{P}(\mathbb{G}_{n,\omega p^*} \notin \mathcal{P}) \leq 2^{-\omega} = o(1).$$

On the other hand for $p = p^*/\omega$,

$$\frac{1}{2} = \mathbb{P}(\mathbb{G}_{n,p^*} \notin \mathcal{P}) \leq \left[\mathbb{P}(\mathbb{G}_{n,p^*/\omega} \notin \mathcal{P})\right]^{\omega}.$$

So

$$\mathbb{P}(\mathbb{G}_{n,p^*/\omega} \notin \mathcal{P}) \geq 2^{-1/\omega} = 1 - o(1).$$

$\square$

In order to shorten many statements of theorems in the book we say that a sequence of events $\mathcal{E}_n$ occurs *with high probability* (w.h.p.) if

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1.$$

Thus the statement that says p^* is a threshold for a property $\mathcal{P}$ in $\mathbb{G}_{n,p}$ is the same as saying that $\mathbb{G}_{n,p} \notin \mathcal{P}$ w.h.p. if $p \ll p^*$, while $\mathbb{G}_{n,p} \in \mathcal{P}$ w.h.p. if $p \gg p^*$.

In many situations we can observe that for some monotone graph properties more "subtle" thresholds hold. We call them "*sharp thresholds.*" More precisely,

Definition 1.8 A function $m^* = m^*(n)$ is a *sharp threshold* for a monotone increasing property $\mathcal{P}$ in the random graph $\mathbb{G}_{n,m}$ if for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(\mathbb{G}_{n,m} \in \mathcal{P}) = \begin{cases} 0 & \text{if} \quad m/m^* \leq 1-\varepsilon \\ 1 & \text{if} \quad m/m^* \geq 1+\varepsilon. \end{cases}$$

A similar definition applies to the edge probability $p = p(n)$ in the random graph $\mathbb{G}_{n,p}$.