3

Random Graphs

3.1 Introduction

The theory of random graphs began in the late 1950s with the seminal paper by Erdös and Rényi [3], devoted to graphs whose edges are i.i.d, with a fixed probability between any pair of vertices.

As we shall see shortly, a random graph with constant p is not a very interesting object to study for our purposes, as the graph is so richly connected that every node is only two hops away from every other node. In fact, with constant p, the degree (i.e., the number of edges per vertex) grows linearly with n, while many real networks are much sparser. As we will see, interesting behavior (such as phase transitions from many small components to one dominating giant component) occurs within much sparser random graphs.

To focus on such graphs, it is necessary to let $p = p(n)$ depend on n; specifically, we will let $p(n)$ go to zero in different ways, which will give rise to several interesting regimes, separated by phase transitions. Contrast this with percolation theory, where the phase transition occurred for $p = p_c$ independent of the lattice size. For this reason, it was not necessary to work out results on a finite lattice of size n and then to study the limit $n \to \infty$; we could directly study the infinite lattice. In random graph theory, on the other hand, we need to perform the extra step of going to the limit, and we will be interested in properties of RGs whose probability goes to one when $n \to \infty$. Such a property Q is said to occur *asymptotically almost surely (a.a.s.)*, although many authors use the somewhat imprecise term *almost every* graph has property Q (a.e.), or also *property Q occurs with high probability (w.h.p.)*.

Definition 3.1 (Random graph)**.** *Given n and p, a random graph $G(n, p)$ is a graph with labeled vertex set $[n] = \{1, \dots, n\}$, where each pair of vertices has an edge independently with probability p.*

As the node degree has a binomial distribution $\mathrm{Binom}(n-1, p)$, this random graph model is sometimes also referred to as the binomial model. We point out that various other types of random graphs have been studied in the literature; we will discuss *random regular graphs*, another class of random graphs, in the next chapter.

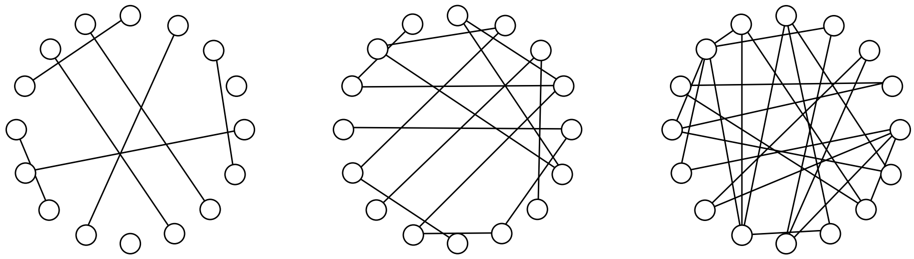

Figure 3.1: Three realizations of $G(16, p)$, with increasing p.

3.2 Preliminaries

Theorem 3.1 (Almost every $G(n, p)$ is connected). *For constant p, $G(n, p)$ is connected* a.a.s.

Proof:

If G is disconnected, then there exists a bypartition of $V(G) = S \cup \bar{S}$ such that there are no edges between S and $\bar{S}$. We can union-bound the probability that there is no such partition $(S, \bar{S})$ by summing over all possible partitions.

Fix $|S| = s$. There are $s(n-s)$ possible edges connecting a node in S to a node in $\bar{S}$, so $\mathbb{P}\left\{S \text{ and } \bar{S} \text{ are disconnected}\right\} = (1-p)^{s(n-s)}$.

The probability that $G(n, p)$ is disconnected is at most $\sum_{s=1}^{n/2} \binom{n}{s} (1-p)^{s(n-s)}$ (note that we do not need to sum beyond $n/2$). Using the bounds $\binom{n}{s} \leq n^s$ and $(1-p)^{n-s} \leq (1-p)^{n/2}$, we find $\mathbb{P}\{G(n,p)\text{disconnected}\} < \sum_{s=1}^{n/2} (n(1-p)^{n/2})^s$.

For n large enough, $x = n(1-p)^{n/2} < 1$, and the sum above is a convergent geometric series $\sum_{k=1}^{n/2} x^s < x/(1-x)$. Since $x \to 0$, the probability that $G(n, p)$ is disconnected $\to 0$ as well.

■

The above union bound is very loose, as graphs with many components are counted several times. We now illustrate two methods that are used frequently to prove results of the above type. Specifically, we often face the task of proving that a graph $G(n, p)$ has some property either with probability going to zero or to one. We assume here that X_n is an integer ≥ 0.

Theorem 3.2 (First Moment Method). *If $\mathbb{E}[X_n] \to 0$, then $X_n = 0$* a.a.s.

Proof:

Apply the Markov inequality $\mathbb{P}\{X \geq x\} \leq \mathbb{E}[X]/x$ with $x = 1$.

■

Theorem 3.3 (Second Moment Method). *If $\mathbb{E}[X_n] > 0$ for n large and $Var[X_n]/(\mathbb{E}[X_n])^2 \to 0$, then $X_n > 0$* a.a.s.

Proof:

Chebyshev's inequality states that if $\text{Var}[X]$ exists, then $\mathbb{P}\{|X - \mathbb{E}[X]| \geq x\} \leq \text{Var}[X]/x^2, x > 0$. The result follows by setting $x = \mathbb{E}[X]$.

■

We now illustrate the use of this approach by deriving the following result that implies the preceding result, but is stronger because it also establishes that the diameter of $G(n,p)$ is very small.

Theorem 3.4 (Almost every $G(n,p)$ has diameter 2)**.** *For constant p, $G(n,p)$ is connected and has diameter* 2 a.a.s.

Proof:

Let X be the number of (unordered) vertex pairs with no common neighbor. To prove the theorem, we need to show that $X = 0$ *a.a.s.*

We apply the first-moment method of Theorem 3.2 above. Let $X_{u,v}$ an indicator variable with $X_{u,v} = 1$ if u and v do not have a common neighbor, and $X_{u,v} = 0$ if they do.

For a vertex pair u, v, $X_{u,v} = 1$ if and only if none of the other $n-2$ vertices is adjacent to both u and v. Therefore, $\mathbb{P}\{X_{u,v} = 1\} = (1-p^2)^{n-2}$, and therefore $\mathbb{E}[X] = \mathbb{E}\left[\sum_{u,v} X_{u,v}\right] = \binom{n}{2}(1-p^2)^{n-2}$. This expression goes to zero with n for fixed p, establishing $\mathbb{P}\{X = 0\} \to 1$.

■

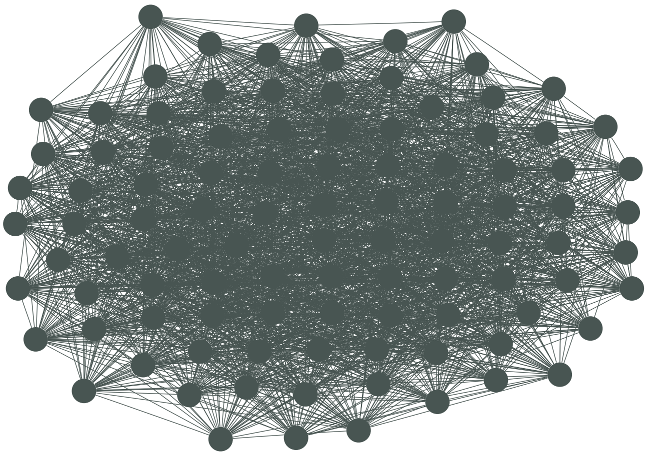

Figure 3.2: An instance of $G(100, 0.5)$, a very dense graph with diameter 2.

Definition 3.2 (Increasing property and threshold function)**.** *An increasing property is a graph property conserved under the addition of edges. A function $t(n)$ is a threshold function for an increasing property if (a) $p(n)/t(n) \to 0$ implies that $G(n,p)$ does not possess the property* a.a.s.*, and if (b) $p(n)/t(n) \to \infty$ implies that it does* a.a.s.

Note that threshold functions are never unique; for example, if $t(n)$ is a threshold function, then so is $ct(n), c > 0$. Examples of increasing properties are:

1. A fixed graph H appears as a subgraph in G.

2. There exists a large component of size $\Theta(n)$ in G.

3. G is connected.

4. The diameter of G is at most d.

A counterexample is the appearance of H as an *induced* subgraph of G (that is, a subset of the vertex set of G together with all the edges of G whose end-vertices are both in this subset): indeed, the addition of edges in G can destroy this property.

Definition 3.3 (Balanced graph)**.** *The ratio $2e(H)/|H|$ for a graph H is called its average vertex degree. A graph G is balanced if its average vertex degree is larger than or equal to the average vertex degree of any of its induced subgraphs.*

Note that trees, cycles, and complete graphs are all balanced.

Definition 3.4 (Automorphism group of a graph)**.** *An automorphism of a graph G is an isomorphism from G to G, i.e., a permutation Π of its vertex set such that $(u, v) \in E(G)$ if and only if $(\Pi(u), \Pi(v)) \in E(G)$.*

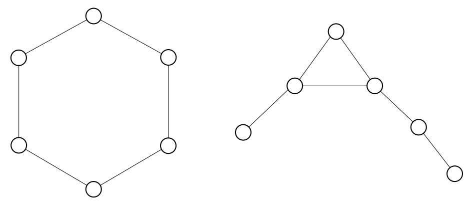

Figure 3.3: Two graphs with 6 vertices and 6 edges, the first with an automorphism group of size $a = 12$, the second with $a = 1$.

Lemma 3.1 (Chernoff bounds for Binomial RVs)**.** *For $X \sim Binom(n, p)$,*

$$\mathbb{P}\{X \geq \mathbb{E}[X] + x\} \leq \exp\left(-\frac{x^2}{2(np + x/3)}\right)$$
$$\mathbb{P}\{X \leq \mathbb{E}[X] - x\} \leq \exp\left(-\frac{x^2}{2np}\right) \tag{3.1}$$

We will also need the following theorem, which counts the number of different trees with n vertices.

Theorem 3.5 (Cayley's Formula)**.** *There are n^{n-2} labeled trees of order n.*

3.3 Appearance of a subgraph

We now study the following problem: given an unlabeled graph H, what is the probability that this graph H is a subgraph of $G(n, p)$ when $n \to \infty$? This question has a surprisingly simple answer: we identify a threshold function for the appearance of H in $G(n, p)$ that only depends on the number of vertices and edges in H, with the caveat that H has to be balanced.

Theorem 3.6 (Threshold function for appearance of balanced subgraph)**.** *For a balanced graph H with k vertices and l edges ($l \geq 1$), the function $t(n) = n^{-k/l}$ is a threshold function for the appearance of H as a subgraph of G.*

To be a bit more precise, a graph H appears in G if and only if there is at least one subgraph of G that is isomorphic to H.

Proof:

The proof has two parts. In the first part, we show that when $p(n)/t(n) \to 0$, the H is not contained in $G(n,p)$ *a.a.s.*In the second part, we show that when $p(n)/t(n) \to \infty$, then the opposite is true.

Part 1: $p(n)/t(n) \to 0$.

Set $p(n) = o_n n^{-k/l}$, where o_n goes to zero arbitrarily slowly. Let X denote the number of subgraphs of G isomorphic to H (i.e., the number of "copies" of H in G).

We need to show that $\mathbb{P}\{X = 0\} \to 1$. Let A denote the set of labeled graphs H' isomorphic to H, whose vertex label set $[n]$ is the same as that of G.

$$|A| = \binom{n}{k} \frac{k!}{a} \le n^k, \tag{3.2}$$

where a is the size of H's automorphism group.

$$\mathbb{E}[X] = \sum_{H' \in A} \mathbb{P}\{H' \subset G\}. \tag{3.3}$$

As H' is a labeled graph, the probability of it appearing in G is simply the probability that all edges of H' are present, i.e.,

$$\mathbb{P}\{H' \subset G\} = p^l. \tag{3.4}$$

$$\mathbb{E}[X] = |A| p^l \le n^k p^l = n^k (o_n n^{-k/l})^l = o_n^l. \tag{3.5}$$

By the First Moment Method,

$$\mathbb{P}\{H \in G\} = \mathbb{P}\{X \ge 1\} \le \mathbb{E}[X] \le o_n^l \to 0. \tag{3.6}$$

Therefore, H does not appear in $G(n,p)$ *a.a.s.*

Part 2: $p(n)/t(n) \to \infty$.

Set $p(n) = \omega_n n^{-k/l}$, where ω_n goes to ∞ arbitrarily slowly.

We need to show that $\mathbb{P}\{X = 0\} \to 0$. For this, we bound the variance of X.

$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[\left(\sum_{H' \in A} \mathbf{1}_{\{H' \subset G\}}\right)^2\right] = \sum_{(H', H'') \in A^2} \mathbb{P}\{H' \cup H'' \subset G\}. \tag{3.7}$$

As the labeled graph $H' \cup H''$ has $2l - e(H' \cap H'')$ links, so $\mathbb{P}\{H' \cup H'' \subset G\} = p^{2l - e(H' \cap H'')}$.

As H is balanced, we know that any subgraph (induced or not) of H, including $H' \cap H''$, has $e(H' \cap H'')/|H' \cap H| \le e(H)/|H| = l/k$. Therefore, if $|H' \cap H''| = i$, then $e(H' \cap H'') \le il/k$. We partition the set A^2 into classes A_i^2 with identical order of the intersection, i.e.,

$$A_i^2 = \{(H', H'') \in A^2 : |H' \cap H''| = i\} \tag{3.8}$$

$$S_i = \sum_{(H', H'') \in A_i^2} \mathbb{P}\{H' \cup H'' \subset G\}. \tag{3.9}$$

We will show that $\mathbb{E}[X]$ is dominated by $i = 0$, i.e., H' and H'' are disjoint. In this case, the events $\{H' \in G\}$ and $\{H'' \in G\}$ are independent, as they have no edges in common. Thus,

$$\begin{aligned} S_0 &= \sum_{(H',H'')\in A_0^2} \mathbb{P}\{H' \cup H'' \subset G\} \\ &= \sum_{(H',H'')\in A_0^2} \mathbb{P}\{H' \subset G\}\,\mathbb{P}\{H'' \subset G\} \qquad (H' \text{ and } H'' \text{ disjoint}) \\ &\leq \sum_{(H',H'')\in A^2} \mathbb{P}\{H' \subset G\}\,\mathbb{P}\{H'' \subset G\} \\ &= (\mathbb{E}[X])^2. \end{aligned} \tag{3.10}$$

We now examine the contribution to $\mathbb{E}\left[X^2\right]$ of the terms $i \geq 1$. For this, note that for a fixed graph H', the number of H'' such that $|H' \cap H''| = i$ is given by

$$\binom{k}{i}\binom{n-k}{k-i}\frac{k!}{a}, \tag{3.11}$$

as we need to select i nodes from H' to form the intersection, and $k - i$ nodes from the vertices outside H' for the rest; then there are $k!/a$ labelings for H'' isomorphic to H. Also note that it is easy to see that this expression is $O(n^{k-i})$.

We now use this to compute S_i.

$$\begin{aligned} S_i &= \sum_{(H',H'')\in A_i^2} \mathbb{P}\{H' \cup H'' \subset G\} \\ &= \sum_{H'\in A}\ \sum_{H''\in A:|H'\cap H''|=i} \mathbb{P}\{H' \cup H'' \subset G\} \\ &\leq \sum_{H'\in A} \binom{k}{i}\binom{n-k}{k-i}\frac{k!}{a}p^{2l}p^{-il/k} \qquad (\text{as } e(H'\cap H'') \leq il/k) \\ &= |A|\binom{k}{i}\binom{n-k}{k-i}\frac{k!}{a}p^{2l}(\omega_n n^{-k/l})^{-il/k} \\ &\leq |A|p^l c_1 n^{k-i}\frac{k!}{a}p^l\omega_n^{-il/k}n^i \qquad \left(\text{because } \binom{k}{i}\binom{n-k}{k-i} = O(n^{k-i})\right) \\ &= \mathbb{E}[X]\,c_1 n^k\frac{k!}{a}p^l\omega_n^{-il/k} \qquad (\text{using (3.5)}) \\ &\leq (\mathbb{E}[X])c_2\binom{n}{k}\frac{k!}{a}p^l\omega_n^{-il/k} \qquad \left(\text{because } n^k = \Theta\binom{n}{k}\right) \\ &= (\mathbb{E}[X])^2c_2\omega_n^{-il/k} \qquad (\text{using (3.2)}) \\ &\leq (\mathbb{E}[X])^2c_2\omega_n^{-l/k} \end{aligned} \tag{3.12}$$

for n large enough.

$$\mathbb{E}\left[X^2\right]/(\mathbb{E}[X])^2 = S_0/(\mathbb{E}[X])^2 + \sum_{i=1}^{k} S_i/(\mathbb{E}[X])^2 \leq 1 + kc_2\omega_n^{-l/k}, \tag{3.13}$$

and therefore $\mathrm{Var}[X]/\mathbb{E}[X]^2 \to 0$. Therefore, by Theorem 3.3, $X > 0$ *a.a.s.*

■

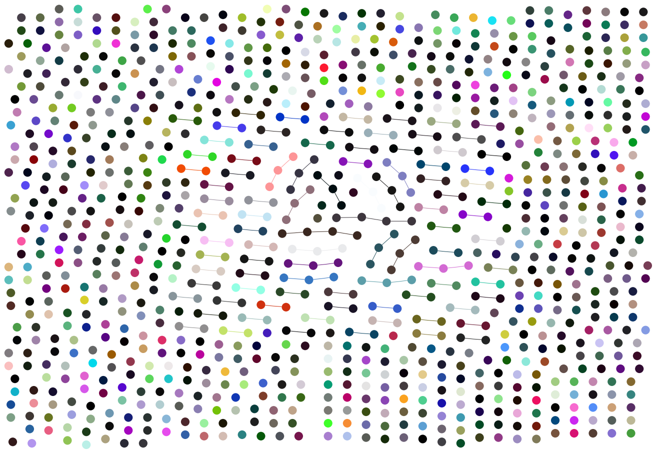

Figure 3.4: An instance of $G(1000, 0.2/1000)$. The graph consists only of small trees.

Corollary 3.1 (Appearance of trees of order k). *The function $t(n) = n^{-k/(k-1)}$ is a threshold function for the appearance of trees of order k.*

Proof:

A tree of order k has k nodes and $k-1$ edges, and it is balanced. The result follows directly from Theorem 3.6 and the fact that there are finitely many trees of order k.

■

Corollary 3.2 (Appearance of cycles of all orders). *The function $t(n) = 1/n$ is a threshold function for the appearance of cycles of any fixed order.*

Corollary 3.3 (Appearance of complete graphs). *The function $t(n) = n^{-2/(k-1)}$ is a threshold function for the appearance of the complete graph K_k with fixed order k.*

3.4 The giant component

After studying the class of threshold functions of the form $n^{-k/l}$ for the appearance of subgraphs, we now focus in more detail on $p(n) = c/n$. Note that this edge probability corresponds to the threshold function for the appearance of cycles of all orders, which suggests that something special is happening at this $p(n)$.

We set $p(n) = c/n$, and study the structure of $G(n, p)$ as a function of c. Specifically, we consider the set of components and their sizes that make up $G(n, p)$. As it turns out, a phase transition occurs at $c = 1$: when c goes from $c < 1$ to $c > 1$, the largest component jumps from $O(\log n)$ to $\Theta(n)$ (this actually occurs as a "double jump", which we do not consider in more detail); the largest component for $c > 1$ is unique.

Let C_v denote the component that vertex v belongs to.

Theorem 3.7 (Small components for $c < 1$). *If $c < 1$, then the largest component of $G(n, p)$ has at most*

$$\frac{3}{(1-c)^2} \log n \tag{3.14}$$

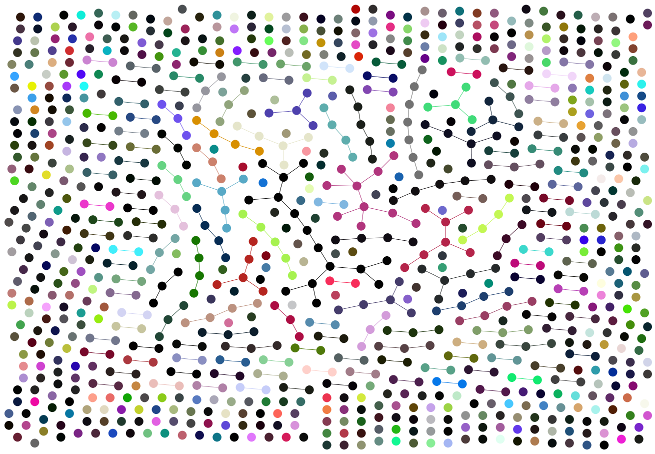

Figure 3.5: An instance of $G(1000, 0.5/1000)$. The graph still only consists of trees, but trees of higher order are appearing.

vertices a.a.s.

Proof:

Let $c = 1 - \epsilon$. We consider a vertex v in $G(n, p)$ and study the following iterative process to successively discover all the vertices of the component that v belongs to. At step i, let the set A_i denote *active* vertices, and the set S_i denote *saturated* vertices, with $A_0 = \{v\}$, and $S_0 = \emptyset$. At the ith step, we select an arbitrary vertex u from A_i. We move u from the active to the saturated set, and mark all neighbors of u that have not been touched yet as active. In this manner, we touch all the vertices in v's component until $A_i = \emptyset$, which is equivalent to $|C_v| = |S_i| = i$.

Let $Y_i = |A_i \cup S_i| = |A_i| + i$ denote the total number of vertices visited by step i, and define $T = \min\{i : Y_i = i\}$, i.e., we have visited all nodes and $A_i = \emptyset$. Then Y_i is a Markov chain with $Y_{i+1} - Y_i \sim \text{Binom}(n - Y_i, p)$, because an edge exists from u to each of the $n - Y_i$ vertices not in $A_i \cup S_i$ independently with probability p, and T is a stopping time for this Markov chain.

We can stochastically upper-bound the process (Y_i) with a random walk (Y_i^+) with increments $X_i^+ \sim \text{Binom}(n, p)$. The corresponding stopping time T^+ for the random walk stochastically dominates T.

We want to bound the probability that vertex v belongs to a component of size at least k. As $|C_v| \geq k \Leftrightarrow Y_k = |A_k \cup S_k| \geq k$,

$$\mathbb{P}\{|C_v| \geq k\} = \mathbb{P}\{T \geq k\} \leq \mathbb{P}\left\{T^+ \geq k\right\} \leq \mathbb{P}\left\{Y_k^+ \geq k\right\}. \quad (3.15)$$

The random walk has $Y_k^+ \sim B(kn, p)$. Using the Chernoff bound (Lemma 3.1) for the binomial distribution and setting $k = (3/\epsilon^2) \log n$, we find

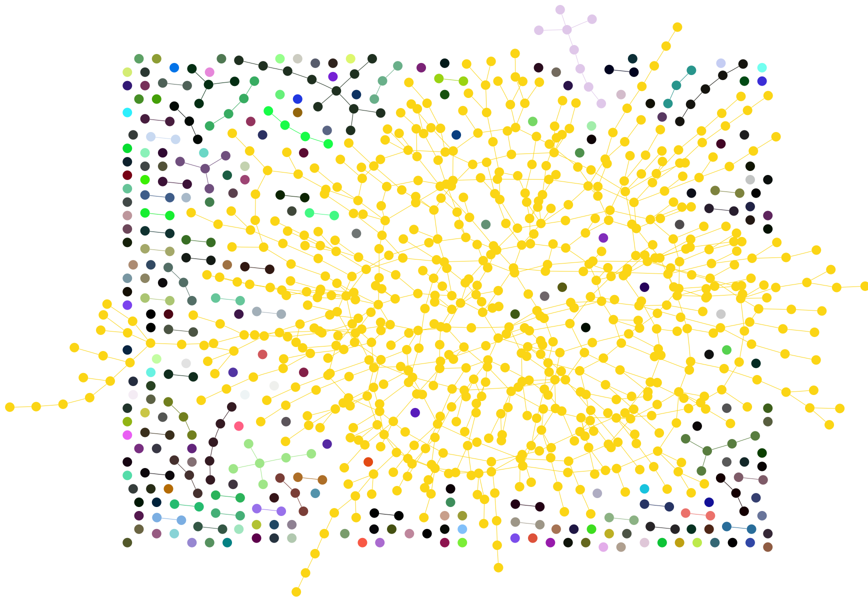

Figure 3.6: An instance of $G(1000, 1.5/1000)$, slightly above the threshold for the appearance of the giant component.

$$
\begin{aligned}
\mathbb{P}\left\{\max_v |C_v| \geq k\right\} &\leq n\mathbb{P}\left\{Y_k^+ \geq k\right\} \\
&= n\mathbb{P}\left\{Y_k^+ \geq ck + \epsilon k\right\} \\
&\leq n \exp\left(-\frac{(\epsilon k)^2}{2(ck + \epsilon k/3)}\right) \\
&\leq n \exp\left(-\frac{\epsilon^2}{2} k\right) = n^{-1/2} = o(1). \qquad (3.16)
\end{aligned}
$$

■

Theorem 3.8 (Unique giant component for $c > 1$)**.** *If $c > 1$, then the largest component of $G(n, p)$ has $\Theta(n)$ vertices, and the second-largest component has $O(\log n)$ vertices* a.a.s.

Proof:

We will study the same process as in the previous proof, starting at an arbitrary vertex v. The proof has three parts. In the first part, we show that it is very likely that the process either dies out early, i.e., $T \leq a_n$ *a.a.s.*, resulting in a small component, or continues for at least b_n steps, resulting in a large component. In the second part, we show that there is only one large component with $k \geq b_n$. In the third part, we confirm that the size of the largest component is of order n.

Part 1: each component is either small or large. Let $c = 1 + \epsilon$, $a_n = \frac{16c}{\epsilon^2} \log n$, and $b_n = n^{2/3}$. We wish to show that the Markov chain Y_k either dies out before a_n (i.e., $T < a_n$), or that for any $a_n < k < b_n$, we have a large number of active nodes A_k left to continue the process, specifically $(\epsilon/2)k$ nodes.

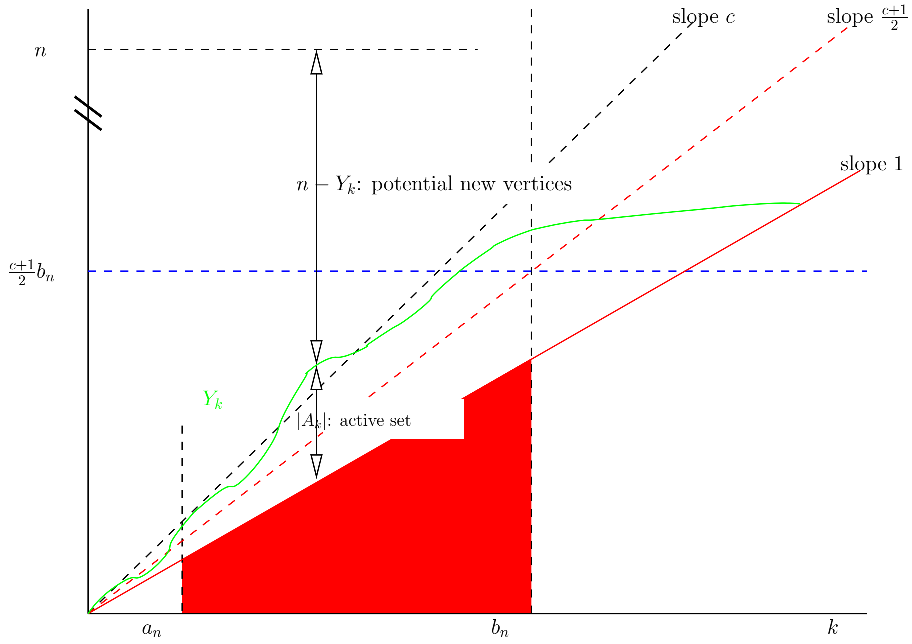

Figure 3.7: An illustration of the different variables involved in the proof for $c > 1$.

The event that we have many vertices left at stage k is

$$\left\{|A_k| \geq \frac{c-1}{2}k\right\} = \left\{Y_k \geq \frac{c+1}{2}k\right\}. \tag{3.17}$$

Fix a starting vertex v, and consider a step $a_n \leq k \leq b_n$. Conditional on $Y_k < \frac{c+1}{2}b_n$ (otherwise the event "many vertices left" occurs), there remain at least $n - \frac{c+1}{2}b_n$ unexplored vertices. Therefore, we can stochastically lower-bound Y_i with a random walk Y_i^- with increments $X_i^- \sim \text{Binom}(n - \frac{c+1}{2}b_n, p)$. Therefore,

$$\mathbb{P}\left\{Y_k \geq \frac{c+1}{2}k\right\} \geq \mathbb{P}\left\{Y_k^- \geq \frac{c+1}{2}k\right\} \tag{3.18}$$

Using this bound, we find

$$\begin{aligned}
\mathbb{P}\left\{\exists v : a_n \leq |C_v| \leq b_n\right\} &\leq n\sum_{k=a_n}^{b_n} \mathbb{P}\left\{Y_k^- < k + \frac{\epsilon}{2}k\right\} \\
&\leq n\sum_{k=a_n}^{b_n} \exp\left(-\frac{\epsilon^2 k^2}{9ck}\right) \\
&\leq nb_n \exp\left(-\frac{\epsilon^2}{9c}a_n\right) = o(1).
\end{aligned} \tag{3.19}$$

Part 2: large component is unique. We now show that the largest component is unique, by considering two vertices u and v that both belong to a component of size larger than b_n, and showing that the probability that they lie in different components in asymptotically zero.

Assume that we run the above process twice, starting from u and from v. We had shown in Part 1 that starting at v, the set $A_{b_n}(v)$ will be of size at least $\epsilon b_n/2$. The same holds for the set of active vertices starting at u.

Now assume that the two processes have not "touched" yet, i.e., have no vertices in common. The probability that they touch at a later step (after b_n) is larger than the probability that they touch in the next step, i.e., that there exists at least one vertex that is adjacent to both active sets $A_{b_n}(u), A_{b_n}(v)$.

$$\begin{aligned}\mathbb{P}\{\text{processes do not touch in next step}\} &= (1-p)^{|A_{b_n}(u)||A_{b_n}(v)|} \\ &\leq (1-p)^{(\epsilon b_n/2)^2} \\ &\leq \exp\left(-\frac{\epsilon^2}{4}cn^{1/3}\right) = o(n^{-2}). \qquad (3.20)\end{aligned}$$

Taking the union bound over all pairs of vertices (u, v) shows that the probability that any two vertices lie in different large components goes to zero, i.e., the giant component is unique *a.a.s.*

Part 3: large component has size $\Theta(n)$. Recall that $b_n = n^{2/3}$. Therefore, to show that the unique giant component is of size $\Theta(n)$, we need to show that all the other vertices only make up at most a constant fraction of all the vertices. For this, we consider a vertex v, and find an upper bound of the probability that it belongs to a small component.

Let N be the number of vertices in small components. Then the size of the giant component is $n - N$. By definition, each small component is smaller than a_n. The probability ρ that C_v is small, i.e., that the process dies before a_n vertices have been reached, is smaller than $\rho^+ = \mathbb{P}\{\text{BP}(\text{Binom}(n - a_n, p)) \text{ dies}\}$, and larger than $\rho^- = \mathbb{P}\{\text{BP}(\text{Binom}(n, p)) \text{ dies}\} - o(1)$, where the $o(1)$ term corresponds to the probability that the process dies too late (after more than a_n vertices have been discovered).

Note that $\text{Binom}(n - o(n), c/n) \to \text{Poisson}(c)$ in distribution. The probability that the process dies out before a_n vertices have been reached is asymptotically equal to $\mathbb{P}\{\text{BP}(\text{Poisson}(c)) \text{ dies}\}$, which is given by $\rho = 1 - \beta$, with $\beta < 1$ the solution of $\beta + e^{-\beta c} = 1$.

Therefore,

$$\mathbb{P}\{C_v \text{small}\} \to \rho, \qquad (3.21)$$

and $\mathbb{E}[N]/n \to \rho$.

To show the result, we need to show that $\mathbb{E}\left[N^2\right] \sim (\mathbb{E}[N])^2$ and use Chebyshev's inequality to bound the probability that N is small. For this, write

$$\begin{aligned}\mathbb{E}\left[N^2\right] &= \mathbb{E}\left[\left(\sum_v \mathbf{1}_{\{C_v \text{ small}\}}\right)^2\right] \\ &= \sum_{u,v} \mathbb{P}\{C_u, C_v \text{ both small}\} \\ &= \sum_v \mathbb{P}\{C_v \text{ small}\} + \sum_{u \neq v} \mathbb{P}\{C_u,\, C_v \text{ small}\} \\ &\leq n\rho + n(n-1)\left(\rho\frac{a_n}{n} + \rho^2\right) \\ &\leq n\rho + n\rho a_n + n^2\rho^2 \qquad (3.22)\end{aligned}$$

and therefore $\mathbb{E}\left[N^2\right] \leq (1 + o(1))n^2\rho^2$. Thus, $\text{Var}[N] = o((\mathbb{E}[N])^2)$, which shows (through Chebyshev) that N is concentrated around its mean. This completes the proof.

■

3.5 Connectivity

We have seen in the preceding section that a unique giant component appears around $np = 1$. It is much harder to achieve full connectivity, such that the graph possesses a single component encompassing all vertices. We will now show that the threshold for connectivity is $t(n) = \log n/n$, i.e., when the average node degree hits $\log n$.

It is also interesting to understand what happens between the critical probability for the giant component and the threshold for full connectivity. In fact, it can be shown that as we increase $p(n)$, the giant component consumes the remaining smaller components in descending order. The small components are in fact small trees, and there are threshold functions for the disappearance of trees of given order between the thresholds for giant component and that for full connectivity, in analogy with the appearance of trees in order below the threshold for the giant component.

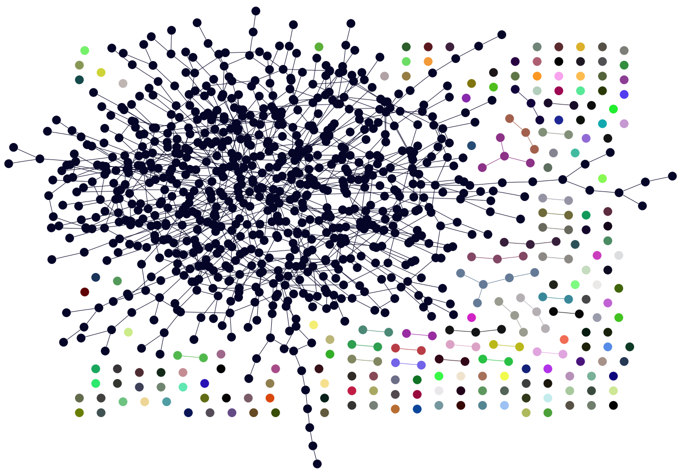

Figure 3.8: An instance of $G(1000, 2/1000)$, in between the critical average degree for the giant component and for full connectivity.

Just before we hit full connectivity, the only remaining small components are isolated vertices, which we establish next. Theorem 3.10 below then shows that $t(n) = \log n/n$ is a threshold function for $G(n, p)$ to be connected.

Theorem 3.9. *The function $t(n) = \log n/n$ is a threshold function for the disappearence of isolated vertices in $G(n, p)$.*

Proof:

Let X_i denote the indicator for vertex i to be isolated, and X is the sum of all X_i. We have $\mathbb{E}[X] = n(1-p)^{n-1}$.

First, let $p(n) = \omega_n \log n/n$, with $\omega_n \to \infty$. We have

$$\begin{aligned} \mathbb{E}[X] &\leq ne^{-\omega_n \log n} \\ &= n^{1-\omega_n} \to 0. \end{aligned} \tag{3.23}$$

By the First Moment Method, there are *a.a.s.* no isolated vertices.

Second, let $p(n) = o_n \log n/n$, with $o_n \to 0$.

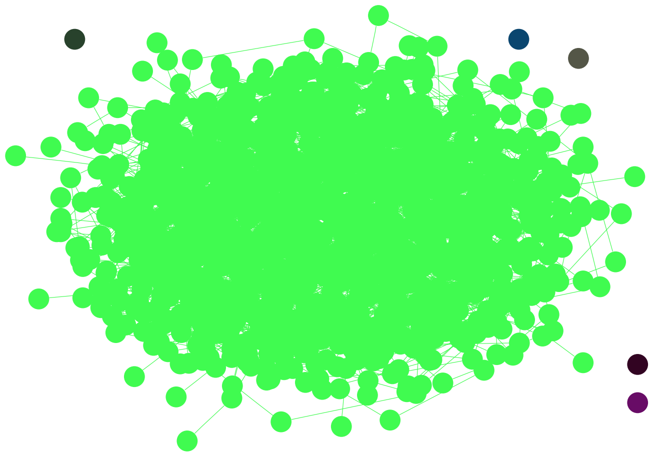

Figure 3.9: An instance of $G(1000, 5/1000)$, slightly below the threshold for full connectivity. The only remaining small components are isolated vertices.

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= \mathbb{E}\left[\sum_{i,j} X_i X_j\right] \\
&= \mathbb{E}\left[X\right] + \mathbb{E}\left[\sum_{i \neq j} X_i X_j\right] \\
&= \mathbb{E}\left[X\right] + n(n-1)(1-p)^{2(n-2)+1} \\
&\sim \mathbb{E}\left[X\right] + n^{2(1-o_n)} \sim \mathbb{E}\left[X\right] + \mathbb{E}\left[X\right]^2 . \qquad (3.24)
\end{aligned}$$

As $\mathbb{E}[X] \to \infty$, this shows that $\mathbb{E}\left[X^2\right] \sim \mathbb{E}[X]^2$, proving the result through the Second Moment Method.

■

Note that this result can actually easily be sharpened by setting $p(n) = c \log n/n$, and studying the phase transition as a function of c, in analogy to the way the giant component appeared.

Theorem 3.10. *The function $t(n) = \log n/n$ is a threshold function for connectivity in $G(n, p)$.*

Proof:

Theorem 3.9 shows immediately that if $p(n) = o(t(n))$, then the graph is not connected, because it still has isolated vertices.

To show the converse, set $p(n) = \omega_n \log n/n$, with $\omega_n \to \infty$ arbitrarily. We know that the RG does not contain isolated vertices. We now show that it does not contain any other small components either, i.e., specifically components of size k at most $n/2$.

We bound the probability that a small component of size between $2 \leq k \leq n/2$ appears. The case $k = 2$ is left as an exercise. For $k > 2$, we note that such a component contains necessarily a tree of order k, which means it contains at least $k - 1$ edges, and none of the the $k(n - k)$ possible edges to other vertices exist.

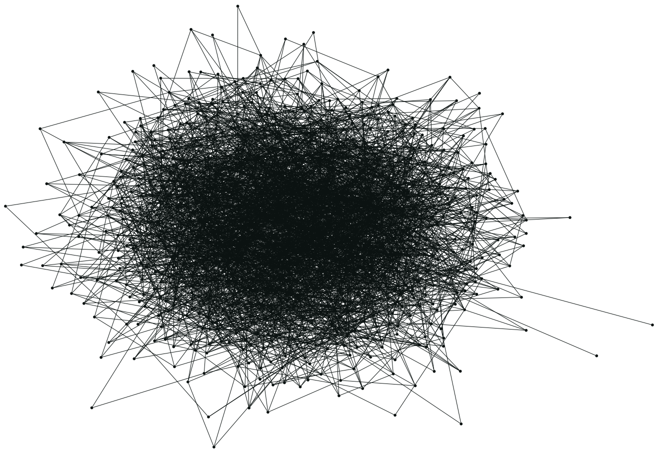

Figure 3.10: An instance of $G(1000, 8/1000)$, slightly above the threshold for full connectivity.

$$
\begin{aligned}
&\mathbb{P}\{G(n,p) \text{ contains component of order } k\} \\
&\leq \binom{n}{k} k^{k-2}p^{k-1}(1-p)^{k(n-k)} \qquad \text{(using Cayley's Formula, c.f. Theorem 3.5)} \\
&\leq k^{-2}\exp\left[k(\log n+1)+(k-1)\left(\log(\omega_n\log n)-\log n\right)-k\omega_n\log n+\frac{k^2}{n}\omega_n\log n\right] \\
&\leq nk^{-2}\exp\left[k+k\log(\omega_n\log n)-\frac{1}{2}k\omega_n\log n\right] \\
&\leq nk^{-2}\exp\left[-(1/3)k\omega_n\log n\right] \qquad \text{(for } n \text{ large enough)} \\
&= k^{-2}n^{1-k\omega_n/3},
\end{aligned}
\tag{3.25}
$$

using $\binom{n}{k} = (n)_k/(k)_k \leq n^k/(k/e)^k = (ne/k)^k$ (from Stirling's approximation), and using $\log(1-p) \leq -p$.

$$
\begin{aligned}
&\mathbb{P}\{G(n,p) \text{ contains components of order } 3 \leq k \leq n/2\} \\
&\leq \sum_{k=2}^{n/2} k^{-2}n^{1-k\omega_n/3} = O(n^{2-2\omega_n/3}) = o(1).
\end{aligned}
\tag{3.26}
$$

Therefore, the RG contains no components smaller than $n/2$.

■

The following figures illustrate the evolution between $p(n) = 1/n$ and $p(n) = \log n/n$. Figure 3.8 shows an instance of $G(1000, 0.002)$, roughly halfway between the two thresholds.

Figures 3.9 and 3.10 show instances of $G(n,p)$ just below and just above the threshold (of approx. 6.9/1000) for full connectivity.