Chapter 1

Erdős-Rényi Random Graphs

In this chapter we will introduce and study the random graph model introduced by Erdős and Rényi in the late 1950’s. This example has been extensively studied and a very nice account of many of the results can be found in the classic book of Bollobás (2001), so here we will give a brief account of the main results on the emergence of a giant component, in order to prepare for the analysis of more complicated examples. In contrast to other treatments, we mainly rely on methods from probability and stochastic processes rather than combinatorics.

To define the model, we begin with the set of vertices $V = \{1, 2, \ldots, n\}$. For $1 \leq x < y \leq n$ let $\eta_{x,y}$ be independent $= 1$ with probability $p$ and 0 otherwise. Let $\eta_{y,x} = \eta_{x,y}$. If $\eta_{x,y} = 1$ there is an edge from $x$ to $y$. Here, we will be primarily concerned with situation $p = \lambda/n$ and in particular with showing that when $\lambda < 1$ all of the components are small, with the largest $O(\log n)$, while for $\lambda > 1$ there is a giant component with $\sim g(\lambda)n$ vertices. The intuition behind this result is that a site has a $\text{Binomial}(n - 1, \lambda/n)$ number of neighbors, which has mean $\approx \lambda$. Suppose that we start with $I_0 = 1$, and for $t \geq 1$ let $I_t$ be the set of vertices not in $\bigcup^{t-1}_{s=0} I_s$ that are connected to some site in $I_{t-1}$. Then when $t$ is not too large the number of points in $I_t$, $Z_t = |I_t|$, is approximately a branching process in which each individual in generation $t$ has an average of $\lambda$ children. If $\lambda < 1$ the branching process dies out quickly and all components are small. When $\lambda > 1$, the branching process survives with probability $g(\lambda)$, and all sites with surviving branching processes combine to make the giant component.

1.1 Branching processes

In this section we define branching processes and gives their basic properties. Since $n$ will be the number of vertices in our graph, we will use $t$ and $s$ for our discrete time parameter. Let $\xi_i, i, t \geq 0$, be i.i.d. nonnegative integer-valued random variables. Define a sequence $Z_t, t \geq 0$ by $Z_0 = 1$ and

$$Z_{t+1} = \begin{cases} \xi_1^{t+1} + \cdots + \xi_{Z_t}^{t+1} & \text{if } Z_t > 0 \\ 0 & \text{if } Z_t = 0 \end{cases}$$
Z_t is called a Galton-Watson process. The idea behind the definition is that Z_t is the number of individuals in the tth generation, and each member of the tth generation gives birth independently to an identically distributed number of children. p_k = P(\xi_i^t = k) is called the offspring distribution.

**Lemma 1.1.1.** Let \( F_t = \sigma(\xi_i^s : i \geq 1, 1 \leq s \leq t) \) and \( \mu = E\xi_i^t \in (0, \infty) \). Then \( Z_t/\mu^t \) is a martingale w.r.t. \( F_t \).

**Proof.** Clearly, \( Z_t \) is measurable with respect to \( F_t \), or \( Z_t \in F_t \). Recall, see Theorem 4.1.2 in PTE5, that if \( X = Y \) on \( B \in F \) then \( E(X|F) = E(Y|F) \) on \( B \). On \( \{Z_t = k\} \),

\[
E(Z_{t+1}|F_t) = E(\xi_i^{t+1} + \cdots + \xi_k^{t+1}|F_t) = k\mu = \mu Z_t
\]

Dividing both sides by \( \mu^{t+1} \) now gives the desired result. \( \Box \)

**Theorem 1.1.2.** If \( \mu < 1 \) then \( Z_t = 0 \) for all \( t \) sufficiently large.

**Proof.** \( E(Z_t/\mu^t) = E(Z_0) = 1 \), so \( E(Z_t) = \mu^t \). Now \( Z_t \geq 1 \) on \( \{Z_t > 0\} \) so

\[
P(Z_t > 0) \leq E(Z_t; Z_t > 0) = E(Z_t) = \mu^t \rightarrow 0
\]

exponentially fast if \( \mu < 1 \).

The last answer should be intuitive. If each individual on the average gives birth to less than one child, the species will die out. The next result shows that after we exclude the trivial case in which each individual has exactly one child, the same result holds when \( \mu = 1 \).

**Theorem 1.1.3.** If \( \mu = 1 \) and \( P(\xi_i^t = 1) < 1 \) then \( Z_t = 0 \) for all \( t \) sufficiently large.

**Proof.** When \( \mu = 1 \), \( Z_t \) is itself a nonnegative martingale, so the martingale convergence theorem, Theorem 4.2.12 in PTE5, implies that \( Z_t \) converges to an a.s. finite limit \( Z_\infty \). Since \( Z_t \) is integer valued, we must have \( Z_t = Z_\infty \) for large \( t \). If \( P(\xi_i^t = 1) < 1 \) and \( k > 0 \) then \( P(Z_t = k \text{ for all } t \geq T) = 0 \) for any \( T \), so we must have \( Z_\infty \equiv 0 \).

**Theorem 1.1.4.** If \( \mu > 1 \) then \( P(Z_t > 0 \text{ for all } t) > 0 \).

**Proof.** For \( \theta \in [0, 1] \), let \( \phi(\theta) = \sum_{k \geq 0} p_k \theta^k \) where \( p_k = P(\xi_i^t = k) \). \( \phi \) is the generating function for the offspring distribution \( p_k \). Differentiating gives for \( \theta < 1 \)

\[
\phi'(\theta) = \sum_{k=1}^{\infty} k p_k \theta^{k-1} \geq 0
\]

\[
\phi''(\theta) = \sum_{k=2} \sum_{k=1}^{\infty} k(k-1)p_k \theta^{k-2} \geq 0
\]

So \( \phi \) is increasing and convex, and \( \lim_{\theta \uparrow 1} \phi'(\theta) = \sum_{k=1}^{\infty} kp_k = \mu \). Our interest in \( \phi \) stems from the following facts.
(a) If $\theta_t = P(Z_t = 0)$ then $\theta_t = \sum_{k=0}^{\infty} p_k \theta_{t-1}^k = \phi(\theta_{t-1})$

Proof of (a). If $Z_1 = k$, an event with probability $p_k$, then $Z_t = 0$ if and only if all $k$ families die out in the remaining $t - 1$ units of time, independent events with probability $\theta_{t-1}^k$. Summing over the disjoint possibilities for each $k$ gives the desired result.

(b) If $\phi'(1) = \mu > 1$ there is a unique $\rho < 1$ so that $\phi(\rho) = \rho$.

Proof of (b). $\phi(0) \geq 0$, $\phi(1) = 1$, and $\phi'(1) > 1$, so $\phi(1 - \epsilon) < 1 - \epsilon$ for small $\epsilon$. The last two observations imply the existence of a fixed point. To see it is unique, observe that $\mu > 1$ implies $p_k > 0$ for some $k > 1$, so $\phi''(\theta) > 0$ for $\theta > 0$. Since $\phi$ is strictly convex, it follows that if $\rho < 1$ is the smallest fixed point, then $\phi(x) < x$ for $x \in (\rho, 1)$.

(c) As $t \uparrow \infty$, $\theta_t \uparrow \rho$.

Proof of (c). $\theta_0 = 0$, $\phi(\rho) = \rho$, and $\phi$ is increasing, so induction implies $\theta_t$ is increasing and $\theta_t \leq \rho$. Let $\theta_\infty = \lim \theta_t$. Taking limits in $\theta_t = \phi(\theta_{t-1})$, we see $\theta_\infty = \phi(\theta_\infty)$. Since $\theta_\infty \leq \rho$, it follows that $\theta_\infty = \rho$.

Combining (a)–(c) shows $P(Z_t = 0$ for some $t) = \lim_{t \to \infty} \theta_t = \rho < 1$.

Example. Consider the Poisson distribution with mean $\lambda$, i.e,

$$P(\xi = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$
In this case \( \phi(s) = \sum_{k=0}^{\infty} e^{-\lambda s^k} \lambda^k / k! = \exp(\lambda(s - 1)) \) so the fixed point equation is

\[
\rho = \exp(\lambda(\rho - 1)) \tag{1.1.1}
\]  

Theorem 1.1.4 shows that when \( \mu > 1 \), the limit of \( Z_t / \mu^t \) has a chance of being nonzero. The next theorem is not the best possible but it suffices for or purposes. For a proof see Example 4.4.9 in PTE5.

**Theorem 1.1.5.** If \( \sum_k p_k > 1 \) and \( \sum k^2 p_k < \infty \) then \( W = \lim Z_t / \mu^t \) is not \( \equiv 0 \).

The ultimate result about this question is due to Kesten and Stigum. For a “conceptual proof” see Lyons, Pemantle, and Peres (1995).

**Theorem 1.1.6.** Suppose \( \sum_k p_k > 1 \) then \( W = \lim Z_t / \mu^t \) is not \( \equiv 0 \) if and only if \( \sum_{k=1}^{\infty} k \log k p_k < \infty \).

Our next result shows that when \( W \) is not \( \equiv 0 \) it is positive on the set where the branching process does not die out.

**Theorem 1.1.7.** If \( P(W = 0) < 1 \) then \( \{W > 0\} = \{Z_n > 0 \text{ for all } n \} \), i.e., the symmetric difference of the two sets has probability 0.

**Proof.** Let \( \rho = P(W = 0) \). In order for \( Z_t / \mu^t \) to converge to 0 this must also hold for the branching process started by each of the children in the first generation. Breaking things down according to the number of children in the first generation

\[
\rho = \sum_{k=0}^{\infty} p_k \rho^k = \phi(\rho)
\]

so \( \rho < 1 \) is a fixed point of the generating function and hence \( \rho = P(Z_t = 0 \text{ for some } t) \). Clearly, \( \{W > 0\} \subset \{Z_t > 0 \text{ for all } t\} \). Since the two sets have the same probability \( P(\{Z_t > 0 \text{ for all } t\} - \{W > 0\}) = 0 \), which is the desired result.

The limit theorems above describe the growth of the process when it does not die out. Our next question is: what happens in a supercritical branching process when it dies out?

**Theorem 1.1.8.** A supercritical branching process conditioned to become extinct is a subcritical branching process. If the original offspring distribution is Poisson(\( \lambda \)) with \( \lambda > 1 \) then the conditioned one is Poisson(\( \lambda \rho \)) where \( \rho \) is the extinction probability.

**Proof.** Let \( T_0 = \inf\{t : Z_t = 0\} \) and consider \( \tilde{Z}_t = (Z_t|T_0 < \infty) \). To check the Markov property for \( \tilde{Z}_t \) note that the Markov property for \( Z_t \) implies:

\[
P(Z_{t+1} = z_{t+1}, T_0 < \infty | Z_t = z_t, \ldots, Z_0 = z_0) = P(Z_{t+1} = z_{t+1}, T_0 < \infty | Z_t = z_t)
\]
To compute the transition probability for $\bar{Z}_t$, observe that if $\rho$ is the extinction probability then $P_x(T_0 < \infty) = \rho^x$. Let $p(x, y)$ be the transition probability for $Z_t$. Note that the Markov property implies

$$\bar{p}(x, y) = \frac{P_x(Z_1 = y, T_0 < \infty)}{P_x(T_0 < \infty)} = \frac{P_x(Z_1 = y)P_y(T_0 < \infty)}{P_x(T_0 < \infty)} = \frac{p(x, y)\rho^y}{\rho^x}$$

Taking $x = 1$ and computing the generating function

$$\sum_{y=0}^{\infty} \bar{p}(1, y)\theta^y = \rho^{-1}\sum_{y=0}^{\infty} p(1, y)(\theta \rho)^y = \rho^{-1}\phi(\theta \rho)$$

where $p_y = p(1, y)$ is the offspring distribution.

$\bar{p}_y = \bar{p}(1, y)$ is the distribution of the size of the family of an individual, conditioned on the branching process dying out. If we start with $x$ individuals then in $Z_n$ each gives rise to an independent family. In $\bar{Z}_n$ each family must die out, so $\bar{Z}_n$ is a branching process with offspring distribution $\bar{p}(1, y)$. To prove this formally observe that

$$p(x, y) = \sum_{j_1, \ldots, j_x \geq 0, j_1 + \cdots + j_x = y} p_{j_1} \cdots p_{j_x}$$

Writing $\sum_*$ as shorthand for the sum in the last display

$$\frac{p(x, y)\rho^y}{\rho^x} = \sum_* \frac{p_{j_1} \rho^{j_1} \cdots p_{j_x} \rho^{j_x}}{\rho} = \sum_* \bar{p}_{j_1} \cdots \bar{p}_{j_x}$$

In the case of the Poisson distribution $\phi(s) = \exp(\lambda(s - 1))$ so if $\lambda > 1$, using the fixed point equation (1.1.1)

$$\frac{\phi(s\rho)}{\rho} = \frac{\exp(\lambda(s\rho - 1))}{\exp(\lambda(\rho - 1))} = \exp(\lambda\rho(s - 1))$$

which completes the proof. \[\square\]

Geometrically, a supercritical branching process conditioned to die out is a branching process with a generating function that is obtained by taking the graph of $\phi$ over $[0, \rho]$ and rescaling to make the domain and range $[0, 1]$. In the next result we take the graph of $\phi$ over $[\rho, 1]$ and rescale to make the domain and range $[0, 1]$. 

Theorem 1.1.9. Consider a supercritical branching process with offspring distribution $p_k$ and generating function $\phi$. If we condition on nonextinction and look only at the individuals that have an infinite line of descent then the number of individuals in generation $t$, $\tilde{Z}_t$ is a branching process with offspring generating function

$$\tilde{\phi}(\theta) = \frac{\phi((1 - \rho)\theta + \rho) - \rho}{1 - \rho}$$

where $\rho$ is the extinction probability, i.e., the smallest solution of $\phi(\rho) = \rho$ in $[0, 1]$.

Proof. There is nothing to prove if $\rho = 0$ so suppose $0 < \rho < 1$. If $Z_0 = 1$ and we condition on survival of the branching process, then the number of individuals in the first generation who have an infinite line of descent has distribution

$$\tilde{p}_j = \frac{1}{1 - \rho} \sum_{k=j}^{\infty} p_k \binom{k}{j} (1 - \rho)^j \rho^{k-j}$$

Multiplying by $\theta^j$, summing, and interchanging the order of summation

$$\sum_{j=1}^{\infty} \tilde{p}_j \theta^j = \frac{1}{1 - \rho} \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} p_k \binom{k}{j} (1 - \rho)^j \rho^{k-j} \theta^j$$

$$= \frac{1}{1 - \rho} \sum_{k=1}^{\infty} p_k \sum_{j=1}^{k} \binom{k}{j} (1 - \rho)^j \theta^j \rho^{k-j}$$

Using the binomial theorem and noticing that the $j = 0$ term is missing the above

$$= \frac{1}{1 - \rho} \sum_{k=1}^{\infty} p_k \{(1 - \rho)\theta + \rho)^k - \rho^k\}$$
1.1. BRANCHING PROCESSES

We can add the $k = 0$ term to the sum since its value is 0. Having done this the result is

$$\frac{\phi((1 - \rho)\theta + \rho) - \phi(\rho)}{1 - \rho}$$

Since $\phi(\rho) = \rho$ the result follows. ∎
1.2 Cluster growth as an epidemic

In this section we use branching process results to study the growth of the connected component, or cluster, containing 1, a process which is the same as a discrete time epidemic. To begin the construction, we let $S_0 = \{2, 3, \ldots, n\}$, $I_0 = \{1\}$, and $R_0 = \emptyset$. The letters are motivated by the epidemic interpretation of the growing cluster. $S_t$ are the susceptibles, $I_t$ are infected, and $R_t$ are removed (recovered and immune to further infection). In graph terms, we have already examined the connections of all sites in $R_t$, $I_t$ are the sites to be investigated on this turn, and $S_t$ are unexplored. These sets evolve as follows:

\begin{align*}
R_{t+1} &= R_t \cup I_t \\
I_{t+1} &= \{y \in S_t : \eta_{x,y} = 1 \text{ for some } x \in I_t\} \\
S_{t+1} &= S_t - I_{t+1}
\end{align*}

where $\eta_{x,y} = \eta_{y,x} = 1$ if and only if there is an edge between $x$ and $y$. The cluster containing 1, $C_1 = \bigcup_{t=0}^{\infty} I_t$.

Kendall (1956) was the first to suggest a branching process approximation for epidemics. We begin with the easy result

**Theorem 1.2.1.** If $\lambda < 1$ then $E|C_1| \leq 1/(1 - \lambda) < \infty$.

**Proof.** To define a comparison branching process we introduce a new independent set of variables $\zeta_{t,x,y}, x, t \geq 1, 1 \leq x, y \leq n$ that are independent, $= 1$ with probability $\lambda/n$, and 0 otherwise. Let $Z_0 = 1, S_t^c = \{1, 2, \ldots, n\} - S_t$ and

\begin{align*}
Z_{t+1} = \sum_{x \in I_t, y \in S_t} \eta_{x,y} + \sum_{x \in I_t, y \in S_t^c} \zeta_{x,y} + \sum_{x=n+1}^{n+|I_t|} \sum_{y=1}^{n} \zeta_{x,y} \tag{1.2.2}
\end{align*}

Here, the second term is the set of extra births in the branching process due to the fact that $|S_t| < n$. The third term produces births for these individuals that are in the branching process but not in the epidemic.

It is immediate from the construction that $Z_t$ is a branching process with offspring distribution Binomial$(n, \lambda/n)$ and

\begin{equation*}
Z_t \geq |I_t|
\end{equation*}

It is easy to take care of the case $\lambda < 1$. $EZ_t = \lambda^t$, so the mean cluster size $E|C_1| = E(\sum_{t=0}^{\infty} |I_t|) \leq \sum_{t=0}^{\infty} \lambda^t = 1/(1 - \lambda)$.

**Theorem 1.2.2.** Suppose $\lambda > 1$ and let $\rho > 0$ be the probability the branching process does not die out. Then with high probability there is a cluster of size $\sim \rho n$ as $n \to \infty$.

**Proof.** To prove survival for $\lambda > 1$ we want to lower bound the growing cluster by a branching process. If we let the branching process run without supervision then there will be lots of collisions between children, so we will instead run the branching process using a
1.2. CLUSTER GROWTH AS AN EPIDEMIC

breadth-first search. That is, we first generate the children of the ancestor which are born at time \( t = 1 \), let the number be \( Y_1 \), and write them in an ordered list. We generate the offspring of the first individual in the list which are born at time 2, then the offspring of the second individual that are born at time 3, and so until we have generated all the children of the first generation. We then order the \( Y_2 \) individuals in the second generation and generate their offspring one at a time. Note that there are two time scales here. \( t \) counts the number of individuals that have given birth, \( m \rightarrow Y_m \) gives the size of the \( m \)th generation, which is known at time \( T_m \) on the birth counting time scale.

Lemma 1.2.3. Let \( \delta \) be chosen so that \( \bar{\lambda} = \lambda(1 - \delta) > 1 \). If at time \( T_m \) the growing cluster does not yet \( \delta n \) vertices in \( V_k \) the process \( Y_\ell, \ell \leq m \) dominates a branching process with a binomial\((1 - \delta)n, \lambda/n\) offspring distribution and hence mean \( \bar{\lambda} > 1 \).

Proof. Suppose that at the \( t \)th step of the construction the growing cluster does not yet contain \( \delta n \) vertices. Then we can pick a set \( G_t \) of susceptible vertices with \( |G_t| = (1 - \delta)n \), that is, they are not yet part of the growing cluster. We allow the vertex \( x \) to give birth onto each \( y \in G_k \) with probability \( \lambda/n \) independent of everything that has happened so far in the construction. This completes the proof of the lemma.

There is no need to reduce the number of possible births at time \( t = 1 \) to \( (1 - \delta)n \) but if we do then \( Y_m \) is exactly a branching process with a binomial\((1 - \delta)n, \lambda/n\) offspring distribution. We say that the process \( Y_m \) survives if the total number of offspring reaches \( n^{2/3} \) at a time we call \( T_n \). Note that if \( n \) is large it follows that the growing cluster does not contain \( \delta n \) vertices by time \( T_n \) so the branching process comparison is valid.

At this point we need to argue that if the branching process survives then there are a large number of particles at time \( T_n \). When the branching process does not die out, \( Y_t/\bar{\lambda}^t \to W \) as \( t \to \infty \), where \( W \) is a random variable that is positive on the survival set. The total progeny at time \( t \)

\[
X_t = \sum_{s=0}^{t} Y_s \approx \frac{W\bar{\lambda}^t}{1 - 1/\bar{\lambda}} = \frac{Y_t}{1 - 1/\bar{\lambda}}
\]

Taking \( t = T_n \), we have that with high probability

\[
Y(T_n) \geq n^{2/3}(1 - 1/\bar{\lambda})
\]

Suppose now that we let \( Y_m \) and \( Y'_m \) be lower bounds on two growing clusters that do not die out and run to times \( T_n \) and \( T'_n \). There are two cases: (a) they have already intersected or (b) the sets \( Y(T_m) \) and \( Y'(T'_m) \) are disjoint sets \( V_1 \) and \( V_2 \) of active sites of size \( \geq \epsilon \delta n^{2/3} \). The probability of no edge connecting \( V_1 \) and \( V_2 \) is

\[
= \left(1 - \frac{\lambda}{n}\right)^\epsilon \delta n^{4/3} \leq \exp(-\eta_4 n^{1/3})
\]

In the next section we will have better estimates that will give a stronger result. However here we will content ourself to observe that if the cluster growth survives starting from \( x \)
then with high probability each other \( y \) that survives will intersect with it. If we let \( \rho_\delta \) be the survival probability of the lower bounding branching process then we can conclude that in the limit the fraction of vertices in the giant component is \( \geq \rho_\delta \). Letting \( \delta \to 0 \) gives the desired lower bound.

The upper bound is easy. Fix an integer \( k \) and declare that a vertex is part of the giant component if its branching process survives to time \( k \). An easy second moment argument. See Step 4 in the next section shows that the limiting fraction of vertices in the giant components is \( \leq \rho_k \) and letting \( k \to \infty \) completes the proof.

The proof here requires much more work than the Erdős-Rényi case deserves but as we will see in Section 3.1, this proof extends immediately to inhomogeneous graphs with finitely many types.
1.3 Cluster growth as a random walk

Although the connection with branching processes is intuitive, it is more convenient technically to expose the cluster one site at a time to obtain something that can be approximated by a random walk. In this section we will introduce that connection and use it to prove the existence of a giant component with $\sim \theta(\lambda)n$ vertices when $p = \lambda/n$ with $\lambda > 1$.

Since we have an emotional attachment to using $S_t$ for a random walk, we will change the previous notation and let

$$R_0 = \emptyset, \ U_0 = \{2, 3, \ldots, n\}, \text{ and } A_0 = \{1\}. \ \ \ \ \text{At time } \tau = \inf\{t : A_t = \emptyset\} \ \text{the process stops. If } A_t \neq \emptyset, \text{ pick } i_t \text{ from } A_t \text{ according to some rule that is measurable with respect to } A_t = \sigma(A_0, \ldots A_t) \text{ and let}$$

$$R_{t+1} = R_t \cup \{i_t\} \ \ \ A_{t+1} = A_t - \{i_t\} \cup \{y \in U_t : \eta_{i_t,y} = 1\} \ \ \ U_{t+1} = U_t - \{y \in U_t : \eta_{i_t,y} = 1\}$$

(1.3.1)

This time $|R_t| = t$ for $t \leq \tau$, so the cluster size is $\tau$.

**Upper bound for $\lambda < 1$.** To define a comparison random walk, we introduce a new independent set of variables $\zeta_y^t$, $t \geq 1$, $y \leq n$ that are independent, $= 1$ with probability $\lambda/n$, and 0 otherwise. Let $S_0 = 1$ and for $t \geq 0$, let $U^c_t = \{1, 2, \ldots, n\} - U_t$

$$S_{t+1} = S_t - 1 + \begin{cases} \sum_{y \in U_t} \eta_{i_t,y} + \sum_{y \in U^c_t} \zeta_y^t & \text{if } A_t \neq \emptyset \\ \sum_{y=1}^n \zeta_y^t & \text{if } A_t = \emptyset \end{cases}$$

$S_t$ is a random walk with $S_t \geq |A_t|$ if $t \leq \tau$, so if $T = \inf\{t : S_t = 0\}$ then $\tau \leq T$.

The increments $X_t$ of the random walk are $-1 + \text{Binomial}(n, \lambda/n)$. If $\lambda < 1$ stopping the martingale $S_t - (\lambda - 1)t$ at the bounded stopping time $T \wedge t$ gives

$$ES_{T \wedge t} - (\lambda - 1)E(T \wedge t) = ES_0 = 1$$

Since $ES_{T \wedge t} \geq 0$, it follows that $E(T \wedge t) \leq 1/(1-\lambda)$. Letting $t \to \infty$ we have $ET \leq 1/(1-\lambda)$. Having verified that $ET < \infty$ we can now use Wald’s equation, see Theorem 4.8.6 in PTE5 to conclude $E(S_T - S_0) = (\lambda - 1)ET$ and hence

$$ET = 1/(1-\lambda) \quad (1.3.2)$$

We can get a much better result by using the exponential martingale:

**Theorem 1.3.1.** Suppose $\lambda < 1$ and let $\alpha = \lambda - 1 - \log(\lambda) > 0$. If $a > 1/\alpha$ then

$$P \left( \max_{1 \leq x \leq n} |C_x| \geq a \log n \right) \to 0$$
Remark. This bound is very accurate. Corollary 5.11 of Bollobás(2001) shows that the size of the largest component is asymptotically

\[ \frac{1}{\alpha} \left( \log n - \frac{5}{2} \log \log n \right) + O(1) \]

Proof. We begin by computing the moment generating function:

\[ \psi(\theta) = E \exp(\theta X_i) = e^{-\theta(1 - \lambda/n + (\lambda/n)e^\theta)} \]

\[ \leq \exp(-\theta + \lambda(e^\theta - 1)) = \psi(\theta) \tag{1.3.3} \]

since \( 1 + x \leq e^x \). Note that the right-hand side is the moment generating function of \(-1 + \text{Poisson mean } \lambda\). \( \psi'(0) = E X_i = 1 - \lambda \) so if \( \lambda < 1 \) then \( \psi(\theta) < 1 \) when \( \theta > 0 \) is small. To optimize we set the derivative

\[ \frac{d}{d\theta}(-\theta + \lambda(e^\theta - 1)) = -1 + \lambda e^\theta = 0 \]

which occurs for \( \theta_1 = -\log \lambda \). At this point \( \psi(\theta_1) = \exp(\log(\lambda) + 1 - \lambda) \equiv e^{-\alpha} < 1 \).

\[ \exp(\theta_1 S_t)/\phi(\theta_1)^t \]

is a nonnegative martingale, so using the optional stopping theorem for the nonnegative supermartingale \( M_t = \exp(\theta_1 S_t)/\psi(\theta_1)^t \), see e.g., Theorem 4.8.4 in PTE5.

\[ 1/\lambda = e^{\theta_1} \geq E(\psi(\theta_1)^{-T}) = E(e^{\alpha T}) \]

Using Chebyshev’s inequality now

\[ P(T \geq k) \leq e^{-k\alpha}/\lambda \tag{1.3.4} \]

Letting \( C_x \) denote the cluster containing \( x \), noting that \( T \geq |C_x| \) in distribution, and taking

\[ k = (1 + \epsilon)(\log n)/\alpha \]

\[ P(|C_x| \geq (1 + \epsilon)(\log n)/\alpha) \leq n^{-(1+\epsilon)/\lambda} \]

from which the desired result follows. \( \square \)

**Lower Bound for** \( \lambda > 1 \). To get a lower bound on the growth of the cluster let \( \hat{U}_t^{\delta} \) consists of the \((1 - \delta)n\) vertices in \( \hat{U}_t \) with the smallest indices. As long as \( \hat{A}_t \neq \emptyset \) and \( \hat{U}_t^{\delta} \geq (1 - \delta)n \) which corresponds to \(|\hat{A}_t| + t \leq n\delta\), we can define

\[ \hat{R}_{t+1} = \hat{R}_t \cup \{j_t\} \]

\[ \hat{A}_{t+1} = \hat{A}_t - \{j_t\} \cup \{y \in \hat{U}_t^{\delta} : \eta_{j_t,y} = 1\} \]

\[ \hat{U}_{t+1} = \hat{U}_t - \{y \in \hat{U}_t^{\delta} : \eta_{j_t,y} = 1\} \tag{1.3.5} \]

where in this construction \( i_t = \min \hat{A}_t \). Since we have reduced the size of \( U_t \), we have \(|A_t| \geq |\hat{A}_t|\). To define a comparison random walk, we let \( W_0 = 1 \), let

\[ T_W = \inf\{s : W_s = 0, \text{ or } W_s + s \geq n\delta\} \]
1.3. CLUSTER GROWTH AS A RANDOM WALK

define

\[ W_{t+1} = W_t - 1 + \begin{cases} 
\sum_{y \in U_t} \eta_{t,y} & \text{if } t < T_W \\
\sum_{y=1}^{n(1-\delta)} \zeta^t_y & \text{if } t \geq T_W 
\end{cases} \]

It is easy to see that for \( t \leq T_W \), \( |\hat{A}_t| = W_t \) so \( \tau \geq T_W \).

We will use the new lower bound \( W_t \) and the earlier upper bound \( S_t \) to prove

**Theorem 1.3.2.** Suppose \( \lambda > 1 \). There is a constant \( \beta \) so that with high probability, there is only one component of the random graph with more than \( \beta \log n \) vertices. The size of this component \( \sim (1 - \rho(\lambda))n \) where \( \rho(\lambda) \) is the extinction probability for the Poisson(\( \lambda \)) branching process.

Here with high probability means that the probability tends to 1 as \( n \to \infty \).

**Proof.** There are four steps.

**Step 1.** There is a constant \( \gamma \) so that if \( W_s \geq \gamma \log n \) then the probability \( W_t \) hits 0 at some time \( t \geq s \) is \( \leq n^{-2} \).

**Step 2.** There is a constant \( \beta \) so that \( P(0 < |A(\beta \log n)| < \gamma \log n) = o(n^{-1}) \).

**Step 3.** If \( A(\beta \log n) > \gamma \log n \) then with probability \( \geq 1 - \exp(-\eta n^{2/3}) \) we have \( \epsilon n^{2/3} \leq |A(\beta \log n)| \leq 2\lambda n^{2/3} \). Combined with the first two steps, this shows that with probability \( \to 1 \), all clusters reaching size \( \beta \log n \) will grow to size \( O(n^{2/3}) \) and will intersect producing a giant component.

**Step 4.** We show that the number of sites \( x \) with clusters \( C_x \) of size \( \geq \beta \log n \) is asymptotically \((1 - \rho)n\). Here the fact that we only have to run the cluster exploration process to time \( \beta \log n \) reduces the dependence between the events \( \{|C_x| \geq \beta \log n\} \) so that we can prove the result by computing second moments and using Chebyshev’s inequality.

Turning to the details of the proofs.

**Step 1.** The increments of \( W \) have the distribution \(-1 + \text{Binomial}((1 - \delta)n, \lambda/n)\).

By (1.3.3) the moment generating function of an increment

\[ \phi_\delta(\theta) \leq \exp(-\theta + \lambda(1 - \delta)(e^\theta - 1)) \] (1.3.6) [mgfbdf2]

Choose \( \delta > 0 \) so that \( \lambda(1 - \delta) > 1 \). \( \phi'_\delta(0) = -1 + \lambda(1 - \delta) > 0 \) so \( \psi_\delta(-\theta) < 1 \) when \( \theta > 0 \) is small. Since \( \psi_\delta(\theta) \) is convex and tends to \( \infty \) as \( \theta \to \infty \) there is a unique positive solution of \( \phi_\delta(-\theta_\delta) = 1 \). \( M_t = \exp(-\theta_\delta W_t) \) is a nonnegative (super)martingale. We put super in parentheses since it is a martingale but we will only use that it is a supermartingale. Suppose \( W_0 = k \) and let \( T_0 = \inf \{ t : W_t = 0 \} \). Using Theorem 4.8.4 in PTE5 (the stopping theorem for nonnegative supermartingales) and noting that \( W_t \to \infty \) on \( \{ T_0 = \infty \} \) we have

\[ e^{-k\theta_\delta} \geq P_k(T_0 < \infty) \] (1.3.7) [extbd1]

If we take \( k_\delta = (2/\theta_\delta) \log n \) we make the left-hand side \( = n^{-2} \).

**Step 2.** To control the behavior of \( W_t \) and \( S_t \) with good error bounds we use the following:
Lemma 1.3.3. Let \( Z = X_1 + \cdots + X_t \) where the \( X_i \) are i.i.d. and have \( E \exp(\theta X_i) < \infty \) for \( |\theta| < \theta_0 \). Let \( \mu = EX_i \). In Section 2.7 in PTE5 it is shown that if \( b < \mu < a \) then there are constant \( \eta(b), \eta(a) > 0 \) so that
\[
P(Z \geq ta) \leq \exp(-t\eta(a))
\]
\[
P(Z \leq tb) \leq \exp(-t\eta(b))
\]

Let \( X_1, \ldots, X_t \) be i.i.d. \(-1 + \text{binomial}(n, \lambda)\), write \( S_t \) instead of \( Z \). Let \( a = 2\lambda - 1 > \mu = \lambda - 1 \). Using Lemma 1.3.3 there are \( \eta_1, \eta_2 > 0 \) so that
\[
P(S_t \geq (2\lambda - 1)t) \leq \exp(-\eta_1 t)
\] (1.3.8)

Let \( r = \beta \log n \) where \( \beta(\lambda - 1)/2 > 2\gamma \) and \( \exp(-\eta_1 \beta \log n) \leq n^{-2} \). When \( |A_r| > 0 \), \( S_t \) has not hit 0 at times \( t \leq r \) so it provides an upper bound on \( A_t \). When \( S_r + r \leq 2\lambda r \) we have \( |U_s| \geq n - 2\lambda r \) for all \( t \leq r \). The number of times that the upper bound process adds a vertex not added to \( A_t \) is at most \( B(r, \lambda) \) which is the sum of \( r \times 2\lambda r \) Bernoulli random variable with success probability \( \lambda/n \). From this it follows that
\[
P(|A_r| > 0, S_r - |A_r| \geq 2) \leq \left( \frac{2\lambda r^2}{2} \right) (\lambda/n)^2 \leq C(\log n)^2
\]

Combining our results
\[
P(0 < |A(\beta \log n)| \leq \gamma \log n) \leq C(\log n)^2/n^2
\]

Step 3. Let \( \epsilon_\delta = (\lambda(1-\delta)-1)/2 \). Using Lemma 1.3.3 twice with \( X_i = \text{binomial}((1-\delta)n, \lambda/n) \) we have
\[
P_1(W(n^{2/3}) - W(0) \leq \epsilon_\delta n^{2/3}) \leq \exp(-\eta_2 n^{2/3})
\] (1.3.9)
\[
P_1(W(n^{2/3}) - W(0) + n^{2/3} \geq 2\lambda n^{2/3}) \leq \exp(-\eta_3 n^{2/3})
\] (1.3.10)

We take \( W(0) = |A_r| \leq 2\lambda \beta \log n \) in (1.3.10). Since \( W_t + t \) is nondecreasing this shows that with probability \( 1 - O(n^{-2}) \), \( W_s + s \leq \delta n \) for all \( s \leq n^{2/3} \), and the coupling between \( W_s \) and \( |A(s+r)| \) remains valid for \( 0 \leq s \leq n^{2/3} \).

The bound in (1.3.9) implies that if a cluster reaches size \( r = \beta \log n \) then the set of active sites at time \( r + n^{2/3} \) is \( \geq \epsilon_\delta n^{2/3} \) with high probability. Thus if we have two such clusters of size \( \geq \beta \log n \) then either (a) they will intersect by time \( r + n^{2/3} \) or (b) at time \( r + n^{2/3} \) they have disjoint sets \( V_1 \) and \( V_2 \) of active sites of size \( \geq \epsilon_\delta n^{2/3} \). The probability of no edge connecting \( V_1 \) and \( V_2 \) is
\[
= \left( 1 - \frac{\lambda}{n} \right)^{\epsilon_\delta n^{2/3}} \leq \exp(-\eta_4 n^{1/3})
\] (1.3.11)

This proves the first assertion in Theorem 1.3.2.
Step 4. To prove the second assertion it suffices to show that
\[
\left| \left\{ x : |C_x| \leq \beta \log n \right\} \right| / n \to \rho(\lambda) \tag{1.3.12}
\]

The first step is to note that the branching process approach to cluster growth implies
\[
P(|C_x| \leq \beta \log n) \to \rho(\lambda).
\]

To complete the proof we bound the correlation of the random variables \(Y_x = 1\) if \(|C_x| \leq \beta \log n\) and 0 otherwise and use Chebyshev’s inequality. We isolate the reasoning as

**Lemma 1.3.4.** Let \(F\) be an event that involves exposing \(J\) vertices starting at 1, and let \(G\) be an event that involves exposing \(K\) vertices starting at 2. Then
\[
|P(F \cap G) - P(F)P(G)| \leq JK \cdot \frac{2\lambda}{n}
\]

*Proof.* Let \(R_t, U_t\) and \(A_t\) be the process of exposing the cluster of 1. Introduce independent copies of the basic indicator random variables \(\eta_{x,y}'\). Let \(R'_0 = \emptyset, A'_0 = \{2\}\) and \(U'_0 = \{1,2,\ldots,n\} - \{2\}\). If \(A'_t \neq \emptyset\), pick \(i'_t = \min A'_t\). If \(i'_t \notin R_{\beta \log n}\) let
\[
\begin{align*}
R'_{t+1} &= R'_t \cup \{i'_t\} \\
A'_{t+1} &= A'_t - \{i'_t\} \cup \{y \in U'_t : \eta_{i'_t,y}' = 1\} \\
U'_{t+1} &= U'_t - \{y \in U'_t : \eta_{i'_t,y}' = 1\}
\end{align*}
\]

(1.3.13)

However if \(i'_t \in R_J\), an event we call a collision, we use \(\eta_{i'_t,y}\) instead of \(\eta_{i'_t,y}'\). In words if while growing cluster 2 we choose a site that was used in the growth of cluster 1, we use the original random variables \(\eta_{x,y}\). Otherwise we use independent random variables. It should be clear from the construction that
\[
|P(F \cap G) - P(F)P(G)| \leq P(R_J \cap R'_K \neq \emptyset) \leq JK \cdot \frac{\lambda}{n}
\]

which proves the desired result.

Using Lemma 1.3.4 with \(J = K = \beta \log n\), the probability of a collision is at most \(\lambda(\beta \log n)^2/n\). Using our bound on the covariance with the formula for the variance of a sum of random variables
\[
\text{var} \left( \sum_{x=1}^n Y_x \right) \leq n + n(n-1)\frac{\lambda(\beta \log n)^2}{n} \leq Cn \log n
\]

so it follows from Chebyshev’s inequality that
\[
P \left( \sum_{x=1}^n (Y_x - EY_x) \geq n^{2/3} \right) \leq \frac{Cn \log n}{n^{1/3}} \to 0
\]

This proves (1.3.12) and completes the proof of Theorem 1.3.2.
1.4 Diameter of the giant component

Having proved the existence of the giant component, we can use the branching process results from Section 1.2 to study the typical distance between two points on the giant component.

**Theorem 1.4.1.** Suppose \( \lambda > 1 \) and pick two points \( x \) and \( y \) at random from the giant component. Then \( d(x, y) / \log n \to 1 / \log \lambda \) in probability.

The answer in Theorem 1.4.1 is intuitive. The branching process approximation grows at rate \( \lambda^t \), so the average distance is given by solving \( \lambda^t = n \), i.e., \( t = (\log n) / \log \lambda \).

**Proof.** We begin with a small detail. The growth of the cluster containing \( x \) in the graph on \( n \) can be approximated by a branching process \( Z^n_t \). Unfortunately the offspring distribution \( \text{Binomial}(n, \lambda/n) \) depends on \( n \), so if we are concerned with the behavior of \( Z^n_t(n) \) with \( t(n) \to \infty \) we have a triangular array, not a single sequence. To deal with this, we use (3.6.1) in PTE5 to conclude that the total variation distance between \( \text{Binomial}(n, \lambda/n) \) and \( \text{Poisson}(\lambda) \) is \( \leq 2\lambda^2/n \). It follows from the definition of the total variation distance that \( \xi = \text{Binomial}(n, \lambda/n) \) and \( \xi' = \text{Poisson}(\lambda) \) can be constructed on the same space in such a way that \( P(\xi \neq \xi') \leq 2\lambda^2/n \). In what follows we will run the branching process until the total number of birth events is \( o(n) \) so with a probability that \( \to 1 \) there are no differences between the Binomial and Poisson versions. Thus we can use the growth results for the Poisson process and make conclusions about the Binomial one.

The size of \( Z_t(n) \geq |I_t| \) at time \( (1 - \epsilon) \log n / \log \lambda \) is \( \sim n^{1-\epsilon} W \), so most of the points are at least distance \( \geq (1 - \epsilon) \log n / \log \lambda \). As Theorem 1.3.2 shows, membership in the giant component is asymptotically equivalent to the cluster size being larger than \( \beta \log n \). Let \( Y^n_\infty \) be the total progeny of the branching process. When \( \lambda > 1 \), \( P(\beta \log n < Y^n_\infty < \infty) \to 0 \) as \( n \to \infty \), so points in the infinite cluster are with high probability associated with branching processes that don’t die out. If \( \delta \) is small then the size of one of these processes at time \( (1 + \epsilon) \log n / (2 \log \lambda) \) is \( \sim n^{(1+\epsilon)/2} \delta \) with high probability. If we consider two growing clusters at this time then either (a) they have already intersected or (b) they will have \( n^{1+\epsilon} \delta^2 \) chances to intersect at the next time, and the probability they will fail is

\[
\leq \left( 1 - \frac{\lambda}{n} \right) n^{1+\epsilon} \delta^2 \leq \exp(-n^\epsilon \delta^2) \to 0
\]

and the proof is complete. \( \square \)

Our next task is to show that the diameter \( D = \max d(x, y) \) of the giant component has a different limiting behavior than the average distance between two randomly chosen points. A dangling end is path \( v_0, v_1, \ldots, v_{k+1} \) with the degree \( d(v_0) = 1 \) and \( d(v_i) = 2 \) for \( 1 \leq i \leq k \).

**Theorem 1.4.2.** Let \( p_1 = e^{-\lambda} \lambda \) and choose \( k(n) \) so that \( c_n = np_1^{k(n)+1} \) stays bounded away from 0 and \( \infty \). When \( n \) is large the probability of a dangling end of length \( k(n) \) is approximately \( 1 - \exp(-c_n) \).
Here \( k(n) = \log n / (\log p_1) + O(1) \). The statement is made complicated by the fact that increasing \( k(n) \) by 1 decreases \( np_1^{k(n)+1} \) by a factor \( p_1 \), so we cannot choose \( k(n) \) to make \( c_n \) converge to \( c \in (0, \infty) \).

**Proof.** Let \( A_{(x,y)} \) be the event that there is dangling end \( v_0 = x, v_1, \ldots, v_{k+1} = y \)

\[
P(\bigcup_{(x,y)} A_{(x,y)}) \leq \sum_{(x,y)} P(A_{(x,y)}) \sim n(n-1) \cdot p_1^{k+1} \cdot \frac{1}{n-1} = c_n
\]

To see this note that there are \( n(n-1) \) values for the ordered pair \((x, y)\) with \( y \neq x \). \( v_0 \) has degree 1, an event of probability \( p_1 \). As for the \( v_i \) with \( i > 0 \), we arrive there along the edge from \( v_{i-1} \) and we have \( n-i \) uninspected outgoing edges so the probability of finding exactly one open is \( \sim p_1 \). When \( i < k \) we don’t care what vertex we connect to, but when \( i = k \) we want \( v_k \) to connect to \( y \), an event that by symmetry has probability \( 1/(n-1) \).

By the inclusion-exclusion formula we can get a lower bound by subtracting the sum of \( P(A_{(x,y)} \cap A_{(w,z)}) \) for all \((x, y) \neq (w, z)\). It is easy to see that if \( y \neq z \), \( A_{(x,y)} \cap A_{(x,z)} = \emptyset \). In all the other cases the paths for these two events cannot intersect, except perhaps at \( y = z \). This gives the second event less room to occur, so in all cases

\[
P(A_{(x,y)} \cap A_{(w,z)}) \leq P(A_{(x,y)}) \cdot P(A_{(w,z)})
\]

Combining the two estimates we have

\[
\sum_{(x,y) \neq (w,z)} P(A_{(x,y)} \cap A_{(w,z)}) \leq \frac{1}{2} (n(n-1))^2 P(A_{(x,y)})^2 \sim c_n^2 / 2
\]

Before progressing to the third level of approximation, we need to improve (1.4.1) by observing that the occurrence of \( A_{x,y} \) in effect removes \( k+1 \) vertices from the graph so unless \( x = w \) and \( y \neq z \),

\[
P(A_{(x,y)} \cap A_{(w,z)}) / P(A_{(x,y)}) \cdot P(A_{(w,z)}) \to 1
\]

as \( n \to \infty \). This result generalizes easily to a fixed finite number of events. Using this with the \( m \)th approximation to the inclusion exclusion formula we get an approximation

\[
c_n - c_n^2 / 2! + \ldots + (-1)^{m+1} c_n^m / m!
\]

which is an upper bound or a lower bound depending on whether the last term is + or -. The infinite series is \( 1 - \exp(-c_n) \) and the desired result follows.

Suppose we have two dangling ends, one from \( x \) to \( y \) and the other from \( w \) to \( z \), each of length \( a \log n \). Conditioning on their existence removes some vertices from the graph but does not significantly change the growth of clusters from \( y \) and \( z \). Thus in \( ER(n, \lambda/n) \), we have positive probability of \( y \) and \( z \) belonging to the giant component. When they do, they will have distance \( \approx (\log n)/(\log \lambda) \), and \( x \) and \( w \) will have distance \( 2a \log n + (\log n)/(\log \lambda) \).
Chung and Lu (2001) have obtained upper bounds on the diameter of the giant component, see Theorem 6 on their page 272.

Looking at the last paragraph, a mathematician might conclude that the diameter is what we should be studying, because it is harder to understand than the typical distance. While the diameter is the more traditional notion, there are two good reasons to study the distance between two randomly chosen points. The first is that the diameter is a number while the distribution of the typical distance contains substantially more information. In addition, the diameter is rather sensitive to small changes in the graph: adding a dangling end can substantially change the diameter, but will not affect the typical distance. and this completes the proof.
1.5 CLT for the giant component

Up to this point we have been content to study the growth of clusters while they are $o(n)$. In this section we will use an idea of Martin-Löf (1986) to follow the random walk approach all of the way to the end of the formation of the giant component and prove a central limit theorem for the size of the giant component.

To avoid the problem caused by the process dying out, it is convenient to modify the rules so that if $A_t = \emptyset$ we pick $i_t \in U_t$, and rewrite the recursion as

$$
R_{t+1} = R_t \cup \{i_t\} \\
A_{t+1} = A_t - \{i_t\} \cup \{y \in U_t : \eta_{i_t,y} = 1\} \\
U_{t+1} = U_t - (\{i_t\} \cup \{y \in U_t : \eta_{i_t,y} = 1\})
$$

In words, when one cluster is finished we pick a new vertex and start exposing its cluster.

When $A_t = \emptyset$ we subtract $1 + \text{Binomial}(|U_t|, \lambda/n)$ points from $U_t$ versus $\text{Binomial}(|U_t|, \lambda/n)$ points when $A_t \neq \emptyset$. However, we will experience only a geometrically distributed number of failures before finding the giant component, so this difference can be ignored. Let $F_t$ be the $\sigma$-field generated by the process up to time $t$. Let $u^n_t = |U_t|$.

**Lemma 1.5.1.** $u^n_{[ns]}/n$ converges in distribution to $u_s$ the solution of

$$
\frac{du_s}{ds} = -\lambda u_s \quad u_0 = 1
$$

and hence $u_s = \exp(-\lambda s)$.

*Proof.* Let $\Delta u^n_t = u^n_{t+1} - u^n_t$. If $A_t \neq \emptyset$ then

$$
E(\Delta u^n_t | F_t) = -\frac{u^n_t \lambda}{n} \\
\text{var}(\Delta u^n_t | F_t) = \frac{u^n_t \lambda}{n} \left(1 - \frac{\lambda}{n}\right)
$$

If we let $t = [ns]$ for $0 \leq s \leq 1$ and divide by $n$ then

$$
E \left( \frac{\Delta u^n_{[ns]}}{n} \bigg| F_{[ns]} \right) = -\frac{u^n_{[ns]}}{n} \cdot \lambda \cdot \frac{1}{n} \\
\text{var} \left( \frac{\Delta u^n_{[ns]}}{n} \bigg| F_{[ns]} \right) = \frac{u^n_{[ns]}}{n} \cdot \lambda \left(1 - \frac{\lambda}{n}\right) \cdot \frac{1}{n^2}
$$

(1.5.1)

Dividing each right-hand side by $1/n$, the time increment in the rescaled process, we see that $\Delta u^n_{[ns]}$ has

$$
\text{infinitesimal mean} = -\frac{u^n_{[ns]}}{n} \lambda
$$
infinitesimal variance = \frac{u^n_{[ns]}}{n} \lambda \left( 1 - \frac{\lambda}{n} \right) \cdot \frac{1}{n}

Letting \( n \to \infty \), the infinitesimal variance \( \to 0 \), so the result follows from (7.1) in Durrett (1996).

The last proof is simple and intuitive, but may be too sophisticated for some reader’s tastes, so we now give

**Alternative Proof.** The calculations above show that

\[ M^n_t = \left( 1 - \frac{\lambda}{n} \right)^{-t} \frac{u^n_t}{n} \]

is a martingale with

\[
E(M^n_t - M^n_0)^2 = \sum_{s=0}^{t-1} E(M^n_{s+1} - M^n_s)^2 \\
\leq \sum_{s=0}^{t-1} \left( 1 - \frac{\lambda}{n} \right)^{-s+1} \lambda/n^2 \to 0
\]

so by Kolmogorov’s maximal inequality

\[
E\left( \max_{0 \leq s \leq n} (M^n_s - M^n_0)^2 \right) \to 0
\]

Since \( M^n_0 = 1 \), this says that when \( n \) is large \( M^n_s \approx 1 \) uniformly in \( s \), so \( u^n_{[ns]}/n \approx (1 - \lambda/n)^{[ns]} \to e^{-\lambda s} \).

To determine the size of the giant component, we note that when \( u^n_t + r^n_t = n \), \( A_t = \emptyset \). This may occur several times for small \( t \) while we are searching for the giant component, but the solution we are looking for is the first occurrence after an excursion of \( O(n) \). To locate roughly the time at which this occurs, we note that scaling \( r^n_t = |R_t| \equiv t \) as we did \( u^n_t \), \( r^n_{[ns]}/n \to s \). [Here and in what follows we will use \( t \) for the original integer time scale and \( s \in [0, 1] \) for rescaled time.] After scaling

\[ u^n_t + r^n_t = n \quad \Rightarrow \quad e^{-\lambda s} + s = 1 \]

Solving we have \( 1 - s = \exp(\lambda((1 - s) - 1)) \), which is the fixed point equation for the extinction probability, \( 1 - s \). As the graph below shows \( e^{-\lambda s} + s > 1 \) for \( s > 1 - \rho \), so we are interested only in \( u^n_{[ns]}/n \) for \( 0 \leq s \leq 1 - \rho + \epsilon \). In this part of the process we first generate a geometrically distributed number of small clusters and then expose the giant component.
1.5. CLT FOR THE GIANT COMPONENT

Consider now \( y^n_{[ns]} = (u^n_{[ns]} - n \exp(-\lambda s))/\sqrt{n} \) for \( 0 \leq s \leq 1 - \rho \).

**Lemma 1.5.2.** As \( n \to \infty \), \( y^n_{[ns]} \) converges in distribution to a normal with mean \( 0 \) and variance \( e^{-\lambda s} - e^{-2\lambda s} \).

**Proof.** If \( A_{[ns]} \neq \emptyset \) then using the formulas in (1.5.1)

\[
E(\Delta y^n_{[ns]}|\mathcal{F}_{[ns]}) = -\frac{1}{\sqrt{n}} \left( -u^n_{[ns]} \cdot \frac{\lambda}{n} - n \exp(-\lambda s)(\exp(-\lambda/n) - 1) \right) \\
\sim -\frac{\lambda}{n} \left( u^n_{[ns]} - n \exp(-\lambda s) \right) = -\lambda y^n_{[ns]} \cdot \frac{1}{n}
\]

\[
\text{var}(\Delta y^n_{[ns]}|\mathcal{F}_{[ns]}) = \text{var} \left( \frac{\Delta u^n_{[ns]}}{\sqrt{n}} \bigg| \mathcal{F}_{[ns]} \right) = \frac{1}{n} \cdot u^n_{[ns]} \cdot \frac{\lambda}{n} \left( 1 - \frac{\lambda}{n} \right) \sim \lambda e^{-\lambda s} \cdot \frac{1}{n}
\]

Using (7.1) in Chapter 8 of Durrett (1996) again, we see that \( y^n_{[ns]} \) converges in distribution to the solution of the following stochastic differential equation.

\[
dy_s = -\lambda y_s \, ds + \sqrt{\lambda e^{-\lambda s}} \, dB_s \quad y_0 = 0
\]

The solution to this equation is

\[
y_s = \int_0^s e^{-\lambda(s-r)} \sqrt{\lambda e^{-\lambda r}} \, dB_r \quad (1.5.2) \]

\[\text{soleq}\]
To check this, recall that if one continuously invests an amount \( g_s \) in an exponentially decaying stock market then your net wealth \( x_s \) satisfies

\[
\frac{dx_s}{ds} = -\lambda x_s + g_s
\]

Since computation of interest is linear, each amount decays exponentially from its date of investment, and this differential equation has solution

\[
x_s = \int_0^s e^{-\lambda(s-r)} g_r \, dr
\]

Readers who want a more rigorous proof can use stochastic calculus to check this.

Since the integrand in (1.5.2) is deterministic, \( y_s \) has a normal distribution with mean 0 and variance

\[
\int_0^t \exp(-2\lambda(s-r)) \lambda e^{-\lambda r} \, dr = e^{-2\lambda s} \int_0^s \lambda e^{\lambda r} \, dr = e^{-\lambda s} - e^{-2\lambda s}
\]

which proves the result.

Remark. Again if one wants to avoid stochastic calculus, the theorem can be proved by applying the martingale central limit theorem to

\[
M^n_t - M^n_0 = \left( 1 - \frac{\lambda}{n} \right)^{-t} u^n_t / n - 1
\]

The key observation is that

\[
\sum_{r=0}^{[ns]-1} E((M^n_{r+1} - M^n_r)^2 | F_r) \to \int_0^s \lambda e^{\lambda u} \, du = e^{\lambda s} - 1
\]

i.e., the variance process has a deterministic limit. See e.g., Section 7.2 in Durrett (2004). Multiplying the martingale by \( e^{-\lambda s} \), multiplies the variance by \( e^{-2\lambda s} \) and we arrive at the same limit as before.

We have analyzed the fluctuations of \( u^n_t \). To determine the fluctuations of the point where \( u^n_t + t = n \), we can now prove the result as we do the central limit theorem of renewal theory. To briefly recall that approach, let \( \xi_1, \xi_2, \ldots \) be i.i.d. positive random variables with \( E \xi_i = \mu \) and var \( (\xi_i) = \sigma^2 \in (0, \infty) \). Let \( T_n = \xi_1 + \cdots + \xi_n \) and \( N(t) = \inf\{n : T_n > t\} \). The central limit theorem implies

\[
T_n \approx n\mu + \sigma \sqrt{n} \chi
\]

where \( \chi \) is a standard normal. Setting \( n = t/\mu \)

\[
T_{t/\mu} \approx t + \sigma \sqrt{\frac{t}{\mu}} \chi
\]
If $\chi > 0$ then $N(t) < t/\mu$. The law of large numbers implies $T_n - T_m \approx (n - m)\mu$ when $n - m$ is large so we will have

$$\frac{t}{\mu} - N(t) \approx \frac{\sigma}{\mu} \sqrt{\frac{t}{\mu}} \chi$$

The same reasoning applies in the current situation. Taking $s = 1 - \rho$ in Lemma 1.5.2 and letting $Z$ denote a normal with variance $e^{-\lambda(1-\rho)} - e^{-2\lambda(1-\rho)}$ we have

$$u^n_{[\mu(1-\rho)]} \approx n \exp(-\lambda(1 - \rho)) + \sqrt{n} Z$$

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)

(1.5.3)
1.6 Combinatorial approach

Combinatorial methods give more refined results about the Erdős-Rényi model. However, we first need some graph theoretic preliminaries. It is easy to see that a tree with \( k \) vertices has \( k - 1 \) edges. We call a graph with \( k \) vertices and \( k \) edges, a unicyclic graph, since it will have exactly one cycle, i.e., a path of adjacent vertices \( x_0, x_1, \ldots, x_k = x_0 \), and \( x_j \neq x_0 \) for \( 1 \leq j < k \). We call a graph with \( k \) vertices and \( k + \ell \) edges with \( \ell \geq 1 \) a complex component with complexity \( \ell \). To simplify discussion we will consider unicyclic components to be complex components with \( \ell = 0 \).

Our first result shows that complex components are rare, except possibly near the critical value \( \lambda = 1 \).

**Lemma 1.6.1.** Let \( A < \infty \) be a constant, and consider only components with \( \leq A \log n \) vertices. The expected number of unicyclic components is \( \leq \lambda (A \log n)^2 \). The probability of having at least one complex component is \( \leq (A \log n)^4 \lambda^2 / n \).

**Remark.** Theorem 1.3.1 implies that if \( \lambda < 1 \) there is a \( C_\lambda \) so that with high probability the largest component is \( \leq C_\lambda \log n \). Theorem 1.3.2 implies that if \( \lambda > 1 \) there is a \( \beta_\lambda \) so that with high probability the second largest component is \( \leq \beta_\lambda \log n \).

**Proof.** We look at the growth of the cluster from the random walk viewpoint. An increase in complexity is caused by a “self-intersection”, i.e., a connection from \( i \) to some point in \( A_t \). Our assumption implies that there are \( \leq A \log n \) times at which we expose the neighbors of a vertex and \( |A_t| \leq A \log n \) for all \( t \) so the number of self-intersections is \( \leq \text{Binomial}((A \log n)^2, \lambda / n) \). The result for unicyclic components follows by recalling there are \( n \) possible starting points and taking expected value. The result for complex components follows by computing the probability of two or more self-intersections as we did in (??).

Having ruled out complex components, we can, for \( \lambda \neq 1 \), restrict our attention to tree and unicyclic components. In 1989 Cayley showed that there are \( k^{k-2} \) trees with \( k \) labeled vertices. When \( p = \lambda / n \) the expected number of trees of size \( k \) present is

\[
\binom{n}{k} k^{k-2} \left( \frac{\lambda}{n} \right)^{k-1} \left( 1 - \frac{\lambda}{n} \right)^{k(n-k)+\binom{k}{2}-(k-1)}
\]

since each of the \( k - 1 \) edges in the tree needs to be present and there can be no edges connecting its \( k \) vertices to its complement or any other edges connecting the \( k \) vertices. For fixed \( k \), we can drop \( -k^2 + \binom{k}{2} - k + 1 \) from the exponent of the last term and the above is asymptotic to

\[
n \frac{k^{k-2}}{k!} \lambda^{k-1} e^{-\lambda k} \equiv nq_k
\]

where the second equality defines \( q_k \).

Recalling that in the subcritical regime cluster sizes have the same distribution as the total progeny in a Poisson(\( \lambda \)) branching process, we get the following corollary, which is “well-known,” but not easy to prove directly from the definition of the branching process:
Corollary 1.6.2. The probability distribution of the total progeny \( \tau \) of a Poisson(\( \lambda \)) branching process with \( \lambda < 1 \) is given by

\[
P(\tau = k) = k q_k = \frac{1}{\lambda} \frac{k^{k-1}}{k!} (\lambda e^{-\lambda})^k
\]

(1.6.3)

There is an extra factor of \( k \) due to the fact that a tree of size \( k \) is \( C_k \) for \( k \) values of \( x \).

This distribution was first discovered by Borel in 1942. It is called the Borel-Tanner distribution, since Tanner (1961) showed that when \( \lambda < 1 \) it gave the distribution of the total number of customers served in the first busy period of a queue with Poisson rate \( \lambda \) arrivals and service times always equal to 1. Of course, this becomes a branching process if we think of the customers that arrive during a person’s service time as their children.

Duality. Suppose \( \lambda > 1 \) and let \( \rho \) be the extinction probability. The title of this subsection refers to the fact that there is a close relationship between Erdős-Rényi random graphs with mean degrees \( \lambda > 1 \) and \( \lambda \rho < 1 \). Using the fixed point equation \( \rho = e^{\lambda(\rho-1)} \) for the extinction probability \( \rho \) gives

\[
\lambda \rho e^{-\lambda \rho} = \lambda e^{\lambda(\rho-1)} e^{-\lambda \rho} = \lambda e^{-\lambda}
\]

(1.6.4)

Let \( ER(n, p) \) denote an Erdős-Rényi graph with \( n \) vertices and edges present with probability \( p \). Let \( m = n \rho \) and consider \( ER(m, \lambda \rho / m) \), an Erdős-Rényi graph with number of vertices equal to the number of vertices in non-giant components of \( ER(n, \lambda / n) \). Changing variables in (1.6.2) we see that

\[
\frac{m}{\lambda \rho} \frac{k^{k-2}}{k!} (\lambda e^{-\lambda \rho})^k = \frac{n}{\lambda} \frac{k^{k-2}}{k!} (\lambda e^{-\lambda})^k
\]

In words, the expected number of trees of size \( k \) is the same in \( ER(m, \lambda \rho / m) \) and \( ER(n, \lambda / n) \). Changing variables in the same way in (1.6.3)

\[
\frac{1}{\lambda \rho} \frac{k^{k-1}}{k!} (\lambda e^{-\lambda \rho})^k = \frac{1}{\lambda} \frac{k^{k-1}}{k!} (\lambda e^{-\lambda})^k \cdot \frac{1}{\rho}
\]

In words, the total progeny of a Poisson(\( \lambda \)) branching process conditioned on extinction is the same as that of a Poisson(\( \lambda \rho \)) branching process, which is Theorem 1.8.1.

(1.6.2) is a result about the expected number of trees. The next result is a law of large numbers, which says that the actual number is close to the expected value.

Theorem 1.6.3. Let \( T^n_k \) be the number of tree components of size \( k \) in the Erdős-Rényi graph with \( n \) vertices. As \( n \to \infty \), \( T^n_k / n \to q_k \) in probability, where \( q_k \) is defined in (1.6.2)

Proof. This proof comes from Bollobás (2001), 106–107. The expected number of ordered pairs of tree components of size \( k \) (with the second tree different from the first) is

\[
\binom{n}{k} k^{k-2} \left( \frac{\lambda}{n} \right)^{k-1} \left( 1 - \frac{\lambda}{n} \right)^{k(n-k)+\binom{k}{2} - k + 1} \cdot \binom{n-k}{k} k^{k-2} \left( \frac{\lambda}{n} \right)^{k-1} \left( 1 - \frac{\lambda}{n} \right)^{k(n-2k)+\binom{k}{2} - k + 1}
\]

(1.6.5)
The second formula differs from the first only in two places: first we have only \( n - k \) vertices to choose from, and the first term already takes into account the fact that there are no connections from the first tree to the second. Since \( \binom{n-k}{k} \leq \binom{n}{k} \) the above is

\[
\leq (ET_k)^2 \left(1 - \frac{\lambda}{n}\right)^{-k^2} \leq (ET_k)^2 e^{\lambda k^2/n}
\]

From this we get

\[
\text{var}(T_k) = E(T_k(T_k - 1)) + ET_k - (ET_k)^2 \leq ET_k + (ET_k)^2(e^{\lambda k^2/n} - 1)
\]

Using Chebyshev’s inequality

\[
P(|T_k - ET_k| \geq n^{2/3}) \leq \frac{ET_k + (ET_k)^2(e^{\lambda k^2/n} - 1)}{n^{4/3}} \to 0
\]

since \( ET_k \sim nq_k \) and \( e^{\lambda k^2/n} - 1 \sim \lambda k^2/n \). This gives the desired result. Note that we could replace \( n^{2/3} \) in the last display by \( \omega(n)n^{1/2} \) where \( \omega(n) \to \infty \) as \( n \to \infty \).

The results above allow us to verify the remark we made about the largest non-giant component for \( \lambda > 1 \).

**Theorem 1.6.4.** Suppose \( \lambda > 1 \) and let \( C^{(2)} \) be the second largest component. If \( \alpha = \lambda - 1 - \log \lambda \) and \( a > 1/\alpha \) then as \( n \to \infty \)

\[
P\left(|C^{(2)}| \geq a \log n\right) \to 0
\]

**Proof.** For simplicity we will do our calculations for the limit (1.6.2) rather than for the exact formula (1.6.1). Stirling’s formula tells us that

\[
k! \sim k^{k+1/2}e^{-k}\sqrt{2\pi} \quad \text{as} \quad k \to \infty
\]

so we have (Lemma 1.7.1 will show this is valid for \( k = o(n^{1/2}) \))

\[
q_k = \frac{1}{\lambda} \cdot \frac{k^{k-2}}{k!} (\lambda e^{-\lambda})^k \sim \frac{1}{\lambda\sqrt{2\pi}} k^{-5/2} (\lambda e^{1-\lambda})^k
\]

Now \( g(\lambda) \equiv \lambda e^{1-\lambda} = 1 \) when \( \lambda = 1 \) and \( g'(\lambda) = (1 - \lambda)e^{1-\lambda} \). Thus \( g(\lambda) \) is increasing for \( \lambda < 1 \), decreasing for \( \lambda > 1 \), and has \( g(\lambda) < 1 \) when \( \lambda \neq 1 \). Summing and using the fact that \( k^{-5/2} \) is decreasing and \( \lambda e^{1-\lambda} < 1 \)

\[
Q_K = \sum_{k=K}^{\infty} q_k \sim \frac{1}{\lambda\sqrt{2\pi}} K^{-5/2} (\lambda e^{1-\lambda})^K \frac{(\lambda e^{1-\lambda})^K}{1 - \lambda e^{1-\lambda}}
\]

Taking \( K = a \log n \)

\[
(\lambda e^{1-\lambda})^a \log n = \exp((\log \lambda - \lambda + 1)a \log n) = n^{-(1+\epsilon)}
\]

when \( a = (1 + \epsilon)/\alpha \), which proves the desired result. \( \square \)
1.7 Critical regime

In this section we will look at component sizes when \( \lambda = 1 + \theta n^{-1/3} \) where \(-\infty < \theta < \infty\), which corresponds to the critical regime for the Erdős-Rényi random graph. We begin with a calculation that is simple and gives the right answer but is not completely correct. (1.6.2) tells us that the expected number of trees of size \( k \) is, for large \( k \),

\[
n \frac{k^{k-2}}{k!} \lambda^{k-1} e^{-\lambda k} \sim \frac{n}{\sqrt{2\pi \lambda}} k^{-5/2} (\lambda e^{1-\lambda})^k
\]

(1.7.1) by Stirling’s formula

\[
k! \sim k^k e^{-k} \sqrt{2\pi k}
\]

(1.7.2)

When \( \lambda = 1 \), \( \lambda e^{1-\lambda} = 1 \) so summing from \( k = K \) to \( \infty \), the expected number of tree of size \( \geq K \) is

\[
\sum_{k=K}^{\infty} \frac{n}{\sqrt{2\pi \lambda}} k^{-5/2} \sim \frac{2}{3\sqrt{2\pi}} nK^{-3/2}
\]

This is small when \( K \gg n^{2/3} \) suggesting that the largest tree components are of order \( n^{2/3} \).

Having figured out what to guess, we will now go back and do the calculation carefully. For the moment \( \lambda \) is a general parameter value, which you should of as being close to 1.

Lemma 1.7.1. Let \( \alpha(\lambda) = \lambda - 1 - \log(\lambda) \). If \( k \to \infty \) and \( k = o(n^{3/4}) \) then the expected number of tree components of size \( k \) in an Erdős-Rényi graph with \( n \) vertices

\[
\gamma_{n,k}(\lambda) \sim n \cdot \frac{k^{-5/2}}{\lambda^{\sqrt{2\pi}}} \exp \left( -\alpha(\lambda) k + (\lambda - 1) \frac{k^2}{2n} - \frac{k^3}{6n^2} \right)
\]

(1.7.3)

Proof. (1.6.1) tells us that

\[
\gamma_{n,k}(\lambda) = \left( \frac{n}{k} \right)^{k-2} \left( \frac{\lambda}{n} \right)^{k-1} \left( 1 - \frac{\lambda}{n} \right)^{k(n-k)+(\frac{k}{2})-k}
\]

Using Stirling’s formula (1.7.2) and \( k = o(n) \) in the last exponent gives

\[
\gamma_{n,k}(\lambda) \sim n \left[ \prod_{j=1}^{k-1} \left( 1 - \frac{j}{n} \right) \right] \cdot \frac{k^{-5/2}}{e^{-\lambda k} \sqrt{2\pi}} \cdot \lambda^{k-1} \left( 1 - \frac{\lambda}{n} \right)^{kn-k^2/2}
\]

Using the expansion \( \log(1-x) = -x - x^2/2 - x^3/3 - \ldots \) we see that if \( k = o(n) \) then

\[
\left( 1 - \frac{\lambda}{n} \right)^{kn-k^2/2} \sim \exp(-\lambda k + \lambda k^2/2n)
\]

while if \( k = o(n^{3/4}) \) we have

\[
\prod_{j=1}^{k-1} \left( 1 - \frac{j}{n} \right) = \exp \left( -\frac{1}{n} \sum_{j=1}^{k-1} j - \frac{1}{2n^2} \sum_{j=1}^{k-1} j^2 + O \left( \frac{k^4}{n^3} \right) \right) \sim \exp \left( -\frac{k^2}{2n} - \frac{k^3}{6n^2} \right)
\]

Combining our calculations gives the desired formula.
Taking $\lambda=1$ in Lemma 1.7.1 we have

$$\gamma_{n,k}(\lambda) \sim \frac{n k^{-5/2}}{\sqrt{2\pi}} e^{-k^3/6n^2} \quad (1.7.4)$$

Note that the right-hand side behaves like a power law when $k = o(n^{2/3})$ but decays to 0 exponentially fast on $[n^{2/3}, \infty)$.

### 1.7.1 Aldous’ theorem

Returning from the land of combinatorics, we go back to using the stochastic process that exposes the neighbors of one vertex at a time. Let $R_t$ be the removed sites at time $t$, $A_t$ the active sites, and $U_t$ the unexplored sites. In our first use of this process here we will stop at $\tau = \inf\{t : A_t = \emptyset\}$, but to make sure we find a large component while we will start with $|A_0|$ large. The ideas here are from Martin-Löf (1998), but we carry out the details somewhat differently.

As in our study of the giant component, we will speed up time and rescale the size of our sets to get a limit. To see what to guess, note that the combinatorial calculations suggest that the largest components are of order $n^{2/3}$. Since $R_t = t$ and $R_\tau$ is the size of the clusters containing $A_0$, we will scale time by $n^{2/3}$. When $\lambda = 1 + \theta n^{-1/3}$, $|A_t|$ will be almost a mean zero random walk. In this case $|A_t| - |A_0|$ will be $O(t^{1/2})$ so we will scale the number by $n^{1/3}$.

Having decided on the scaling, we compute the infinitesimal mean and variance. Let $a_t = |A_t|$, $\Delta a_t = |A_{t+1}| - |A_t|$, and note that $u_t = |U_t| = n - t - a_t$. Since the number of new vertices is binomial$(u_t, \lambda/n)$ with $\lambda = 1 + \theta n^{-1/3}$ we have

$$E(\Delta a_t | F_t) = -1 + \frac{t + a_t}{n} + \frac{\theta n^{-1/3}(1 - (t + a_t)/n)}{n}$$

$$\var(\Delta a_t | F_t) = \frac{(n - t - a_t)}{n} \frac{1 + \theta n^{-1/3}}{n} \left( \frac{1}{n} - \frac{1 + \theta n^{-1/3}}{n} \right)$$

Speeding up time by $n^{2/3}$, dividing by $n^{1/3}$, and using $a_{[sn^{2/3}]} = O(n^{1/3})$,

$$E \left( \frac{\Delta a_{[sn^{2/3}]/n^{1/3}}}{ \left[ F_{[sn^{2/3}]/n^{1/3}} \right] } \right) = \frac{-[sn^{2/3}] - a_{[sn^{2/3}]/n^{1/3}}}{n \cdot n^{1/3}} + \frac{\theta n^{-1/3}(1 + \theta n^{-1/3})}{n^{1/3}}$$

$$= -sn^{-2/3} + O(n^{-1}) + \theta n^{-2/3} + O(n^{-1})$$

The variance is much easier

$$\var \left( \frac{\Delta a_{[sn^{2/3}]/n^{1/3}}}{ \left[ F_{[sn^{2/3}]/n^{1/3}} \right] } \right) \sim \frac{1}{n^{2/3}}$$
Letting $n \to \infty$ we see that $a((sn^2)/3)/n^{1/3}$ converges in distribution to the solution of
\[ dB_\theta^s = (-s + \theta) \, ds + dB_s \]
which is simply $B_\theta^s = B_s + \theta s - s^2/2$ (run until the first time it hits zero).

This calculation above leads to a remarkable result of Aldous (1997) that gives the joint distribution of the sizes of the large clusters divided by $n^{2/3}$. Consider now the version of the exploration process in which we choose $i_t \in U_t$ when $A_t = \emptyset$. This adds one to $|A_t|$ each time it hits zero which has the effect of creating a reflecting barrier at 0 but is perhaps better seen through the eyes of Paul Lévy:

\[ W_\theta^t = W_\theta^0 - \min_{0 \leq s \leq t} W_\theta^s \]

We say that $(u, v)$ is an excursion interval of $W_\theta^t$ if $W_\theta^u = W_\theta^v = 0$ but $W_\theta^t \neq 0$ for $t \in (u, v)$. During excursions $|A_t|$ does not hit 0. Since we expose one vertex at a time, the lengths of the excursion intervals represent cluster sizes in the random graph.

**Theorem 1.7.2.** Let $K^\theta_1 \geq K^\theta_2 \geq \ldots$ be the ordered component sizes of $ER(n, (1+\theta n^{-1/3})/n)$. Then as $n \to \infty$, \((n^{-2/3}K^\theta_j : j \geq 1)\) converges in distribution to \(\{L_j : j \geq 1\}\) where $L_1 > L_2 > L_3 > \ldots$ are the ordered lengths of excursion intervals in \(\{W_\theta^s : s \geq 0\}\).

By working harder in this framework it is possible to prove a result about the “topology” of the clusters. A cluster of size $k$ must have $k-1$ edges. We define the number of edges in excess of this number to be the surplus of the cluster. When the surplus is 1 the cluster is unicyclic, i.e., it has one cycle. Let $N^\theta(t)$ be a counting process with intensity $W_\theta^t$. That is $N^\theta$ has jumps of size 1 and

\[ N^\theta(t) - \int_0^t W_\theta^s \, du \text{ is a martingale} \]

The arrivals in this point process are called marks. Intuitively connections between vertices in the cluster are produced at a rate proportional to the size of the active set.

**Theorem 1.7.3.** Let $\sigma_1^n \geq \sigma_2^n \geq \ldots$ be the surpluses ordered component sizes defined in Theorem 1.7.2. Then as $n \to \infty$,

\[ \{(n^{-2/3}K^\theta_j, \sigma_j^n) : j \geq 1\} \]

converges in distribution to \(\{(|\gamma_j|, \mu(\gamma_j)) : j \geq 1\}\) where the $\gamma_j$ are the excursion intervals listed in order of decreasing length and $\mu(\gamma_j)$ is the number of marks in $\gamma_j$.

Much is known about the number of complex components. See Janson (1993), Janson, Knuth, Luczak, and Pittel (1993), Luckzak, Pittel, and Weirman (1994), and the second edition of Bollobás’ *Random Graphs*.
**Multiplicative coalescent.** While Theorem 1.7.3 is a nice limit theorem, the truly remarkable part of Aldous’ contribution was to note that if we divide the sizes of clusters in \( ER(n, (1 + tn^{-1/3})/n) \) by \( n^{2/3} \) to create a process indexed by \(-\infty < t < \infty\) then there is a simple and intuitive description of the limit. Each pair of clusters of sizes \((x, y)\) merges at rate \( xy \) to a cluster of size \( x + y \). To see why this is true consider two clusters of sizes \( xn^{2/3} \) and \( yn^{2/3} \). In a short interval of time \((t, t + h)\), an edge is added between two vertices with probability \( hn^{-4/3} \), so the probability of making a connection between clusters of size \( x \) and \( y \) in time \( h \) is \( \approx xyh \).

To have an honest Markov process, we need a state space. Aldous chose \( \ell^2_\alpha \), the collection of decreasing sequences \( x_1 \geq x_2 \geq x_3 \geq \ldots \) with \( \sum_k x_k^2 < \infty \). The other thing that should be noticed is that the time interval is \(-\infty < t < \infty\) so there is no initial distribution. In Aldous’ original paper he solved this problem by showing that there was only one “standard multiplicative coalescent” that had the one dimensional distributions consistent with Theorem 1.7.2. Aldous and Limic later characterized all processes on \(-\infty < t < \infty\) in which each pair of clusters of sizes \((x, y)\) merges at rate \( xy \) to a cluster of size \( x + y \). (In terms of the theory of Markov processes one is finding all of the entrance laws.) In addition to the constant process \((x_1 = v > 0, x_i = 0, i \geq 2)\) there are some nonstandard ones, which are irrelevant for the following application.

**Theorem 1.7.4.** Let \( K_1^n(t) \geq K_2^n(t) \geq \ldots \) be the ordered component sizes of \( ER(n, (1 + tn^{-1/3})/n) \). As \( n \to \infty \), \( \{K_j^n(t)/n^{2/3} : j \geq 1\}, -\infty < t < \infty \) converges in distribution to the standard multiplicative coalescent.

The convergence of rescaled large components to the multiplicative coalescent, provides a nice intuitive process of the growth of clusters in the critical regime.
1.8 Critical Exponents

Let \( p(s, \mu) \) be the limit as \( n \to \infty \) of the probability a randomly chosen vertex in an Erdős-Rényi graph with mean degree \( \mu \) belongs to a cluster of size \( s \). The critical exponents for the Erdős-Rényi random graph are defined much like they are for ordinary percolation on \( \mathbb{Z}^d \). See e.g., page 236 in Grimmett’s (1999) book.

- The percolation probability \( \theta = 1 - \sum_{s=1}^{\infty} P(s, \mu) \) has

\[
\theta \approx (\mu - \mu_c)^\beta \quad \text{as } \mu \downarrow \mu_c \quad (1.8.1) \]

In the physics literature the meaning of \( \approx \) is not precisely defined. It could be something as weak as

\[
\frac{\log \theta}{\log(\mu - \mu_c)} \to \beta \quad \text{as } \mu \downarrow \mu_c
\]

In order to derive relations between exponents we will suppose \( \theta \sim C(\mu - \mu_c)^\beta \) where \( a(t) \sim b(t) \) means \( a(t)/b(t) \to 1 \).

- Let \( C_x \) be the cluster containing \( x \). To have the mean cluster size finite for \( \mu \neq \mu_c \) we exclude the giant component

\[
\chi = E(|C_x|; |C_x| < \infty) \approx |\mu - \mu_c|^{-\gamma} \quad (1.8.2)
\]

- For higher moments, we let \( \chi_k = E(|C_x|^k; |C_x| < \infty) \) for integers \( k \geq 2 \), for real numbers \( r > 0 \)

\[
\chi_k \approx |\mu - \mu_c|^{-\Gamma(k)} \quad (1.8.3)
\]

and let \( \Delta_k = \Gamma(k) - \Gamma(k-1) \). As in the case of percolation it is expected that \( \Delta_k = \Delta \) is independent of \( k \).

- There is an exponent for the decay of the cluster size at criticality

\[
p(s, \mu_c) \approx s^{1-\tau} \quad (1.8.4)
\]

In the first three cases the Erdős-Rényi Greek letters are the same as those for percolation. In this case the percolation exponent is

\[
P_{p_c}(|C| \geq n) \approx n^{-1/\delta}
\]

- Finally there is another exponent for the behavior near criticality

\[
p(s, \mu) \approx s^{1-\tau} f(s^{\delta^{1/\sigma}}) \quad \delta = \mu - \mu_c \quad (1.8.5)
\]

although it is probably more accurate to call this a scaling relationship.
The most significant difference between the critical exponents for Erdős-Rényi random graphs and those for $d$-dimensional percolation is that the Erdős-Rényi graph is not a spatial model, so we do not have a correlation length $\xi(p)$ that gives the spatial size of a typical finite cluster. This quantity is often defined in terms of the exponential decay of probability $\tau^f(0, x) = P_p(0 \leftrightarrow x, |C_0| < \infty)$.

For example, if $e_1$ is the first unit vector

$$\xi(p) = \lim_{n \to \infty} -\frac{1}{n} \log \tau^f(0, ne_1)$$

A simpler approach taken by Kesten (1987) is to define the correlation length by

$$\xi(p) = \left( \frac{1}{\chi_1(p)} \sum_y |y|^2 P(0 \to y; |C_0| < \infty) \right)^{1/2}$$

and the critical exponent $\nu$ by $\xi(p) \approx |p - p_c|^{-\nu}$. Finally on a $d$-dimensional graph we have another exponent $\eta$ called the anomalous dimension

$$P_{cr}(0 \to x) \approx |x|^{2-d-\eta}$$

The relationship between the Erdős-Rényi critical exponents and those for percolation is more than an analogy: the values are equal for percolation in dimensions $d > d_c$. The word that makes the connection is “mean-field theory.” In the Isian model this refers to a version in which each spin interacts equally with all the others. If one makes this modification to percolation then the Erdős-Rényi model results. Physicists tell us that above the upper critical dimension, which is $d_c = 6$ for percolation then the critical exponents take on their mean-field values: $\beta = 1, \gamma = 1, \Delta = 2$. To compute the others we need a spatial object which is usually taken to be percolation on trees. This leads to $\nu = 1/2$ and $\eta = 0$. Thanks to work of Aizenmann and Newman (1984) and Hara and Slade (1989, 1994), etc there are a number of rigorous results confirming this picture. See Section 10.3 of Grimmett (1999) for more details. The key words are tree graph inequalities and the lace expansion.

### 1.8.1. Survival probability

Let $x$ be a randomly chosen vertex and let $Z_m$ be the number of vertices at distance $m$ from $x$. When $m = o(\log n)$ the cluster containing $x$ is whp a tree and $Z_m$ is a branching process in which each individual in generation $m$ has a Poisson($\mu$) number of offspring. In Section 1.3 we show that a vertex is in the giant component if and only if it is in a cluster of size $\geq \beta \log n$, so the probability of this event can be computed using the branching process.
Consider a branching process with offspring distribution $r_k$ with mean $\mu > 1$ and finite second moment. Let $\phi(z) = \sum_{k=0}^{\infty} r_k z^k$ be the generating function. If $r_K$ is Poisson($\mu$) then

$$
\phi(z) = \sum_{k=0}^{\infty} e^{-\mu} \mu^k \frac{1}{k!} z^k \exp(-\mu(1 - z)) \quad (1.8.6)
$$

Let $\rho$ be probability the system dies out. Breaking things down according to the number of children in the first generation

$$
\rho = \sum_{k=0}^{\infty} r_k \rho^k = \phi(\rho)
$$

$\rho(1) = 1$ is a trivial solution. $\rho$ is the unique solution of $\phi(\rho) = \rho$ in $[0, 1)$.

$$
\phi(1) = 1 \quad \phi'(1) = \sum_{k=0}^{\infty} kr_k = \mu \quad \phi''(1) = \sum_{k=0}^{\infty} k(k-1)r_k = \mu_2
$$

If $\mu$ is close to 1 then $\rho$ will be close to 1. Ignoring a few details, if $x$ is close to 1 expanding $\phi$ in power series around 1 gives

$$
\phi(1 - x) = 1 - \mu x + \mu_2 x^2 / 2
$$

so for a fixed point we want

$$
(\mu - 1)x = \mu_2 x^2 / 2
$$

or $x = 2(\mu - 1)/\mu_2$. If we let $\theta(\mu) = P(|C_x| = \infty)$ which is the same as the fraction of vertices in the giant component

$$
\theta(\mu) \sim \frac{2}{\mu_2} (\mu - 1) \quad (1.8.7)
$$

so the critical exponent $\beta = 1$.

### 1.8.2. Mean cluster size

Let $C_x$ be the cluster containing $x$. If $\mu < 1$

$$
E|C_x| = 1 + \mu + \mu^2 + \cdots = \frac{1}{1 - \mu}. \quad (1.8.8)
$$

so $\gamma = 1$. In the supercritical regime we consider

$$
E(|C_x| \mid |C_x| < \infty)
$$

The cluster size is the same as the total progeny in a supercritical branching process conditioned to die out.
Theorem 1.8.1. A supercritical branching process conditioned to become extinct is a subcritical branching process. If the original offspring distribution is Poisson(µ) with µ > 1 then the conditioned one is Poisson(µρ) where ρ is the extinction probability.

Proof. Let T₀ = inf{t : Zₜ = 0} and consider ¯Zₜ = (Zₜ|T₀ < ∞). To check the Markov property for ¯Zₜ note that the Markov property for Zₜ implies:

\[ P(Z_{t+1} = z_{t+1}, T₀ < ∞|Zₜ = zₜ, \ldots Z₀ = z₀) = P(Z_{t+1} = z_{t+1}, T₀ < ∞|Zₜ = zₜ) \]

To compute the transition probability for ¯Zₜ, observe that if ρ is the extinction probability then

\[ P(\varphi(x) = 1|T₀ < ∞) = \rho. \]

Let p(x, y) be the transition probability for Zₜ. Note that the Markov property implies

\[ \bar{p}(x, y) = \frac{P_x(Z_1 = y, T₀ < ∞)}{P_x(T₀ < ∞)} = \frac{P_x(Z_1 = y)P_y(T₀ < ∞)}{P_x(T₀ < ∞)} = \frac{p(x, y)ρ^y}{ρ^x}. \]

Taking x = 1 and computing the generating function

\[ \sum_{y=0}^{∞} \bar{p}(1, y)z^y = ρ^{-1} \sum_{y=0}^{∞} p(1, y)(zρ)^y = ρ^{-1}φ(zρ) \]

where p_y = p(1, y) is the offspring distribution.

\[ \bar{p}_y = \bar{p}(1, y) \]

is the distribution of the size of the family of an individual, conditioned on the branching process dying out. If we start with x individuals then in Zₙ each gives rise to an independent family. In Zₙ each family must die out, so Zₙ is a branching process with offspring distribution \( \bar{p}(1, y) \). To prove this formally observe that

\[ p(x, y) = \sum_{j₁, j₂ ≥ 0, j₁ + \cdots + jₓ = y} p_{j₁} \cdots p_{jₓ} \]

Writing Σ* as shorthand for the sum in the last display

\[ \frac{p(x, y)ρ^y}{ρ^x} = \sum* \frac{p_{j₁}ρ^{j₁} \cdots p_{jₓ}ρ^{jₓ}}{ρ} = \sum* \bar{p}_{j₁} \cdots \bar{p}_{jₓ} \]

In the case of the Poisson(µ) distribution φ(z) = exp(µ(z − 1)) so if µ > 1, so using (1.8.9)

\[ \frac{φ(zρ)}{ρ} = \frac{exp(µ(zρ - 1))}{exp(µ(ρ - 1))} = exp(µρ(z - 1)) \]

which completes the proof that the conditioned process is a Poisson(µρ) branching process.

Using the result for the subcritical case in (1.8.8)

\[ E(|Cₓ|||Cₓ| < ∞) = \frac{1}{1 - µρ} \]
1.8. CRITICAL EXPONENTS

Since \( \mu_2 = \mu^2 \) for Poisson, (1.8.7) shows that if \( \mu \) is close to 1

\[
\rho \approx 1 - \frac{2(\mu - 1)}{\mu^2}
\]

so we have

\[
1 - \mu \rho = 1 - \mu + \frac{2(\mu - 1)}{\mu} = \frac{2(\mu - 1) - \mu(\mu - 1)}{\mu}
\]

From this we see that

\[
E(|C_x| | |C_x| < \infty) = \frac{1}{1 - \mu \rho} \sim \frac{1}{1 - \mu}
\]

thus the asymptotic behavior of the mean cluster size as \( \mu \downarrow 1 \) is the same as as \( \mu \uparrow 1 \), i.e., the exponent \( \gamma = 1 \) and the constant \( C = 1 \).

1.8.3. Higher moments

To compute higher moments it is convenient to use the random walk approach developed in Section 1.3. Let \( R_t \) be the set of removed sites, \( U_t \) be the unexplored sites and \( A_t \) the set of active sites. At time \( \tau = \inf\{t : A_t = \emptyset\} \) we have found all the sites in the cluster and the process stops. \(|R_t| = t\) for all \( t \leq \tau \), so the cluster size is \( \tau \). If \(|A_t| > 0\) and the number of removed sites is small

\[
S_{t+1} \approx S_t - 1 + \text{Poisson}(\mu)
\]

To have the process defined for all time let \( \xi_1, \xi_2, \ldots \) be i.i.d. \(-1 + \text{Poisson}(\mu)\) and \( S_{t+1} = S_t + \xi_{t+1} \).

We begin by computing the moment generating function of \( \xi_i \)

\[
\psi(\theta) = \sum_{m=0}^{\infty} e^{-\mu} \frac{\mu^m}{m!} e^{\theta m} = \exp(\mu(e^\theta - 1))
\]

exp(\( \theta S_t \))/\( \phi(\theta)^t \) is a nonnegative martingale, so using the optional stopping theorem for the nonnegative supermartingale \( M_t = \exp(\theta S_t) / \psi(\theta)^t \),

\[
M_0 = e^\theta \geq E(\psi(\theta)^{-\tau})
\]

\[
\psi'(0) = E\xi_i = 1 - \mu \text{ so if } \mu < 1 \text{ then } \psi(\theta) < 1 \text{ when } \theta > 0 \text{ is small. To optimize we note that the derivative}
\]

\[
\frac{d}{d\theta} (-\theta + \mu(e^\theta - 1)) = -1 + \mu e^\theta = 0
\]

when \( \theta_1 = -\log \mu \). At this point \( e^{\theta_1} = 1/\mu \) and

\[
\psi(\theta_1) = \exp(\log(\mu) + 1 - \mu) \equiv e^{-\alpha} < 1
\]

Since \( \psi(\theta_1^{-1}) = e^\alpha > 1 \), using Chebyshev’s inequality with (1.8.11)

\[
e^{\alpha m} P(\tau \geq m) = \psi(\theta_1)^{-m} P(\tau \geq m) \leq E(\psi(\theta_1)^{-\tau}) \leq e^{\theta_1}
\]
One particle dies on each time step so \( 1 + \xi_1 + \cdots + \xi_\tau = \tau \) and we have

\[
P \left( \max_{0 \leq n \leq \tau} S_n \geq m \right) \leq e^{-m\alpha/\mu} \tag{1.8.12} \]

There are several other martingales associated with a random walk. Perhaps the simplest is \( S_n - n(\mu - 1) \). Using the domination in (1.8.12) we can conclude that

\[
1 \equiv S_0 = E(S_\tau - (\mu - 1)\tau) = (1 - \mu)E\tau
\]

so the expected cluster size is \( E\tau = 1/(1 - \mu) \).

If \( T_n \) is a random walk in which steps have mean 0 and variance \( \sigma^2 \) then \( T_n - n\sigma^2 \) is a martingale. Applying this result to \( T_n = S_n - n(\mu - 1) \) and recalling that \(-1 + \text{Poisson}(\mu)\) has variance \( \mu \) we see that

\[
(S_n - n(\mu - 1))^2 - \mu n
\]

is a martingale. Using the domination we can stop at time \( \tau \) and conclude

\[
1 = E(S_\tau - \tau(\mu - 1))^2 - \mu E\tau
\]

\[
= (\mu - 1)^2 E\tau^2 - \frac{\mu}{1 - \mu}
\]

Rearranging we have

\[
E\tau^2 = \frac{1}{(1 - \mu)^3} \tag{1.8.13}
\]

Since \( E\tau = 1/(1 - \mu) \) the critical exponent \( \Delta_2 = 2 \).

1.8.4. Cluster size at (and near) criticality

(1.7.4) implies that when \( \mu = 1 \) the probability of a cluster of size \( k \) is

\[
\sim \frac{nk^{-5/2}}{\sqrt{2\pi}} e^{-k^3/6n^2}
\]

This says that the largest components are of size \( n^{2/3} \). The critical exponent \( \tau \) defined in (1.8.4)

\[
P(s, t_c) \sim f(0)s^{1-\tau}
\]

has the value \( \tau = 5/2 \)

To get a result for \( \mu \) close to 1, we use (1.7.3)

\[
P(s, \mu) \approx \frac{s^{1-5/2}}{\sqrt{2\pi}} \cdot \exp(-s(\mu - 1)^2/2) \tag{1.8.14}
\]

Thus (1.8.5) holds with \( \tau = 5/2, \sigma = 1/2 \) and

\[
f(x) = e^{-x/2} \tag{1.8.15}
\]
1.9 Scaling theory

In the previous section we showed that the cluster size distribution for the Erdős-Rényi graph satisfied

\[ p(s, \mu) \approx s^{1-\tau} f(s|\mu - \mu_c|^{1/\sigma}) \]  

where \( f(x) = e^{-x/2} \)

It has long been known for percolation (see Stauffer (1979)) that (1.9.1) implies

\[ \beta = (\tau - 2)/\sigma \]  
\[ \gamma = (3 - \tau)/\sigma \]  
\[ \Delta = 1/\sigma \]

Since (1.9.1) gives the asymptotics for the cluster size distribution as \( \mu \to \mu_c \) it should not be surprising that we can compute the asymptotic behavior of the moments and of the percolation probability. Kesten (1984) has proved a number of rigorous results concerning scaling relations concentrating on the two dimensional case.

**Lemma 1.9.1.** If \( 2 < \tau < 3 \) and \( f \) is bounded, and is Lipschitz continuous at 0.

\[ \beta = (\tau - 2)/\sigma \]

**Proof.** Using (1.9.1) and replacing sum by integration.

\[ \theta(t) \approx \int_1^\infty s^{1-\tau}[f(0) - f(s\delta^{1/\sigma})] \, ds \]

Changing variables \( s = x\delta^{-1/\sigma}, \, ds = \delta^{-1/\sigma} \, dx \) the above

\[ = \delta^{(\tau-2)/\sigma} \int_{\delta^{1/\sigma}}^\infty x^{1-\tau}[f(0) - f(x)] \, dx \]

Since \( f \) is bounded and \( \tau > 2 \) the integral over \([1, \infty)\) is finite. Since \( f \) is Lipschitz continuous the integrand is \( \leq Cx^{2-\tau} \) near 0. Since \( \gamma < 3 \), the integral over \([0, 1]\) is finite, and it follows that \( \theta(t) \sim C\delta^{(\tau-2)/\sigma} \).

The next result establishes (1.9.3) and (1.9.4)

**Lemma 1.9.2.** If \( 2 < \gamma < 3 \) and \( \int_1^\infty x^mf(x) \, dx < \infty \) for all \( m \) then for all \( r \geq 1 \)

\[ \Gamma(r) = (r + 2 - \tau)/\sigma \]

It follows that for all integers \( k \geq 2 \),

\[ \Delta_k = \Gamma(k) - \Gamma(k - 1) = 1/\sigma. \]
Proof.

\[ E|\mathcal{C}_x|^r = \int_1^\infty s^{r+1-\tau} f(s^{\delta^{1/\sigma}}) \, ds \quad (1.9.5) \]

Changing variables \( s = x^{\delta^{1/\sigma}} \), \( ds = \delta^{1/\sigma} \, dx \) the above

\[ = \delta^{(\tau-2)/\sigma} \int_{\delta^{1/\sigma}}^\infty x^{\rho+1-\tau} f(x) \, dx \]

The assumption \( \int_1^\infty x^m f(x) \, dx < \infty \) for all \( m \) implies that the integral over \([1, \infty)\) is finite. Since \( \tau < 3 \) and \( \rho \geq 1 \) the integral over \([0, 1]\) is finite.
1.10 Threshold for connectivity

In this section we will investigate the question: How large does \( \lambda \) have to be so that the probability \( \text{ER}(n, \lambda/n) \) is connected (i.e., ALL vertices in ONE component) tends to 1. Half of the answer is easy. Let \( d_x \) be the degree of \( x \).

\[
P(d_x = 0) = \left( 1 - \frac{\lambda}{n} \right)^n
\]

Using the series expansion \( \log(1 - x) = -x - x^2/2 - x^3/3 - \ldots \) it is easy to see that if \( \lambda = o(n^{1/2}) \) then

\[
n \log \left( 1 - \frac{\lambda}{n} \right) = -\lambda - \lambda^2/2n - \lambda^3/3n^2 - \ldots
\]

and hence

\[
\left( 1 - \frac{\lambda}{n} \right)^n e^\lambda \to 1
\]  

Thus when \( \lambda = a \log n \) we have \( P(d_x = 0) \sim n^{-a} \), and if \( a < 1 \) the number of isolated vertices \( I_n = |\{x : d_x = 0\}| \) has

\[
EI_n = nP(d_x = 0) \sim n^{1-a} \to \infty
\]

To show that the actual value of \( I_n \) is close to the mean we note that if \( x \neq y \)

\[
P(d_x = 0, d_y = 0) = \left( 1 - \frac{\lambda}{n} \right)^{2n-1} = \left( 1 - \frac{\lambda}{n} \right)^{-1} P(d_x = 0)P(d_y = 0)
\]

so we have

\[
\text{var} (I_n) = nP(d_1 = 0)(1 - P(d_1 = 0)) + n(n-1) \left( \left( 1 - \frac{\lambda}{n} \right)^{-1} - 1 \right) P(d_1 = 0)P(d_2 = 0)
\]

When \( \lambda = a \log n \)

\[
\text{var} (I_n) \sim n^{1-a} + n^2 \left( \frac{\lambda}{n} \right) n^{-2a} \sim EI_n
\]

Using Chebyshev’s inequality it follows that if \( a < 1 \)

\[
P \left( |I_n - EI_n| > (\log n)(EI_n)^{1/2} \right) \leq \frac{1}{\log^2 n}
\]

The last result shows that if \( \lambda = a \log n \) with \( a < 1 \) then with high probability there are about \( n^{1-a} \) isolated vertices, and hence the graph is not connected. Showing that the graph is connected is more complicated because we have to consider all possible ways in which the graph can fail to be connected. (1.6.2) tells us that the expected number of trees of size \( k \) is

\[
\sim n \frac{k^{k-2}}{k!} \lambda^{k-1} e^{-\lambda k}
\]
When \( k = 2 \) and \( \lambda = a \log n \) this is

\[
\frac{n}{2} (a \log n)^{-2a}
\]

Thus if \( 1/2 < a < 1 \) there are isolated vertices, but no components of size 2. It is easy to generalize the last argument to conclude that when \( \lambda = a \log n \) and \( 1/(k+1) < a < 1/k \) there are trees of size \( k \) but not of size \( k+1 \). Bollobás (2001), see Section 7.1, uses this observation with the fact we know that the largest component is \( O(\log n) \) to sum the expected values and prove:

**Theorem 1.10.1.** Consider \( G = ER(n, \lambda/n) \) with \( \lambda = a \log n \). The probability \( G \) is connected tends to 0 if \( a < 1 \) and to 1 if \( a > 1 \).

**Proof.** We will use the approach of Section 1.3 to show that the probability a vertex fails to connect to the giant component is \( o(1/n) \). Since we have a large \( \lambda \) we can use the lower bound process with \( \delta = 1/2 \). The constant \( \theta_{1/2} \) that appears in (1.3.7) is defined by

\[
\theta + (\lambda/2)(e^{-\theta} - 1) = 0
\]

This is hard to compute for fixed \( \lambda \), so instead we decide we want \( \theta = 1 \) and see that this means \( \lambda = 2e/(e - 1) \). Using monotonicity we see that if \( \lambda \geq 2e/(e - 1) \), (??) implies that for our comparison random walk

\[
P_{2\log n}(T_0 < \infty) \leq n^{-2}
\]

To reach size \( 2\log n \) we use the large deviations bound in Lemma ??, with \( \delta = 0 \) and \( x = 1/2 \) to conclude that if \( S_K \) is a sum of \( K \) independent Binomial\((n, \lambda/n)\) random variables then \( \mu = K\lambda \) and

\[
P(S_K \leq \mu/2) \leq \exp(-\gamma(1/2)\mu)
\]

where \( \gamma(1/2) = (1/2) \log(1/2) - 1/2 + 1 = 1/2(1 - \log(2)) \geq .15 \). If \( \lambda = (1 + \epsilon) \log n \) with \( \epsilon \geq 0 \) and \( K = 14 \) then

\[
P(S_{14} \leq 7 \log n) \leq n^{-2.1}
\]

The last calculation shows that if \( x \) has at least 14 neighbors, then with high probability at distance 2 there are at least \( 7 \log n \) vertices. The next step is to bound

\[
P(d_x \leq 13) = \sum_{k=0}^{13} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \leq \sum_{k=0}^{13} \frac{\lambda^k}{k!} e^{-\lambda(n-k)/n} \leq 14(a \log n)^{13} n^{-a} e^{(13a \log n)/n}
\]

which is \( o(n^{-1}) \) if \( a > 1 \).
1.10. THRESHOLD FOR CONNECTIVITY

To finish up now (and to prepare for the next proof), we apply the large deviations result
lemma ?? to lower bounding random walk $W_t$ twice to conclude that if $-1 + (a \log n) \geq (a/2) \log n$ then there are positive constants $\eta_i$ so that

$$
P(W(n^{1/2}) - W(0) \leq (a/2)n^{1/2} \log n) \leq \exp(-\eta_1 n^{1/2} \log n)
$$

$$
P(W(n^{1/2}) + n^{1/2} - W(0) \geq 2an^{1/2} \log n) \leq \exp(-\eta_2 n^{1/2} \log n)
$$

Combining our results we see that with probability $1 - o(n^{-1})$ the RAU process will not expose $n/2$ vertices and have at least $0.1n^{1/2} \log n$ active vertices at time $n^{1/2}$. When this occurs for $x$ and for $y$ the probability that their clusters fail to intersect is at most

$$
\left(1 - \frac{\log n}{n}\right)^{0.01n(\log n)^2} \leq e^{-(\log n)^3/100}
$$

and the proof is complete. \hfill \qed

The next result is an easy extension of the previous argument and will allow us to get a sharper result about the transition to connectivity.

**Theorem 1.10.2.** Consider $G = ER(n, \lambda/n)$ with $\lambda = a \log n$. If $a > 1/2$ then with probability tending to 1, $G$ consists only of a giant component and isolated vertices.

**Proof.** It follows from the previous argument that if $n$ is large $P_{2\log n}(T_0 < \infty) \leq n^{-2}$. Using (1.10.3) it is easy to see that if $S_{28}$ is a sum of 28 independent Binomial($n, \lambda/n$) and $\lambda \geq (1/2) \log n$ then

$$
P(S_{28} \leq 7 \log n) \leq n^{-2.1}
$$

Consider now a branching process $Z^x_k$ with offspring distribution Binomial($n, \lambda/n$). If $Z^x_1 = 0$ the cluster containing $x$ is a singleton. We might be unlucky and have $Z^x_1 = 1$ but in this case

$$
P(Z^x_1 = 1, Z^x_2 \leq 27) = \frac{n^\lambda}{n} \left(1 - \frac{\lambda}{n}\right)^{n-1} \sum_{k=0}^{27} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}
$$

$$
\leq \sum_{k=0}^{27} \frac{\lambda^{k+1}}{k!} e^{-\lambda(2n-k-1)/n} \leq 28(a \log n)^{28} n^{-2a} e^{(28a \log n)/n}
$$

which is $o(n^{-1})$ if $a > 1/2$. If we are lucky enough to find two neighbors on the first try then

$$
P(Z^x_2 \leq 28|Z^x_1 \geq 2) \leq \sum_{k=0}^{28} \binom{2n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{2n-k}
$$

$$
\leq \sum_{k=0}^{28} \frac{(2\lambda)^k}{k!} e^{-2\lambda(n-k)/n} \leq 29(a \log n)^{28} n^{-2a} e^{(56a \log n)/n}
$$

and the proof is complete.
Here we have replaced 27 by 28 to take care of collisions. One collision has probability \( c/n \), but by a now familiar estimate, the probability of two collisions in one step of the branching process is \( O(1/n^2) \).

In the previous proof we only used \( a > 1/2 \) in the W-estimates and the final estimate so the proof is complete.

We are now ready to more precisely locate the connectivity threshold.

**Theorem 1.10.3.** Consider \( G = ER(n, \lambda/n) \) with \( \lambda = \log n + b + o(1) \). The number of isolated vertices \( I_n \) converges to a Poisson distribution with mean \( e^{-b} \) and hence the probability \( G \) is connected tends to \( \exp(-e^{-b}) \).

**Proof.** By the previous result \( G \) will be connected if and only if there are no isolated vertices. Using (1.10.1), the probability \( x \) is isolated is

\[
\left(1 - \frac{\lambda}{n}\right)^n \sim \exp(-\log n - b) \sim e^{-b}/n
\]

so \( EI_n \). The expected number of ordered \( k \)-tuples of isolated vertices is

\[
(n \cdot (n-1) \cdots (n-k+1)) \left(1 - \frac{\lambda}{n}\right)^{n+(n-1)+\cdots+(n-k+1)} \to e^{-bk}
\]

so the Poisson convergence follows from the method of moments.

Having shown that the probability \( ER(n, c(\log n)/n) \) is connected tends to 1 for \( c > 1 \), our next question is to determine its diameter. The heuristic is the same as in Theorem 1.4.1 for \( ER(n, c/n) \). We set \( (np)^m = n \) and solve to conclude that the average distance between two points is \( \log n/(\log np) \). For \( ER(n, c(\log n)/n) \) this is also the diameter. We begin with a large deviations result for the Binomial distribution.

**Lemma 1.10.4.** Let \( X = \text{Binomial}(n, p) \) and let \( H(a) = a \log(a/p) + (1-a) \log((1-a)/(1-p)) \). Then for \( b < p < c \)

\[
P(X \leq nb) \leq \exp(-H(b)n) \quad P(X \geq nc) \leq \exp(-H(c)n)
\]

**Remark.** To see why this is the answer, note that from the definition of the Binomial distribution and Stirling’s formula, \( n! \sim n^n e^{-n} \sqrt{2\pi n} \),

\[
P(X = na) = \binom{n}{na} p^{na} (1-p)^{n(1-a)} \approx \frac{n^n}{(na)^{na} n(1-a))^{n(1-a)}} p^{na} (1-p)^{n(1-a)}
\]

Now cancel the \( n^n \)’s in the fraction and take \( (1/n) \log \) of both sides.
1.10. THRESHOLD FOR CONNECTIVITY

Proof. Suppose \( a > p \). Then for any \( \theta > 0 \)
\[
e^{a} P(X \geq na) \leq ((1 - p) + pe^{\theta})^{n}
\]
which we can rewrite as
\[
P(X \geq na) \leq [\exp\{-\theta a + \log((1 - p) + pe^{\theta})\}]^{n}
\]
To optimize we differentiate the term in set braces with respect to \( \theta \) and set
\[
0 = -a + \frac{pe^{\theta}}{(1 - p) + pe^{\theta}}
\]
Solving we have \( e^{\theta} = (a(1 - p))/(p(1 - a)) \). Using \((1 - p) + pe^{\theta} = pe^{\theta}/a = (1 - p)/(1 - a)\) and plugging in we have
\[
= \exp\left(-a \log\left(\frac{a(1 - p)}{p(1 - a)}\right) + \log((1 - p)/(1 - a))\right) = \exp(-H(a)
\]
For \( a < p \) the first equality is valid for \( \theta < 0 \) but the remaining calculations are the same. \(\square\)

In many cases the following simplification is useful

**Lemma 1.10.5.** If \( X = \text{Binomial}(n, p) \) then
\[
P(X \leq n(p - z)) \leq \exp(-nz^{2}/2p) \quad P(X \geq n(p + z)) \leq \exp(-nz^{2}/2(p + z))
\]

Taking \( z = py \) this becomes
\[
P(X \leq np(1 - y)) \leq \exp(-npy^{2}/2) \quad P(X \geq np(1 + y)) \leq \exp(-npy^{2}/2(1 + y))
\]

**Remark.** If \( X/n \) were Gaussian with variance \( p(1 - p)/n \) these probabilities would be \( \leq \exp(-nz^{2}/2p(1 - p)) \)

**Proof.** We begin with the second inequality. The function defined in Lemma 1.10.4 has \( H(p) = 0 \) and
\[
H'(a) = \log(a/p) - \log((1 - a)/(1 - p))
\]
so \( H'(p) = 0 \). If \( p < a \leq 1 \), Taylor’s theorem implies that there is a \( y \in [p, a] \) so that \( H(a) = H''(y)(a-p)^{2}/2 \). Differentiating again \( H''(a) = 1/a(1-a) \) which is minimized at 1/2. Therefore if \( a \leq 1/2 \), \( H(a) \geq (a - p)^{2}/2a \), while if \( a \geq 1/2 \), \( H(a) \geq 2(a - p)^{2} \geq (a-p)^{2}/2a \). Substituting \( a = p + z \) gives the second result. The argument for the first result is almost identical, but when we substitute \( a = p - z \) we can let \( a = p \) in the denominator. \(\square\)

**Theorem 1.10.6.** If \( \lim \inf np/(\log n) > c > 1 \) and \( (log p)/(\log n) \to 0 \) then the diameter of \( ER(n, p) \), \( D(n, p) \sim (\log n)/(\log np) \).
Remarks. The first condition guarantees that the probability that the graph is connected tends to 1 as \( n \to \infty \). To explain the second suppose \( p = n^{-15/17} \). In this case \( np = n^{2/7} \) and it is not hard to show that the diameter will be 9 with probability approaching 1, but the formula gives 17/2. A more delicate situation occurs when

\[
p = n^{(1/d)-1} (\log(n^2/c))^{1/d}
\]

In this case in the limit the diameter is \( d \) with probability \( e^{-c/2} \) and \( d + 1 \) with probability \( 1 - e^{-c/2} \). See Theorem 10.10 in Bollobás (2001). Chung and Lu (2001) have shown, see Theorem 5 on page 272, that the conclusion of Theorem 1.10.6 holds for the giant component if \( \log n > np \to \infty \).

Proof. Fix \( x \) and let \( B_k(x) = \{ y : d(x,y) = k \} \). \( |B_k| \) is dominated by \( B_k \) a branching process with a Binomial\((n,p)\) offspring distribution. Using Lemma 1.10.5 with \( y = 3 \) we have

\[
P(Z_1 \geq 4np) \leq \exp(-9np/8) \leq n^{-9c/8}
\]

for large \( n \). Using Lemma 1.10.5 again with \( n \) replaced by \( (k+2)(np)^{k-1}n \) and \( y = 1/(k+2) \) we have

\[
P(Z_k > (k+3)(np)^k | Z_{k-1} \leq (k+2)(np)^{k-1}) \leq \exp(-(np)^k/(k+3))
\]

Since \( np \geq c \log n \) for large \( n \) the right-hand side converges to 0 very rapidly as \( n \to \infty \) and we have

\[
P(Z_k \leq (k+3)(np)^k \text{ for } 1 \leq k \leq (\log n)/(\log np)) = 1 - O(n^{-9c/8}) \quad (1.10.4)
\]

To get a lower bound we note that by the proof of Theorem 1.10.1 \( P(|B_1| \leq 13) = 1 - o(n^{-1}) \) and

\[
P(|B_2| \leq 7np | |B_1| \geq 14) \leq n^{-2.1}
\]

To control the growth of \( |B_k| \) for \( k \geq 3 \) we note that as long as \( \sum_{j=0}^k |B_j| \leq n^{2/3} \), \( |B_k| \) dominates a branching process with a Binomial \((n',p)\) offspring distribution where \( n' = n - n^{2/3} \) and \( n'p \geq c \log n \) for large \( n \). Define a sequence of constants by \( a_2 = 7 \) and for \( k \geq 3, a_k = a_{k-1}(1 - 1/k^2) \). Since \( a_k \) is decreasing, at each iteration we subtract less than \( 7/k^2 \). \( \sum_{k=3}^{\infty} 1/k^2 = \pi^2/6 - 1 - 1/4 \leq 0.4 \) so \( a_k \geq 4 \) for all \( k \). Using Lemma 1.10.5 again with \( n \) replaced by \( a_{k-1}(n'p)^{k-2}n' \) and \( y = 1/k^2 \) we have for \( k \geq 3 \)

\[
P(|B_k| < a_k(n'p)^{k-1} | Z_{k-1} \geq a_{k-1}(n'p)^{k-2}) \leq \exp(-(n'p)^{k-1}/2k^4)
\]

Since \( n'p \geq c \log n \) for large \( n \) the right-hand side converges to 0 very rapidly as \( n \to \infty \) and we have

\[
P(|B_k| \geq 4(n'p)^{k-1} \text{ for } 2 \leq k \leq (0.6)(\log n)/(\log np)) = 1 - o(n^{-1}) \quad (1.10.5)
\]

Here 0.6 is chosen so that (1.10.4) implies that for large \( n \) we have \( \sum_{j=0}^k |B_j| \leq n^{2/3} \) the indicated range of \( k \)'s.
The bound in (1.10.4) implies that if $\epsilon > 0$ then with probability $1 - O(n^{-9c/8})$

$$
\sum_{k=0}^{(1-\epsilon)\log n/(\log np)} Z_k < n
$$

for large $n$ so $\liminf D(n, p)/(\log n/(\log np)) \geq 1 - \epsilon$. Let $k$ be the first integer larger than $1 + ((1 + \epsilon)(\log n))/(2(\log np))$ Using (1.10.5) we see that if $\epsilon > 0$ then

$$
P(|B_k| \geq 4n^{(1+\epsilon)/2}) = 1 - o(n^{-1})
$$

By a now familiar estimate, when this occurs for two different starting points then either the two clusters have already intersected or with probability $\geq \exp(-16n^{(1+\epsilon)})$ they will do so on the next step. This shows that $\limsup D(n, p)/(\log n/(\log np)) \leq 1 + \epsilon$ and completes the proof. \qed
CHAPTER 1. ERDÖS-RÉNYI RANDOM GRAPHS

References


Bollobás, B. (1984) ?, see page 26


1.10. THRESHOLD FOR CONNECTIVITY


