Chapter 1

Erdös-Rényi Random Graphs

In this chapter we will analyze the random graph model introduced by Erdös and Rényi in the late 1950's. This example has been extensively studied. A very nice account of many of the results can be found in the classic book of Bollobás (2001), where the analysis has been done primarily by combinatorial methods. In contrast to many other treatments, we mainly rely on methods from probability.

To define the model, we begin with the set of vertices $V = \{1, 2, \ldots n\}$. For $1 \le x < y \le n$ 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 connecting x and y. Here, we will be primarily concerned with situation $p = \lambda/n$ and in particular with showing that there is a **phase transition**

- When $\lambda < 1$ all of the components are small, with the largest $O(\log n)$,.
- When $\lambda > 1$ there is a **giant component** with $\sim g(\lambda)n$ vertices, and the second largest component is of size $O(\log n)$.
- At the critical value $\lambda = 1$ the largest components have size $n^{2/3}$,.
- As the reader will learn in sections 1.6 and 1.7, the critical value is surrounded by a critical regime $\lambda = 1 + tn^{-1/3}$, in which cluster growth can be approximated by a multiplicative coalescent.

In words, these results show that there is a **double jump transition**: the size of the largest compoent goes from $\log n$ to $n^{2/3}$ to n. The intuition behind this result is that a site has a binomial$(n-1, \lambda/n)$ number of neighbors, which has mean $\approx \lambda$. To compute the cluster (connected component) containing 1, we set $I_0 = \{1\}$ let I_1 be the neighbors of 1, and for $t \ge 2$ let I_t be the vertices in for which the distance $d(1, x) = t$. In Section 1.2 we will show that when t is not too large $Z_t = |I_t|$, is approximately a branching process in which each individual has an average of λ 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 describe 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 ξ_i^t, $i, t \geq 0$, 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 to an independent and identically distributed number of children. $p_k = P(\xi_i^t = k)$ is called the **offspring distribution**.

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

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

$$E(Z_{t+1}|\mathcal{F}_t) = E(\xi_1^{t+1} + \cdots + \xi_k^{t+1}|\mathcal{F}_t) = k\mu = \mu Z_t.$$

Dividing both sides by μ^{t+1} now gives the desired result. □

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

$$P(Z_t > 0) \leq E(Z_t; Z_t > 0) = E(Z_t) = \mu^t \to 0$$

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_∞. 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 $Z_\infty \equiv 0$. □

To state the next result we need a defintion. For $\theta \in [0, 1]$, let $\phi(\theta) = \sum_{k \geq 0} p_k \theta^k$ be the **generating function** for the offspring distribution p_k.

Theorem 1.1.4. *If $\mu > 1$ then $\sigma = P(Z_t > 0 \text{ for all } t) > 0$. The* **extinction probability** *$\rho = 1 - \sigma$ is the unique solution of $\phi(x) = x$ with $x \in [0, 1)$.*

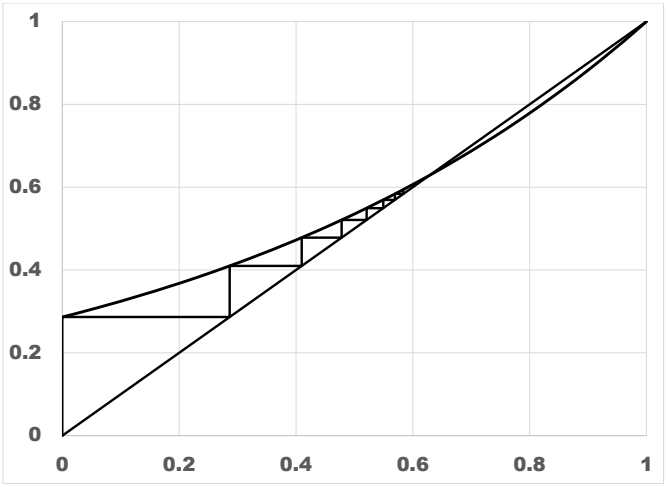

Figure 1.1: Generating function for the Poisson distribution with mean 1.25. The alternating vertical and horizontal lines illustrate the iteration of the gerating function: $(0,0) \to (0,\theta_1) \to (\theta_1,\theta_1) \to (\theta_1,\theta_2) \ldots$ Fixed point $\rho = 0.62861$.

Proof. Differentiating gives for $\theta < 1$

$$\phi'(\theta) = \sum_{k=1}^{\infty} k\, p_k \theta^{k-1} \geq 0,$$
$$\phi''(\theta) = \sum_{k=2}^{\infty} k(k-1) p_k \theta^{k-2} \geq 0.$$

So ϕ is increasing and convex, and $\lim_{\theta\uparrow 1} \phi'(\theta) = \sum_{k=1}^{\infty} k p_k = \mu$. Our interest in ϕ 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 θ_{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 ϵ. 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 ϕ 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 ϕ is increasing, so induction implies θ_t is increasing and $\theta_t \le \rho$. Let $\theta_\infty = \lim \theta_t$. Taking limits in $\theta_t = \phi(\theta_{t-1})$, we see $\theta_\infty = \phi(\theta_\infty)$. Since $\theta_\infty \le \rho$, it follows that $\theta_\infty = \rho$.

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

Example 1.1.5. Poisson distribution with mean λ.

$$P(\xi = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

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/μ^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 1.1.6. *If $\sum_k kp_k > 1$ then $W = \lim Z_t/\mu^t$ exists. If $\sum_k k^2 p_k < \infty$ then $W \not\equiv 0$.*

The ultimate result about this question is due to Kesten and Stigum (1966). For a "conceptual proof" see Lyons, Pemantle, and Peres (1995).

Theorem 1.1.7. *Suppose that $\mu = \sum_k kp_k > 1$ then $W = \lim Z_t/\mu^t \not\equiv 0$ if and only if $\sum_{k=1}^{\infty} p_k\, k \log k < \infty$.*

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

Theorem 1.1.8. *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 $r = P(W = 0)$. In order for Z_t/μ^t to converge to 0 the limit must be 0 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

$$r = \sum_{k=0}^{\infty} p_k r^k = \phi(r).$$

This shows $r < 1$ is a fixed point of the generating function and hence $r = \rho = P(Z_t = 0 \text{ for some } t)$. 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. □

1.1.1 Conditioning

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 whenwe condition it to die out?

Theorem 1.1.9. *A supercritical branching process conditioned to become extinct is a subcritical branching process. If the original offspring distribution has generating function $\phi(z)$ then the conditioned process has generating function $\rho^{-1}\phi(\rho z)$. Thus If the original offspring distribution is Poisson(λ) with $\lambda > 1$ then the conditioned one is Poisson($\lambda\rho$) where ρ is the extinction probability.*

Formula (1.1.2) shows the generating function of a supercritical branching process conditioned to die out is obtained by taking the graph of ϕ over $[0, \rho]$ and rescaling to make the domain and range $[0, 1]$. It is ckear from Figure 1.2 that the conditioned branching process is subcritical.

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

$$P(Z_{t+1} = z_{t+1}, T_0 < \infty | Z_t = z_t, \dots Z_0 = z_0) = P(Z_{t+1} = z_{t+1}, T_0 < \infty | Z_t = z_t).$$

To compute the transition probability for $\bar{Z}_t$, observe that if ρ is the extinction probability then $P_x(T_0 < \infty) = \rho^x$. Let $p(x, y)$ be the transition probability for Z_t. 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 the starting state $x = 1$ and computing the generating function

$$\bar{\phi}(\theta) = \sum_{y=0}^{\infty} \bar{p}(1, y)\theta^y = \rho^{-1} \sum_{y=0}^{\infty} p(1, y)(\theta\rho)^y = \rho^{-1}\phi(\theta\rho) \tag{1.1.2}$$

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

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

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

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

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. □

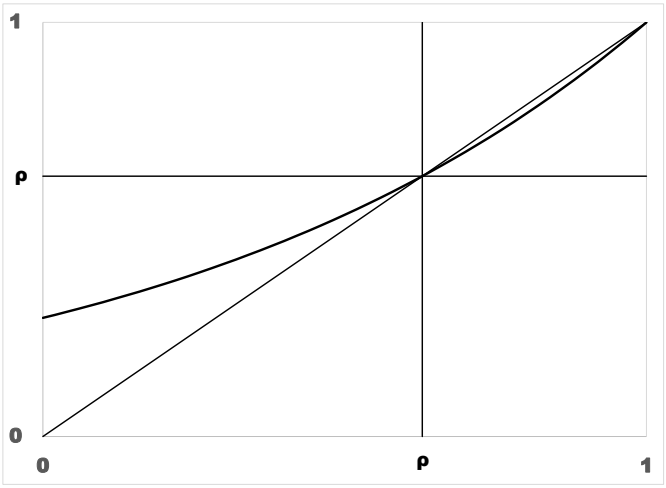

Figure 1.2: Generating function for the Poisson distribution with mean 1.25, illustrating Theorem 1.1.9 (g.f. in lower left quadrant) and Theorem 1.1.10 (g.f. in upper right). See formulas (1.1.2) and (1.1.3).

Theorem 1.1.10. *Consider a supercritical branching process with offspring distribution p_k and generating function ϕ. 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*

$$\tilde{\phi}(\theta) = \frac{\phi((1-\rho)\theta + \rho) - \rho}{1 - \rho} \tag{1.1.3}$$

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

In words, we take the graph of ϕ over $[\rho, 1]$ and rescale to make the domain and range $[0, 1]$.

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 θ^j, summing, and interchanging the order of summation

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

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

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

We can add the $k=0$ term to the sum since its value is $1-1=0$. Doing 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 a branching process

In this section we use branching process results to study the connected component, or cluster, containing 1. To do this we introduce a process that is the same as a discrete time epidemic. To begin the process, 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 theory 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:

$$R_{t+1} = R_t \cup I_t \tag{1.2.1}$$

$$I_{t+1} = \{y \in S_t : \eta_{x,y} = 1 \text{ for some } x \in I_t\} \tag{1.2.2}$$

$$S_{t+1} = S_t - I_{t+1} \tag{1.2.3}$$

where $\eta_{x,y} = \eta_{y,x} = 1$ if and only if there is an edge between x and y. The cluster containing 1, $\mathcal{C}_1 = \cup_{t=0}^{\infty} I_t = R_\infty$.

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 $E|\mathcal{C}_1| \le 1/(1-\lambda) < \infty$.*

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

$$Z_{t+1} = \sum_{x \in I_t, y \in S_t} \eta_{x,y} + \sum_{x \in I_t} \sum_{y \in S_t^c} \zeta_{x,y}^t + \sum_{x=n+1}^{n+Z_t-|I_t|} \sum_{y=1}^{n} \zeta_{x,y}^t. \tag{1.2.4}$$

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 it provides an upper bound on the growing cluster

$$Z_t \ge |I_t|.$$

When $\lambda < 1$. $EZ_t = \lambda^t$, so the mean cluster size

$$E|\mathcal{C}_1| = E\left(\sum_{t=0}^{\infty} |I_t|\right) \le \sum_{t=0}^{\infty} \lambda^t = 1/(1-\lambda)$$

which proves the desired result. □

Theorem 1.2.2. *Suppose $\lambda > 1$ and let $\rho < 1$ be the probability the branching process Z_t dies out. Then the probability vertex 1 belongs to a giant component, one with size $\geq \epsilon n$ for some $\epsilon > 0$, tends to $1 - \rho$ as $n \to \infty$.*

In the next section, we will give a simpler proof that there is a unique giant component and it has $\sim (1 - \rho)n$ members.

Proof. Since Z_t is an upper bound on I_t the probability of a giant component is $\leq 1 - \rho$. To prove the result in the theorem, we want to define a branching process that lower bounds the growing cluster. To be able to control the growth of the branching process we build it using a **breadth-first search**. If the word search sounds strange note that the construction gives an algorithm that can be used to compute the cluster. Modifying the notation used above,

At time 0, $A_0 = \{1\}$ is an ordered list of **active sites**, $U_0 = \{2, \ldots n\}$ are **unexplored sites**, and $R_0 = \emptyset$ are **removed sites** whose neighbors are known.

At time 1, $R_0 = \{1\}$, $A_1 = \{y : \eta_{1,y} = 1\}$ and $U_1 = U_0 - A_1$.

At time 2, we let i_2 be the smallest member of the ordered set A_1, let $N_2 = \{y \in U_1 : \eta_{i_2,y} = 1\}$, and let $U_2 = U_1 - N_2$.

At time m, if $A_m = \emptyset$ we are done and R_m is the cluster containing 1. If $A_m \neq \emptyset$ we pick the smallest element i_m in this ordered set, let $N_m = \{y \in U_{m-1} : \eta_{i_m,y} = 1\}$, let $U_m = U_{m-1} - N_m$ and $n_m = |N_m|$. To create A_m we put the elements of N_m in increasing order and add them after the current elements of A_{m-1},

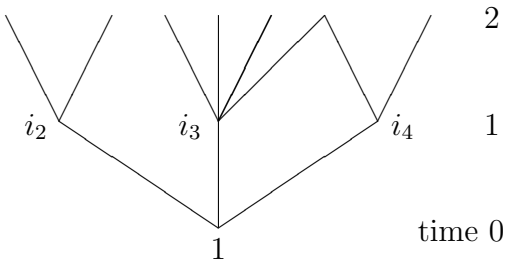

Figure 1.3: Example of breadth-first search. The number of vertices addedon the the ith step Here $n_1 = 3$, $n_2 = 2$, $n_3 = 4$, and $n_4 = 1$ since i_4 has a child that is also a child of i_3.

The branching process is embedded in the breadth first search. A picture (see Figure 1.3) is useful to help explain the definitions. Let $Z_0 = 1$ and $Z_1 = n_1$. Define $Y_0 = 1$ and for $m \geq 2$

$$Y_{m-1} = Z_0 + \cdots + Z_{m-1} = Y_{m-2} + Z_{m-1}$$

be the total number of children up to generation $m-1$, and

$$Z_m = n_{Y_{m-2}+1} + \cdots + n_{Y_{m-1}}.$$

To check the definitions in the example, $Y_0 + 1 = 2$ and $Y_1 = Z_0 + Z_1 = 4$, so

$$Z_2 = n_2 + n_3 + n_4 = 2 + 4 + 1 = 7$$

This means $Y_2 = 4 + 7 = 11$ and hence $Z_3 = n_5 + n_6 + \cdots + n_{11}$.

Turning to the proof of Theorem 1.2.2, let δ be chosen so that $\bar{\lambda} = \lambda(1-\delta) > 1$. While the total number of offspring $Y_m \le \delta n$ the number of offspring n_m dominates a binomial$(n, (1-\delta)\lambda/n) \approx$ Poisson$(\lambda(1-\delta))$. The extinction probability of the Poisson branching process satisfies

$$\rho_\delta = \exp(\lambda(1-\delta)(\rho_\delta - 1)).$$

It is easy to see that $\rho_\delta \to \rho$ as $\delta \to 0$. Completing the proof is a little slippery because we need to interchange the limits $\delta \to 0$ and $n \to \infty$ Fix a small value of δ so that $\rho_\delta < \rho + \eta$. If we let $m(n, \delta)$ be the the number of generations that are completed before the total number of births reaches δn then $m(n, \delta) \to \infty$. Since the probability the branching process dies out in the first k generations can be computed by iterating the generating function, it follows that the probability that the Poisson$((1-\delta)\lambda)$ survives until generation k and then dies out out is $< \eta$ if k is large. □

An important application of breadth first search is the qualitative property "Erdős-Rényi clusters are locally tree like." Let $T_\alpha = \min\{t : U_t \le n - n^\alpha\}$, and note that at time m when we make connections from i_m to its neighbors we cannot make a connection to a vertex in R_m since we have identified all of their neighbors, but we can make a connection to another member of the set of active vertices A_m. If this occurs, an event we call a **collision**, then we create a loop in the graph and it is no longer a tree.

Theorem 1.2.3. *The expected number of collisions at times $t < T_\alpha$ is $\lambda n^{2\alpha-1}$. Taking $\alpha = 1/2 - \epsilon$ this result implies that the cluster will with high probability be a tree until it contains $n^{1/2-\epsilon}$ vertices.*

Proof. At times $t < T_\alpha$, $|A_t| \le n^\alpha$. If $T_\alpha < \infty$ then $T_\alpha \le n^\alpha$ since at each time we add a vertex to the removed set. These two observations imply that the number of pairs of verties that can be connected is $n^{2\alpha}$ while the probability a connection is made is $\le \lambda/n$. □

Remark 1.2.4. A better upper bound on the probbaility results if we note that the number of collisions is dominated by a binomial$(n^{2\alpha}, \lambda/n)$ random variable. In some cases we will need a conclusion that holds for every cluster in the graph. We can do that for Erdős-Rényi graphs by chosing α to be smaller and using the domination by a binomial.

1.2.1 Distances on the giant component

The branching process results derived earlier in this section can be used to study the typical distance between two points on the giant component.

Theorem 1.2.5. *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 1.2.5 is intuitive. Consider the cluster growing from 1. The size of the branching process at time t is $\sim \lambda^t W$. When $t = 1-\epsilon)(\log n)/(\log\lambda)$ only $n^{1-\epsilon}$ can be reached, but when $\epsilon = 0$ the number is $\Theta(n)$.

Proof. The lower bound on the distance comes from the last observation and the fact that branching process is an upper bound on the growth of the cluster. To prove the upper bound requires more work since the branching process approximation is not valid out to time $t = (\log n)/\log\lambda$. To get around this difficulty we will grow clusters starting from x and y for time $t(\epsilon,n) = (1/2+\epsilon)(\log n)/\log\lambda$. If there were no collisions then the number of offspring is of order $n^{1/2+\epsilon}$. Lemma 1.2.3 implies that the expected number of collisions up to time $t(\epsilon,n)$ is $\lambda n^{2\epsilon}$. When we have a collision we lose not only that vertex but also all of its children. Theorem 1.2.3 shows that with high probability there is no collision before time $(1/2-\epsilon)(\log n)/\log\lambda$. I we are conservative and suppose that the clusters of the vertices lost to collisions all have time $2\epsilon(\log n)/\log\lambda$ to grow the number of lost particles is $n^{4\epsilon} \ll n^{1/2+\epsilon}$ if ϵ is small.

At time $t(\epsilon,n)$ there are two cases to consider. (i) The clusters grown from x and y have already intersected. In this case the distance form x to y is $\le 2t(\epsilon)$ or (ii) they have not intersected. If the clusters growth from x and y have not died out and δ is small then with high probability they will be $\ge \delta n^{1/2+\epsilon}$ active vertices in the growing cluster at time $t(\epsilon,n)$, so the probability they will not intersect on the next step is

$$\le \left(1-\frac{\lambda}{n}\right)^{\delta^2 n^{1+2\epsilon}} \to 0$$

which completes the proof. □

Dangling ends and the diameter

Our next goal is to show that the diameter $D = \max d(x,y)$ of the giant component in an Erdős-Rényi random graph is larger by a constant factor than the distance between two randomly chosen points on it. To do this we define a **dangling end** of length k to be a self-avoiding path $v_0, v_1, \ldots v_k$ with the degree $d(v_0) = 1$ and $d(v_i) = 2$ for $1 \le i < k$, and $d(v_k) \ge 3$.

These requirements are imposed so that two dangling ends of length k cannot intersect, except possibly at the final vertex v_k. To explain, note that since $d(v_0) = 1$ and $d(v_i) = 2$ for $1 \le i < k$ only one dangling end can start at a given vertex. The vertices v_i with $1 \le i < k$

cannot be in two dangling ends since if we look at the first vertex to be in two then it would have to have degree 3. The limited ability of dangling ends to overlap will lead to the conclusion that the limiting number of dangling ends with a fixed length k will be Poisson. To prove this we use the following.

Lemma 1.2.6. *Let $x^{(k)} = x(x-1)\cdots(x-k+1)$ and let Z_n be a sequence of nonnegative integer valued random variables with $EZ_n^{(k)} \to \lambda^k$ for all positive integers k. Then Z_n converges in distribution to $Z_\infty = Poisson(\lambda)$.*

Proof. It is easy to chek that if $Z = \text{Poisson}(\lambda)$ then $EZ_\infty^{(k)} = \lambda^k$. Since there are constants $c_{k,j}$ so that

$$x^k = \sum_{j=1}^{k} c_{k,j} x^{(j)}$$

we have $EZ_n^k \to EZ_\infty^k$. The Poisson distribution is determined by its moments so the conclusion follows, see e.g., Section 3.3.5 in PTE5. □

Let $\lfloor x \rfloor$ be the largest integer $\le x$.

Theorem 1.2.7. *Let $p_1 = e^{-\lambda}\lambda$ and choose $k(n) = \lfloor \alpha \cdot \log n / \log(1/p_1) \rfloor$ so that $c_n = np_1^{k(n)}$ stays bounded away from 0 and ∞. The number of dangling ends of length $k(n)$, $M_n \approx Poisson(c_n)$. If α is small then $M_n > 0$ with high probability.*

Proof. To use Lemma 1.2.6 we need to compute the moments of $M = M_n$. Let $A_{(x,y)}$ be the event that there is dangling end $v_0 = x, v_1, \ldots v_{k+1} = y$. Let $P_{n,k+1} = n \cdot (n-1) \cdots (n-k)$ (where P is for permutations) be the number of ways to pick the vertices $v_0, \ldots v_k$, which are all distinct

$$EM = \sum_{(x,y)} P(A_{(x,y)}) \sim P_{n,k+1}(\lambda/n)^k (1-\lambda/n)^{nk} \cdot q_2 \sim n(\lambda e^{-1})^k$$

where $q_2 = P(\text{Poisson}(\lambda) \ge 2)$.

Here $(\lambda/n)^k$ accounts for the k connections that exist between the v_i, $0 \le i \le k$, and the factor $(1-\lambda/n)^{nk}$ gives the probability that no other connections that do not exist. To explain the second formula, note that 0 is not connected to $n-1$ vertices and the v_i with $1 \le i < k$ are not connected to $n-2$ vertices. However in this list of nonexisting connections $\binom{k}{2}$ of them appear twice in the list, so the exact exponent is $(n-1)+(n-2)(k-1)-\binom{k}{2}$. This differs from nk by $\le k^2$ so if $k = O(\log n)$ the difference is insignificant.

Our next step is to investigate the second factorial moment $E[M(M-1)]$. To simplify the calculation we will first consider the case in which the two dangling ends do not end at the same vertex. We will then check that probability for two ending at the same place is smaller. Thus our goal is to compute

$$\sum_{x \ne w, y \ne z} P(A_{(x,y)} \cap A_{(w,z)}).$$

Let $v_0 = x, v_1, \ldots v_k = y$ be the path that produces the first one and let $u_0 = w, u_1, \ldots u_k = z$ be the path that produces the second one. Note that the degrees of the vertices in the dangling end imply that $x \neq z$ and $y \neq w$.

- Since the paths are disjoint and self-avoiding, there are $P_{m,2(k+1)}$ ways of picking the two sequences (including the choice of x and w). Note that if $k = O(\log n)$,
$$P_{m,2(k+1)}/P^2_{m,(k+1)} \to 1.$$
- As in the formula for the mean, a factor of $(\lambda/n)^{2k}$ accounts for the $2k$ connections that exist between the vertices in the dangling ends and a factor $(1-\lambda/n)^{2kn}$ gives the probability (up to a factor $1 + o(1)$) that no other connections that do not exist.
- Given the existence of a path satisfying the first two conditions, we have $d(v_k) \geq 3$ with probability $\to q_2$.
- Finally note that the number of ways of picking vertices for two dangling ends that intersect in the final vertex is $P_{m,2k+1}$ so the probability of this event is of smaller order than that of two disjoint dangling ends.

It is straightforward to extend the last argument to show asymptotic independence of m disjoint dangling ends for any m.

If α is small then EM is large, so with high probability there is at least one dangling end. To show that it is connected to the giant component with positive probability run the cluster growth process starting from v_k and notice that the only effect of the dangling end is to reduce the number of vertices. If one wants the construction to succeed with high probability replace 3 by a much larger number.

□

The result we have just proved is not very precise. Riordan and Wormald (2010) have found the exact order. To state the result we will need a fact from Section 1.6, see (1.6.4). If ρ is the extinction probability of a Poisson(λ) branching process then
$$\lambda\rho e^{-\lambda\rho} = \lambda e^{-\lambda}.$$

Theorem 1.2.8. *Let $\lambda > 1$ and $\lambda_* = \lambda\rho$.*
$$\textit{diameter}(G(n, \lambda/n)) = \frac{\log n}{\log \lambda} + \frac{2\log n}{\log(1/\lambda_*)} + O_p(1)$$
where $O_p(1)$ denotes a term X_n with $P(X_n > K_n) \to 0$ for any $K_n \to \infty$.

The first term on the right is the typical distance betweeen two points while the second gives the contribution of two dangling ends.

There are a large number of results about what happens when $\lambda \to \infty$ at different rates. See Chung and Lu (2001) for details. For example if $\lambda/(\log n) \to \infty$ the diameter is concentrated on at most 2 values while if $\lambda/\log n = c$ then it is concentrated on at most 2, 3, or 4 values if $c > 8$, $c \in (2, 8]$, or $c \in (1, 2]$. In addition there are four other results.

1.3 Cluster growth as a random walk

Although the connection with branching processes is intuitive, it is more convenient technically to grow the cluster by adding the neighbors of one site at a time in order to obtain something that can be approximated by a random walk. In this section we will introduce that approach, use it to prove the existence of a giant component when $p = \lambda/n$ with $\lambda > 1$ and show.that as $n \to \infty$ has $\sim (1 - \rho(\lambda))n$ vertices where $\rho(\lambda)$ is the extinction probability for the Poisson(λ) branching process.

As in the previous section, we let R_t be the set of removed sites, U_t the unexplored sites and A_t the active sites. However, here we do not need a close connection with a branching process, so we do not need to impose an order on A_t. The initial conditions are $R_0 = \emptyset$, $U_0 = \{2, 3, \ldots, n\}$, and $A_0 = \{1\}$. At time $\tau = \inf\{t : A_t = \emptyset\}$ the process stops, and R_τ is the cluster contianing 1. Since one vertex is added to the removed set at each time, there are τ points in the cluster. If $A_t \neq \emptyset$, pick i_t from A_t according to some rule that is measurable with respect to $\mathcal{A}_t = \sigma(A_0, \ldots A_t)$ and let

$$\begin{aligned} R_{t+1} &= R_t \cup \{i_t\}, \\ A_{t+1} &= A_t - \{i_t\} \cup \{y \in U_t : \eta_{i_t,y} = 1\}, \\ U_{t+1} &= U_t - \{y \in U_t : \eta_{i_t,y} = 1\}. \end{aligned} \tag{1.3.1}$$

We call this the **exploration process**.

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

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

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 $\tau \leq T$. We continue the definition after $A_t = \emptyset$ because it is convenient to have the random walk defined for all time.

The increments $X_t = S_t - S_{t-1}$, $t \geq 1$ of the random walk are $-1 +$ binomial$(n, \lambda/n)$. If $\lambda < 1$ stopping the martingale $S_t - (\lambda - 1)t$ at the bounded stopping time $T \wedge t$ (where $x \wedge y = \min\{x, y\}$) 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

$$ET = 1/(1 - \lambda). \tag{1.3.2}$$

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*

$$P\left(\max_{1\le x\le n} |\mathcal{C}_x| \ge a \log n\right) \to 0.$$

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 of $X_i = -1+\text{binomial}(n, \lambda/n)$

$$E\exp(\theta X_i)) = e^{-\theta}(1 - \lambda/n + (\lambda/n)e^{\theta})^n \tag{1.3.3}$$

$$\le \exp(-\theta + \lambda(e^{\theta} - 1)) = \psi(\theta) \tag{1.3.4}$$

since $1 + x \le e^x$. Note that the right-hand side is the moment generating function of $-1 +$ Poisson(λ). The mean $\psi'(0) = EX_i = \lambda - 1$, 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.

$$1/\lambda = e^{\theta_1} \ge E(\psi(\theta_1)^{-T}) = E(e^{\alpha T}).$$

Using Chebyshev's inequality now

$$e^{k\alpha} P(T \ge k) \le 1/\lambda. \tag{1.3.5}$$

Letting $\mathcal{C}_x$ denote the cluster containing x, noting that $T \ge |\mathcal{C}_x|$ in distribution, and taking $k = (1 + \epsilon)(\log n)/\alpha$

$$P(|\mathcal{C}_x| \ge (1 + \epsilon)(\log n)/\alpha) \le n^{-(1+\epsilon)}/\lambda$$

from which the desired result follows. □

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 $\hat{A}_t \ne \emptyset$ and $\hat{U}^t \ge (1 - \delta)n$ which corresponds to $|\hat{A}_t| + t \le n\delta$, we can define

$$\begin{aligned}
\hat{R}_{t+1} &= \hat{R}_t \cup \{j_t\}\\
\hat{A}_{t+1} &= \hat{A}_t - \{j_t\} \cup \{y \in \hat{U}_t^{\delta} : \eta_{j_t,y} = 1\}\\
\hat{U}_{t+1} &= \hat{U}_t - \{y \in \hat{U}_t^{\delta} : \eta_{j_t,y} = 1\}
\end{aligned}$$

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

$$T_W = \inf\{s : W_s = 0, \text{ or } W_s + s \geq n\delta\}$$

and define

$$W_{t+1} = W_t - 1 + \begin{cases} \sum_{y \in \hat{U}_t^\delta} \eta_{i_t,y} & \text{if } t < T_W, \\ \sum_{y=1}^{n(1-\delta)} \zeta_y^t & \text{if } t \geq T_W. \end{cases}$$

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 β 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(λ) branching process.*

Here *with high probability* means that the probability tends to 1 as $n \to \infty$.

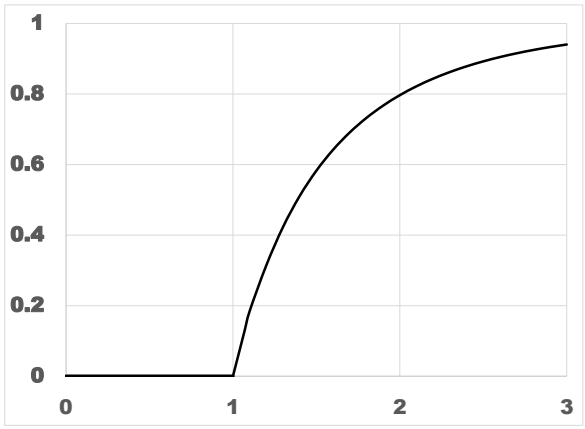

Figure 1.4: Graph of the surival probability $1 - \rho(\lambda)$.

Proof. There are four steps.

Step 1. There is a constant γ 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 β 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 $\epsilon n^{2/3} \leq |A(n^{2/3})| \leq 2\lambda n^{2/3}$. 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 $\mathcal{C}_x$ 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 $\{|\mathcal{C}_x| \geq \beta \log n$ 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 (1.3.4) the moment generating function of an increment

$$\phi_\delta(\theta) \leq \exp(-\theta + \lambda(1-\delta)(e^\theta - 1)). \tag{1.3.6}$$

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 ∞ 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 nonewgative supermartingales) and noting that $W_t \to \infty$ on $\{T_0 = \infty\}$ we have

$$e^{-k\theta_\delta} \geq P_k(T_0 < \infty). \tag{1.3.7}$$

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 $\mu = EX_s$. if $b < \mu < a$ then there are constant $\eta(b), \eta(a) > 0$ so that*

$$P(Z \geq ta) \leq \exp(-t\eta(a)),$$
$$P(Z \leq tb) \leq \exp(-t\eta(b)).$$

Proof. These are standard results on large deviations, see Theorem A.1.8 □

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 is an $\eta > 0$ so that

$$P(S_t \geq (2\lambda - 1)t) \leq \exp(-\eta t). \tag{1.3.8}$$

Let $r = \beta \log n$ where $\beta(\lambda - 1)/2 > 2\gamma$ and $\exp(-\eta\beta \log n) \leq n^{-2}$. 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 ist he sum of $r \times 2\lambda r$ Bernoulli random variable with success probability λ/n. From this it follows that

$$P(|A_r| > 0, S_r - |A_r| \geq 2) \leq \binom{2\lambda r^2}{2} (\lambda/n)^2 \leq C\frac{(\log n)^2}{n}.$$

Combining our results

$$P(0 < |A(\beta \log n)| \le \gamma \log n) \le C\frac{(\log n)^2}{n^2}.$$

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

$$P(W(n^{2/3}) - W(0) \le \epsilon_\delta n^{2/3}) \le \exp(-\eta_2 n^{2/3}), \tag{1.3.9}$$

$$P(W(n^{2/3}) - W(0) + n^{2/3} \ge 2\lambda n^{2/3}) \le \exp(-\eta_3 n^{2/3}). \tag{1.3.10}$$

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

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 $\ge \epsilon_\delta n^{2/3}$ with high probability. Thus if we have two such clusters of size $\ge \beta \log n$ then either (a) they will intersect by time $r + n^{2/3}$ of (b) at time $r + n^{2/3}$ they have disjoint sets V_1 and V_2 of active sites of size $\ge \epsilon_\delta n^{2/3}$. The probability of no edge connecting V_1 and V_2 is

$$= \left(1 - \frac{\lambda}{n}\right)^{\epsilon_\delta^2 n^{4/3}} \le \exp(-\eta_4 n^{1/3}). \tag{1.3.11}$$

This proves the first assertion in Theorem 1.3.2.

Step 4. To prove the second assertion it suffices to show that

$$|\{x : |\mathcal{C}_x| \le \beta \log n\}|/n \to \rho(\lambda). \tag{1.3.12}$$

The first step is to note that the branching process approach to cluster growth implies

$$P(|\mathcal{C}_x| \le \beta \log n) \to \rho(\lambda).$$

To comlete the proof we bound the correlation of the random variables $Y_x = 1$ if $|\mathcal{C}_x| \le \beta \log n$ 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*

$$|P(F \cap G) - P(F)P(G)| \le JK \cdot \frac{2\lambda}{n}$$

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 $A'_t \ne \emptyset$, pick $i'_t = \min A'_t$. If $i'_t \notin R_{\beta \log n}$ let

$$R'_{t+1} = R'_t \cup \{i'_t\}, \tag{1.3.13}$$

$$A'_{t+1} = A'_t - \{i'_t\} \cup \{y \in U'_t : \eta'_{i'_t,y} = 1\}, \tag{1.3.14}$$

$$U'_{t+1} = U'_t - \{y \in U'_t : \eta'_{i'_t,y} = 1\}, \tag{1.3.15}$$

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

$$|P(F \cap G) - P(F)P(G)| \le P(R_J \cap R'_K \neq \emptyset) \le JK \cdot \frac{\lambda}{n}$$

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

$$\text{var}\left(\sum_{x=1}^{n} Y_x\right) \le n + n(n-1)\frac{\lambda(\beta \log n)^2}{n} \le Cn \log n$$

so it follows from Chebyshev's inequality that

$$P\left(\sum_{x=1}^{n}(Y_x - EY_x) \ge n^{2/3}\right) \le \frac{Cn \log n}{n^{4/3}} \to 0.$$

This proves (1.3.12) and completes the proof of Theorem 1.3.2.

1.4 Long paths

At this point we know that the giant component exists for $\lambda > 1$ and is unique, but we do not know much about what it looks like. Ajtai, Komlpos, and Szememredi (1981) proved that there was a function $\alpha : (1, \infty) \to (0, 1)$ so that the Erdös-Renyi random graph with $\lambda = c/n$ contains a path of length at least $\alpha(c)n$. When we study the contact process on graph this result can be used to give a very easy proof that when λ is large enough the contact survives for time $\exp(\theta n)$.

1.4.1. Krivelevich and Sudakov (2012)

have given a simple proof based on **depth-first search.** Given a graph $G = (V, E)$, which we will asume is undirected the algorithm uses an ordering σ on the vertices of V and divides them into three disjoint sets that change over time. Here we have changed the notation in the paper to match that in the next section.

- R_t is the set of **retired vertices** for which the exploration is complete.
- S_t is the set of **sleeping vertices** that have not yet been touched.
- $A_t = V - (R_t \cup S_t)$ are the **active vertices** kept in a last in, first out data structure.

It follows from the construction given below and induction, that at any time t

(B1) There are no edges between R_t and S_t.

(B2) The vertices in A_t are a path.

On each step of the algorithm we might have a vertex moves from S to A or from A to R, but on most steps nothing happens. To be precise, a step consists of one query - when we inspect the graph to see if there is an edge from i to j. The edge is present if $X_{ij} = 1$ where $X_{ij} = X_{ji}$ are random variables that indicate the presence of the edges.

- If $A \neq \emptyset$, we let v be the vertex at the top of the stack. If v does not have a neighbor in S (which we know because we have queried all the possible connections to S) then v is removed from A and added to R. Once vertices enter R they never leave.
- If we have not queried all the possible connections from v to S then we query the next vertex u in S in the ordering. If our query finds that u is a neighbor of v then u is deleted from S and added to A and replaces v as the top vertex in the stack.
- If $A = \emptyset$, then we have just completed computing one of the components of the graph. The algorithm chooses the first vertex in S according to the ordering σ. When $A \cup S = \emptyset$ all the connected components have been determined.

The following estimate is an important part of the proof.

Lemma 1.4.1. *Let X_{ij} with $1 \le i < j \le n$ be i.i.d. Bernoulli$((1+\epsilon)/n)$ indexed by the edges of the graph. Let F be a set of edges. If $|F| = cn^2$ then whp*

$$\left| \sum_{ij \in F} X_{ij} - c(1+\epsilon)n \right| \le n^{2/3}.$$

Proof. The sum has mean $c(1+\epsilon)n$ and variance $\le c(1+\epsilon)n$ so

$$P\left(\left| \sum_{ij \in F} X_{ij} - c(1+\epsilon)n \right| \le n^{2/3} \right) \le \frac{c(1+\epsilon)n}{n^{4/3}}$$

by Chebyshev's inequality, proving the desired result. □

Theorem 1.4.2. *If $b < 1/4$ and $\eta > 0$ is small enough then when $\epsilon < \epsilon_0(b, \eta)$ whp Erdős-Rényi $(n, (1+\epsilon)/n)$ contains a path of length $b\epsilon^2 n$.*

Improved lower bounds. The result in the next section will give a lower bound $\sim \epsilon^2$. Ding, Lubetzky and Peres (2014) and Robinson and Wormald (1992) have proved $(4/3)\epsilon^2$ as a lower bound.

Upper bounds. In Section 1.8 we will see that the size of the giant component when $\lambda = 1 + \epsilon$ is $\sim C\epsilon$ as $\epsilon \to 0$. To get a better upper bound we note that Pittel (1990) has shown that the fraction vertices in the 2-core, which is the maximal subgraph in which all vertices have degree ≥ 2 is $\le C\epsilon^2$. A long path may not have any subgraph in which all vertices have degree ≥ 2. We leave it to the reader to show that if there is path of length cn when $\lambda = 1 + \epsilon$ then with high probability there is a circle of length $cn - n^{0.6}/\sqrt{ep}$ when $\lambda = 1 + 2\epsilon$.

Proof. We will lower bound the size of A_N at time $N = \epsilon n^2/2$ and then the result follows from (B2). We first show that if η is small and $\eta > 2\epsilon$ then $|R_N| \le \eta n$. Suppose not, and let t be the first time $|R_t| = \eta n$. At that moment

$$|A_t| \le 1 + \sum_{s-1}^{t} X_s \le \epsilon(1+\epsilon)n/2 + n^{2/3} < \eta n$$

by the Lemma 1.4.1 with $c = \epsilon/2$ and the choice of η. This implies

$$|S_t| = n - |R_t| - |A_t| \ge n(1 - 2\eta) \ge 2n/3$$

if η is small enough. At this point the algorithm must have queried all of the

$$|R_t| \cdot |S_t| \ge \frac{2n}{3} \dot{\eta} n n \geq^2 /3 > N$$

edges connection R_t to S_t and found them to be missing, but this is impossible. since there has not been enough time to do that.

We now return to time N. If

$$|R_N| \le \eta n \quad \text{and} \quad |A_N| \le b\epsilon^2 n$$

then we have $S_N \neq \emptyset$. This means that we are still revealing connected components of G. Each positive answer we get resulted in moving a vertex from S to A (some of which may have been later moved to R). By Lemma 1.4.1 with $c = \epsilon/2$, the number of positive answers to queries at time $N = \epsilon n^2/2$ must be at least $\epsilon(1+\epsilon)n/2 - n^{2/3}$ and hence we have

$$|R_N \cup A_N| \ge \frac{\epsilon n}{2} + \frac{\epsilon^2 n}{2} - n^{2/3}. \tag{1.4.1}$$

The next step is where we use that $\lambda = 1 + \epsilon > 1$. If $|A_N| \le b\epsilon^2 n$ then using (1.4.1) and the fact that all vertices are in A_N, R_N or S_N we have

$$|R_N| \ge \frac{\epsilon n}{2} + (0.5 - b)\epsilon^2 n - n^{2/3}, \tag{1.4.2}$$

$$|S_N| = n - |R_N| - |A_N| \ge n - |R_N| - b\epsilon^2 n. \tag{1.4.3}$$

Using $R_N \le \eta n$, $|A_N| \le b\epsilon^2 n$, and $|S_N| = n - |R_N| - |A_N|$ gives

$$\begin{aligned} |R_N| \cdot |S_N| &= |R_N|(n - |R_N| - |A_N|) \\ &\ge |R_N|(n - |R_N|) - \eta n \cdot b\epsilon^2 n. \end{aligned}$$

Using the fact that $m \to m(n-m)$ is increasing on $[0, n/2]$ together with (1.4.2) we have that the above is

$$\begin{aligned} &\ge \left(\frac{\epsilon n}{2} + (0.5 - b)\epsilon^2 n - n^{2/3}\right) \cdot \left(n - \frac{\epsilon n}{2} - (0.5 - b)\epsilon^2 n + n^{2/3}\right) - \eta b\epsilon^2 n^2 \\ &\ge \frac{\epsilon n^2}{2} + (0.5 - b)\epsilon^2 n^2 - \frac{\epsilon^2 n^2}{4} - \eta b\epsilon^2 n^2 - O(\epsilon^3 n^2) - O(n^{5/3}). \end{aligned}$$

To check the second inequality call the terms in the first parenthesis 1, 2, and 3, and those in the second w, x, y, and z. The first term in the second equation is 1w, the second 2w and , the third 1x. The terms 2x, 1y, 2y are $O(\epsilon^3 n^2)$ while terms involving 3 or z are $O(n^{5/3})$ The sum of the first four terms in the last lin of the display above is

$$N + [(0.5 - b) - 0.25 - b\eta] \cdot \epsilon^2 n.$$

If $b < 1/4$ and η is small, then the second term is positive, so the right hand side is again $> N$ and we have a contradiction. □

1.4.2. Enriquez, Faraud, and Ménard (2017)

To quote their abstract, "we show that the profile of the tree constructed by the depth first search algorithm on the giant component of an Erdős-Rényi graph with n vertices and connection probability c/n with $c > 1$ converges to an explicit deterministic shape." This makes it possible to prove a very precise result about longest paths. Following our usual style the number of the results in the paper are in parentheses.

Theorem 1.4.3. *(Theorem 1) Let H_n be the length of the longest simple path.*

$$\liminf_{N\to\infty} H_N/N \geq \rho_c - Li_2(\rho_c)/c.$$

Here $Li_2(x)$ is the **dilogairthm function** which can be defined as $\sum_{k=1}^{\infty} z^k/k^2$ or

$$Li_2(z) = -\int_0^z \frac{\log(1-t)}{t}\, dt. \tag{1.4.4}$$

To explain the remark after Theorem 1.4.2 recall that $\log(1-t) = -t + t^2/2 - t^3/3 + \cdots$ so

$$Li(z) \approx \int_0^z 1 - t/2 + O(t^2)\, dt = z - z^2/4 + O(z^3)$$

and $z - Li_2(z) \sim z^2/4$.

Their definition of depth-first search is a little different from the previous discussion. At each step we define the following objects.

- A_m is an ordered set of **active vertices**.
- a_m is the last element in A_m
- S_m is a set of **sleeping vertices**.
- $R_m = \{1, 2 \ldots n\} - (A_m \cup S_m)$ is a set of **retired vertices**.

As beforet (i) the vertices in R_m have no neighbors in S_m, and (ii) at any time A_m is a path.

Initially $A_0 = \{1\}$, $S_0 = \{2, 3, \ldots n\}$, $R_0 = \emptyset$. The algorithm is the same as the previous one once we omit the steps that do nothing.

- If a_m has a neighbor in S_m we set $a_{m+1} = \min\{b \in S_m : \{a_m, b\} \in E_n\}$ where E_n is the set of edges of the graph. $A_{m+1} = A_m \cup \{b\}$ i.e., b is added as the last element of A_{m+1}, $S_{m+1} = S_m - \{a_{m+1}\}$, $R_{m+1} = R_m$.
- If a_m has no neighbor in S_m we set $A_{m+1} = A_m - \{a_m\}$. $R_{m+1} = R_m \cup \{a_m\}$.

The process stops at $\tau = \min\{m : A_m = \emptyset\}$. The sequence of vertices a_m is a nearest neighbor walk on $\mathcal{C}_1$ the connected component containing 1. The current height of the walker is denoted by $X_m = |A_m| - 1$. This defines a path that starts at 0, has increments in $\{-1, 1\}$ and is nonnegative except at its final value, which is -1. This occurs at time$\tau = 2|\mathcal{C}_1| - 1$ where $|\mathcal{C}_1|$ is the number of vertices in the component containing 1.

Scaling limit of DFS

Since we are primarily interested in the geometry of the giant component, we study the behavior of X_m conditioned on **S** the event that 1 belongs to the giant component.

Theorem 1.4.4. *(Theorem 2). Conditional on* **S**

$$\lim_{n\to\infty} X_{[tN]}/N = h(t)$$

where h is continuous and defined on $[0, 2\rho_c]$, where ρ_c is the fraction of vertices in the giant component.

To define h we introduce functions $f(\rho)$ and $g(\rho)$ defined on $[0, \rho_c]$ by

$$f(\rho) = \frac{1}{c}\left[Li_2(\rho_c) - Li_2(\rho) + \log\frac{1-\rho_c}{1-\rho} - 2\left(\frac{\log(1-\rho_c)}{\rho_c} - \frac{\log(1-\rho)}{\rho}\right)\right], \tag{1.4.5}$$

$$g(\rho) = \frac{1}{c}\left[Li_2(\rho) - Li_2(\rho_c) + \log\frac{1-\rho}{1-\rho_c}\right]. \tag{1.4.6}$$

The graph of h has an increasing part and a decreasing part which are defined by

$$(t, h(t))_{0\le t\le f(0)} = (f(\rho), g(\rho))_{0\le\rho\le\rho_c}, \tag{1.4.7}$$

$$(t, h(t))_{f(0)\le t\le 2\rho_c} = (f(\rho) + 2\rho[1 - (f(\rho) + g(\rho))/2], g(\rho))_{0\le\rho\le\rho_c}. \tag{1.4.8}$$

Theorem 1 is obtained from this by computing the maximum height of the curve given in Theorem 2 which is

$$g(0) = \frac{1}{c}(-\log(1-\rho_c) - Li_2(\rho_c)) = \rho_c - \frac{Li_2(\rho_c)}{c}.$$

where in the second equlaity we have use the equation satisifed by the survival probability

$$1 - \rho_c = \exp(-c\rho_c) \quad \text{which implies} \quad \frac{\log(1-\rho_c)}{\rho_c} = -c \tag{1.4.9}$$

To understand the formulas for f and g it is useful to have a concrete example. We choose $c = 1.5$ since there is a graph in the paper for this value of c. A little computation shows that

$$\rho_c = 0.58282, \quad Li_2(\rho_c) = 0.701584, \quad \text{and} \quad \log(1-\rho_c) = -0.874227.$$

As a check on our computations we note that $\log(1-\rho_c)/\rho_c = -1.5$

$$f(0) = \frac{1}{c}[Li_2(\rho_c) + \log(1-\rho_c) - 2(-c+1)] = 0.551565$$

$$g(0) = \frac{1}{c}[-Li_2(\rho_c) - \log(1-\rho_c)] = 0.115102$$

while $f(\rho_c) = g(\rho_c) = 0$.

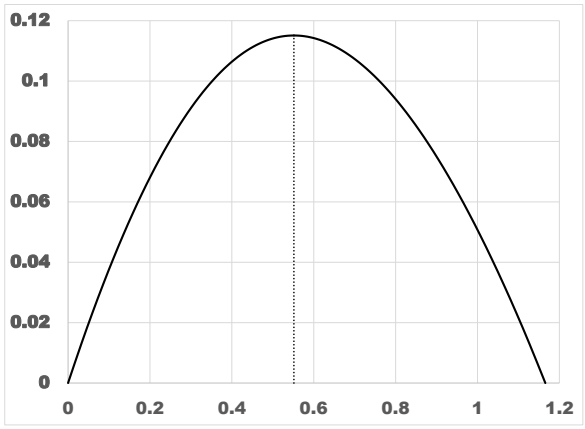

Figure 1.5: Graph of $(t, h(t)$ for $0 \le f(0) \le 2\rho_c$. The graph looks symmetric but $f(0) = 0.551565$ while $2\rho_c = 2(0.58282)$

Main ideas that lead to the proof

Let $\alpha_m = |A_m \cup R_m|/n$ be the fraction of the vertices explored up to time m.

$$\alpha_m = (X_m + m + 2)/2n. \tag{1.4.10}$$

Proof. To prove the revised formula by induction we note that $|A_0| = 1$, $|R_0| = 0$, $X_0 = 0$ so

$$\alpha_0 = 1/n = (0 + 0 + 2)/2n = (X_m + m + 2)/2n.$$

If a_m has a neighbor in S_m, then $|R_m|$ does not change while $|A_m|$, X_m, m increase by 1. If a_m does not have a neighbor in S_m, a_m moves from A_m to R_m, so α_m does not change while m increases by 1 and X_m decreases by 1. □

When A_m shrinks, the evolution of X_m is quite complicated since the walk has returned to a vertex whose neighbors have been partially explored. In contrast when the process visits a vertex a_m for the first time, things are very simple. At such a time the algorithm has never examined the connections between a_m and the vertices of S_m, so the number of neighbors of a_m in S_m,

$$\mathcal{N}(a_m) = \text{binomial}((1 - \alpha_m)n, c/n) \approx \text{Poisson}((1 - \alpha_m)c).$$

The reader should note that as the algorithm progresses the neighborhood of a_m looks like an Erdős-Rényi graph but the mean changes, This motivates the division of the process into two phases

If $(1 - \alpha_m)c > 1$ then we are **"on the way up."** i.e., we are in the increasing part of h.

If $(1 - \alpha_m) < c$ then we are **"on the way down,"** i.e., we are in the decreasing part of h.

Since we will be happy with lower bounds on the path length, we will only concern ourselves with the way up. In addition, as we will explain in a moment, we will not go all the way to the end of the upward phase.

Renewal times

The evolution on the way up is simplified by the fact that if a_m is part of the giant component then the path before a_M will not be erased until we are in the second part of the construction. Let $\tau_0 = 0$, and

$$\tau_{i+1} = \inf\{m > \tau_i : X_m = i+1, \inf\{k : X_{m+k} = i\} > \sqrt{n}\} \wedge 2n.$$

It is important to note that at time τ_{i+1} the process is at height $i+1$. Since we ar eonly concerned with what occurs when $(1-\alpha_m)c > 1+\eta$ ignore the second part of the definition ($\wedge 2n$) which is needed to take care of problems that can occur when the graph is almost. For this reason we will study only the behavior while $(1-\alpha_m)c > 1+\eta$ with η positive, or what is the same while $\alpha_m < 1-(1+\eta)/c$.

The next result explains our interest in the elapsed time being $> n^{1/2}$. This result is obvious if we notice that the graph with vertex set S_n is Erdős-Rényi$((1-\alpha_n)N, c/N)$. See Theorem 1.3.2 for a crude result or Theorem 1.6.5 for one with an optimal constant

Lemma 1.4.5. *(Lemma 3). Let* $\mathbf{G}$ *be the event that at every time with* $\alpha_n < 1-(1+\eta)/c$ *the graph* S_m *has no components between* $n^{1/10}$ *and* $n^{9/10}$ *and maximum degree vertex is at most* $\log n$. $\lim_{N\to\infty} P(\mathbf{G}) = 1$.

Denote by $\mathbf{F}_n$ the event that this is the first visit to a_n. Conditional on $\mathbf{F}_n$ the number of neighbors $C(a_m)$ and clusters of the neighbors $x_1 < \cdots < x_{C(a_m)}$ are independent of the excursions outside S_m. When we move from a_m to x_i then we have to explore the entire component of x_i before we return to a_m. If the component containing x_i is small ($< n^{1/10}$ this takes time $< 2n^{1/10}$) while if x_i is part of the giant component in S_m the exploration will take time $> 2n^{9/10} > n^{1/2}$ so at the times τ_i we are exploring the giant component of S_m and we will not return to x_i until we begin to go back down.

Grouping the times τ_i based on the height of $X_{\tau(i)}$. Let

$$h_k = \inf\{i : \alpha_{\tau_i} > k\epsilon\} \tag{1.4.11}$$

For the proof of their Lemma 5 it is important to know that

$$h_{k+1} - h_k \text{ is the hitting time of } 2\epsilon n \text{ by the sequence } \tau_{h_k+m} - \tau_{h_k} + m. \tag{1.4.12}$$

To check this note that α_n only changes by $\pm 1/N$ on each step so

$$\begin{aligned} 2\epsilon n &= 2[\alpha(\tau(h_{k+1})) - \alpha(\tau(h_k))] \\ &= X(\tau(h_{k+1})) + \tau(h_{k+1})) - X(\tau(h_k)) - \tau(h_k) \\ &= \tau(h_{k+1})) + h_{k+1} - (\tau(h_k) + h_k). \end{aligned}$$

In order to have $(1-\alpha_m)c > 1+\eta$ we only look at

$$k \le K = [1-(1+\eta)/c]/\epsilon.$$

The parameter α varies only slightly between two successive h_k so $\tau_{i+1}-\tau_i$, $h_k \le i \le h_{k+1}$ are almost i.i.d.

The part of the proof that takes the most work is

Lemma 1.4.6. *(Lemma 4). There is a constant C so that if n is large enough then for all $i \in [h_k, h_{k+1}]$ with $k \le K$*

$$\frac{2}{\rho_{(1-k\epsilon)c}} - 1 - C\epsilon \le E(\tau_{i+1}-\tau_i|\tau_i) \le \frac{2}{\rho_{(1-k\epsilon)c}} - 1 + C\epsilon.$$

At $\tau_i = m$ we have just visited a vertex a_m that is part of the giant component. We have to explore its neighbors one at a time until we find one that is part of the giant component and this occurs at time τ_{i+1}. Being lazy we refer the reader to their paper for the details, and we content ourselves to show how it implies the desired conclusion.

Lemma 1.4.7. *(Lemma 5). There is a constant C depending only on η so that for every $k \le K$ with high probability*

$$\epsilon\rho_{(1-k\epsilon)c} - C\epsilon^2 \le \liminf_{n\to\infty} \frac{h_{k+1}-h_k}{n} \le \limsup_{n\to\infty} \frac{h_{k+1}-h_k}{n} \ le\epsilon\rho_{(1-k\epsilon)c} + C\epsilon^2.$$

Proof. Fix $k \le K$. We are going to construct a martingale involving the sequence $\{\tau_i\}_{h_k \le i \le h_{k+1}}$. Recall than on $\mathbf{G}$ the sequence τ_i is a Markov chain and that h_k is a stopping time for it. Indeed as we saw earlier, $\tau_{i+1}-\tau_i$ has an explicit distribution depending onlu on $\tau_i + i$.

$$M_m^k = \begin{cases} \tau_{h_k+m} - \tau_{h_k} - \sum_{k=0}^{m-1} E(\tau_{h_k+i+1}-\tau_{h_k+i}|\tau_{h_k+i}) & m \le h_{k+1}-h_k, \\ \tau_{h_{k+1}} - \tau_{h_k} - \sum_{k=0}^{(h_{k+1}-h_k)-1} E(\tau_{h_k+i+1}-\tau_{h_k+i}|\tau_{h_k+i}) & m \ge h_{k+1}-h_k. \end{cases} \tag{1.4.13}$$

On $\mathbf{G}$ the difference $|\tau_{h_k+i+1}-\tau_{h_k+i}| \le n^{1/10}\log n$ By Lemma 1.4.6 for every $i \ge 0$

$$\frac{2}{\rho_{(1-k\epsilon)c}} - 1 - C\epsilon \le E(\tau_{h_k+i+1}-\tau_{h_k+i}|\tau_{h_k+i}) \le \frac{2}{\rho_{(1-k\epsilon)c}} - 1 + C\epsilon.$$

Since the increments of our martingale are bounded we can use Theorem A.2.2

Theorem 1.4.8 (Azuma-Hoeffding inequality.)**.** *If M_ℓ is a martingale and $|M_\ell - M_{\ell-1}| \le c_\ell$ then*

$$P(M_m - M_0 \ge y) \le \exp\left(-y \Big/ 2\sum_{\ell=1}^m c_\ell^2\right).$$

Taking $c_k = n^{1/10}\log n$ and using $m \le n$ gives

$$P(M_m^k > n^{3/4}) \le \exp\left(-\frac{(n^{3/4})^2}{2n(n^{1/10}\log n)^2}\right) \le C\exp(-n^{1/4}).$$

Therefore by a union bound

$$P\left(\sup_{m\le \epsilon n} M_m^k > n^{3/4}\right) \le \epsilon n \cdot C\exp(-n^{1/4})$$

and since the number of martingales $K \le C/\epsilon$

$$P\left(\sup_{k\le K}\sup_{m\le \epsilon n} M_m^k > n^{3/4}\right) \le Cn\exp(-n^{1/4}).$$

Recalling the definition of the martingales in (1.4.13), we see that with high probability

$$\left|\tau_{h_k+n} - \tau_{h_k} - \sum_{i=0}^{n-1} E[\tau_{h_k+i+1} - \tau_{h_k+i}|\tau_{h_k+i}]\right| < n^{3/4}$$

so using Lemma 1.4.6

$$n\left(\frac{2}{\rho_{(1-k\epsilon)c}} - C\epsilon\right) \le \tau_{h_k+n} - \tau_{h_k} + n \le n\left(\frac{2}{\rho_{(1-k\epsilon)c}} + C\epsilon\right).$$

Using (1.4.12) we have

$$(h_{k+1} - h_k)\left(\frac{2}{\rho_{(1-k\epsilon)c}} - C\epsilon\right) \le 2n\epsilon \le (h_{k+1} - h_k)\left(\frac{2}{\rho_{(1-k\epsilon)c}} + C\epsilon\right)$$

or rearranging

$$2\epsilon\left(\frac{2}{\rho_{(1-k\epsilon)c}} + C\epsilon\right)^{-1} \le \frac{(h_{k+1} - h_k)}{n} \le 2\epsilon\left(\frac{2}{\rho_{(1-k\epsilon)c}} - C\epsilon\right)^{-1}.$$

which gives the desired conclusion. □

Proof of Theorem 1.4.4. Introducing a superscript ϵ to allow us to keep track fo the dependence on this parameter, and applying Lemma 1.4.7

$$\sum_{i=0}^{k-1} \epsilon\rho_{(1-k\epsilon)c} - C\epsilon^2 \le \liminf_{n\to\infty} \frac{h_k^\epsilon}{n} \le \limsup_{n\to\infty} \frac{h_k^\epsilon}{n} \le \sum_{i=0}^{k-1} \epsilon\rho_{(1-k\epsilon)c} + C\epsilon^2.$$

Taking $k = \lceil u/\epsilon \rceil$

$$\int_0^u \rho_{(1-x)c}\,dx - C\epsilon^2 \le \liminf_{n\to\infty} \frac{h_k^\epsilon}{n} \le \limsup_{n\to\infty} \frac{h_k^\epsilon}{n} \le \int_0^u \rho_{(1-x)c}\,dx + C\epsilon^2.$$

The k points of the **normalized profile** can be written as

$$\left(\frac{\tau_{h_k^\epsilon}}{n}, \frac{X(\tau_{h_k^\epsilon})}{n}\right) = \left(\frac{\tau_{h_k^\epsilon}}{n}, \frac{h_k^\epsilon}{n}\right) = \left(2k\epsilon + O(n^{-4/5}) - \frac{h_k^\epsilon}{n}, \frac{h_k^\epsilon}{n}\right)$$

the last equality coming from the fact that on the event $\mathbf{G}$ each increment $\tau_{i+1} - \tau_i$ is bounded by $n^{1/10} \log n$. Combining the last two equations we see that as $n \to \infty$ these K points are uniformly within $C\epsilon$ of the following parametrized curve

$$\begin{aligned} t(u) &= 2u - \int_0^u \rho_{(1-x)\epsilon}\, dx, \\ x(u) &= \int_0^u \rho_{(1-x)\epsilon}\, dx. \end{aligned} \tag{1.4.14}$$

These are called x and y in the paper but here we would like the emphasize that the first coordinate is time and the the second is the height. Finally, (1.4.12) implies

$$\left(\frac{\tau_{h_{k+1}^\epsilon}}{n} - \frac{\tau_{h_{k+1}^\epsilon}}{n}\right) + \left(\frac{X(\tau_{h_{k+1}^\epsilon})}{n} - \frac{X(\tau_{h_k^\epsilon})}{n}\right) = 2\epsilon$$

and that the slope of the renormalized profile is smaller than 1 in absolute value, we can conclude that the whole renormailized profile stays within $C\epsilon$ of the curve defined in (1.4.14).

Denouement

At this point the interesting part of the proof is done and it only remains to relate the curve defined in (1.4.14) to the one in (1.4.7), which is an exercise in calculus. To do this, we parametrize the curve by $\rho_{(1-u)c}$ instead of by u and note that since $\rho_{(1-u)c}$ is the extinction probability of a branching process with a Poisson$((1-u)c)$ distribution then recalling the formula for the generating function

$$1 - \rho_{(1-u)c} = \exp(-(1-u)c\rho_{(1-u)c}).$$

Taking log's we have

$$\log(1 - \rho_{(1-u)c}) = -(1-u)c\rho_{(1-u)c} \tag{1.4.15}$$

which rearranges to

$$u = 1 + \frac{\log(1-\rho_{(1-u)c})}{c\rho_{(1-u)c}} = \frac{\log(1-\rho_{(1-u)c})}{c\rho_{(1-u)c}} - \frac{\log(1-\rho_c)}{c\rho_c}$$

where in the secdond step we have used (1.4.15) with $u = 0$ to conclude $\log(1-\rho_c) = -c\rho_c$. The last calculation shows that $2u$ is the last term in (1.4.5).

To obtain the formula in terms of $f(\rho)$ and $g(\rho)$, we need to prove

Lemma 1.4.9. $\int_{1-u}^1 \rho_{x\epsilon}\, dx = \frac{1}{c}\left[Li_2(\rho_{(1-u)\rho_c}) - Li_2(\rho_c) + \log(1-\rho_{(1-u)c}) - \log(1-\rho_c)\right]$.

Proof. To do this we note that replacing $(1-u)$ by x in (1.4.15) gives

$$\frac{c\rho'_{xc}}{1-\rho_{xc}} = c\rho_{xc} + c^2 x\rho_{xc}.$$

Rearranging the last equation and dividing by c gives

$$\begin{aligned}\int_{1-u}^{1} \rho_{x\epsilon}\, dx &= \int_{1-u}^{1} \frac{\rho'_{xc}}{1-\rho_{xc}} - \int_{1-u}^{1} cx\rho_{xc} \\ &= \frac{1}{c}\left(\ln(1-\rho_{(1-\mu)c}) - \ln(1-\rho_c)\right) - \int_{1-\mu}^{1} \frac{\log(1-\rho_{xc})}{\rho_{xc}}\rho'_{xc}\, dx\end{aligned}$$

where we have used (1.4.15) in the second term. Integrating using the definition of Li_2 given iin (1.4.4) shows that the last term is

$$-\frac{1}{c}\left(Li_2(\rho_c) - Li_2(\rho_{(1-\mu)c}\right).$$

We would like to thank Laurent Ménard for taking time to write out the proof and email it to us. □

1.5 CLT for the size of 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

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

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 $O(1)$ visits to $\emptyset$ before finding the giant component, so this difference can be ignored. Let $\mathcal{F}_t$ be the σ-field generated by the process up to time t. Let $u_t^n = |U_t|$.

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

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

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

Proof. Let $\Delta u_t^n = u_{t+1}^n - u_t^n$. If $A_t \neq \emptyset$ then

$$\begin{aligned} E(\Delta u_t^n | \mathcal{F}_t) &= -u_t^n \frac{\lambda}{n}, \\ \text{var}\,(\Delta u_t^n | \mathcal{F}_t) &= u_t^n \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right). \end{aligned}$$

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

$$E\left(\frac{\Delta u_{[ns]}^n}{n}\middle| \mathcal{F}_{[ns]}\right) = -\frac{u_{[ns]}^n}{n} \cdot \lambda \cdot \frac{1}{n}, \tag{1.5.1}$$

$$\text{var}\left(\frac{\Delta u_{[ns]}^n}{n}\middle| \mathcal{F}_{[ns]}\right) = \frac{u_{[ns]}^n}{n} \cdot \lambda\left(1 - \frac{\lambda}{n}\right) \cdot \frac{1}{n^2}. \tag{1.5.2}$$

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

$$\begin{aligned} \text{infinitesimal mean} &= -\frac{u_{[ns]}^n}{n}\lambda, \\ \text{infinitesimal variance} &= \frac{u_{[ns]}^n}{n}\lambda\left(1 - \frac{\lambda}{n}\right) \cdot \frac{1}{n}. \end{aligned}$$

Letting $n \to \infty$, the infinitesimal variance $\to 0$, so the result follows from the degenerate case of Theorem 4.1 in Section 7.4 of Ethier and Kurtz (1986) in which the infinitesimal varince is 0. □

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

$$M_t^n = \left(1 - \frac{\lambda}{n}\right)^{-t} u_t^n/n$$

is a martingale with

$$\begin{aligned} E(M_t^n - M_0^n)^2 &= \sum_{s=0}^{t-1} E(M_{s+1}^n - M_s^n)^2 \\ &\le \sum_{s=0}^{t-1} \left(1 - \frac{\lambda}{n}\right)^{-s+1} \lambda/n^2 \to 0 \end{aligned}$$

so by Kolmogorov's maximal inequality

$$E\left(\max_{0\le s\le n} (M_s^n - M_0^n)^2\right) \to 0.$$

Since $M_0^n = 1$, this says that when n is large $M_s^n \approx 1$ uniformly in s, so $u_{[ns]}^n/n \approx (1 - \lambda/n)^{[ns]} \to e^{-\lambda s}$. □

To determine the size of the giant component, we note that when $u_t^n + r_t^n = n$, $A_t = \emptyset$. This may occur several times for small t while we are seaching 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 $r_t^n = |R_t| \equiv t$ as we did u_t^n, $r_{[ns]}^n/n \to s$. (Here and for the rest of this section we will use t for the original integer time scale and $s \in [0, 1]$ for rescaled time.]) After scaling

$$u_t^n + r_t^n = n \quad \Rightarrow \quad e^{-\lambda s} + s = 1.$$

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 in Figure 1.7 shows $e^{-\lambda s} + s > 1$ for $s > 1 - \rho$, so we are interested only in $u_{[ns]}^n/n$ for $0 \le s \le 1 - \rho + \epsilon$. In this part of the process we first generate a geometrically distributed number of small clusters and then find the giant component.

Consider now $y_{[ns]}^n = (u_{[ns]}^n - n\exp(-\lambda s))/\sqrt{n}$ for $0 \le s \le 1 - \rho$.

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

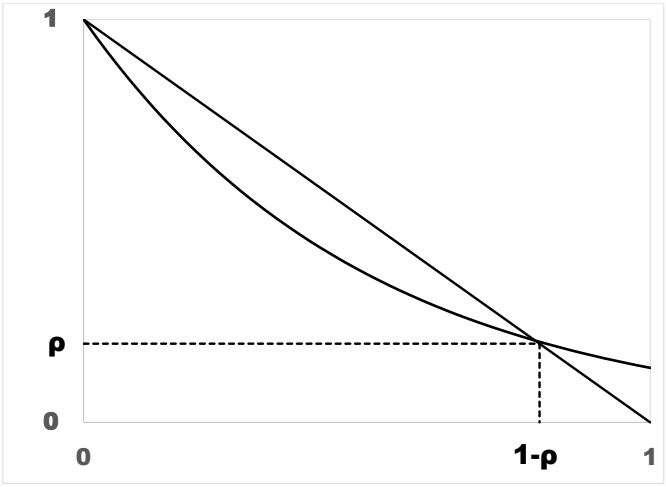

Figure 1.6: Example of Lemma 1.5.1 with $\lambda = 2$.

Proof. If $A_{[ns]} \neq \emptyset$ then using the formulas in (1.5.2)

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

Using Theorem 4.1 in Section 7.4 of Ethier and Kurtz (1986), 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

$$y_s = \int_0^s e^{-\lambda(s-r)}\sqrt{\lambda e^{-\lambda r}}\, dB_r. \tag{1.5.3}$$

To check this, notel 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 the stochastic integral (1.5.3) 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. To prove the fact about the stochastic integral note that any approximating sum has a normal distribution and then pass to the limit. □

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

$$M_t^n - M_0^n = \left(1 - \frac{\lambda}{n}\right)^{-t} u_t^n/n - 1.$$

The key observation is that

$$\sum_{r=0}^{[ns]-1} E((M_{r+1}^n - M_r^n)^2 | \mathcal{F}_r) \to \int_0^s \lambda e^{\lambda u}\, du = e^{\lambda s} - 1,$$

i.e., the quadratic variation has a deterministic limit. 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 $u_{[ns]}^n$. To determine the fluctutations of the point where $u_t^n + t = n$, we can now prove the result as we do the central limit theorem of renewal theory. To briefly recall that appraoch, let $\xi_1, \xi_2, \ldots$ be i.i.d. positive random variables with $E\xi_i = \mu$ and $\text{var}\,(\xi_i) = \sigma^2 \in (0, \infty)$. 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 χ 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 $\mathcal{Z}$ denote a normal with variance $e^{-\lambda(1-\rho)} - e^{-2\lambda(1-\rho)}$ we have

$$u^n_{[n(1-\rho)]} \approx n \exp(-\lambda(1-\rho)) + \sqrt{n}\mathcal{Z}. \tag{1.5.4}$$

$A_{[ns]} = \emptyset$ when $u^n_{[ns]} = n - [ns]$. To find out when this occurs, we suppose equality holds at $s_0 = (1-\rho) + \mathcal{Y}/\sqrt{n}$. Using (1.5.4) and noting $s_0 \to (1-\rho)$ as $n \to \infty$

$$n \exp(-\lambda\{(1-\rho) + \mathcal{Y}/\sqrt{n}\}) + \sqrt{n}\mathcal{Z} = u^n_{[ns_0]} = n - [ns_0],$$

or rearranging

$$\exp(-\lambda s_0) - 1 + s_0 = -\sqrt{\mathcal{Z}}/\sqrt{n}.$$

Let $h(t) = e^{-\lambda t} - 1 + t$ which is $= 0$ at $t = 1 - \rho$. $h'(t) = -\lambda e^{-\lambda t} + 1$, so we can write the above as

$$h'(1-\rho)\mathcal{Y}/\sqrt{n} \approx -\sqrt{\mathcal{Z}}/\sqrt{n},$$

or $\mathcal{Y} \approx \mathcal{Z}/h'(1-\rho)$. $h'(1-\rho) = 1 - \lambda\rho$. Putting the pieces together.

Theorem 1.5.3. *Suppose $\lambda > 1$. The size of the largest component $\mathcal{C}^{(1)}$ satisfies*

$$\frac{|\mathcal{C}^{(1)}| - n(1-\rho)}{\sqrt{n}} \Rightarrow \chi$$

where $\Rightarrow$ means convergence in distribution and χ has a normal distribution with mean 0 and variance $(\rho - \rho^2)/(1-\lambda\rho)^2$.

For other approaches to this result see Pittel (1990) and Barraez, Boucherno, and Fernandez de la Vega (2000). To compare variances note that Pittel's $c = \lambda$ and $T = \rho/\lambda$.

1.6 Combinatorial approach

Combinatorial methods give more refined results about the Erdös-Rényi model. However, we first need some graph theoretic preliminaries about the nature of components. It is easy to see by using induction 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 ℓ. Section 5.4 of Bollobás (2001) gives combinatorial results for unicyclic and complex components. In Section 1.7 we will use Adlous' (1997) result give some results about much less explicit results about complex components in the critical regime. See Theorem 1.7.5. Repeating the proof of Theorem 1.2.3

Lemma 1.6.1. *Let $A < \infty$ be a constant, and consider only "small" components with $\leq A \log n$ vertices. The probability that a small component is unicyclic is $\leq (A \log n)^2/n$. The probability a small component has complexity $\ell \geq 1$ is $\leq (A \log n)^3/n^2$.*

Proof. We run the exploration process defined in (**??**) for $\leq A \log n$ steps and have a probability of collision of $\leq (A \log n)/n$ on each step. The probability of at least one collision is $\leq (A \log n)^2/n$, while the probability of at least two collisions is bounded above by the expected number of pairs of steps on which collisions occur which is

$$\leq \binom{A \log n}{2} (A \log n/n)^2$$

proving the desired result. □

Theorem 1.6.2. *Complex components are rare unless λ is close to 1.*

Proof. Theorem 1.3.1 implies that if $\lambda < 1$ there is a C_λ 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 β_λ so that with high probability the second largest component is $\leq \beta_\lambda \log n$. Lemma 1.6.1 implies that when the largest com[ponent is $\leq A \log n$ then with high probaiblity there are no complex components. The expected number of unicyclic components is $\leq (A \log n)^2$.In this section we are interested in the faction of clusters that have size k so we can also ignore unicyclic components. □

Theorem 1.6.2 implies that for fixed $\lambda \neq 1$, we can restrict our attention to tree components. In 1889 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

$$\binom{n}{k} k^{k-2} \left(\frac{\lambda}{n}\right)^{k-1} \left(1 - \frac{\lambda}{n}\right)^{k(n-k)+\binom{k}{2}-(k-1)} \tag{1.6.1}$$

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 nq_k where

$$q_k = \frac{k^{k-2}}{k!}\lambda^{k-1}e^{-\lambda k}. \tag{1.6.2}$$

Recalling that in the subcritical regime cluster sizes have the same distribution as the total progeny in a Poisson(λ) 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 1.6.3. *The probability distribution of the total progeny τ of a* Poisson(λ) *branching process with $\lambda < 1$ is given by*

$$P(\tau = k) = kq_k = \frac{1}{\lambda}\frac{k^{k-1}}{k!}(\lambda e^{-\lambda})^k \tag{1.6.3}$$

There is an extra factor of k due to the fact that a tree of size k is $\mathcal{C}_x$ 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 λ 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.

1.6.1 Super/subcritical duality

Suppose $\lambda > 1$ and let ρ be the extinction probability for the branching process with a Poisson(λ) offspring distribution. The title of this subsection refers to the fact that there is a close relationship between Erdös-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 ρ gives

$$\lambda\rho e^{-\lambda\rho} = \lambda e^{\lambda(\rho-1)}e^{-\lambda\rho} = \lambda e^{-\lambda}. \tag{1.6.4}$$

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

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

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

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

In words, the total progeny of a Poisson(λ) branching process conditioned on extinction is the same as that of a Poisson($\lambda\rho$) branching process, which is Theorem 1.1.9.

(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 $1 + o(1)$ times the expected value.

Theorem 1.6.4. *Let T_k^n be the number of tree components of size k in the Erdos-Renyi graph with n vertices. As $n \to \infty$, $T_k^n/n \to q_k$ 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

$$\binom{n}{k} k^{k-2} \left(\frac{\lambda}{n}\right)^{k-1} \left(1-\frac{\lambda}{n}\right)^{k(n-k)+\binom{k}{2}-k+1} \tag{1.6.5}$$

$$\cdot \binom{n-k}{k} k^{k-2} \left(\frac{\lambda}{n}\right)^{k-1} \left(1-\frac{\lambda}{n}\right)^{k(n-2k)+\binom{k}{2}-k+1}. \tag{1.6.6}$$

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 $\binom{n-k}{k} \le \binom{n}{k}$ the above is

$$\le (ET_k)^2 \left(1-\frac{\lambda}{n}\right)^{-k^2} \le (ET_k)^2 e^{\lambda k^2/n}.$$

From this we get

$$\text{var}\,(T_k) = E(T_k(T_k-1)) + ET_k - (ET_k)^2 \le ET_k + (ET_k)^2(e^{\lambda k^2/n}-1).$$

Using Chebyshev's inequality

$$P(|T_k - ET_k| \ge n^{2/3}) \le \frac{ET_k + (ET_k)^2(e^{\lambda k^2/n}-1)}{n^{4/3}} \to 0,$$

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

$$P\left(|\mathcal{C}^{(2)}| \ge a \log n\right) \to 0.$$

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

$$k! \sim k^{k+1/2} e^{-k} \sqrt{2\pi} \quad \text{as } k \to \infty, \tag{1.6.7}$$

so we have (Lemma 1.6.6 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 $\lambda e^{1-\lambda} < 1$

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

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. □

1.6.2 Cluster sizes near criticality

In the next section we will use weak convergence to look at component sizes near $\lambda = 1$. Here we will use combinatorics to derive some results. We begin with a calculation that is simple and gives the right answer, but is not completely rigorous. (1.6.2) tell us that the expected number of trees of size k is, for large k,

$$n \frac{k^{k-2}}{k!} \lambda^{k-1} e^{-\lambda k} \sim \frac{n}{\lambda\sqrt{2\pi}} k^{-5/2} (\lambda e^{1-\lambda})^k. \tag{1.6.8}$$

By **Stirling's formula**

$$k! \sim k^k e^{-k} \sqrt{2\pi k} \tag{1.6.9}$$

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

$$\sum_{k=K}^{\infty} \frac{n}{\sqrt{2\pi}} k^{-5/2} \sim \frac{2}{3\sqrt{2\pi}} n K^{-3/2}$$

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 λ is a general parameter value, which you should think of as close to 1.

Lemma 1.6.6. *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*

$$\gamma_{n,k}(\lambda) \sim n \cdot \frac{k^{-5/2}}{\lambda\sqrt{2\pi}} \exp\left(-\alpha(\lambda) k + (\lambda - 1)\frac{k^2}{2n} - \frac{k^3}{6n^2} \right). \tag{1.6.10}$$

Proof. (1.6.1) tells us that

$$\gamma_{n,k}(\lambda) = \binom{n}{k} k^{k-2} \left(\frac{\lambda}{n}\right)^{k-1} \left(1 - \frac{\lambda}{n}\right)^{k(n-k)+\binom{k}{2}-k}.$$

Using Stirling's formula (1.6.9) and $k = o(n)$ in the last exponent gives

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

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 1.6.6 we have

$$\gamma_{n,k}(\lambda) \sim \frac{nk^{-5/2}}{\sqrt{2\pi}} e^{-k^3/6n^2}. \tag{1.6.11}$$

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 Critical regime

In contrast to the previous section, probabilistic methods, specifically martingales and weak convergence will take center stage. We begin with a result for the critical case $\lambda = 1$ that is simple to prove but then move on to result which describe the emergence of the giant component in the critical regime $\lambda = 1 + \theta n^{-1/3}$ where $-\infty < \theta < \infty$

1.7.1 An inequality of Nachmias and Peres (2010)

We will prove what they call the easy upper bound.

Theorem 1.7.1. *Suppose $\lambda = 1$. Let $\mathcal{C}_1$ be the largest component. If $x > 1$*

$$P(|\mathcal{C}_1| > xn^{2/3}) \le \frac{6}{x^{3/2}}.$$

By working harder and using exponential martingales one can show that

Theorem 1.7.2. *For any $n > 1000$ and $x > 8$*

$$P(|\mathcal{C}_1| > xn^{2/3}) \le \frac{4}{x}\exp(-x^2(x-4)/32).$$

As the authors say in their paper, upper bounds were proved by Bolobás (1984) and Luczak (1990) but their methods are remarkable for their simplicity, especially in comparison to the machinery that will be used in Section 1.7.2.

Proof of Theorem 1.7.1. To begin, we recall the recursion defined in (1.3.1) in Section 1.3 for the set of active sites when we add the neighbors of one site at each step:

$$|A_{t+1}| = |A_t| - 1 + \xi_{t+1}, \tag{1.7.1}$$

where ξ_{t+1} is the number of vertices added on step $t+1$, and we start with $A_0 = 1$. Since the number of vertices to which a connection can be made decreases with time, $\xi_i \le_d X_i =$ binomial$(n, 1/n)$, where $\le_d$ us a stochastic ordering between the distributions. Before we forget we would like to record the facts

$$EX = n \cdot (1/n) = 1 \quad \text{and} \quad \text{var}\,(X) = n(1/n)(1-1/n) = 1 - 1/n. \tag{1.7.2}$$

The size of the cluster is $\tau = \min\{i \ge 1 : A_i = 0\}$. If we let $S_t = 1 + \sum_{k=1}^{t}(X_k - 1)$ then $S_t \ge_d A_t$ and

$$\sigma = \min\{i \ge 1 : S_i = 0\}$$

has $\sigma \ge_d \tau$. If we arrange things so that $\xi_i \le X_i$ almost surely for $i \le \tau$ then we have $\tau \le \sigma$ almost surely.

We are interested in how large the cluster can get so we let

$$\gamma = \min\{i \ge 1 : S_i = 0 \text{ or } S_i \ge H\}.$$

Since $E(X_i - 1) = 0$, $S_{i\wedge\sigma}$ is a nonnegative martingale. Using the optional stopping theorem for nonnegative supermartingles (Theorem 4.8.4 in PTE5) we conclude that

$$1 = ES_0 \geq ES_\gamma \geq HP(S_\gamma \geq H),$$

so the probability S_i reaches $\geq H$ befor hitting 0 is

$$P(S_\gamma \geq H) \leq 1/H. \tag{1.7.3}$$

Lemma 1.7.3. *Let $p \in (0,1)$, let X_i be i.i.d. random variables with binomial(n,p) distributions, and let $S_t = 1 + \sum_{i=1}^t X_i$. Fix an integer H and define*

$$\gamma = \min\{t \geq 1 : S_t = 0 \text{ or } S_t \geq H\}.$$

Given $S_\gamma \geq H$ the overshoot $S_\gamma - H$ is dominated by the binomial(n,p) distributions.

Proof. Suppose we condition instead on $\{S_\gamma \geq H, \gamma = k\}$. Let $r = H - S_{k-1} + 1$. Let $I_1, \ldots I_n$ be i.i.d. Bernoulli(p) so that $I_1 + \cdots + I_n = X_k$. If $S_k \geq H$ let J be the smallest value of m so that $I_1 + \cdots + I_m = r$. The overshoot $S_\gamma - H = I_{J+1} + \cdot + I_n$ which is clearly smaller in distribution than binomial(n,p). Since the result holds if we condition on the time $\gamma = k$ then it also holds if we only condition on $S_\gamma \geq H$. □

Using Lemma 1.7.3 with $f(x) = 2Hx + x^2$ we see that if $X = \text{binomial}(n, 1/n)$ then

$$\begin{aligned} E(2H(S_\gamma - H) + (S_\gamma - H)^2 | S_\gamma \geq H) &\leq E(2HX + X^2) \\ &\leq 2H + 1 + (1 - 1/n) \leq 2H + 2, \end{aligned}$$

where we have used $EX = 1$ and $EX^2 = (EX)^2 + \text{var}\,(X)$ and (1.7.2), Writing $S_\gamma^2 = H^2 + 2H(S_\gamma - H) + (S_\gamma - H)^2$ we have for $H \geq 2$

$$E(S_\gamma^2 | S_\gamma \geq H) \leq H^2 + 2H + 2 \leq H^2 + 3H. \tag{1.7.4}$$

We now introduce a seccond martingale. Consulting Section 4.8.1 in PTE5 we see that

$$S_t^2 - (1 - 1/n)t$$

is the **quadratic martingale** associated with the random walk S_t. Since Lemma 1.7.3 controls how far the martingale jumps over H we can use the optional stopping theorem to conclude that

$$E\left(S_\gamma^2 - (1 - 1/n)\gamma\right) = 1.$$

To prove this, we note that since $\gamma \wedge m$ is a bounded stopping time so

$$E\ S_{\gamma\wedge m}^2 = 1 + (1 - 1/n)E(\gamma \wedge m).$$

Using the dominated convergence theorem on the left and the monotone convergence theorem on the right we get

$$1 + (1 - 1/n)E(\gamma) = ES_\gamma^2 = P(S_\gamma \geq H)E((S_\gamma)^2 | S_\gamma \geq H) \leq H + 3$$

by (1.7.3) and (1.7.4). Rearranging gives

$$E\gamma \le \frac{H+2}{1-1/n} \le H+3 \qquad \text{if } H \le n-3.$$

We conclude that for $2 \le H \le n-3$

$$P(\gamma > H^2) \le \frac{H+3}{H^2} \le \frac{2}{H}.$$

Define $\bar{\gamma} = \gamma \wedge H^2$. If $\gamma < H^2$ and $S_\gamma < H$ then $S_{\bar{\gamma}} = S_\gamma = 0$ so we have

$$P(S_{\bar{\gamma}} > 0) \le P(S_\gamma \ge H) + P(\gamma \ge H^2) \le 3/H. \tag{1.7.5}$$

Let $K = H^2$ and note that if $\mathcal{C}(v) > H^2$ we must have $S_{\bar{\gamma}} > 0$, because if not then the cluster size would be $\le \bar{\gamma} \le H^2$. From this we deduce that if $9 \le K \le (n-3)^2$ then

$$P(|\mathcal{C}(v)| > K) \le \frac{3}{\sqrt{K}}.$$

Let N_K be the total number of vertices contained in components of size larger than K. Then three trivial inequalities give

$$P(|\mathcal{C}_1| > K) \le P(N_K > K) \le \frac{EN_K}{K} \le \frac{nP(|\mathcal{C}(v)| > K)}{K}.$$

Putting $K = (\lfloor \sqrt{xn^{2/3}} \rfloor)^2$ where $x > 1$ yields

$$P(\mathcal{C}_1 > xn^{2/3}) \le P(\mathcal{C}_1 > K) \le \frac{3n}{(\lfloor \sqrt{xn^{2/3}} \rfloor)^3} \le \frac{6}{x^{3/2}}$$

which completes the proof of Theorem 1.7.1.

1.7.2 Aldous' theorem for critical regime cluster sizes

In this section we continue to use the recursion in (1.7.1) but we combine it with some less elementary probability. In our first use of this recursion 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 in Section 1.5, we will speed up time and rescale the sizes 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_τ 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

$$\begin{aligned} E(\Delta a_t|\mathcal{F}_t) &= -1 + (n - t - a_t)(1 + \theta n^{-1/3})/n \\ &= -\frac{t + a_t}{n} + (\theta n^{-1/3})(1 - (t + a_t)/n), \\ \text{var}\,(\Delta a_t|\mathcal{F}_t) &= (n - t - a_t)\frac{1 + \theta n^{-1/3}}{n}\left(1 - \frac{1 + \theta n^{-1/3}}{n}\right). \end{aligned}$$

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

$$\begin{aligned} E\left(\frac{\Delta a_{[sn^{2/3}]}}{n^{1/3}}\middle|\mathcal{F}_{[sn^{2/3}]}\right) &= \frac{-[sn^{2/3}] - a_{[sn^{2/3}]}}{n \cdot n^{1/3}} + \frac{\theta n^{-1/3}(1 + \theta n^{-1/3})}{n^{1/3}} \\ &= -sn^{-2/3} + O(n^{-1}) + \theta n^{-2/3} + O(n^{-1}). \end{aligned}$$

The variance is much easier

$$\text{var}\left(\frac{\Delta a_{[sn^{2/3}]}}{n^{1/3}}\middle|\mathcal{F}_{[sn^{2/3}]}\right) \sim \frac{1}{n^{2/3}}.$$

Letting $n \to \infty$ we see that $a_{[sn^{2/3}]}/n^{1/3}$ converges in distribution to the solution of

$$dB_s^\theta = (-s + \theta)\, ds + dB_s,$$

which is simply $B_s^\theta = B_s + \theta s - s^2/2$ (run until the first time it hits zero).

The calculation above leads to a remarkable result of Aldous (1997), stated here as two theorems, 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:

$$W_t^\theta = B_t^\theta - \min_{0 \le s \le t} B_s^\theta$$

We say that (u, v) is an excursion interval of W_t^θ if $W_u^\theta = W_v^\theta = 0$ but $W_t^\theta \neq 0$ 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 1.7.4. *Let $K_1^n \ge K_2^n \ge \ldots$ be the ordered component sizes of Erdős-Rényi$(n, (1 + \theta n^{-1/3})/n)$. Then as $n \to \infty$, $\{n^{-2/3}K_j^n : j \ge 1\}$ converges in distribution to $\{L_j : j \ge 1\}$ where $L_1 > L_2 > L_3 > \ldots$ are the ordered lengths of excursion intervals in $\{W_s^\theta : s \ge 0\}$.*

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 at least $k - 1$ edges. Changing from the notation used in the previous section, we define the number of edges in excess of this number to be

the **surplus** of the cluster, since that is the number of "collisions" needed to create such a component. Let $N^\theta(t)$ be a counting process with intensity W_t^θ. That is N^θ has jumps of size 1 and

$$N^\theta - \int_0^t W_u^\theta \, du \quad \text{is a martingale.}$$

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 1.7.5. *Let $\sigma_1^n \ge \sigma_2^n \ge \dots$ be the surpluses of the ordered component sizes defined in Theorem 1.7.4. Then as $n \to \infty$,*

$$\{(n^{-2/3}K_j^n, \sigma_j^n) : j \ge 1\}$$

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

Much is known about the number of complex components. See Janson (1993), Janson, Knuth, Luczak, and Pittel (1993), Luczak, Pittel, and Weirman (1994), and the (2001) second edition of Bollobás' book on *Random Graphs.*

1.7.3 Multiplicative coalescent.

While Theorem 1.7.5 is a nice limit theorem, the truly remarkable part of multiplicative contribution was to note that if we divide the sizes of clusters in Erdős-Rényi$(n, (1+tn^{-1/3})/n)$ 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 $\ell^2_\searrow$ the collection of decreasing sequences $x_1 \ge x_2 \ge x_3 \ge \dots$ 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 1.7.4. Aldous and Limic (1998) 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 $(x_1 = v > 0,\ x_i = 0,\ i \ge 2)$ there are some nonstandard ones, which are irrelevant for the following application.

Theorem 1.7.6. *Let $K_1^n(t) \ge K_2^n(t) \ge \dots$ be the ordered component sizes of Erdős-Rényi$(n, (1 + tn^{-1/3})/n)$. As $n \to \infty$, $\{K_j^n(t)/n^{2/3} : j \ge 1\}$, $-\infty < t < \infty$ 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 μ 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 $\theta(\mu) = 1 - \sum_{s=1}^{\infty} p(s,\mu)$ has

$$\theta \approx (\mu - \mu_c)^\beta \qquad \text{as } \mu \downarrow \mu_c. \tag{1.8.1}$$

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

$$\frac{\log \theta}{\log(\mu - \mu_c)} \to \beta \qquad \text{as } \mu \downarrow \mu_c.$$

  In order to derive relations between exponents, we will suppose $\theta \sim C(\mu - \mu_c)^\beta$ where $a(t) \sim b(t)$ means $a(t)/b(t) \to 1$. This stronger result holds for Erdős-Rényi graphs, but even the weaker definition has not been shown for ordinary percolation in $d > 2$.

- Let $\mathcal{C}_x$ be the cluster containing x. To have the mean cluster size finite for $\mu \neq \mu_c$, we exclude the giant component

$$\chi(\mu) = E(|\mathcal{C}_x|; |\mathcal{C}_x| < \infty) \approx |\mu - \mu_c|^{-\gamma}. \tag{1.8.2}$$

  It is not clear (to me at least) why the exponent should be the same when $\mu \uparrow \mu_c$ and $\mu \downarrow \mu_c$, but it is true and the constants are the same as well, see (1.8.8).

- We let $\chi_k = E(|\mathcal{C}_\rho|^k; |\mathcal{C}_x| < \infty)$ for integers $k \geq 2$,

$$\chi_k(\mu) \approx |\mu - \mu_c|^{-\Gamma(k)} \tag{1.8.3}$$

  and let $\Delta_k = \Gamma(k) - \Gamma(k-1)$. As in the case of ordinary percolation it is expected that $\Delta_k = \Delta$ is independent of $k \geq 2$. These numbers will turn out to be different from $\Delta_1 = \gamma$.

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

$$p(s, \mu_c) \approx s^{1-\tau} \tag{1.8.4}$$

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

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

  It would be sensible to use $p(s, \mu_c) \approx s^{-1-1/\delta_r}$ for the Erdős-Rényi case but to avoid having to rewrite Section 1.8.4, we will stick with definition given above and trust that the reader will not be confused by having two meanings for τ.

- Finally there is another exponent for the behavior near criticality

$$p(s,\mu) \approx s^{1-\tau} f(s\delta^{1/\sigma}) \quad \text{where} \quad \delta = \mu - \mu_c. \tag{1.8.5}$$

although it is probably more accurate to call this a **scaling relationship**, see Section 1.8.4.

***d*-dimensional percolation.** The most significant difference between the critical exponents for ordinary percolation and those for the Erdős-Rényi random graphs is that the Erdős-Rényi gragh is not a spatial model, so we do not have a **correlation length** $\xi(p)$ that gives the length (and height and width) of a typical finite cluster. This quantity is often defined in terms of the exponential decay of probability 0 and x are in the same finite cluster

$$\tau^f(0,x) = P_p(0 \leftrightarrow x, |\mathcal{C}_0| < \infty).$$

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

$$\xi(p) = \left(\frac{1}{\chi_1(p)} \sum_y |y|^2 P(0 \to y; |\mathcal{C}_0| < \infty) \right)^{1/2}.$$

and the critical exponent ν by $\xi(p) \approx |p - p_c|^{-\nu}$.

Finally on a d-dimensional graph we have another exponent η called the **anomalous dimension**

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

Note that when $\eta = 0$ this has the behavior of the Green's function of Brownian motion. In this case summing give

$$\sum_{x:|x|\le R} P_{cr}(0 \to x) \approx \int_1^R r^{2-d} \cdot r^{d-1}\, dr \sim CR^2.$$

Mean-field theory makes the connection between the Erdős-Rényi critical exponents and those for percolation in high dimensions. In the case of Ising model this term 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 ordinary percolation then the critical exponents take on their mean-field values:

$$\beta = 1, \qquad \gamma = 1, \qquad \Delta = 2.$$

To compute the others we need a spatial object which is usually taken to be percolation on trees. This leads to

$$\nu = 1/2 \qquad \eta = 0.$$

Thanks to work of Aizenmann and Newman (1984), Barsky and Aizenman (1991), 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, $\beta = 1$

Let G be Erdős-Rényi$(n, \mu/n)$ with $\mu > 1$. Let x be a randomly chosen $x \in G$ and let Z_m be the number of vertices at distance m from x. When the growing cluster containing x has size $\leq n^{1/2-\epsilon}$ the cluster is whp a tree (see Theorem 1.2.3), so Z_m is a branching process in which each individual in generation m has a Poisson(μ) number of offspring. In Section 1.3 we showed that if $\mu > 1$ then there is a constant $C(\mu)$ so that a vertex is in the giant component if and only if it is in a cluster of size $\geq C(\mu) \log n$, 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 $\phi(z) = \sum_{k=0}^{\infty} r_k z^z$ be the generating function. If r_K is Poisson(μ) then

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

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

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

$\phi(1) = 1$ is a trivial solution. In Theorem 1.1.4 we showed ρ is the unique solution of $\phi(\rho) = \rho$ in $[0, 1)$. If μ is close to 1 then ρ will be close to 1. To find an approximation to ρ in this case we note that

$$\phi(1) = 1, \qquad \phi'(1) = \sum_{k=0}^{\infty} k r_k = \mu, \qquad \phi''(1) = \sum_{k=0}^{\infty} k(k-1) r_k = \mu_2.$$

If x is close to 1 expanding ϕ in power series around 1 and ignoring the error gives

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

so for a fixed point at $(1-x)$ we want

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

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

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

so the critical exponent $\beta = 1$. Note that this holds whenever the offspring distribution has finite variance. If the distribution is a power law with infinite variance then the power changes. See Section 2.7 for results for the configuration model with a power-law degree distribution..

1.8.2 Mean cluster size, $\gamma = 1$

Let $\mathcal{C}_x$ be the cluster containing x. Taking the branching process viewpoint, if $\mu < 1$

$$E|\mathcal{C}_x| = \sum_{n=0}^{\infty} EZ_m = \sum_{m=0}^{\infty} \mu^m = \frac{1}{1-\mu}. \tag{1.8.8}$$

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

$$E(|\mathcal{C}_x|; |\mathcal{C}_x| < \infty).$$

The cluster size is the same as the total progeny in a supercritical branching process conditioned to die out. In Section 1.1,we proved the following result

A supercritical branching process conditioned to become extinct is a subcritical branching process. If the original offspring distribution is Poisson(μ) with $\mu > 1$ then the conditioned one is Poisson($\mu\rho$) where ρ is the extinction probability.

Using the result for the subcritical case in (1.8.8)

$$E(|\mathcal{C}_x| \,|\, |\mathcal{C}_x| < \infty) = \frac{1}{1-\mu\rho}.$$

Since $\mu_2 = \mu^2$ for Poisson, (1.8.7) shows that if μ 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

$$E(|\mathcal{C}_x| \,|\, |\mathcal{C}_x| < \infty) = \frac{1}{1-\mu\rho} \sim \frac{1}{(1-\mu)} \tag{1.8.9}$$

thus the asymptotic behavior of the mean cluster size as $\mu \downarrow 1$ is the same as as $\mu \uparrow 1$, not just the exponent $\gamma = 1$ but also the constant $C = 1$.

1.8.3 Higher moments, $\Delta = 2$

To compute higher moments of the custer size, it is convenient to convert the branching process into a random walk in which only one particle is allowed to give birth at each time. If S_m is the number of living particles after m births then as long as $S_m > 0$ and $m = o(n)$

$$S_{m+1} \approx S_m - 1 + \text{Poisson}(\mu)$$

When the number of living particles hits 0 it stays 0. To have the process defined for all time let $\xi_1, \xi_2, \ldots$ be i.i.d. $-1 + \text{Poisson}(\mu)$ and $S_{m+1} = S_m + \xi_{m+1}$.

We will use martingales to compute moments of the cluster size. To do this rigorously we need bounds on the tail of the distribution. To obtain the necessary bounds, we begin by computing the moment generating function of ξ_i

$$\psi(\theta) = e^{-\theta} \sum_{m=0}^{\infty} e^{-\mu} \frac{\mu^m}{m!} e^{\theta m} = e^{-\theta} \exp(\mu(e^\theta - 1)) \tag{1.8.10}$$

$\bar{M}_t = \exp(\theta S_t)/\psi(\theta)^t$ is a nonnegative martingale, so using the optional stopping theorem for the nonnegative supermartingales, Theorem 4.8.4 in PTE5,

$$\bar{M}_0 = e^\theta \geq E(\psi(\theta)^{-\tau}). \tag{1.8.11}$$

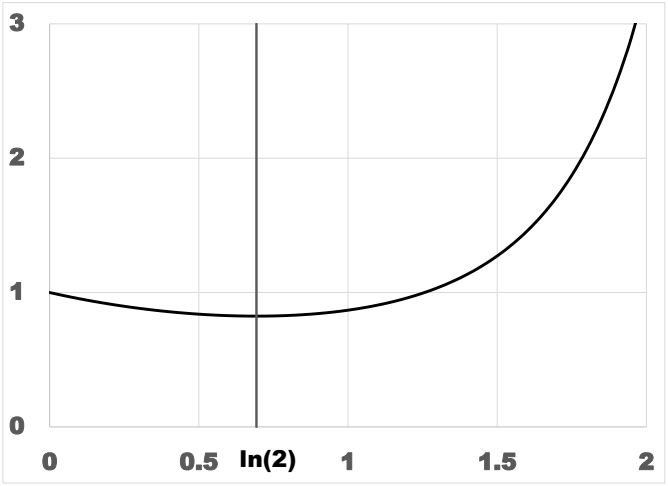

Figure 1.7: Graph of $\psi(\theta)$ when $\mu = 1/2$ showing minimum at $\theta = \ln(2)$.

The mean $\psi'(0) = E\xi_i = 1 - \mu$ so if $\mu < 1$ then $\psi(\theta) < 1$ when $\theta > 0$ is small. To prepare to optimize the bound we will derive later, we note that the derivative of $\log \psi(\theta)$

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

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

$$P\left(\max_{0\leq t\leq\tau} S_t \geq m\right) \leq e^{-m\alpha}/\mu. \tag{1.8.12}$$

Computing moments of the cluster size. Recall that the random walk of interest is

$$S_t = S_0 + \xi_1 + \cdots \xi_t,$$

where ξ_i have distribution $-1 + \text{Poisson}(\mu)$ so we will let $\nu = E\xi_i = \mu - 1$. The simplest martingale is the **linear martingale**

$$M_n^{(1)} = S_n - n\nu.$$

Using the domination in (1.8.12) we can conclude that

$$1 = S_0 = E(S_\tau - (\mu - 1)\tau) = (1 - \mu)E\tau,$$

so the expected cluster size is $E\tau = 1/(1 - \mu)$, as already computed in (1.8.8)

If T_m is a random walk in which steps have mean 0 and variance σ^2 then $T_m - m\sigma^2$ is a martingale, called the **quadratic martingale**. Applying this result to $T_m = S_m - m\nu$ and recalling that $-1 + \text{Poisson}(\mu)$ has variance $\nu_2 = \mu$ we see that

$$M_m^{(2)} = (S_m - m\nu)^2 - \mu m$$

is a martingale. Using the domination in (1.8.12) we can stop at time τ and conclude

$$1 = E(S_\tau - \tau\nu)^2 - \mu E\tau = (\mu - 1)^2 E\tau^2 - \frac{\mu}{1 - \mu}.$$

Rearranging we have

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

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

To compute all the moments we require an infinite sequence of martingales, and hence need to take a more systematic approach to finding them. The **cumulant generating function** is

$$\kappa(\theta) = \log E\exp(\theta\xi_i) = -\theta + \mu(e^\theta - 1). \tag{1.8.14}$$

$\bar{M}_m = \exp(uS_m)/\phi(u)^m = \exp(uS_m - m\kappa(u))$ is a martingale so differentiating we see that if

$$g_k(u; m, x) = \frac{\partial^k}{\partial u^k} \exp(ux - m\kappa(u))$$

and $f_k(m, x) = g_k(0; m, x)$ then $f_k(m, S_m)$ is a martingale. For more details see page 80 of Neveu (1975). To compute the f_k note that

$$g_k(u; m, x) = h_k(u; m, x) \exp(ux - m\kappa(u)),$$

where the h_k satisfy the recursion

$$h_k(u; m, x) = h_{k-1}(u; m, x) \cdot (x - m\kappa'(u)) + (\partial/\partial u)h_{k-1}(u; m, x). \tag{1.8.15}$$

From this we get

$$\begin{aligned}
h_1(u; m, x) &= x - m\kappa'(u),\\
h_2(u; m, x) &= (x - m\kappa'(u))^2 - m\kappa''(u),\\
h_3(u; m, x) &= (x - m\kappa'(u))^3 - 3m\kappa''(u)(x - m\kappa'(u)) - m\kappa'''(u).
\end{aligned}$$

The derivatives of the cumulant generating function $\kappa(u) = -u + \mu(e^u - 1)$ are $\kappa'(u) = -1 + \mu e^u$ and $\kappa'^{(k)}(u) = \mu e^u$ for $k \geq 2$. Writing $\nu = \mu - 1$ we have

$$\begin{aligned}
f_1(m, x) &= x - m\nu,\\
f_2(m, x) &= (x - m\nu)^2 - m\mu,\\
f_3(m, x) &= (x - m\nu)^3 - 3m\mu(x - m\nu) - m\mu.
\end{aligned}$$

We have already seen the martingales that result from the first two functions. The third one is given by

$$M_m^{(3)} = (S_m - m\nu)^3 - 3m\mu(S_m - m\nu) - m\mu.$$

Using the domination in (1.8.12) we can stop at time τ and conclude

$$\begin{aligned}
1 &= E(S_\tau - \tau\nu)^3 - 3\mu E[\tau(S_\tau - \tau\nu)] - \mu E\tau\\
&= -(\mu - 1)^3 E\tau^3 + 3\mu(\mu - 1)E\tau^2 - \mu E\tau.
\end{aligned}$$

Since $E\tau = O(1/(1-\mu))$ while $E\tau^2 = O(1/(1-\mu)^3)$, the 1 and the $\mu E\tau$ are of much smaller order than the other terms in the equation so we have

$$\frac{E\tau^3}{E\tau^2} \sim \frac{3\mu}{(\mu - 1)^2} \qquad \text{and } \Delta_3 = 2.$$

From the last computation we see that we don't need all of the terms, just the largest ones. Using the recursion in (1.8.15) when $k = 4$ and letting $y = x - n\kappa'(u)$ to reduce the size of the formula

$$h_4(u; n, x) = y^4 - 3n\kappa''(y)y^2 + \cdots - 3n\kappa''(y)y^2 \ldots$$

so repeating the last argument $E\tau^3/E\tau^2 \sim C(\mu-1)^{-2}$ and $\Delta_4 = 2$. The pattern persists for higher moments, proving the result for Poisson branching processes. The method requires

($\star$) the moment generating function is finite for small θ,

but the simple formula for the cumulants is not important, so it will hold for all offspring distribution satisfying ($\star$).

1.8.4 Scaling theory

We have introduced five critical exponets ofor the phase transition in Erdős-Rényi random graphs: β, γ, Δ, τ and σ. The values of these exponents are not independent. In this section we will show that they satisfy three equalities, so if σ and τ are known then the other three can be computed. See (1.8.18), (1.8.19), and (1.8.20).

Formula (1.6.11) 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 τ 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 μ close to 1, we use (1.6.10)

$$P(s, \mu) \approx \frac{s^{1-5/2}}{\sqrt{2\pi}} \cdot \exp(-s(\mu - 1)^2/2). \tag{1.8.16}$$

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

$$f(x) = e^{-x/2}. \tag{1.8.17}$$

It has long been known for percolation (see Stauffer (1979)) that (1.8.5) implies

$$\beta = (\tau - 2)/\sigma \tag{1.8.18}$$
$$\gamma = (3 - \tau)/\sigma \tag{1.8.19}$$
$$\Delta = 1/\sigma \tag{1.8.20}$$

Since (**??**) gives the asymptotics for the cluster size distribution as $\mu \to \mu_c$ it should not be surprising that we can compute the symptotic behavior of the moments and of the percolation probability. Kesten (1987) has proved a number of rigorous results concerning scaling relations concentrating on the two dimensional case.

Lemma 1.8.1. *If $2 < \tau < 3$ and f is bounded, and is Lipschitz continuous at 0.*

$$\beta = (\tau - 2)/\sigma.$$

Proof. Using (**??**) 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 $\le Cx^{2-\tau}$ 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 (1.8.19) and (1.8.20).

Lemma 1.8.2. *If $2 < \gamma < 3$ and $\int_1^\infty x^m f(x)\, dx < \infty$ for all m then for all $r \ge 1$*

$$\Gamma(r) = (r + 2 - \tau)/\sigma.$$

It follows that for all integers $k \ge 2$,

$$\Delta_k = \Gamma(k) - \Gamma(k-1) = 1/\sigma.$$

Proof.

$$E|\mathcal{C}_x|^r = \int_1^\infty s^{r+1-\tau} f(s\delta^{1/\sigma})\, ds. \tag{1.8.21}$$

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

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

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 $\rho \ge 1$ the integral over $[0, 1]$ is finite □

1.9 Threshold for being connected

In this section we will ansewer the question: How large does λ have to be so that the probability Erdős-Rényi$(n, \lambda/n)$ is connected (i.e., ALL vertices in ONE component) tends to 1. I learned the following slick proof from the book by Frieze and Karonski who attribute it to Erdös and Rényi in 1959.

Theorem 1.9.1. *Let $p = (c_n + \log n)/n$ and let $H(n,p)$ be the event that Erdős-Rényi(n,p) is connected.*

$$P(H(n,p)) \to \begin{cases} 0 & \text{if } c_n \to -\infty, \\ \exp(-e^{-c}) & \text{if } c_n \to c \in (-\infty, \infty), \\ 1 & \text{if } c_n \to \infty. \end{cases}$$

The proof will be completed by proving the next two lemmas. By monotonicity of the event $H(n.p)$ in p it suffices to prove the result when $c_n \to c \in (-\infty, \infty)$.

Lemma 1.9.2. *Consider $G =$ Erdős-Rényi$(n, \lambda/n)$ with $\lambda = \log n + c_n + o(1)$ and $c_n \to c$. Then the number of isolated vertices I_n converges to a Poisson distribution with mean e^{-c}.*

Proof. The probability x is isolated is

$$\left(1 - \frac{\lambda}{n}\right)^n \sim \exp(-\log n - c_n) \sim e^{-c}/n,$$

so $EI_n \to e^{-c}$. The expected number of ordered k-tuples of isolated vertices is

$$EI_n^{(k)} = (n \cdot (n-1) \cdots (n-k+1)) \left(1 - \frac{\lambda}{n}\right)^{n+(n-1)+\cdots+(n-k+1)} \to e^{-ck}$$

so the Poisson convergence follows from Lemma 1.2.6. □

The next result completes the proof. For simplicity we suppose $c_n \equiv c$.

Lemma 1.9.3. *Let X_k be the number of clusters with k vertices. Let $p = (c + \log n)/n$ and let $H(n,p)$ be the event that Erdős-Rényi(n,p) is connected. As $n \to \infty$*

$$P(H(n,p)) = P(X_1 = 0) + o(1).$$

Proof. If the graph is not connected then there is at least one component of size $\leq n/2$, so $P(H(n,p)^c) = P(\cup_{k=1}^{n/2} \{X_k > 0\})$ and we have

$$P(X_1 > 0) \leq P(H(n,p)^c) \leq P(X_1 > 0) + \sum_{k=2}^{n/2} P(X_k > 0).$$

Computing as in Section 1.6 (see (1.6.1))

$$\sum_{k=2}^{n/2} P(X_k > 0) \le \sum_{k=2}^{n/2} \cdot \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \equiv \sum_{k=2}^{n/2} u_k$$

Now for $2 \le k \le 10$ using $\binom{n}{k} \le n^k/k!$ and $e^k \ge k^k/k!$

$$\begin{aligned} u_k &\le e^k n^k \left(\frac{c + \log n}{n} \right)^{k-1} e^{-k(n-10)(c+\log n)/n} \\ &\le e^k n^k \left(\frac{c + \log n}{n} \right)^{k-1} e^{-k(c+\log n)} e^{10k(c+\log n)/n} \\ &\le e^{k(1-c)} \left(\frac{c + \log n}{n} \right)^{k-1} \cdot (1 + o(1)) \end{aligned}$$

while for $10 < k \le n/2$

$$\begin{aligned} u_k &\le (ne/k)^k k^{k-2} \left(\frac{c + \log n}{n} \right)^{k-1} e^{-k(\log n + c)/2} \\ &\le n \left(\frac{e^{1-c/2}(c + \log n)}{n^{1/2}} \right)^k \end{aligned}$$

so we have $\sum_{k=2}^{n/2} u_k \to 0$ which completes the proof. □

Our final result which we state without proof gives a formula for the diameter.

Theorem 1.9.4. *If* $\liminf np/(\log n) > c > 1$ *and* $(\log p)/(\log n) \to 0$ *then the diameter of Erdős-Rényi*(n, p)*,* $D(n, p) \sim (\log n)/(\log np)$*.*

Remark 1.9.5. The first condition guarantees that the probability that the graph is connected tends to 1 as $n \to \infty$. The formula for the diameter is the same as in the sparse case when $np = \lambda$. To explain the second suppose $p = n^{-15/17}$. In this case $np = n^{2/17}$. 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). We mentioned other work by Chung and Lu (2001) and Riordan and Wormald (2010) at the end of Section 1.2. We leave it to the interested reader to explore this topic further.

1.10 References

Aizenmann, M., and Newman, C.M. (1984) Tree graph inequalities and critical behavior in percolation models. *J. Stat. Phys.* 36, 107–143

Ajtal, M., Komlós, J., and Szemerédi, E (1981) The longest path in a random graph. *Combinatorica.* 1, 1–17

Aldous, D. (1997) Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.* 25, 812–854

Aldous, D., and Limic, V. (1998) The entrance boundary of the multiplicative coalescent. *Electro. J. Probab.* 3, paper3, 59 pages

Barrez, D., Boucheron, S., and Fernandez de la Vega, W. (2000) On the fluctuations of the gaint component. *Comb. Probab. Comput.* 9, 287–304

Barsky and Aizenman (1991) Percolation critical exponents under the triangle condition. *Ann. Probab.* 19, 1520–1536

Bollobás, B. (1984) The evolution of random graphs. *Trans. Amer. Math. Soc.* 286, 257–274

Borel, E. (1942). Sur l'emploi du théorème de Bernoulli pour faciliter le calcul d'une infinité de coefficients. Application au probléme de l'attente á un guichet. *C. R. Acad. Sci.* 214: 452—456.

Chung, F., and Lu, L. (2001) The diameter of sparse random graphs. *Adv. Appl. Math.* 26, 257–279

Ding, J., Lubetzky, E., and Peres, Y. (2014) Anatomy of a giant component: the strictly supercritical regime. *European J. Combinatorics.* 35, 155–168

Enriquez, N., Faraud, G., and Ménard, L. (2020) Limiting shape of depth first search on an Erdős-Rényi graph. *Rand Struct. Alg.* 56, 501–516

Hara, T., and Slade, G. (1989) Mean-field critical behavior for percolation in high dimensions. Pages 450–453 in *Proceedings of the IXth International Congress of Mathematics Physics.* Adam Hilger, Boston

Hara, T., and Slade, G. (1994) Mean-field critical behavior and the lace expansion. *Probability and Phase Transition.* Edited by G.R. Grimmett. Kluwer, Dordrecht

Janson, S. (1993) Multicyclic components in a random graph process. *Rand. Struct. Alg.* 4, 71–84

Janson, S., Knuth, D.E., Luczak, T., and Pittel, B. (1993) The birth of the giant component. *Rand. Struct. Alg.* 4, 233–358

Kendall, D.G. (1956) Deterministic and stochastic epidemics in closed populations. *Proc.* 3^{rd} *Berekely Symp. Staisti. Prob.* Vol. 4, 149–165

Kesten, H. (1987) Scaling relations for 2D-percolation *Commun. Math. Phys.* 109, 109-156

Kesten, H., and Stigum, B. (1966) A limit theorem for multidimensional Gallton-Watson processes. *Ann. Math. Stat..* 37, 1211–1223

Krivelevich, M., and Sudakov, B. (2013) The phase transition in random graphs: A simple proof. *Rand. Struct. Alg.* 43, 131–138

Luczak, T. (1990) Component behavior near the critical point of a random graph process. *Rand. Struct. Alg.* 1, 287–310

Luczak, T., Pittel, B., and Weirman, J.C. (1994) The structure of a random graph at the point of phase transition. *Trans. Amer. Math. Soc.* 341, 721–748

Lyons, R. Pemantle,R., and Peres, Y. (1995) Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes *Ann. Probab.* 23, 1125–1138

Martin-Löf, A. (1986) Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Prob.* 23, 265–282

Martin-Löf, A. (1998) The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *J. Appl. Prob.* 35, 671–682

Nachmias, A., and Peres, Y. (2010) The critical random graph with martingales. *Israel J. Math.* 176, 29–41

Pittel, B. (1990) On tree census and the giant component in sparse random graphs. *Rand. Struct. Alg.* 1, 311–342

Riordan, O., and Wormald, N. (2010) The diameter of sparse random graphs. *Comb. Probab. Comput.* 19, 835–926

Robinson, R.W., and Wormald, N.C. (1992) Almost all cubic graphs are Hamiltonian. *Rand. Struct. Alg.* 3, 117–125

Stauffer, D. (1979) Scaling theory for percolation clusters. *Physics Reports.* 54, 1–79

Tanner, J. C. (1961). A derivation of the Borel distribution. *Biometrika* 48, 222–224.