Chapter 6

Branching processes

Branching processes, which are the focus of this chapter, arise naturally in the study of stochastic processes on trees and locally tree-like graphs. Similarly to martingales, finding a hidden (or not-so-hidden) branching process within a probabilistic model can lead to useful bounds and insights into asymptotic behavior. After a review of the basic *extinction theory* of branching processes in Section 6.1 and of a fruitful *random-walk perspective* in Section 6.2, we give a couple examples of applications in discrete probability in Section 6.3. In particular we analyze the height of a binary search tree, a standard data structure in computer science. We also give an introduction to phylogenetics, where a "multitype" variant of the Galton-Watson branching process plays an important role; we use the techniques derived in this chapter to establish a phase transition in the reconstruction of ancestral molecular sequences. We end this chapter in Section 6.4 with a detailed look into the *phase transition of the Erdős-Rényi graph model*. The random-walk perspective mentioned above allows one to analyze the "exploration" of a largest connected component, leading to information about the "evolution" of its size as edge density increases. Tools from all chapters come to bear on this final, marquee application.

6.1 Background

We begin with a review of the theory of Galton-Watson branching processes, a standard stochastic model for population growth. In particular we discuss extinction theory. We also briefly introduce a multitype variant, where branching process

Version: December 20, 2023
Modern Discrete Probability: An Essential Toolkit

and Markov chain aspects interact to produce interesting new behavior.

6.1.1 Basic definitions

Recall the definition of a Galton-Watson process.

Definition 6.1.1. *A* Galton-Watson branching process *is a Markov chain of the following form:*

Galton-Watson process

- *Let* $Z_0 := 1$.
- *Let* $X(i,t)$, $i \geq 1$, $t \geq 1$, *be an array of i.i.d.* $\mathbb{Z}_+$*-valued random variables with finite mean* $m = \mathbb{E}[X(1,1)] < +\infty$, *and define inductively,*

$$Z_t := \sum_{1 \leq i \leq Z_{t-1}} X(i,t).$$

We denote by $\{p_k\}_{k\geq 0}$ *the law of* $X(1,1)$. *We also let* $f(s) := \mathbb{E}[s^{X(1,1)}]$ *be the corresponding probability generating function. To avoid trivialities we assume* $\mathbb{P}[X(1,1) = i] < 1$ *for all* $i \geq 0$. *We further assume that* $p_0 > 0$.

In words, Z_t models the size of a population at time (or generation) t. The random variable $X(i,t)$ corresponds to the number of offspring of the i-th individual (if there is one) in generation $t-1$. Generation t is formed of all offspring of the individuals in generation $t-1$.

By tracking genealogical relationships, that is, who is whose child, we obtain a tree T rooted at the single individual in generation 0 with a vertex for each individual in the progeny and an edge for each parent-child relationship. We refer to T as a *Galton-Watson tree*.

Galton-Watson tree

A basic observation about Galton-Watson processes is that their growth (or decay) is exponential in t.

Lemma 6.1.2 (Exponential growth I)**.** *Let*

$$W_t := m^{-t} Z_t. \tag{6.1.1}$$

Then (W_t) *is a nonnegative martingale with respect to the filtration*

$$\mathcal{F}_t = \sigma(Z_0, \ldots, Z_t).$$

In particular, $\mathbb{E}[Z_t] = m^t$.

Proof. We use Lemma B.6.17. Observe that on $\{Z_{t-1} = k\}$

$$\mathbb{E}[Z_t \,|\, \mathcal{F}_{t-1}] = \mathbb{E}\left[\sum_{1 \leq j \leq k} X(j,t) \,\middle|\, \mathcal{F}_{t-1}\right] = mk = mZ_{t-1}.$$

This is true for all k. Rearranging shows that (W_t) is a martingale. For the second claim, note that $\mathbb{E}[W_t] = \mathbb{E}[W_0] = 1$. ∎

In fact, the martingale convergence theorem (Theorem 3.1.47) gives the following.

Lemma 6.1.3 (Exponential growth II). *We have $W_t \to W_\infty < +\infty$ almost surely for some nonnegative random variable $W_\infty \in \sigma(\cup_t \mathcal{F}_t)$ with $\mathbb{E}[W_\infty] \leq 1$.*

Proof. This follows immediately from the martingale convergence theorem for nonnegative martingales (Corollary 3.1.48). ∎

6.1.2 Extinction

Observe that 0 is a fixed point of the process. The event

$$\{Z_t \to 0\} = \{\exists t \,:\, Z_t = 0\},$$

is called *extinction*. Establishing when extinction occurs is a central question in branching process theory. We let η be the probability of extinction. Recall that, to avoid trivialities, we assume that $p_0 > 0$ and $p_1 < 1$. Here is a first observation about extinction.

extinction

Lemma 6.1.4. *Almost surely either $Z_t \to 0$ or $Z_t \to +\infty$.*

Proof. The process (Z_t) is integer-valued and 0 is the only fixed point of the process under the assumption that $p_1 < 1$. From any state k, the probability of never coming back to $k > 0$ is at least $p_0^k > 0$, so every state $k > 0$ is transient. So the only possibilities left are $Z_t \to 0$ and $Z_t \to +\infty$, and the claim follows. ∎

In the critical case, that immediately implies almost sure extinction.

Theorem 6.1.5 (Extinction: critical case). *Assume $m = 1$. Then $Z_t \to 0$ almost surely, that is, $\eta = 1$.*

Proof. When $m = 1$, (Z_t) itself is a martingale. Hence (Z_t) must converge to 0 by Lemma 6.1.3. ∎

We address the general case using probability generating functions. Let $f_t(s) = \mathbb{E}[s^{Z_t}]$, where by convention we set $f_t(0) := \mathbb{P}[Z_t = 0]$. Note that, by monotonicity,

$$\eta = \mathbb{P}[\exists t \geq 0 \,:\, Z_t = 0] = \lim_{t\to+\infty} \mathbb{P}[Z_t = 0] = \lim_{t\to+\infty} f_t(0). \tag{6.1.2}$$

Moreover, by the tower property (Lemma B.6.16) and the Markov property (Theorem 1.1.18), f_t has a natural recursive form

$$\begin{aligned} f_t(s) &= \mathbb{E}[s^{Z_t}] \\ &= \mathbb{E}[\mathbb{E}[s^{Z_t} \,|\, \mathcal{F}_{t-1}]] \\ &= \mathbb{E}[f(s)^{Z_{t-1}}] \\ &= f_{t-1}(f(s)) = \cdots = f^{(t)}(s), \end{aligned} \tag{6.1.3}$$

where $f^{(t)}$ is the t-th iterate of f. The subcritical case below has an easier proof (see Exercise 6.1).

Theorem 6.1.6 (Extinction: subcriticial and supercritical cases)**.** *The probability of extinction η is given by the smallest fixed point of f in $[0,1]$. Moreover:*

(i) (Subcritical regime) *If $m < 1$ then $\eta = 1$.*

(ii) (Supercritical regime) *If $m > 1$ then $\eta < 1$.*

Proof. The case $p_0 + p_1 = 1$ is straightforward: the process dies almost surely after a geometrically distributed time. So we assume $p_0 + p_1 < 1$ for the rest of the proof.

We first summarize without proof some properties of f which follow from standard power series facts.

Lemma 6.1.7. *On $[0,1]$, the function f satisfies:*

(i) $f(0) = p_0$*,* $f(1) = 1$*;*

(ii) *f is infinitely differentiable on $[0,1)$;*

(iii) *f is strictly convex and increasing; and*

(iv) $\lim_{s\uparrow 1} f'(s) = m < +\infty$.

We first characterize the fixed points of f. See Figure 6.1 for an illustration.

Lemma 6.1.8. *We have the following.*

(i) If $m > 1$ then f has a unique fixed point $\eta_0 \in [0,1)$.

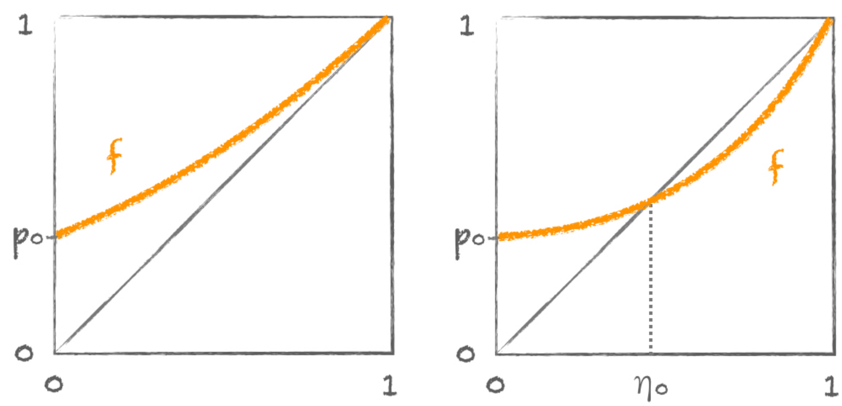

Figure 6.1: Fixed points of f in subcritical (left) and supercritical (right) cases.

(ii) If $m < 1$ then $f(t) > t$ for $t \in [0, 1)$. Let $\eta_0 := 1$ in that case.

Proof. Assume $m > 1$. Since $f'(1) = m > 1$, there is $\delta > 0$ such that $f(1-\delta) < 1 - \delta$. On the other hand $f(0) = p_0 > 0$ so by continuity of f there must be a fixed point in $(0, 1-\delta)$. Moreover, by strict convexity and the fact that $f(1) = 1$, if $x \in (0, 1)$ is a fixed point then $f(y) < y$ for $y \in (x, 1)$, proving uniqueness.

The second part follows by strict convexity and monotonicity. ■

It remains to prove convergence of the iterates to the appropriate fixed point. See Figure 6.2 for an illustration.

Lemma 6.1.9. *We have the following.*

(i) If $x \in [0, \eta_0)$, then $f^{(t)}(x) \uparrow \eta_0$.

(ii) If $x \in (\eta_0, 1)$ then $f^{(t)}(x) \downarrow \eta_0$.

Proof. We only prove (i). The argument for (ii) is similar. By monotonicity, for $x \in [0, \eta_0)$, we have $x < f(x) < f(\eta_0) = \eta_0$. Iterating

$$x < f^{(1)}(x) < \cdots < f^{(t)}(x) < f^{(t)}(\eta_0) = \eta_0.$$

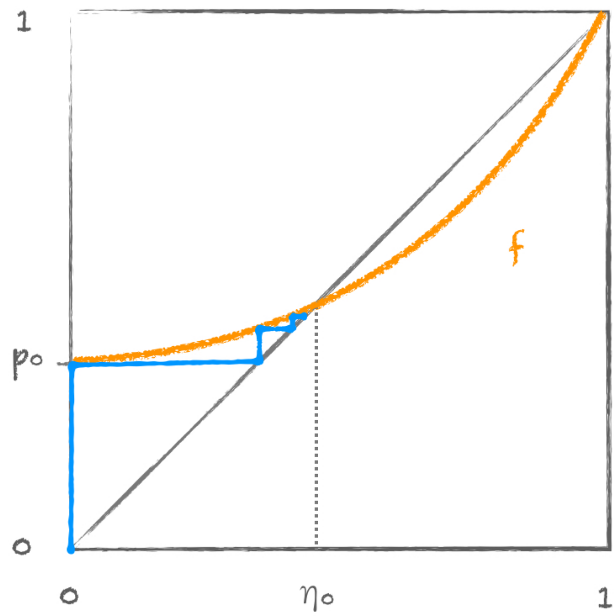

Figure 6.2: Convergence of iterates to a fixed point.

So $f^{(t)}(x) \uparrow L \leq \eta_0$ as $t \to \infty$. By continuity of f, we can take the limit $t \to \infty$ inside of f on the right-hand side of the equality

$$f^{(t)}(x) = f(f^{(t-1)}(x)),$$

to get $L = f(L)$. So by definition of η_0 we must have $L = \eta_0$. ■

The result then follows from the above lemmas together with Equations (6.1.2) and (6.1.3). ■

Example 6.1.10 (Poisson branching process)**.** Consider the offspring distribution $X(1,1) \sim \mathrm{Poi}(\lambda)$ with mean $\lambda > 0$. We refer to this case as the *Poisson branching process*. Then

$$f(s) = \mathbb{E}[s^{X(1,1)}] = \sum_{i \geq 0} e^{-\lambda} \frac{\lambda^i}{i!} s^i = e^{\lambda(s-1)}.$$

So the process goes extinct with probability 1 when $\lambda \leq 1$. For $\lambda > 1$, the probability of extinction η is the smallest solution in $[0,1]$ to the equation

$$e^{-\lambda(1-x)} = x.$$

The survival probability $\zeta_\lambda := 1 - \eta$ satisfies $1 - e^{-\lambda\zeta_\lambda} = \zeta_\lambda$. ◀

We can use these extinction results to obtain more information on the limit in Lemma 6.1.3. Recall the definition of (W_t) in (6.1.1). Of course, conditioned on extinction, $W_\infty = 0$ almost surely. On the other hand:

Lemma 6.1.11 (Exponential growth III)**.** *Conditioned on nonextinction, either $W_\infty = 0$ almost surely or $W_\infty > 0$ almost surely.*

As a result, $\mathbb{P}[W_\infty = 0] \in \{\eta, 1\}$.

Proof of Lemma 6.1.11. A property of rooted trees is said to be *inherited* if all finite trees satisfy the property and whenever a tree satisfies the property then so do all subtrees rooted at the children of the root. The property $\{W_\infty = 0\}$, as a property of the Galton-Watson tree T, is inherited, seeing that Z_t is a sum over the children of the root of the number of descendants at the corresponding generation $t-1$. The result then follows from the following 0-1 law.

Lemma 6.1.12 (0-1 law for inherited properties)**.** *For a Galton-Watson tree T, an inherited property A has, conditioned on nonextinction, probability 0 or 1.*

Proof. Let $T^{(1)}, \ldots, T^{(Z_1)}$ be the descendant subtrees of the children of the root. We use the notation $T \in A$ to mean that tree T satisfies A. By the tower property, the definition of inherited, and conditional independence,

$$\begin{aligned}
\mathbb{P}[A] &= \mathbb{E}[\mathbb{P}[T \in A \,|\, Z_1]] \\
&\leq \mathbb{E}[\mathbb{P}[T^{(i)} \in A, \forall i \leq Z_1 \,|\, Z_1]] \\
&= \mathbb{E}[\mathbb{P}[A]^{Z_1}] \\
&= f(\mathbb{P}[A]).
\end{aligned}$$

So $\mathbb{P}[A] \in [0, \eta] \cup \{1\}$ by the proof of Lemma 6.1.8.

Moreover since A holds for finite trees, we have $\mathbb{P}[A] \geq \eta$, where recall that η is the probability of extinction. Hence, in fact, $\mathbb{P}[A] \in \{\eta, 1\}$. Conditioning on nonextinction gives the claim. ■

That concludes the proof. ■

A further moment assumption provides a more detailed picture.

Lemma 6.1.13 (Exponential growth IV)**.** *Let (Z_t) be a Galton-Watson branching process with $m = \mathbb{E}[X(1,1)] > 1$ and $\sigma^2 = \mathrm{Var}[X(1,1)] < +\infty$. Then, (W_t) converges in L^2 and, in particular, $\mathbb{E}[W_\infty] = 1$. Further, $\mathbb{P}[W_\infty = 0] = \eta$.*

Proof. We bound $\mathbb{E}[W_t^2]$ by computing it explicitly by induction. From the orthogonality of increments (Lemma 3.1.50), it holds that

$$\mathbb{E}[W_t^2] = \mathbb{E}[W_{t-1}^2] + \mathbb{E}[(W_t - W_{t-1})^2].$$

Since $\mathbb{E}[W_t \,|\, \mathcal{F}_{t-1}] = W_{t-1}$ by the martingale property,

$$\begin{aligned}
\mathbb{E}[(W_t - W_{t-1})^2 \,|\, \mathcal{F}_{t-1}] &= \mathrm{Var}[W_t \,|\, \mathcal{F}_{t-1}] \\
&= m^{-2t}\,\mathrm{Var}[Z_t \,|\, \mathcal{F}_{t-1}] \\
&= m^{-2t}\,\mathrm{Var}\left[\sum_{i=1}^{Z_{t-1}} X(i,t) \,\middle|\, \mathcal{F}_{t-1}\right] \\
&= m^{-2t} Z_{t-1}\sigma^2.
\end{aligned}$$

Hence, taking expectations and using Lemma 6.1.2, we get

$$\mathbb{E}[W_t^2] = \mathbb{E}[W_{t-1}^2] + m^{-t-1}\sigma^2.$$

Since $\mathbb{E}[W_0^2] = 1$, induction gives

$$\mathbb{E}[W_t^2] = 1 + \sigma^2 \sum_{i=2}^{t+1} m^{-i},$$

which is uniformly bounded from above when $m > 1$.

By the convergence theorem for martingales bounded in L^2 (Theorem 3.1.51), (W_t) converges almost surely and in L^2 to a finite limit W_∞ and

$$1 = \mathbb{E}[W_t] \to \mathbb{E}[W_\infty].$$

The last statement follows from Lemma 6.1.11. ■

Remark 6.1.14. *A theorem of Kesten and Stigum gives a necessary and sufficient condition for $\mathbb{E}[W_\infty] = 1$ to hold [KS66b]. See, e.g., [LP16, Chapter 12].*

6.1.3 ▷ *Percolation: Galton-Watson trees*

Let T be the Galton-Watson tree for an offspring distribution with mean $m > 1$. Now perform bond percolation on T with density p (see Definition 1.2.1). Let $\mathcal{C}_0$ be the open cluster of the root in T. Recall from Section 2.3.3 that the critical value is

$$p_c(T) = \sup\{p \in [0,1] \,:\, \theta(p) = 0\},$$

where the percolation function (conditioned on T) is $\theta(p) = \mathbb{P}_p[|\mathcal{C}_0| = +\infty \,|\, T]$.

Theorem 6.1.15 (Bond percolation on Galton-Watson trees)**.** *Assume $m > 1$. Conditioned on nonextinction of T,*

$$p_c(T) = \frac{1}{m},$$

almost surely.

Proof. We can think of $\mathcal{C}_0$ (or more precisely, its size on each level) as being itself generated by a Galton-Watson branching process, where this time the offspring distribution is the law of $\sum_{i=1}^{X(1,1)} I_i$ where the I_is are i.i.d. $\mathrm{Ber}(p)$ and $X(1,1)$ is distributed according to the offspring distribution of T. In words, we are "thinning" T. By conditioning on $X(1,1)$ and then using the tower property (Lemma B.6.16), the offspring mean under the process generating $\mathcal{C}_0$ is mp.

If $mp \leq 1$ then by the extinction theory (Theorems 6.1.5 and 6.1.6)

$$1 = \mathbb{P}_p[|\mathcal{C}_0| < +\infty] = \mathbb{E}[\mathbb{P}_p[|\mathcal{C}_0| < +\infty \,|\, T]],$$

and we must have $\mathbb{P}_p[|\mathcal{C}_0| < +\infty \,|\, T] = 1$ almost surely. Taking $p = 1/m$, we get $p_c(T) \geq \frac{1}{m}$ almost surely. That holds, in particular, on the nonextinction of T which happens with positive probability.

For the other direction, fix p such that $mp > 1$. The property of trees $\{\mathbb{P}_p[|\mathcal{C}_0| < +\infty \,|\, T] = 1\}$ is inherited. So by Lemma 6.1.12, conditioned on nonextinction of T, it has probability 0 or 1. That probability is of course 1 on extinction. By Theorem 6.1.6,

$$1 > \mathbb{P}_p[|\mathcal{C}_0| < +\infty] = \mathbb{E}[\mathbb{P}_p[|\mathcal{C}_0| < +\infty \,|\, T]],$$

and, conditioned on nonextinction of T, we must have $\mathbb{P}_p[|\mathcal{C}_0| < +\infty \,|\, T] = 0$—i.e., $p_c(T) < p$—almost surely. Repeating this argument for a sequence $p_n \downarrow 1/m$ simultaneously (i.e., on the same T) and using the monotonicity of $\mathbb{P}_p[|\mathcal{C}_0| < +\infty \,|\, T]$, we get that $p_c(T) \leq 1/m$ almost surely conditioned on nonextinction of T. That proves the claim. ∎

6.1.4 Multitype branching processes

Multitype branching processes are a useful generalization of Galton-Watson processes (Definition 6.1.1). Their behavior combines aspects of branching processes (exponential growth, extinction, etc.) and Markov chains (reducibility, mixing, etc.). We will not develop the full theory here. In this section, we define this class of processes and hint (largely without proofs) at their properties. In Section 6.3.2, we illustrate some of the more intricate interplay between the driving phenomena involved in a special example of practical importance.

multitype branching processes

Definition In a multitype branching process, each individual has one of τ types, which we will denote in this section by $1, \ldots, \tau$ for simplicity. Each type $\alpha \in [\tau] = \{1, \ldots, \tau\}$ has its own offspring distribution $\{p_{\mathbf{k}}^{(\alpha)} : \mathbf{k} \in \mathbb{Z}_+^\tau\}$, which specifies the distribution of the number of offspring of each type it has. Just to emphasize, this is a *collection of (typically distinct) multivariate distributions*.

For reasons that will become clear below, it will be convenient to work with row vectors. For each $\alpha \in [\tau]$, let

$$\mathbf{X}^{(\alpha)}(i,t) = \left(X_1^{(\alpha)}(i,t), \ldots, X_\tau^{(\alpha)}(i,t)\right), \quad \forall i, t \geq 1$$

be an array of i.i.d. $\mathbb{Z}_+^\tau$-valued random row vectors with distribution $\{p_{\mathbf{k}}^{(\alpha)}\}$. Let

$$\mathbf{Z}_0 = \mathbf{k}_0 \in \mathbb{Z}_+^\tau,$$

be the initial population at time 0, again as a row vector. Recursively, the population vector

$$\mathbf{Z}_t = (Z_{t,1}, \ldots, Z_{t,\tau}) \in \mathbb{Z}_+^\tau,$$

at time $t \geq 1$ is set to

$$\mathbf{Z}_t := \sum_{\alpha=1}^{\tau} \sum_{i=1}^{Z_{t-1,\alpha}} \mathbf{X}^{(\alpha)}(i,t). \tag{6.1.4}$$

In words, the i-th individual of type α at generation $t-1$ produces $X_\beta^{(\alpha)}(i,t)$ individuals of type β at generation t (before itself dying). Let $\mathcal{F}_t = \sigma(\mathbf{Z}_0, \ldots, \mathbf{Z}_t)$ be the corresponding filtration. We assume throughout that $\mathbb{P}[\|\mathbf{X}^{(\alpha)}(1,1)\|_1 = 1] < 1$ for at least one α (which is referred to as the *nonsingular case*); otherwise the process reduces to a simple finite Markov chain.

nonsingular case

Martingales As in the single-type case, the means of the offspring distributions play a key role in the theory. This time however they form a matrix, the so-called *mean matrix* $M = (m_{\alpha,\beta})$ with entries

mean matrix

$$m_{\alpha,\beta} = \mathbb{E}\left[X_\beta^{(\alpha)}(1,1)\right], \quad \forall \alpha, \beta \in [\tau].$$

That is, $m_{\alpha,\beta}$ is the expected number of offspring of type β of an individual of type α. We assume throughout that $m_{\alpha,\beta} < +\infty$ for all α, β.

To see how M drives the growth of the process, we generalize the proof of Lemma 6.1.2. By the recursive formula (6.1.4),

$$\begin{aligned}\mathbb{E}\left[\mathbf{Z}_t \,|\, \mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\sum_{\alpha=1}^{\tau}\sum_{i=1}^{Z_{t-1,\alpha}} \mathbf{X}^{(\alpha)}(i,t) \,\middle|\, \mathcal{F}_{t-1}\right] \\ &= \sum_{\alpha=1}^{\tau}\sum_{i=1}^{Z_{t-1,\alpha}} \mathbb{E}\left[\mathbf{X}^{(\alpha)}(i,t) \,\middle|\, \mathcal{F}_{t-1}\right] \\ &= \sum_{\alpha=1}^{\tau} Z_{t-1,\alpha}\, \mathbb{E}\left[\mathbf{X}^{(\alpha)}(1,1)\right] \\ &= \mathbf{Z}_{t-1}M, \end{aligned} \tag{6.1.5}$$

where recall that $\mathbf{Z}_{t-1}$ and $\mathbf{Z}_t$ are row vectors. Inductively,

$$\mathbb{E}\left[\mathbf{Z}_t \,|\, \mathbf{Z}_0\right] = \mathbf{Z}_0 M^t. \tag{6.1.6}$$

Moreover any real right eigenvector $\mathbf{u}$ (as a column vector) of M with real eigenvalue $\lambda \neq 0$ gives rise to a martingale

$$U_t := \lambda^{-t}\mathbf{Z}_t\mathbf{u}, \quad t \geq 0, \tag{6.1.7}$$

since

$$\begin{aligned}\mathbb{E}[U_t \,|\, \mathcal{F}_{t-1}] &= \mathbb{E}[\lambda^{-t}\mathbf{Z}_t\mathbf{u} \,|\, \mathcal{F}_{t-1}] \\ &= \lambda^{-t}\, \mathbb{E}[\mathbf{Z}_t \,|\, \mathcal{F}_{t-1}]\, \mathbf{u} \\ &= \lambda^{-t}\mathbf{Z}_{t-1}M\mathbf{u} \\ &= \lambda^{-t}\mathbf{Z}_{t-1}\lambda\mathbf{u} \\ &= U_{t-1}. \end{aligned}$$

Extinction The classical *Perron-Frobenius Theorem* characterizes the direction of largest growth of the matrix M. We state a version of it without proof in the case where all entries of M are strictly positive, which is referred to as the *positive regular case*. Note that, unlike the case of simple finite Markov chains, the matrix M is not in general stochastic, as it also reflects the "growth" of the population in addition to the "transitions" between types. We encountered the following concept in Section 5.2.5 and Exercise 5.5.

positive regular case

Definition 6.1.16. *The* spectral radius $\rho(A)$ *of a matrix A is the maximum of the eigenvalues of A in absolute value.*

spectral radius

Theorem 6.1.17 (Perron-Frobenius theorem: positive regular case)**.** *Let M be a strictly positive, square matrix. Then $\rho := \rho(M)$ is an eigenvalue of M with algebraic and geometric multiplicities 1. It is also the only eigenvalue with absolute value ρ. The corresponding left and right eigenvectors, denoted by $\mathbf{v}$ (as a row vector) and $\mathbf{w}$ (as a column vector) respectively, are positive vectors. They are referred to as left and right* Perron vector. *We assume that they are normalized so that $\mathbf{1}\mathbf{w} = 1$ and $\mathbf{v}\mathbf{w} = 1$. Here $\mathbf{1}$ is the all-one row vector.*

Perron vector

Because $\mathbf{w}$ is positive, the martingale

$$W_t := \rho^{-t}\mathbf{Z}_t\mathbf{w}, \quad t \geq 0,$$

is nonnegative. Therefore it converges almost surely to a random limit with a finite mean by Corollary 3.1.48. When $\rho < 1$, an argument based on Markov's inequality (Theorem 2.1.1) implies that the process goes extinct almost surely. Formally, let $q^{(\alpha)}$ be the probability of extinction when started with a single individual of type α, that is,

$$q^{(\alpha)} := \mathbb{P}[\mathbf{Z}_t = \mathbf{0} \text{ for some } t \,|\, \mathbf{Z}_0 = \mathbf{e}_\alpha],$$

where $\mathbf{e}_\alpha \in \mathbb{Z}_+^\tau$ is the standard basis row vector with a one in the α-th coordinate, and let $\mathbf{q} := (q^{(1)}, \ldots, q^{(\tau)})$. Then

$$\rho < 1 \implies \mathbf{q} = \mathbf{1}. \tag{6.1.8}$$

Exercise 6.1 asks for the proof. We state the following more general result without proof. We use the notation of Theorem 6.1.17. We will also refer to the generating functions

$$f^{(\alpha)}(\mathbf{s}) := \mathbb{E}\left[\prod_{\beta=1}^{\tau} s_\beta^{X_\beta^{(\alpha)}(1,1)}\right], \quad \mathbf{s} \in [0,1]^\tau$$

with $\mathbf{f} = (f^{(1)}, \ldots, f^{(\tau)})$.

Theorem 6.1.18 (Extinction: multitype case)**.** *Let $(\mathbf{Z})_t$ be a positive regular, nonsingular multitype branching process with a finite mean matrix M.*

(i) If $\rho \leq 1$ then $\mathbf{q} = \mathbf{1}$.

(ii) If $\rho > 1$ then:

- *a- It holds that $\mathbf{q} < \mathbf{1}$.*
- *b- The unique solution to $\mathbf{f}(\mathbf{s}) = \mathbf{s}$ in $[0,1)^\tau$ is $\mathbf{q}$.*

c- *Almost surely*

$$\lim_{t \to +\infty} \rho^{-t} \mathbf{Z_t} = \mathbf{v} W_\infty,$$

where W_∞ *is a nonnegative random variable.*

d- *If in addition* $\mathrm{Var}[X^{(\alpha)}_\beta(1,1)] < +\infty$ *for all* α, β *then*

$$\mathbb{E}[W_\infty \,|\, \mathbf{Z}_0 = \mathbf{e}_\alpha] = w_\alpha,$$

and

$$q^{(\alpha)} = \mathbb{P}[W_\infty = 0 \,|\, \mathbf{Z}_0 = \mathbf{e}_\alpha],$$

for all $\alpha \in [\tau]$.

Remark 6.1.19. *As in the single-type case, a theorem of Kesten and Stigum gives a necessary and sufficient condition for the last claim of Theorem 6.1.18 (ii) to hold [KS66b].*

Linear functionals Theorem 6.1.18 also characterizes the limit behavior of linear functionals of the form $\mathbf{Z_t u}$ for any vector that is *not* orthogonal to $\mathbf{v}$. In contrast, interesting new behavior arises when $\mathbf{u}$ is orthogonal to $\mathbf{v}$. We will not derive the general theory here. We only show through a second moment calculation that a phase transition takes place.

We restrict ourselves to the supercritical case $\rho > 1$ and to $\mathbf{u} = (u_1, \ldots, u_\tau)$ being a real right eigenvector of M with a real eigenvalue $\lambda \notin \{0, \rho\}$. Let U_t be the corresponding martingale from (6.1.7). The vector $\mathbf{u}$ is necessarily orthogonal to $\mathbf{v}$. Indeed $\mathbf{v}M\mathbf{u}$ is equal to both $\rho\mathbf{vu}$ and $\lambda\mathbf{vu}$. Because $\rho \neq \lambda$ by assumption, this is only possible if all three expressions are 0. That implies $\mathbf{vu} = 0$ since we also have $\rho \neq 0$ by assumption.

To compute the second moment of U_t, we mimic the computations in the proof of Lemma 6.1.13. We have

$$\mathbb{E}[U_t^2 \,|\, \mathbf{Z}_0] = \mathbb{E}[U_{t-1}^2 \,|\, \mathbf{Z}_0] + \mathbb{E}[(U_t - U_{t-1})^2 \,|\, \mathbf{Z}_0],$$

by the orthogonality of increments (Lemma 3.1.50). Since $\mathbb{E}[U_t \,|\, \mathcal{F}_{t-1}] = U_{t-1}$ by the martingale property, we get

$$\begin{aligned}
\mathbb{E}[(U_t - U_{t-1})^2 \,|\, \mathcal{F}_{t-1}] &= \mathrm{Var}[U_t \,|\, \mathcal{F}_{t-1}] \\
&= \mathrm{Var}[\lambda^{-t} \mathbf{Z}_t \mathbf{u} \,|\, \mathcal{F}_{t-1}] \\
&= \lambda^{-2t} \,\mathrm{Var}\left[\left(\sum_{\alpha=1}^{\tau} \sum_{i=1}^{Z_{t-1,\alpha}} \mathbf{X}^{(\alpha)}(i,t) \right) \mathbf{u} \,\middle|\, \mathcal{F}_{t-1} \right] \\
&= \lambda^{-2t} \sum_{\alpha=1}^{\tau} Z_{t-1,\alpha} \,\mathrm{Var}\left[\mathbf{X}^{(\alpha)}(1,1)\, \mathbf{u} \right] \\
&= \lambda^{-2t} \mathbf{Z}_{t-1} \mathbf{S}^{(\mathbf{u})},
\end{aligned}$$

where $\mathbf{S}^{(\mathbf{u})} = (\mathrm{Var}[\mathbf{X}^{(1)}(1,1)\,\mathbf{u}], \ldots, \mathrm{Var}[\mathbf{X}^{(\tau)}(1,1)\,\mathbf{u}])$ as a column vector. In the last display, we used (6.1.4) on the third line and the independence of the random vectors $\mathbf{X}^{(\alpha)}(i,t)$ on the fourth line. Hence, taking expectations and using (6.1.6), we get

$$\mathbb{E}[U_t^2 \,|\, \mathbf{Z}_0] = \mathbb{E}[U_{t-1}^2 \,|\, \mathbf{Z}_0] + \lambda^{-2t}\mathbf{Z}_0 M^{t-1}\mathbf{S}^{(\mathbf{u})}.$$

and finally

$$\mathbb{E}[U_t^2 \,|\, \mathbf{Z}_0] = (\mathbf{Z}_0\mathbf{u})^2 + \sum_{s=1}^{t} \lambda^{-2s}\mathbf{Z}_0 M^{s-1}\mathbf{S}^{(\mathbf{u})}. \tag{6.1.9}$$

The case $\mathbf{S}^{(\mathbf{u})} = \mathbf{0}$ is trivial (see Exercise 6.4), so we exclude it from the following lemma.

Lemma 6.1.20 (Second moment of U_t)**.** *Assume $\mathbf{S}^{(\mathbf{u})} \neq \mathbf{0}$ and $\mathbf{Z}_0 \neq \mathbf{0}$. The sequence $\mathbb{E}[U_t^2 \,|\, \mathbf{Z}_0]$, $t = 0, 1, 2, \ldots$, is non-decreasing and satisfies*

$$\sup_{t\geq 0} \mathbb{E}[U_t^2 \,|\, \mathbf{Z}_0] \begin{cases} < +\infty & \text{if } \rho < \lambda^2, \\ = +\infty & \text{otherwise.} \end{cases}$$

Proof. Because $\mathbf{S}^{(\mathbf{u})} \neq \mathbf{0}$ and nonnegative and the matrix M is strictly positive by assumption, we have that

$$\widetilde{\mathbf{S}}^{(\mathbf{u})} := M\mathbf{S}^{(\mathbf{u})} > \mathbf{0}.$$

Since $\mathbf{w}$ is also strictly positive, there is $0 < C^- \leq C^+ < +\infty$ such that

$$C^-\mathbf{w} \leq \widetilde{\mathbf{S}}^{(\mathbf{u})} \leq C^+\mathbf{w}.$$

Moreover, since M is positive, each inequality is preserved when multiplying on both sides by M, that is, for any $s \geq 1$

$$C^-\rho^s\mathbf{w} \leq M^s\widetilde{\mathbf{S}}^{(\mathbf{u})} \leq C^+\rho^s\mathbf{w}. \tag{6.1.10}$$

Now rewrite (6.1.9) as

$$\mathbb{E}[U_t^2 \,|\, \mathbf{Z}_0] = (\mathbf{Z}_0\mathbf{u})^2 + \lambda^{-2}\mathbf{Z}_0\mathbf{S}^{(\mathbf{u})} + \lambda^{-4}\sum_{s=2}^{t} \mathbf{Z}_0(1/\lambda^2)^{s-2}M^{s-2}\widetilde{\mathbf{S}}^{(\mathbf{u})}.$$

There are two cases:

- When $\rho < \lambda^2$, using (6.1.10), the sum on the right-hand side can be bounded *above* by

$$C^+\mathbf{Z}_0\mathbf{w}\sum_{s=2}^{t}\left(\frac{\rho}{\lambda^2}\right)^{s-2} \leq C^+\mathbf{Z}_0\mathbf{w}\frac{1}{1-(\rho/\lambda^2)} < +\infty,$$

  uniformly in t.

- When $\rho \geq \lambda^2$, the same sum can be bounded from *below* by

$$C^- \mathbf{Z}_0 \mathbf{w} \sum_{s=2}^{t} \left(\frac{\rho}{\lambda^2}\right)^{s-2} \to +\infty,$$

as $t \to +\infty$. Indeed, $\mathbf{Z}_0 \neq \mathbf{0}$ implies that the inner product $\mathbf{Z}_0\mathbf{w}$ is strictly positive.

That proves the claim. ■

In the case $\rho < \lambda^2$, the martingale (U_t) is bounded in L^2 and therefore converges almost surely to a limit U_∞ with $\mathbb{E}[U_\infty \,|\, \mathbf{Z}_0] = \mathbf{Z}_0\mathbf{u}$ by Theorem 3.1.51. On the other hand, when $\rho \geq \lambda^2$, it can be shown (we will not do this here) that $\mathbf{Z}_t\mathbf{u}/\sqrt{\mathbf{Z}_t\mathbf{w}}$ satisfies a central limit theorem with a limit independent of $\mathbf{Z}_0$. Implications of these claims are illustrated in Section 6.3.2.

6.2 Random-walk representation

In this section, we develop a random-walk representation of the Galton-Watson process. We give two applications: a characterization of the Galton-Watson process *conditioned on extinction* in terms of a dual branching process; and a formula for the *size of the total progeny*. We illustrate both in Section 6.2.4, where we revisit percolation on the infinite b-ary tree.

6.2.1 Exploration process

We introduce an exploration process where a random-walk perspective will naturally arise.

Exploration of a graph

Because this will be useful again later, we describe it first in the context of a locally finite graph $G = (V, E)$. The exploration process starts at an arbitrary vertex $v \in V$ and has 3 types of vertices:

- $\mathcal{A}_t$: *active* vertices,
- $\mathcal{E}_t$: *explored* vertices,
- $\mathcal{N}_t$: *neutral* vertices.

active
explored
neutral

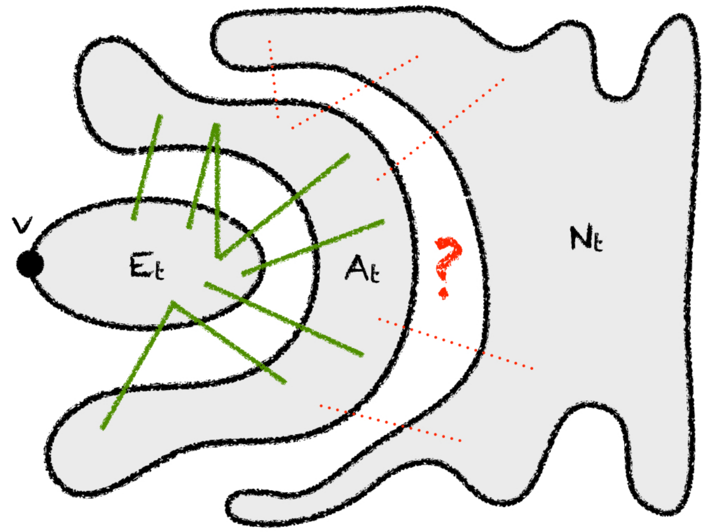

Figure 6.3: Exploration process for $\mathcal{C}_v$.

At the beginning, we have $\mathcal{A}_0 := \{v\}$, $\mathcal{E}_0 := \emptyset$, and $\mathcal{N}_0$ contains all other vertices in G. At time t, if $\mathcal{A}_{t-1} = \emptyset$ (i.e., there are no active vertices) we let $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t) := (\mathcal{A}_{t-1}, \mathcal{E}_{t-1}, \mathcal{N}_{t-1})$. Otherwise, we pick an element, a_t, from $\mathcal{A}_{t-1}$ (say in first-come, first-served basis to be explicit) and set:

- $\mathcal{A}_t := (\mathcal{A}_{t-1} \backslash \{a_t\}) \cup \{x \in \mathcal{N}_{t-1} \,:\, \{x, a_t\} \in E\}$
- $\mathcal{E}_t := \mathcal{E}_{t-1} \cup \{a_t\}$
- $\mathcal{N}_t := \mathcal{N}_{t-1} \backslash \{x \in \mathcal{N}_{t-1} \,:\, \{x, a_t\} \in E\}$

We imagine revealing the edges of G as they are encountered in the exploration process. In words, starting with v, the connected component $\mathcal{C}_v$ of v is progressively grown by adding to it at each time a vertex adjacent to one of the previously explored vertices and uncovering its remaining neighbors in G. In this process, $\mathcal{E}_t$ is the set of previously explored vertices and $\mathcal{A}_t$—the frontier of the process—is the set of vertices who are known to belong to $\mathcal{C}_v$ but whose full neighborhood is waiting to be uncovered. The rest of the vertices form the set $\mathcal{N}_t$. See Figure 6.3.

Let $A_t := |\mathcal{A}_t|$, $E_t := |\mathcal{E}_t|$, and $N_t := |\mathcal{N}_t|$. Note that (E_t)—not to be confused with the edge set—is non-decreasing while (N_t) is non-increasing. Let

$$\tau_0 := \inf\{t \geq 0 \,:\, A_t = 0\},$$

be the first time A_t is 0 (which by convention is $+\infty$ if there is no such t). The process is fixed for all $t > \tau_0$. Notice that $E_t = t$ for all $t \leq \tau_0$, as exactly one vertex is explored at each time until the set of active vertices is empty. The size of the connected component of v can be characterized as follows.

Lemma 6.2.1.

$$\tau_0 = |\mathcal{C}_v|.$$

Proof. Indeed a single vertex of $\mathcal{C}_v$ is explored at each time until all of $\mathcal{C}_v$ has been visited. At that point, $\mathcal{A}_t$ is empty. ■

Random-walk representation of a Galton-Watson tree

Let $(Z_i)_{i\geq 0}$ be a Galton-Watson branching process and let T be the corresponding Galton-Watson tree. We run the exploration process above on T started at the root 0. We will refer to the index i in Z_i as a "generation," and to the index t in the exploration process as "time"—they are *not* the same. Let $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t)$ and $A_t := |\mathcal{A}_t|$, $E_t := |\mathcal{E}_t|$, and $N_t := |\mathcal{N}_t|$ be as above. Let $(\mathcal{F}_t)$ be the corresponding filtration. Because we explore the vertices on first-come, first-served basis, we exhaust all vertices in generation i before considering vertices in generation $i+1$ (i.e., we perform breadth-first search).

The random-walk representation is the following. Observe that the process (A_t) admits a simple recursive form. We start with $A_0 := 1$. Then, conditioning on $\mathcal{F}_{t-1}$:

- If $A_{t-1} = 0$, the exploration process has finished its course and $A_t = 0$.
- Otherwise, (a) one active vertex becomes an explored vertex and (b) its offspring become active vertices. That is,

$$A_t = \begin{cases} A_{t-1} + \big(\underbrace{-1}_{\text{(a)}} + \underbrace{X_t}_{\text{(b)}}\big) & \text{if } t-1 < \tau_0, \\ 0 & \text{otherwise,} \end{cases}$$

  where X_t is distributed according to the offspring distribution.

We let $Y_t := X_t - 1$ and

$$S_t := 1 + \sum_{s=1}^{t} Y_s,$$

with $S_0 := 1$. Then

$$\tau_0 = \inf\{t \geq 0 \,:\, S_t = 0\},$$

and

$$(A_t) = (S_{t \wedge \tau_0}),$$

is a random walk started at 1 with i.i.d. increments (Y_t) stopped when it hits 0 for the first time.

We refer to

$$H = (X_1, \ldots, X_{\tau_0}),$$

as the *history* of the process (Z_i). Observe that, under breadth-first search, the process (Z_i) can be reconstructed from H: $Z_0 = 1$, $Z_1 = X_1$, $Z_2 = X_2 + \ldots + X_{Z_1+1}$, and so forth. (Exercise 6.5 asks for a general formula.) As a result, (Z_i) can be recovered from (S_t) as well. We call $(x_1, \ldots, x_t)$ a *valid history* if

history

valid history

$$1 + (x_1 - 1) + \cdots + (x_s - 1) > 0,$$

for all $s < t$ and

$$1 + (x_1 - 1) + \cdots + (x_t - 1) = 0.$$

Note that a valid history may have probability 0 under the offspring distribution.

6.2.2 Duality principle

The random-walk representation above is useful to prove the following duality principle.

Theorem 6.2.2 (Duality principle)**.** *Let (Z_i) be a branching process with offspring distribution $\{p_k\}_{k \geq 0}$ and extinction probability $\eta < 1$. Let (Z'_i) be a branching process with offspring distribution $\{p'_k\}_{k \geq 0}$ where*

$$p'_k = \eta^{k-1} p_k.$$

Then (Z_i) conditioned on extinction has the same distribution as (Z'_i), which is referred to as the dual branching process.

dual branching process

Let f be the probability generating function of the offspring distribution of (Z_i). Note that

$$\sum_{k \geq 0} p'_k = \sum_{k \geq 0} \eta^{k-1} p_k = \eta^{-1} f(\eta) = 1,$$

because η is a fixed point of f by Theorem 6.1.6. So $\{p'_k\}_{k \geq 0}$ is indeed a probability distribution. Note further that its expectation is

$$\sum_{k \geq 0} k p'_k = \sum_{k \geq 0} k \eta^{k-1} p_k = f'(\eta) < 1,$$

since, by Lemma 6.1.7, f' is strictly increasing, $f(\eta) = \eta < 1$ and $f(1) = 1$ (which would not be possible if $f'(\eta)$ were greater or equal to 1; see Figure 6.1 for an illustration). So the dual branching process is subcritical.

Proof of Theorem 6.2.2. We use the random-walk representation. Let $H = (X_1, \ldots, X_{\tau_0})$ and $H' = (X'_1, \ldots, X'_{\tau'_0})$ be the histories of (Z_i) and (Z'_i) respectively. In the case of extinction of (Z_i), the history H has finite length.

By definition of the conditional probability, for a valid history $(x_1, \ldots, x_t)$ with a finite t,

$$\mathbb{P}[H = (x_1, \ldots, x_t) \,|\, \tau_0 < +\infty] = \frac{\mathbb{P}[H = (x_1, \ldots, x_t)]}{\mathbb{P}[\tau_0 < +\infty]} = \eta^{-1} \prod_{s=1}^{t} p_{x_s}.$$

Because $(x_1 - 1) + \cdots + (x_t - 1) = -1$,

$$\eta^{-1} \prod_{s=1}^{t} p_{x_s} = \eta^{-1} \prod_{s=1}^{t} \eta^{1-x_s} p'_{x_s} = \prod_{s=1}^{t} p'_{x_s} = \mathbb{P}[H' = (x_1, \ldots, x_t)].$$

Since this is true for all valid histories and the processes can be recovered from their histories, we have proved the claim. ∎

Example 6.2.3 (Poisson branching process)**.** Let (Z_i) be a Galton-Watson branching process with offspring distribution $\mathrm{Poi}(\lambda)$ where $\lambda > 1$. Then the dual probability distribution is given by

$$p'_k = \eta^{k-1} p_k = \eta^{k-1} e^{-\lambda} \frac{\lambda^k}{k!} = \eta^{-1} e^{-\lambda} \frac{(\lambda\eta)^k}{k!},$$

where recall from Example 6.1.10 that $e^{-\lambda(1-\eta)} = \eta$, so

$$p'_k = e^{\lambda(1-\eta)} e^{-\lambda} \frac{(\lambda\eta)^k}{k!} = e^{-\lambda\eta} \frac{(\lambda\eta)^k}{k!}.$$

That is, the dual branching process has offspring distribution $\mathrm{Poi}(\lambda\eta)$. ◀

6.2.3 Hitting-time theorem

The random-walk representation also gives a formula for the distribution of the size of the progeny.

Law of total progeny The key is the following claim.

Lemma 6.2.4 (Total progeny and random-walk representation)**.** *Let W be the total progeny of the Galton-Watson branching process (Z_i). Then*

$$W = \tau_0.$$

Proof. Recall that

$$\tau_0 := \inf\{t \geq 0 \,:\, A_t = 0\}.$$

If the process does not go extinct, then $\tau_0 = +\infty$ as there are always more vertices to explore.

Suppose the process goes extinct and that $W = n$. Notice that $E_t = t$ for all $t \leq \tau_0$, as exactly one vertex is explored at each time until the set of active vertices is empty. Moreover, for all t, $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t)$ forms a partition of $[n]$ so

$$A_t + t + N_t = n, \qquad \forall t \leq \tau_0.$$

At $t = \tau_0$, $A_t = N_t = 0$ and we get

$$\tau_0 = n.$$

That proves the claim. ■

To compute the distribution of $W = \tau_0$, we use the following hitting-time theorem, which is proved later in this subsection.

Theorem 6.2.5 (Hitting-time theorem)**.** *Let (R_t) be a random walk started at 0 with i.i.d. increments (U_t) satisfying*

$$\mathbb{P}[U_t \leq 1] = 1.$$

Fix a positive integer ℓ. Let σ_ℓ be the first time t such that $R_t = \ell$. Then

$$\mathbb{P}[\sigma_\ell = t] = \frac{\ell}{t}\mathbb{P}[R_t = \ell].$$

Finally we get:

Theorem 6.2.6 (Law of total progeny)**.** *Let (Z_t) be a Galton-Watson branching process with total progeny W. In the random-walk representation of (Z_t),*

$$\mathbb{P}[W = t] = \frac{1}{t}\mathbb{P}[X_1 + \cdots + X_t = t - 1],$$

for all $t \geq 1$.

Proof. Recall that $Y_t := X_t - 1 \geq -1$ and

$$S_t = 1 + \sum_{s=1}^{t} Y_s,$$

with $S_0 = 1$, and that

$$\begin{aligned}\tau_0 &= \inf\{t \geq 0 \,:\, S_t = 0\}\\ &= \inf\{t \geq 0 \,:\, 1 + (X_1 - 1) + \cdots + (X_t - 1) = 0\}\\ &= \inf\{t \geq 0 \,:\, X_1 + \cdots + X_t = t - 1\}.\end{aligned}$$

Define $R_t := 1 - S_t$ and $U_t := -Y_t$ for all t. Then $R_0 := 0$,

$$\{X_1 + \cdots + X_t = t - 1\} = \{R_t = 1\},$$

and

$$\tau_0 = \inf\{t \geq 0 \,:\, R_t = 1\}.$$

The process (R_t) satisfies the assumptions of the hitting-time theorem (Theorem 6.2.5) with $\ell = 1$ and $\sigma_\ell = \tau_0 = W$. Applying the theorem gives the claim. ■

Example 6.2.7 (Poisson branching process (continued))**.** Let (Z_i) be a Galton-Watson branching process with offspring distribution $\mathrm{Poi}(\lambda)$ where $\lambda > 0$. Let W be its total progeny. By the hitting-time theorem, for $t \geq 1$,

$$\begin{aligned}\mathbb{P}[W = t] &= \frac{1}{t}\mathbb{P}[X_1 + \cdots + X_t = t - 1]\\ &= \frac{1}{t} e^{-\lambda t} \frac{(\lambda t)^{t-1}}{(t-1)!}\\ &= e^{-\lambda t} \frac{(\lambda t)^{t-1}}{t!},\end{aligned}$$

where we used that a sum of independent Poisson is Poisson. ◀

Spitzer's combinatorial lemma Before proving the hitting-time theorem, we begin with a combinatorial lemma of independent interest. Let $u_1, \ldots, u_t \in \mathbb{R}$ and define $r_0 := 0$ and $r_j := u_1 + \cdots + u_j$ for $1 \leq j \leq t$. We say that j is a *ladder index* if $r_j > r_0 \vee \cdots \vee r_{j-1}$. Consider the cyclic permutations of $\boldsymbol{u} = (u_1, \ldots, u_t)$, that is, $\boldsymbol{u}^{(0)} = \boldsymbol{u}$, $\boldsymbol{u}^{(1)} = (u_2, \ldots, u_t, u_1)$, ..., $\boldsymbol{u}^{(t-1)} = (u_t, u_1, \ldots, u_{t-1})$. Define the corresponding partial sums $r_j^{(\beta)} := u_1^{(\beta)} + \cdots + u_j^{(\beta)}$ for $j = 1, \ldots, t$ and $\beta = 0, \ldots, t-1$.

Lemma 6.2.8 (Spitzer's combinatorial lemma)**.** *Assume $r_t > 0$. Let ℓ be the number of cyclic permutations such that t is a ladder index. Then $\ell \geq 1$ and each such cyclic permutation has exactly ℓ ladder indices.*

Proof. We will need the following observation

$$
\begin{aligned}
&(r_1^{(\beta)}, \ldots, r_t^{(\beta)}) \\
&\qquad = (r_{\beta+1} - r_\beta, r_{\beta+2} - r_\beta, \ldots, r_t - r_\beta, \\
&\qquad\qquad [r_t - r_\beta] + r_1, [r_t - r_\beta] + r_2, \ldots, [r_t - r_\beta] + r_\beta) \\
&\qquad = (r_{\beta+1} - r_\beta, r_{\beta+2} - r_\beta, \ldots, r_t - r_\beta, \\
&\qquad\qquad r_t - [r_\beta - r_1], r_t - [r_\beta - r_2], \ldots, r_t - [r_\beta - r_{\beta-1}], r_t).
\end{aligned} \tag{6.2.1}
$$

We first show that $\ell \geq 1$, that is, there is at least one cyclic permutation where t is a ladder index. Let $\beta \geq 1$ be the smallest index achieving the maximum of $r_1, \ldots, r_t$, that is,

$$r_\beta > r_1 \vee \cdots \vee r_{\beta-1} \quad \text{and} \quad r_\beta \geq r_{\beta+1} \vee \cdots \vee r_t.$$

Moreover, $r_t > 0 = r_0$ by assumption. Hence,

$$r_{\beta+j} - r_\beta \leq 0 < r_t, \qquad \forall j = 1, \ldots, t - \beta,$$

and

$$r_t - [r_\beta - r_j] < r_t, \qquad \forall j = 1, \ldots, \beta - 1.$$

From (6.2.1), in $\boldsymbol{u}^{(\beta)}$, t is a ladder index.

For the second claim, since $\ell \geq 1$, we can assume without loss of generality that $\boldsymbol{u}$ is such that t is a ladder index. (Note that $r_t^{(\beta)} = r_t$ for all β.) We show that β is a ladder index in $\boldsymbol{u}$ if and only if t is a ladder index in $\boldsymbol{u}^{(\beta)}$. That does indeed imply the claim as there are ℓ cyclic permutations where t is a ladder index by assumption. We use (6.2.1) again. Observe that β is a ladder index in $\boldsymbol{u}$ if and only if

$$r_\beta > r_0 \vee \cdots \vee r_{\beta-1},$$

which holds if and only if

$$r_\beta > r_0 = 0 \quad \text{and} \quad r_t - [r_\beta - r_j] < r_t, \ \forall j = 1, \ldots, \beta - 1. \tag{6.2.2}$$

Moreover, because $r_t > r_j$ for all j by the assumption that t is ladder index, the last display holds if and only if

$$r_{\beta+j} - r_\beta < r_t, \ \forall j = 1, \ldots, t - \beta, \tag{6.2.3}$$

and

$$r_t - [r_\beta - r_j] < r_t,\ \forall j = 1, \ldots, \beta - 1, \tag{6.2.4}$$

that is, if and only if t is a ladder index in $\boldsymbol{u}^{(\beta)}$ by (6.2.1). Indeed, the second condition (i.e., (6.2.4)) is intact from (6.2.2), while the first one (i.e., (6.2.3)) can be rewritten as $r_\beta > -(r_t - r_{\beta+j})$ where the right-hand side is < 0 for $j = 1, \ldots, t - \beta - 1$ and $= 0$ for $j = t - \beta$. ■

Proof of hitting-time theorem We are now ready to prove the hitting-time theorem. We only handle the case $\ell = 1$ (which is the one we used for the law of the total progeny). Exercise 6.6 asks for the full proof.

Proof of Theorem 6.2.5. Recall that $R_t = \sum_{s=1}^t U_s$ and $\sigma_1 = \inf\{j \geq 0 : R_j = 1\}$. By the assumption that $U_s \leq 1$ almost surely for all s,

$$\{\sigma_1 = t\} = \{t \text{ is the first ladder index in } R_1, \ldots, R_t\}.$$

By symmetry, for all $\beta = 0, \ldots, t-1$,

$$\begin{aligned}&\mathbb{P}[t \text{ is the first ladder index in } R_1, \ldots, R_t]\\ &\qquad = \mathbb{P}[t \text{ is the first ladder index in } R_1^{(\beta)}, \ldots, R_t^{(\beta)}].\end{aligned}$$

Let $\mathcal{E}_\beta$ be the event on the last line. Then,

$$\mathbb{P}[\sigma_1 = t] = \mathbb{E}[\mathbf{1}_{\mathcal{E}_0}] = \frac{1}{t}\mathbb{E}\left[\sum_{\beta=0}^{t-1} \mathbf{1}_{\mathcal{E}_\beta}\right].$$

By Spitzer's combinatorial lemma (Lemma 6.2.8), there is at most one cyclic permutation where t is the *first* ladder index. (There is at least one cyclic permutation where t is a ladder index—but it may not be the *first* one, that is, there may be multiple ladder indices.) In particular, $\sum_{\beta=0}^{t-1} \mathbf{1}_{\mathcal{E}_\beta} \in \{0, 1\}$. So, by the previous display,

$$\mathbb{P}[\sigma_1 = t] = \frac{1}{t}\mathbb{P}\left[\cup_{\beta=0}^{t-1}\mathcal{E}_\beta\right].$$

Finally we claim that $\{R_t = 1\} = \cup_{\beta=0}^{t-1}\mathcal{E}_\beta$. Indeed, because $R_0 = 0$ and $U_s \leq 1$ for all s, the partial sum at the j-th ladder index must take value j. So the event $\cup_{\beta=0}^{t-1}\mathcal{E}_\beta$ implies $\{R_t = 1\}$ since the last partial sum of all cyclic permutations is R_t. Similarly, because there is at least one cyclic permutation such that t is a ladder index, the event $\{R_t = 1\}$ implies that t is in fact the first ladder index in that cyclic permutation, and therefore it implies $\cup_{\beta=1}^{t}\mathcal{E}_\beta$. Hence,

$$\mathbb{P}[\sigma_1 = t] = \frac{1}{t}\mathbb{P}\left[R_t = 1\right],$$

which concludes the proof (for the case $\ell = 1$). ■

6.2.4 ▷ *Percolation: critical exponents on the infinite b-ary tree*

In this section, we use branching processes to study bond percolation (Definition 1.2.1) on the infinite b-ary tree $\widehat{\mathbb{T}}_b$ and derive explicit expressions for quantities of interest. Close to the critical value, we prove the existence of "critical exponents." We illustrate the use of both the duality principle (Theorem 6.2.2) and the hitting-time theorem (Theorem 6.2.6).

Critical value We denote the root by 0. Similarly to what we did in Section 6.1.3, we think of the open cluster of the root, $\mathcal{C}_0$, as the progeny of a branching process as follows. Denote by ∂_n the n-th level of $\widehat{\mathbb{T}}_b$, that is, the vertices of $\widehat{\mathbb{T}}_b$ at graph distance n from the root. In the branching process interpretation, we think of the immediate descendants in $\mathcal{C}_0$ of a vertex v as the offspring of v. By construction, v has at most b children, independently of all other vertices in the same generation. In this branching process, the offspring distribution $\{q_k\}_{k=0}^{b}$ is binomial with parameters b and p; $Z_n := |\mathcal{C}_0 \cap \partial_n|$ represents the size of the progeny at generation n; and $W := |\mathcal{C}_0|$ is the total progeny of the process. In particular $|\mathcal{C}_0| < +\infty$ if and only if the process goes extinct. Because the mean number of offspring is bp, by Theorem 6.1.6, this leads immediately to a second proof of (a rooted variant of) Claim 2.3.9:

Claim 6.2.9.

$$p_c\left(\widehat{\mathbb{T}}_b\right) = \frac{1}{b}.$$

Percolation function The generating function of the offspring distribution is $\phi(s) := ((1-p)+ps)^b$. So, by Theorems 6.1.5 and 6.1.6, the percolation function

$$\theta(p) = \mathbb{P}_p[|\mathcal{C}_0| = +\infty],$$

is 0 on $[0, 1/b]$, while on $(1/b, 1]$ the quantity $\eta(p) := 1 - \theta(p)$ is the unique solution in $[0, 1)$ of the fixed point equation

$$s = ((1-p)+ps)^b. \tag{6.2.5}$$

For $b = 2$, for instance, we can compute the fixed point explicitly by noting that

$$\begin{aligned} 0 &= ((1-p)+ps)^2 - s \\ &= p^2 s^2 + [2p(1-p) - 1]s + (1-p)^2, \end{aligned}$$

whose solution for $p \in (1/2, 1]$ is

$$\begin{aligned} s^* &= \frac{-[2p(1-p)-1] \pm \sqrt{[2p(1-p)-1]^2 - 4p^2(1-p)^2}}{2p^2} \\ &= \frac{-[2p(1-p)-1] \pm \sqrt{1-4p(1-p)}}{2p^2} \\ &= \frac{-[2p(1-p)-1] \pm (2p-1)}{2p^2} \\ &= \frac{2p^2 + [(1-2p) \pm (2p-1)]}{2p^2}. \end{aligned}$$

So, rejecting the fixed point 1,

$$\theta(p) = 1 - \frac{2p^2 + 2(1-2p)}{2p^2} = \frac{2p-1}{p^2}.$$

We have proved:

Claim 6.2.10. *For $b = 2$,*

$$\theta(p) = \begin{cases} 0, & 0 \leq p \leq \frac{1}{2}, \\ \frac{2(p-\frac{1}{2})}{p^2} & \frac{1}{2} < p \leq 1. \end{cases}$$

Since $\eta(p) = (1 - \theta(p))$, we have in that case

$$\eta(p) = \begin{cases} 1, & 0 \leq p \leq \frac{1}{2}, \\ \frac{(1-p)^2}{p^2} & \frac{1}{2} < p \leq 1. \end{cases}$$

Conditioning on a finite cluster The expected size of the population at generation n is $(bp)^n$ by Lemma 6.1.2, so for $p \in [0, \frac{1}{b})$

$$\mathbb{E}_p|\mathcal{C}_0| = \sum_{n \geq 0} (bp)^n = \frac{1}{1-bp}. \tag{6.2.6}$$

For $p \in (\frac{1}{b}, 1)$, the total progeny is infinite with positive probability (and in particular the expectation is infinite), but we can compute the expected cluster size *on the event that* $|\mathcal{C}_0| < +\infty$. For this purpose we use the duality principle.

Recall that $q_k = \binom{b}{k} p^k (1-p)^{b-k}$, $k = 0, \ldots, b$, is the offspring distribution. For $0 \le k \le b$, we let the dual offspring distribution be

$$\begin{aligned}
\hat{q}_k &:= [\eta(p)]^{k-1} q_k \\
&= [\eta(p)]^{k-1} \binom{b}{k} p^k (1-p)^{b-k} \\
&= \frac{[\eta(p)]^k}{((1-p) + p\,\eta(p))^b} \binom{b}{k} p^k (1-p)^{b-k} \\
&= \binom{b}{k} \left(\frac{p\,\eta(p)}{(1-p) + p\,\eta(p)} \right)^k \left(\frac{1-p}{(1-p) + p\,\eta(p)} \right)^{b-k} \\
&=: \binom{b}{k} \hat{p}^k \, (1-\hat{p})^{b-k},
\end{aligned}$$

where we used (6.2.5) and implicitly defined the dual density

$$\hat{p} := \frac{p\,\eta(p)}{(1-p) + p\,\eta(p)}. \tag{6.2.7}$$

In particular $\{\hat{q}_k\}$ is a probability distribution as expected under Theorem 6.2.2—it is in fact binomial with parameters b and $\hat{p}$. Summarizing the implications of Theorem 6.2.2:

Claim 6.2.11. *Conditioned on $|\mathcal{C}_0| < +\infty$, (supercritical) percolation on $\widehat{\mathbb{T}}_b$ with density $p \in (\frac{1}{b}, 1)$ has the same distribution as (subcritical) percolation on $\widehat{\mathbb{T}}_b$ with density defined by* (6.2.7).

Hence, using (6.2.6) with both p and $\hat{p}$ as well as the fact that $\mathbb{P}_p[|\mathcal{C}_0| < +\infty] = \eta(p)$, we have the following.

Claim 6.2.12.

$$\chi^{\mathrm{f}}(p) := \mathbb{E}_p\left[|\mathcal{C}_0| \mathbf{1}_{\{|\mathcal{C}_0| < +\infty\}}\right] = \begin{cases} \frac{1}{1-bp}, & p \in [0, \frac{1}{b}), \\ \frac{\eta(p)}{1-b\hat{p}}, & p \in (\frac{1}{b}, 1). \end{cases}$$

For $b = 2$, $\eta(p) = 1 - \theta(p) = \left(\frac{1-p}{p}\right)^2$ so

$$\hat{p} = \frac{p\left(\frac{1-p}{p}\right)^2}{(1-p) + p\left(\frac{1-p}{p}\right)^2} = \frac{(1-p)^2}{p(1-p) + (1-p)^2} = 1 - p,$$

and

Claim 6.2.13. *For $b = 2$,*

$$\chi^{\mathrm{f}}(p) = \begin{cases} \frac{1/2}{\frac{1}{2}-p}, & p \in [0, \frac{1}{2}), \\ \frac{\frac{1}{2}\left(\frac{1-p}{p}\right)^2}{p-\frac{1}{2}}, & p \in (\frac{1}{2}, 1). \end{cases}$$

Distribution of the open cluster size In fact the hitting-time theorem gives an explicit formula for the distribution of $|\mathcal{C}_0|$. Namely, recall that $|\mathcal{C}_0| \stackrel{\mathrm{d}}{=} \tau_0$ where

$$\tau_0 = \inf\{t \geq 0 \,:\, S_t = 0\},$$

for $S_t = \sum_{\ell \leq t} X_\ell - (t-1)$ where $S_0 = 1$ and the X_ℓs are i.i.d. binomial with parameters b and p. By Theorem 6.2.6,

$$\mathbb{P}[\tau_0 = t] = \frac{1}{t}\,\mathbb{P}[S_t = 0],$$

and we have

$$\mathbb{P}_p[|\mathcal{C}_0| = \ell] = \frac{1}{\ell}\,\mathbb{P}\left[\sum_{i \leq \ell} X_\ell = \ell - 1\right] = \frac{1}{\ell}\binom{b\ell}{\ell-1} p^{\ell-1}(1-p)^{b\ell-(\ell-1)}, \tag{6.2.8}$$

where we used that a sum of independent binomials with the same p is itself binomial. In particular at criticality (where $|\mathcal{C}_0| < +\infty$ almost surely; see Claim 3.1.52), using Stirling's formula (see Appendix A) it can be checked that

$$\mathbb{P}_{p_c}[|\mathcal{C}_0| = \ell] \sim \frac{1}{\ell}\frac{1}{\sqrt{2\pi p_c(1-p_c)b\ell}} = \frac{1}{\sqrt{2\pi(1-p_c)\ell^3}},$$

as $\ell \to +\infty$.

Critical exponents Close to criticality, physicists predict that many quantities behave according to power laws of the form $|p - p_c|^\beta$, where the exponent is referred to as a *critical exponent*. The critical exponents are believed to satisfy certain "universality" properties. But even proving the existence of such exponents in general remains a major open problem. On trees, though, we can simply read off the critical exponents from the above formulas. For $b = 2$, Claims 6.2.10 and 6.2.13 imply for instance that, as $p \to p_c$,

critical exponent

$$\theta(p) \sim 8(p - p_c)\mathbf{1}_{\{p>1/2\}},$$

and

$$\chi^{\mathrm{f}}(p) \sim \frac{1}{2}|p - p_{\mathrm{c}}|^{-1}.$$

In fact, as can be seen from Claim 6.2.12, the critical exponent of $\chi^{\mathrm{f}}(p)$ does not depend on b. The same holds for $\theta(p)$ (see Exercise 6.9). Using (6.2.8), the higher moments of $|\mathcal{C}_0|$ can also be studied around criticality (see Exercise 6.10).

6.3 Applications

We develop two applications of branching processes in discrete probability. First, we prove a result about the height of a random binary search tree. Then we describe a phase transition in an Ising model on a tree with applications to evolutionary biology. In the next section, we also use branching processes to study the phase transition of Erdős-Rényi random graph model.

6.3.1 ▷ *Probabilistic analysis of algorithms: binary search tree*

A *binary search tree (BST)* is a commonly used data structure in computer science. It consists of a rooted binary tree $T_n = (V_n, E_n)$. Each vertex has a "left" and "right" subtree (possibly empty) and a "key" from an input sequence $x_1, \ldots, x_n \in \mathbb{R}$ (which we assume are distinct) that satisfies the BST property: the key at vertex $v \in V$ is greater than all keys in the left subtree below it and less than all keys in the right subtree below it. Such a data structure can be used for a variety of algorithmic tasks, such as searching for keys or sorting them.

binary search tree

The tree is constructed recursively as follows. Assume that the keys $x_1, \ldots, x_i$ have already been inserted and that the current tree T_i satisfies the BST property. To insert x_{i+1}:

- start at the root;
- if the root's key is strictly larger than x_{i+1}, then move to its left descendant, otherwise move to its right descendant;
- if such a descendant does not exist then create it and assign it x_{i+1} as its key;
- otherwise repeat.

Inserting keys (and other operations such as deleting keys, which we do not describe) takes time proportional to the height H_n of the tree T_n, that is, the length of the *longest* path from the root to a leaf. While, in general, the height can be as large as n (if keys are inserted in order for instance), the typical behavior can be much smaller.

Indeed, here we study the case of n keys $X_1, \ldots, X_n$ i.i.d. from a continuous distribution on $\mathbb{R}$ and establish a much better behavior for the random height. Let γ be the unique solution greater than 1 of

$$\left(\frac{1}{e}\right)\left(\frac{2e}{\gamma}\right)^{\gamma} = 1. \tag{6.3.1}$$

See Exercise 6.14 for a proof that γ is well-defined and that the left-hand side is strictly decreasing at γ. We show:

Claim 6.3.1. $H_n / \log n \to_{\mathrm{p}} \gamma$ *as* $n \to \infty$.

Alternative representation of the height

The main idea of the proof is to relate the height H_n of the tree T_n to a product of independent uniform random variables. We make a series of observations about the structure of the tree. First:

> **Observation 1.** Keys affect the construction of the binary search tree *only through their ordering*. Let σ be the corresponding (random) permutation, that is,
>
> $$X_{\sigma(1)} < X_{\sigma(2)} < \cdots < X_{\sigma(n)}.$$
>
> Let $t[\sigma]$ be the binary search tree generated by the permutation σ.

Second, by symmetry:

> **Observation 2.** The permutation σ is *uniformly distributed*.

Denote by S_v the size of the subtree rooted at v (including v itself) in $t[\sigma]$. At the root ρ, we have $S_\rho = n$. What is the size of the subtree rooted at the left descendant ρ' of ρ? Eventually all keys with a rank lower than $\sigma^{-1}(1)$, that is, those keys with indices in $\left\{\sigma(i) : i < \sigma^{-1}(1)\right\}$, find their way into the left subtree of the root. In other words,

$$S_{\rho'} = \sigma^{-1}(1) - 1.$$

Similarly, denoting by ρ'' the right descendant of ρ, we see that

$$S_{\rho''} = n - \sigma^{-1}(1).$$

We refer to $\sigma^{-1}(1)$ as the *rank* of the root. By Observation 2:

> **Observation 3.** The rank $\sigma^{-1}(1)$ of the root is *uniformly distributed* in $[n]$. Moreover it is identically distributed to $\lfloor S_\rho W_\rho \rfloor + 1$, where W_ρ is uniform in $[0, 1]$.

The second part of this last observation can be checked by direct computation. Rename $X'_1, \ldots, X'_{S_{\rho'}}$ the keys in the subtree rooted at ρ' *in the order that they are inserted* and let σ' be the (random) permutation corresponding to their ordering, that is,

$$X'_{\sigma'(1)} < X'_{\sigma'(2)} < \cdots < X'_{\sigma'(S_{\rho'})}.$$

Define σ'' similarly for ρ''. Again by symmetry:

> **Observation 4.** Conditioned on $\sigma^{-1}(1)$ (and therefore on $S_{\rho'}$ and $S_{\rho''}$), the permutations σ' and σ'' are *independent* and *uniformly distributed*.

Finally, recursively:

> **Observation 5.** The binary search tree $t[\sigma]$ is obtained by appending the left subtree $t[\sigma']$ and right subtree $t[\sigma'']$ to the root ρ.

If $S_{\rho'} = 0$, then $t[\sigma'] = \emptyset$ (and there is in fact no ρ'); while, if $S_{\rho'} = 1$, the tree $t[\sigma']$ is comprised of the single vertex ρ'. Similarly for σ''. Hence this recursive process stops whenever we reach a vertex v with $S_v \in \{0, 1\}$. But it will be convenient to extend it indefinitely to produce an *infinite binary tree* $\mathcal{T} = \widehat{\mathbb{T}}_2$, where all additional vertices v are assigned $S_v = 0$.

The upshot of all these observations is that we obtain the following alternative characterization of the height H_n:

- assign an independent $U[0,1]$ (i.e., uniform in $[0,1]$) random variable W_v to each vertex v in the infinite binary tree $\mathcal{T}$;
- at the root ρ, set
$$S_\rho = n;$$
- then recursively from the root down, set
$$S_{v'} := \lfloor S_v W_v \rfloor \quad \text{and} \quad S_{v''} := \lfloor S_v(1 - W_v) \rfloor, \tag{6.3.2}$$
where v' and v'' are the left and right descendants of v in $\mathcal{T}$.

It can be checked that $S_{v'} + S_{v''} = S_v - 1$ almost surely, provided $S_v \geq 1$ (see Exercise 6.15). Moreover notice that, when $S_v = 1$, then $S_{v'} = S_{v''} = 0$ almost surely; while, if $S_v = 0$, then $S_{v'} = S_{v''} = 0$. Finally the height H_n is the highest level containing a vertex with subtree size at least 1, that is,

$$H_n = \sup\{h \,:\, \exists v \in \mathcal{L}_h,\; S_v \geq 1\}, \tag{6.3.3}$$

where $\mathcal{L}_h$ is the set of vertices of $\mathcal{T}$ at graph distance h from the root.

Key technical bound

Because $W \sim U[0,1]$ implies also that $(1-W) \sim U[0,1]$, we immediately get from (6.3.2) that:

Lemma 6.3.2 (Distribution of subtree size)**.** *Let v be a vertex at topological distance ℓ from the root of $\mathcal{T}$. Let $U_1, \ldots, U_\ell$ be i.i.d. $U[0,1]$. Then we have the equality in distribution*

$$S_v \stackrel{\mathrm{d}}{=} \lfloor \cdots \lfloor \lfloor n U_1 \rfloor U_2 \rfloor \cdots U_\ell \rfloor.$$

From Lemma 6.3.2 and the characterization of the height in (6.3.3), we need to control how fast products of independent uniforms decrease. But that is only half of the story: the number of paths of length ℓ from the root *grows exponentially with ℓ*. The following lemma, which takes both effects into account, will play a key role in the analysis. It also explains the definition of γ in (6.3.1). Note that we ignore—for the time being—the repeated rounding in Lemma 6.3.2; it will turn out to have a minor effect.

Lemma 6.3.3 (Product of uniforms)**.** *Let $U_1, U_2, \ldots$ be i.i.d. $U[0,1]$. Then*

$$\lim_{\ell \to +\infty} 2^\ell \, \mathbb{P}\left[U_1 \cdots U_\ell \geq e^{-\ell/c}\right] = \begin{cases} +\infty & \text{if } c < \gamma, \\ 0 & \text{if } c > \gamma. \end{cases}$$

Proof. Taking logarithms turns the product on the left-hand side into a sum of i.i.d. random variables

$$2^\ell \, \mathbb{P}\left[U_1 \cdots U_\ell \geq e^{-\ell/c}\right] = 2^\ell \, \mathbb{P}\left[\sum_{i=1}^{\ell} (-\log U_i) \leq \ell/c\right]. \tag{6.3.4}$$

Now it is elementary to bound the right-hand side.

Lemma 6.3.4 (A tail bound)**.** *Let $U_1, \ldots, U_\ell$ be i.i.d. $U[0,1]$. Then for any $y > 0$*

$$\frac{y^\ell e^{-y}}{\ell!} \leq \mathbb{P}\left[\sum_{i=1}^{\ell} (-\log U_i) \leq y\right] \leq \frac{y^\ell e^{-y}}{\ell!} \left(\frac{1}{1 - \frac{y}{\ell+1}}\right). \tag{6.3.5}$$

Proof. We prove a more general claim, specifically

$$\mathbb{P}\left[\sum_{i=1}^{\ell} (-\log U_i) \leq y\right] = e^{-y} \left\{\sum_{i=\ell}^{+\infty} \frac{y^i}{i!}\right\},$$

from which (6.3.5) follows: the lower bound is obtained by keeping only the first term in the sum; the upper bound is obtained by factoring out $y^\ell e^{-y}/\ell!$ and relating the remaining sum to a geometric series.

So it remains to prove the general claim. First note that $-\log U_1$ is exponentially distributed. Indeed, for any $y \geq 0$,

$$\mathbb{P}[-\log U_1 > y] = \mathbb{P}\left[U_1 < e^{-y}\right] = e^{-y}.$$

So

$$\mathbb{P}[-\log U_1 \leq y] = 1 - e^{-y} = e^{-y}\left\{\sum_{i=0}^{+\infty} \frac{y^i}{i!} - 1\right\} = e^{-y}\left\{\sum_{i=1}^{+\infty} \frac{y^i}{i!}\right\},$$

as claimed in the base case $\ell = 1$.

Proceeding by induction, suppose the claim holds up to $\ell - 1$. Then

$$\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^{\ell}(-\log U_i) > y\right] &= \int_0^{+\infty} e^{-z}\,\mathbb{P}\left[\sum_{i=1}^{\ell-1}(-\log U_i) > y - z\right] \mathrm{d}z \\
&= e^{-y} + \int_0^{y} e^{-z}\,\mathbb{P}\left[\sum_{i=1}^{\ell-1}(-\log U_i) > y - z\right] \mathrm{d}z \\
&= e^{-y} + \int_0^{y} e^{-z} e^{-(y-z)} \left\{\sum_{i=0}^{\ell-2} \frac{(y-z)^i}{i!}\right\} \mathrm{d}z \\
&= e^{-y} + e^{-y} \sum_{i=0}^{\ell-2} \frac{y^{i+1}}{i!(i+1)} \\
&= e^{-y} \sum_{j=0}^{\ell-1} \frac{y^j}{j!}.
\end{aligned}$$

That proves the claim. ■

We return to the proof of Lemma 6.3.3. Plugging (6.3.5) into (6.3.4), we get

$$\begin{aligned}
\frac{2^\ell (\ell/c)^\ell e^{-(\ell/c)}}{\ell!} &\leq 2^\ell\,\mathbb{P}\left[U_1 \cdots U_\ell \geq e^{-\ell/c}\right] \\
&\leq \frac{2^\ell (\ell/c)^\ell e^{-(\ell/c)}}{\ell!} \left(\frac{1}{1 - \frac{(\ell/c)}{\ell+1}}\right). \qquad (6.3.6)
\end{aligned}$$

As $\ell \to +\infty$,

$$\frac{1}{1 - \frac{(\ell/c)}{\ell+1}} \to \frac{1}{1 - \frac{1}{c}}, \qquad (6.3.7)$$

which is positive when $c > 1$. We will use the standard bound (see Exercise 1.3 for a proof)

$$\frac{\ell^\ell}{e^{\ell-1}} \leq \ell! \leq \frac{\ell^{\ell+1}}{e^{\ell-1}}.$$

It implies immediately that

$$\frac{2^\ell(\ell/c)^\ell e^{-(\ell/c)} e^{\ell-1}}{\ell^{\ell+1}} \leq \frac{2^\ell(\ell/c)^\ell e^{-(\ell/c)}}{\ell!} \leq \frac{2^\ell(\ell/c)^\ell e^{-(\ell/c)} e^{\ell-1}}{\ell^\ell},$$

which after simplifying gives

$$(e\ell)^{-1}\left[\left(\frac{1}{e}\right)\left(\frac{2e}{c}\right)^c\right]^{\ell/c} \leq 2^\ell\frac{(\ell/c)^\ell e^{-(\ell/c)}}{\ell!} \leq e^{-1}\left[\left(\frac{1}{e}\right)\left(\frac{2e}{c}\right)^c\right]^{\ell/c}. \quad (6.3.8)$$

By (6.3.1) and the remark following it, the expression in square brackets is > 1 or < 1 depending on whether $c < \gamma$ or $c > \gamma$. Combining (6.3.6), (6.3.7) and (6.3.8) and taking a limit as $\ell \to +\infty$ gives the claim. ■

As an immediate consequence of Lemma 6.3.3, we bound the height from above. Fix any $\varepsilon > 0$ and let $h := (\gamma + \varepsilon)\log n$. We use a union bound as follows

$$\mathbb{P}[H_n \geq h] = \mathbb{P}\left[\bigcup_{v\in\mathcal{L}_h}\{S_v \geq 1\}\right] \leq \sum_{v\in\mathcal{L}_h}\mathbb{P}[S_v \geq 1] = 2^h\,\mathbb{P}[S_v \geq 1], \quad (6.3.9)$$

for any $v \in \mathcal{L}_h$, where the first equality follows from (6.3.3). Since

$$\lfloor\cdots\lfloor\lfloor nU_1\rfloor U_2\rfloor\cdots U_h\rfloor \leq nU_1U_2\cdots U_h,$$

Lemmas 6.3.2 and 6.3.3 imply that

$$\begin{aligned} 2^h\,\mathbb{P}[S_v \geq 1] &\leq 2^h\,\mathbb{P}[nU_1U_2\cdots U_h \geq 1] \\ &= 2^h\,\mathbb{P}\left[U_1U_2\cdots U_h \geq e^{-h/(\gamma+\varepsilon)}\right] \\ &\to 0, \end{aligned} \quad (6.3.10)$$

as $h \to +\infty$. From (6.3.9) and (6.3.10), we obtain finally that for any $\varepsilon > 0$

$$\mathbb{P}[H_n/\log n \geq \gamma + \varepsilon] \to 0,$$

as $n \to +\infty$, which establishes one direction of Claim 6.3.1.

Lower bounding the height: a branching process

Establishing the other direction is where branching processes enter the scene. We will need some additional notation. Fix $c < \gamma$ and let ℓ be a positive integer that will be set later on. For any pair of vertex $v, w \in \mathcal{T}$ with w a descendant of v, let $\mathcal{Q}[v, w]$ be the set of vertices on the path between v and w, including v but excluding w. Further, recalling (6.3.2), define

$$\mathcal{U}[v, w] = \prod_{z \in \mathcal{Q}[u,v]} U_z^{v,w},$$

where $U_z^{v,w} = W_z$ (respectively $1 - W_z$) if the path from v to w takes the left (respectively right) edge upon exiting z. Denote by $\mathcal{L}_\ell[v]$ the set of descendant vertices of v in $\mathcal{T}$ at graph distance ℓ from v and consider the random subset

$$\mathcal{L}_\ell^*[v] = \left\{ w \in \mathcal{L}_\ell[v] \,:\, \mathcal{U}[v, w] \geq e^{-\ell/c} \right\}.$$

Fix a vertex $u \in \mathcal{T}$. We define the following Galton-Watson branching process.

- Initialize $Z_0^{u,\ell} := 1$ and $u_{0,1} := u$.
- For $t \geq 1$, set

$$Z_t^{u,\ell} = \sum_{r=1}^{Z_{t-1}^{u,\ell}} |\mathcal{L}_\ell^*[u_{t-1,r}]|,$$

  and let $u_{t,1}, \ldots, u_{t,Z_t^{u,\ell}}$ be the vertices in $\cup_{r=1}^{Z_{t-1}^{u,\ell}} \mathcal{L}_\ell^*[u_{t-1,r}]$ from left to right.

In words, $Z_1^{u,\ell}$ counts the number of vertices ℓ levels below u whose subtree sizes (ignoring rounding) have not decreased "too much" compared to that of u (in the sense of Lemma 6.3.3). We let such vertices (if any) be $u_{1,1}, \ldots, u_{1,Z_1^{u,\ell}}$. Similarly, $Z_2^{u,\ell}$ counts the same quantity over all vertices ℓ levels below the vertices $u_{1,1}, \ldots, u_{1,Z_1^{u,\ell}}$, and so forth.

Because the W_vs are i.i.d., this process is indeed a Galton-Watson branching process. The expectation of the offspring distribution (which by symmetry does not depend on the choice of u) is

$$m = \mathbb{E}\left[Z_1^{u,\ell}\right] = 2^\ell \, \mathbb{P}\left[U_1 \cdots U_\ell \geq e^{-\ell/c}\right],$$

where we used the notation of Lemma 6.3.3. By that lemma, we can choose ℓ large enough that $m > 1$. Fix such an ℓ for the rest of the proof. In that case, by Theorem 6.1.6, the process survives with probability $1 - \eta$ for some $0 \leq \eta < 1$.

The relevance of this observation can be seen from taking $u = \rho$.

Claim 6.3.5. *Let $c' < c$. Conditioned on survival of $(Z_t^{\rho,\ell})$, for n large enough $H_n \geq c' \log n - \theta_n \ell$ almost surely for some $\theta_n \in [0,1)$.*

Proof. To account for the rounding, we will need the inequality

$$\lfloor \cdots \lfloor \lfloor nU_1 \rfloor U_2 \rfloor \cdots U_s \rfloor \geq nU_1 U_2 \cdots U_s - s, \tag{6.3.11}$$

which holds for all $n, s \geq 1$, as can be checked by induction. Write $s = k\ell$ for some positive integer k to be determined. Conditioned on survival of $(Z_t^{\rho,\ell})$, the population at generation k satisfies

$$Z_k^{\rho,\ell} \geq 1,$$

which implies that, for some $v^* \in \mathcal{L}_s[\rho]$, it holds that

$$n\mathcal{U}[\rho, v^*] \geq n(e^{-\ell/c})^k.$$

Now take $s = c' \log n - \theta_n \ell$ with $c' < c$ and $\theta_n \in [0,1)$ such that s is a multiple of ℓ. Then

$$\begin{aligned} n(e^{-\ell/c})^k = n(e^{-s/c}) = n(n^{-c'/c} e^{-\theta_n \ell/c}) &= n^{1-c'/c} e^{-\theta_n \ell/c} \\ &\geq c' \log n - \theta_n \ell + 1 = s + 1, \end{aligned} \tag{6.3.12}$$

for all n large enough, where we used that $1 - c'/c > 0$, $\theta_n \in [0,1)$ and ℓ is fixed. So, using the characterization of the height in (6.3.2) and (6.3.3) together with inequality (6.3.11), we derive

$$S_{v^*} \geq n\mathcal{U}[\rho, v^*] - s \geq n(e^{-\ell/c})^k - s \geq 1. \tag{6.3.13}$$

That is, $H_n \geq c' \log n - \theta_n \ell$. ∎

But this is not quite what we want: this last claim holds only *conditioned on survival*; or put differently, it holds with probability $1 - \eta$, a value which could be significantly smaller than 1 in general. To handle this last issue, we consider a large number of *independent copies* of the Galton-Watson process above in order to "boost" the probability that *at least one of them survives* to a value arbitrarily close to 1.

Claim 6.3.6. *For any $\delta > 0$, there is a J so that $H_n \geq c' \log n - \theta_n \ell + J\ell$ with probability at least $1 - \delta$ for all n large enough.*

Proof. Let $J\ell$ be a multiple of ℓ and let

$$u_1^*, \ldots, u_{2^{J\ell}}^*$$

be the vertices on level $J\ell$ from left to right. Each process

$$\left(Z_t^{u_i^*,\ell}\right)_{t\geq 0}, \quad i = 1, \ldots, 2^{J\ell},$$

is an independent copy of $(Z_t^{\rho,\ell})_{t\geq 0}$.

We define two "bad events":

- *(No survival)* Let $\mathcal{B}_1$ be the event that all $(Z_t^{u_i^*,\ell})$s go extinct and choose J large enough that this event has probability $< \delta/2$, that is,

$$\mathbb{P}[\mathcal{B}_1] = \eta^{2^{J\ell}} < \delta/2.$$

Under $\mathcal{B}_1^c$, at least one of the branching processes survives; let I be the lowest index among them.

- *(Fast decay at the top)* To bound the height, we also need to control the effect of the first $J\ell$ levels on the subtree sizes. Let $\mathcal{B}_2$ be the event that at least one of the W-values associated with the $2^{J\ell} - 1$ vertices ancestral to the u_i^*s is outside the interval $(\alpha, 1-\alpha)$. Choose α small enough that this event has probability $< \delta/2$, that is,

$$\mathbb{P}[\mathcal{B}_2] \leq (2\alpha)(2^{J\ell} - 1) < \delta/2.$$

Under $\mathcal{B}_2^c$, we have almost surely the lower bound

$$\mathcal{U}[\rho, u_I^*] \geq \alpha^{J\ell}, \tag{6.3.14}$$

since it in fact holds for all u_i^*s simultaneously.

We are now ready to conclude. Assume $\mathcal{B}_1^c$ and $\mathcal{B}_2^c$ hold. Taking

$$s = k\ell = c' \log n - \theta_n \ell,$$

as before, we have $Z_k^{u_I^*,\ell} \geq 1$ so there is $v^* \in \mathcal{L}_s[u_I^*]$ such that

$$n\,\mathcal{U}[\rho, v^*] = n\,\mathcal{U}[\rho, u_I^*]\,\mathcal{U}[u_I^*, v^*] \geq n\alpha^{J\ell}(e^{-\ell/c})^k,$$

where we used (6.3.14). Observe that (6.3.12) remains valid (for potentially larger n) even after multiplying all expressions on the left-hand side of the inequality by $\alpha^{J\ell}$. Arguing as in (6.3.13), we get that $H_n \geq c' \log n - \theta_n \ell + J\ell$. This event holds with probability at least

$$\mathbb{P}[(\mathcal{B}_1 \cup \mathcal{B}_2)^c] = 1 - \mathbb{P}[\mathcal{B}_1] - \mathbb{P}[\mathcal{B}_2] \geq 1 - \delta.$$

We have proved the claim. ∎

For any $\varepsilon > 0$, we can choose $c' = \gamma - \varepsilon$ and $c' < c < \gamma$. Further, δ can be made arbitrarily small (provided n is large enough). Put differently, we have proved that for any $\varepsilon > 0$

$$\mathbb{P}[H_n/\log n \geq \gamma - \varepsilon] \to 1,$$

as $n \to +\infty$, which establishes the other direction of Claim 6.3.1.

6.3.2 ▷ *Data science: the reconstruction problem, the Kesten-Stigum bound and a phase transition in phylogenetics*

In this section, we explore an application of multitype branching processes in statistical phylogenetics, the reconstruction of evolutionary trees from molecular data. Informally, we consider a ferromagnetic Ising model (Example 1.2.5) on an infinite binary tree and we ask: when do the states at level h "remember" the state at the root? We establish the existence of a phase transition. Before defining the problem formally and explaining its connection to evolutionary biology, we describe an equivalent definition of the model. This alternative "Markov chain on a tree" perspective will make it easier to derive recursions for quantities of interest. Equivalence between the two models is proved in Exercise 6.16.

The reconstruction problem

Consider a rooted infinite binary tree $\mathcal{T} = \widehat{\mathbb{T}}_2$, where the root is denoted by 0. Fix a parameter $0 < p < 1/2$, which we will refer to as the *mutation probability* for reasons that will be explained below. We assign a state σ_v in $\mathcal{C} = \{+1, -1\}$ to each vertex v as follows. At the root 0, the state σ_0 is picked uniformly at random in $\{+1, -1\}$. Moving away from the root, the state σ_v at a vertex v, conditioned on the state at its immediate ancestor u, is equal to σ_u with probability $1 - p$ and to $-\sigma_u$ with probability p. In the computational biology literature, this model is referred to as the *Cavender-Farris-Neyman (CFN) model*.

mutation probability

CFN model

For $h \geq 0$, let $\mathcal{L}_h$ be the set of vertices in $\mathcal{T}$ at graph distance h from the root. We denote by $\boldsymbol{\sigma}_h = (\sigma_\ell)_{\ell \in \mathcal{L}_h}$ the vector of states at level h and we denote by μ_h the distribution of $\boldsymbol{\sigma}_h$. The *reconstruction problem* consists in trying to "guess" the state at the root σ_0 given the states $\boldsymbol{\sigma}_h$ at level h. We first note that in general we cannot expect an arbitrarily good estimator. Indeed, rewriting the Markov transition matrix along the edges (i.e., the matrix encoding the probability of the state at a vertex given the state at its immediate ancestor) in its *random cluster* form

reconstruction problem

$$P := \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} = (1-2p)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (2p)\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \qquad (6.3.15)$$

we see that the states $\boldsymbol{\sigma}_1$ at the first level are completely randomized (i.e., independent of σ_0) with probability $(2p)^2$—in which case we cannot hope to reconstruct the root state better than a coin flip. Intuitively the reconstruction problem is solvable if we can find an estimator of the root state which outperforms a random coin flip as h grows to $+\infty$. Let μ_h^+ be the distribution μ_h conditioned on the root state σ_0 being $+1$, and similarly for μ_h^-. Observe that $\mu_h = \frac{1}{2}\mu_h^+ + \frac{1}{2}\mu_h^-$. Recall also that

$$\|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}} = \frac{1}{2}\sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} |\mu_h^+(\mathbf{s}_h) - \mu_h^-(\mathbf{s}_h)|.$$

Definition 6.3.7 (Reconstruction solvability)**.** *We say that the reconstruction problem for* $0 < p < 1/2$ *is solvable if*

reconstruction solvability

$$\liminf_{h \to +\infty} \|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}} > 0,$$

otherwise the problem is unsolvable.

(Exercise 6.17 asks for a proof that $\|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}}$ is monotone in h and therefore has a limit.)

To see the connection with the description above, consider an arbitrary root estimator $\hat{\sigma}_0(\mathbf{s}_h)$. Then the probability of a mistake is

$$\begin{aligned}\mathbb{P}[\hat{\sigma}_0(\boldsymbol{\sigma}_h) \neq \sigma_0] &= \frac{1}{2}\sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h^-(\mathbf{s}_h)\mathbf{1}\{\hat{\sigma}_0(\mathbf{s}_h) = +1\} \\ &\quad + \frac{1}{2}\sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h^+(\mathbf{s}_h)\mathbf{1}\{\hat{\sigma}_0(\mathbf{s}_h) = -1\}.\end{aligned}$$

This expression is minimized by choosing for each $\mathbf{s}_h$ separately

$$\hat{\sigma}_0(\mathbf{s}_h) = \begin{cases} +1 & \text{if } \mu_h^+(\mathbf{s}_h) \geq \mu_h^-(\mathbf{s}_h), \\ -1 & \text{otherwise.} \end{cases}$$

Let $\mu_h(s_0|\mathbf{s}_h)$ be the posterior probability of the root state, that is, the conditional probability of the root state s_0 given the states $\mathbf{s}_h$ at level h. By Bayes' rule,

$$\mu_h(+1|\mathbf{s}_h) = \frac{(1/2)\mu_h^+(\mathbf{s}_h)}{\mu_h(\mathbf{s}_h)},$$

and similarly for $\mu_h(+1|\mathbf{s}_h)$. Hence the choice above is equivalent to

$$\hat{\sigma}_0(\mathbf{s}_h) = \begin{cases} +1 & \text{if } \mu_h(+1|\mathbf{s}_h) \geq \mu_h(-|\mathbf{s}_h), \\ -1 & \text{otherwise.} \end{cases}$$

which is known as the *maximum a posteriori (MAP) estimator*. (We encountered it in a different context in Section 5.1.4.) For short, we will denote it by $\hat{\sigma}_0^{\mathrm{MAP}}$. *MAP estimator*

Now note that

$$\begin{aligned}
&\mathbb{P}[\hat{\sigma}_0^{\mathrm{MAP}}(\boldsymbol{\sigma}_h) = \sigma_0] - \mathbb{P}[\hat{\sigma}_0^{\mathrm{MAP}}(\boldsymbol{\sigma}_h) \neq \sigma_0] \\
&\quad = \frac{1}{2} \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h^+(\mathbf{s}_h)\,[\mathbf{1}\{\hat{\sigma}_0^{\mathrm{MAP}}(\mathbf{s}_h) = +1\} - \mathbf{1}\{\hat{\sigma}_0^{\mathrm{MAP}}(\mathbf{s}_h) = -1\}] \\
&\qquad + \frac{1}{2} \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h^-(\mathbf{s}_h)\,[\mathbf{1}\{\hat{\sigma}_0^{\mathrm{MAP}}(\mathbf{s}_h) = -1\} - \mathbf{1}\{\hat{\sigma}_0^{\mathrm{MAP}}(\mathbf{s}_h) = +1\}] \\
&\quad = \frac{1}{2} \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h^+(\mathbf{s}_h)\,\hat{\sigma}_0^{\mathrm{MAP}}(\mathbf{s}_h) \\
&\qquad - \frac{1}{2} \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h^-(\mathbf{s}_h)\,\hat{\sigma}_0^{\mathrm{MAP}}(\mathbf{s}_h) \\
&\quad = \frac{1}{2} \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} |\mu_h^+(\mathbf{s}_h) - \mu_h^-(\mathbf{s}_h)| \\
&\quad = \|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}},
\end{aligned}$$

where the third equality comes from

$$|a-b| = (a-b)\mathbf{1}\{a \geq b\} + (b-a)\mathbf{1}\{a < b\}.$$

Since $\mathbb{P}[\hat{\sigma}_0(\boldsymbol{\sigma}_h) = \sigma_0] + \mathbb{P}[\hat{\sigma}_0(\boldsymbol{\sigma}_h) \neq \sigma_0] = 1$, the display above can be rewritten as

$$\mathbb{P}[\hat{\sigma}_0^{\mathrm{MAP}}(\boldsymbol{\sigma}_h) \neq \sigma_0] = \frac{1}{2} - \frac{1}{2}\|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}}.$$

Given that $\hat{\sigma}_0^{\mathrm{MAP}}$ was chosen to minimize the error probability, we also have that *for any root estimator* $\hat{\sigma}_0$

$$\mathbb{P}[\hat{\sigma}_0(\boldsymbol{\sigma}_h) \neq \sigma_0] \geq \frac{1}{2} - \frac{1}{2}\|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}}.$$

Since this last inequality also applies to the estimator $-\hat{\sigma}_0$, we have also that

$$\mathbb{P}[\hat{\sigma}_0(\boldsymbol{\sigma}_h) \neq \sigma_0] \leq \frac{1}{2}.$$

The next lemma summarizes the discussion above.

Lemma 6.3.8 (Probability of erroneous reconstruction)**.** *The probability of an erroneous root reconstruction behaves as follows.*

(i) If the reconstruction problem is solvable, then

$$\lim_{h\to+\infty} \mathbb{P}[\hat{\sigma}_0^{\mathrm{MAP}}(\boldsymbol{\sigma}_h) \neq \sigma_0] < \frac{1}{2}.$$

(ii) If the reconstruction problem is unsolvable, then for any root estimator $\hat{\sigma}_0$

$$\lim_{h\to+\infty} \mathbb{P}[\hat{\sigma}_0(\boldsymbol{\sigma}_h) \neq \sigma_0] = \frac{1}{2}.$$

It turns out that the accuracy of the MAP estimator undergoes a phase transition at a critical mutation probability p_*. Our main theorem is the following.

Theorem 6.3.9 (Solvability)**.** *Let* θ_* *be the unique positive solution to*

$$2\theta_*^2 = 1,$$

and set $p^* = \frac{1-\theta^*}{2}$*. Then the reconstruction problem is:*

(i) solvable if $0 < p < p_*$*;*

(ii) unsolvable if $p_* \leq p < 1/2$*.*

We will prove this theorem in the rest of the section.

But first, what does all of this have to do with evolutionary biology? Truncate $\mathcal{T}$ at level h to obtain a finite tree $\mathcal{T}_h$ with leaf set $\mathcal{L}_h$. In phylogenetics, one uses such a tree to depict evolutionary relationships between extant species that are represented by its leaves. Each internal branching corresponds to a past speciation event. Extinctions have been pruned from the tree. The genomes of ancestral species, starting from the most recent common ancestor at the root, are posited to have evolved along the (deterministic) tree $\mathcal{T}_h$ according to a random process of single-site substitutions. To simplify, each position in the genome is assumed to take one of two values, $+1$ or -1, and it evolves independently from all other positions under a CFN model on $\mathcal{T}_h$. That is, on each edge of the tree a mutation occurs with probability p, changing the state of the immediate descendant species at that position. This is of course only a toy model, but it is not far from what evolutionary biologists actually use in practice with great success. One practical problem of interest is to reconstruct the genome of ancestors given access to contemporary genomes. This is, in a nutshell, the reconstruction problem.

Kesten-Stigum bound

Kesten-Stigum bound

The condition in Theorem 6.3.9 is referred to as the *Kesten-Stigum bound*. We explain why next. We showed in Lemma 6.3.8 that the MAP estimator has an error probability bounded away from $1/2$ if and only if the reconstruction problem is solvable. Of course, other estimators may also achieve that same desirable outcome. In fact, from the lemma, to establish reconstruction solvability it *suffices* to exhibit one such "better-than-random" estimator. So, rather than analyzing $\hat{\sigma}_0^{\mathrm{MAP}}$, we look at a simpler estimator first and prove half of Theorem 6.3.9. The other half will be proven below using different ideas.

The key is to notice that a multitype branching process (see Section 6.1.4) hides in the background. For $h \geq 0$, consider the random row vector $\mathbf{Z}_h = (Z_{h,+}, Z_{h,-})$ where the first component records the number of $+1$ states (which we refer to as belonging to the $+$ type) in $\boldsymbol{\sigma}_h$ and, likewise, the second component counts the -1 states (referred to as of $-$ type). Then $(\mathbf{Z}_h)_{h\geq 0}$ is a two-type Galton-Watson process where each individual has exactly two children. Their types depend on the type of the parent. A type $+$ individual has the following offspring distribution:

$$p_{\mathbf{k}}^{(+)} = \begin{cases} (1-p)^2 & \text{if } \mathbf{k} = (2,0), \\ 2p(1-p) & \text{if } \mathbf{k} = (1,1), \\ p^2 & \text{if } \mathbf{k} = (0,2), \\ 0 & \text{otherwise.} \end{cases}$$

Similar expressions hold for $p_{\mathbf{k}}^{(-)}$. The mean matrix is given by

$$\begin{aligned} M &= \begin{pmatrix} 2(1-p)^2 + 2p(1-p) & 2p(1-p) + 2p^2 \\ 2p(1-p) + 2p^2 & 2(1-p)^2 + 2p(1-p) \end{pmatrix} \\ &= 2\begin{pmatrix} (1-p)(1-p+p) & p(1-p+p) \\ p(1-p+p) & (1-p)(1-p+p) \end{pmatrix} \\ &= 2\begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \\ &= 2P, \end{aligned}$$

where (not coincidentally) we have already encountered the matrix P in (6.3.15). As a symmetric matrix, by the spectral theorem (Theorem 5.1.1), P has a real eigenvector decomposition

$$\begin{aligned} P &= \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix} \\ &= \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} + (1-2p) \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{pmatrix} \\ &= \lambda_1 \mathbf{x}_1 \mathbf{x}_1^T + \lambda_2 \mathbf{x}_2 \mathbf{x}_2^T, \end{aligned}$$

where the eigenvalues and eigenvectors are

$$\lambda_1 = 1, \quad \lambda_2 = 1-2p, \quad \mathbf{x}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \quad \mathbf{x}_2 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}.$$

The eigenvalues of M are twice those of P while the eigenvectors are the same. In particular, using the notation and convention of the Perron-Frobenius Theorem (Theorem 6.1.17), we have

$$\rho = 2, \quad \mathbf{w} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}.$$

These should not come entirely as a surprise. In particular, recall from Theorem 6.1.18 that ρ can be interpreted as an "overall rate of growth" of the population, which here is two since each individual has exactly two children (ignoring the types).

Let $\mathbf{u} = (1, -1)$ be a column vector proportional to the second right eigenvector of M. We know from Section 6.1.4 that

$$U_h = (2\lambda_2)^{-h} \mathbf{Z}_h \mathbf{u} = \frac{1}{2^h \theta^h} \sum_{\ell \in \mathcal{L}_h} \sigma_\ell, \quad h \geq 0,$$

is a martingale, where we used the notation

$$\theta := \lambda_2 = 1 - 2p.$$

Upon looking more closely, the quantity U_h has a natural interpretation: its sign is the *majority estimator*, that is, $\mathrm{sgn}(U_h) = +1$ if a majority of individuals at level h are of type $+$ (breaking ties in favor of $+$), and is -1 otherwise. We indicated previously that we only need to find one estimator with an error probability bounded away from $1/2$ to establish reconstruction solvability for a given value of p. The majority estimator *majority estimator*

$$\hat{\sigma}_0^{\mathrm{Maj}} := \mathrm{sgn}(U_h),$$

is an obvious one to try. What is less obvious is that it works—all the way to the threshold. This essentially follows from the results of Section 6.1.4, as we detail next.

We begin with an informal discussion. When can $\hat{\sigma}_0^{\mathrm{Maj}}$ be expected to work? We will not in fact bound the error probability of $\hat{\sigma}_0^{\mathrm{Maj}}$, but instead analyze directly the properties of (U_h). By our modeling assumptions, $\mathbf{Z}_0$ is either $(1,0)$ or $(0,1)$ with equal probability. Hence, by the martingale property, we obtain that

$$\mathbb{E}[U_h \,|\, \mathbf{Z}_0] = \mathbf{Z}_0 \mathbf{u} = \sigma_0. \tag{6.3.16}$$

In words, U_h is "centered" around the root state. Intuitively, its second moment therefore captures how informative it is about σ_0. Lemma 6.1.20 exhibits a phase transition for $\mathbb{E}[U_h^2 \,|\, \mathbf{Z}_0]$. The condition for that lemma to hold is

$$(\mathrm{Var}[\mathbf{X}^{(+)}(1,1)\,\mathbf{u}], \mathrm{Var}[\mathbf{X}^{(-)}(1,1)\,\mathbf{u}]) \neq \mathbf{0},$$

where $\mathbf{X}^{(+)}(1,1) \sim \{p_{\mathbf{k}}^{(+)}\}$ and $\mathbf{X}^{(-)}(1,1) \sim \{p_{\mathbf{k}}^{(-)}\}$. This is indeed satisfied. The lemma then states that $\mathbb{E}[U_h^2 \,|\, \mathbf{Z}_0]$ is uniformly bounded if and only if $\rho < (2\lambda_2)^2$, or after rearranging

$$2\theta^2 > 1. \tag{6.3.17}$$

Note that this is the condition in Theorem 6.3.9. It arises as a tradeoff between the rate of growth $\rho = 2$ and the second largest eigenvalue $\lambda_2 = \theta$ of the Markov transition matrix P. One way to make sense of it is to observe the following:

- On any infinite path out of the root, the process performs a finite Markov chain with transition matrix P. We know from Theorem 5.2.14 (see in particular Example 5.2.8) that the chain mixes—and therefore "forgets" its starting state σ_0—at a rate governed by the spectral gap $1 - \lambda_2$.

- On the other hand, the tree itself is growing at rate $\rho = 2$, which produces an exponentially large number of (overlapping) paths out of the root. That growth helps preserve the information about σ_0 down the tree through the duplication of the state (with mutation) at each branching.

- The condition $\rho < (2\lambda_2)^2$ says in essence that when mixing is slow enough—corresponding to larger values of λ_2—compared to the growth, then the reconstruction problem is solvable. Lemma 6.1.20 was first proved by Kesten and Stigum, and (6.3.17) is thereby known as the Kesten-Stigum bound.

It remains to turn these observations into a formal proof.

Denote by $\mathbb{E}^+$ the expectation conditioned on $\sigma_0 = +1$, and similarly for $\mathbb{E}^-$. The following lemma is a consequence of (6.3.16). We give a quick alternative proof.

Lemma 6.3.10 (Unbiasedness of U_h)**.** *We have*

$$\mathbb{E}^+[U_h] = +1, \qquad \mathbb{E}^-[U_h] = -1.$$

Proof. By applying the Markov transition matrix P on the first level and using the symmetries of the model, for any $\ell \in \mathcal{L}_h$ and $\ell' \in \mathcal{L}_{h-1}$, we have

$$\begin{aligned}
\mathbb{E}^+[\sigma_\ell] &= (1-p)\mathbb{E}^+[\sigma_{\ell'}] + p\,\mathbb{E}^-[\sigma_{\ell'}] \\
&= (1-p)\mathbb{E}^+[\sigma_{\ell'}] + p\,\mathbb{E}^+[-\sigma_{\ell'}] \\
&= (1-2p)\mathbb{E}^+[\sigma_{\ell'}] \\
&= \theta\,\mathbb{E}^+[\sigma_{\ell'}].
\end{aligned}$$

Iterating, we get $\mathbb{E}^+[\sigma_\ell] = \theta^h$. The claim follows by linearity of expectation. ■

Although we do not strictly need it, we also derive an explicit formula for the variance. The proof is typical of how conditional independence properties of this kind of Markov model on trees can be used to derive recursions for quantities of interest.

Lemma 6.3.11 (Variance of U_h)**.** *We have*

$$\mathrm{Var}[U_h] \to \begin{cases} \frac{1/2}{1-(2\theta^2)^{-1}} & \text{if } 2\theta^2 > 1, \\ +\infty, & \text{otherwise.} \end{cases}$$

Proof. By the conditional variance formula

$$\begin{aligned}
\mathrm{Var}[U_h] &= \mathrm{Var}[\mathbb{E}[U_h \,|\, \sigma_0]] + \mathbb{E}[\mathrm{Var}[U_h \,|\, \sigma_0]] \\
&= \mathrm{Var}[\sigma_0] + \mathbb{E}[\mathrm{Var}[U_h \,|\, \sigma_0]] \\
&= 1 + \mathrm{Var}^+[U_h], \qquad (6.3.18)
\end{aligned}$$

where the last line follows from symmetry, with Var^+ indicating the conditional variance given that the root state σ_0 is $+1$. Write $U_h = \dot{U}_h + \ddot{U}_h$ as a sum over the left and right subtrees below the root respectively. Using the conditional independence of those two subtrees given the root state, we get from (6.3.18) that

$$\begin{aligned}
\mathrm{Var}[U_h] &= 1 + \mathrm{Var}^+[U_h] \\
&= 1 + \mathrm{Var}^+\left[\dot{U}_h + \ddot{U}_h\right] \\
&= 1 + 2\mathrm{Var}^+\left[\dot{U}_h\right] \\
&= 1 + 2\left(\mathbb{E}^+\left[\dot{U}_h^2\right] - \mathbb{E}^+\left[\dot{U}_h\right]^2\right). \qquad (6.3.19)
\end{aligned}$$

We now use the Markov transition matrix on the first level to derive a recursion in h. Let $\dot\sigma_0$ be the state at the left child of the root. We use the fact that the random variables $2\theta\dot U_h$ conditioned on $\dot\sigma_0 = +1$ and U_{h-1} conditioned on $\sigma_0 = +1$ are identically distributed. Using $\mathbb{E}^+[\dot U_h] = 1/2$ (by Lemma 6.3.10 and symmetry), we get from (6.3.19) that

$$\begin{aligned}
\mathrm{Var}[U_h] &= 1 - 2\mathbb{E}^+\left[\dot U_h\right]^2 + 2\,\mathbb{E}^+\left[\dot U_h^2\right] \\
&= 1 - 2(1/2)^2 + 2\left[(1-p)\mathbb{E}^+\left[(2\theta)^{-2}U_{h-1}^2\right] + p\,\mathbb{E}^-\left[(2\theta)^{-2}U_{h-1}^2\right]\right] \\
&= 1/2 + (2\theta^2)^{-1}\mathbb{E}^+[U_{h-1}^2] \\
&= 1/2 + (2\theta^2)^{-1}\mathrm{Var}[U_{h-1}], \qquad (6.3.20)
\end{aligned}$$

where we used that

$$\mathrm{Var}[U_{h-1}] = \mathbb{E}[U_{h-1}^2] = \mathbb{E}^+[U_{h-1}^2] = \mathbb{E}^-[U_{h-1}^2],$$

by symmetry and the fact that $\mathbb{E}[U_{h-1}] = 0$. Solving the affine recursion (6.3.20) gives

$$\mathrm{Var}[U_h] = (2\theta^2)^{-h} + (1/2)\sum_{i=0}^{h-1}(2\theta^2)^{-i},$$

where we used that $\mathrm{Var}[U_0] = \mathrm{Var}[\sigma_0] = 1$. The result follows. ■

We can now prove the first part of Theorem 6.3.9.

Proof of Theorem 6.3.9 (i). Let $\bar\mu_h$ be the distribution of U_h and define $\bar\mu_h^+$ and $\bar\mu_h^-$ similarly. We give a bound on $\|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}}$ through a bound on $\|\bar\mu_h^+ - \bar\mu_h^-\|_{\mathrm{TV}}$. Let $\bar{\mathbf{s}}_h$ be the U_h-value associated to $\mathbf{s}_h = (s_{h,\ell})_{\ell\in\mathcal{L}_h} \in \{+1,-1\}^{2^h}$, that is,

$$\bar{\mathbf{s}}_h = \frac{1}{2^h\theta^h}\sum_{\ell\in\mathcal{L}_h} s_{h,\ell}.$$

Then, by marginalizing and the triangle inequality,

$$\begin{aligned}
\sum_z |\bar\mu_h^+(z) - \bar\mu_h^-(z)| &= \sum_z \left|\sum_{\mathbf{s}_h:\bar{\mathbf{s}}_h = z}(\mu_h^+(\mathbf{s}_h) - \mu_h^-(\mathbf{s}_h))\right| \\
&\leq \sum_z \sum_{\mathbf{s}_h:\bar{\mathbf{s}}_h = z} |\mu_h^+(\mathbf{s}_h) - \mu_h^-(\mathbf{s}_h)| \\
&= \sum_{\mathbf{s}_h\in\{+1,-1\}^{2^h}} |\mu_h^+(\mathbf{s}_h) - \mu_h^-(\mathbf{s}_h)|,
\end{aligned}$$

where the first sum is over the support of $\bar{\mu}_h$. So it suffices to bound from below the left-hand side on the first line.

For that purpose, we apply Cauchy-Schwarz and use the variance bound in Lemma 6.3.11. First note that $\frac{1}{2}\bar{\mu}_h^+ + \frac{1}{2}\bar{\mu}_h^- = \bar{\mu}_h$ so that, by the triangle inequality,

$$\frac{|\bar{\mu}_h^+(z) - \bar{\mu}_h^-(z)|}{2\bar{\mu}_h(z)} \leq \frac{\bar{\mu}_h^+(z) + \bar{\mu}_h^-(z)}{2\bar{\mu}_h(z)} = 1. \tag{6.3.21}$$

Hence, we get

$$\begin{aligned}
\sum_z |\bar{\mu}_h^+(z) - \bar{\mu}_h^-(z)| &= \sum_z \frac{|\bar{\mu}_h^+(z) - \bar{\mu}_h^-(z)|}{2\bar{\mu}_h(z)} 2\bar{\mu}_h(z) \\
&\geq 2\sum_z \left(\frac{\bar{\mu}_h^+(z) - \bar{\mu}_h^-(z)}{2\bar{\mu}_h(z)}\right)^2 \bar{\mu}_h(z) \\
&\geq 2\frac{\left(\sum_z z \left(\frac{\bar{\mu}_h^+(z) - \bar{\mu}_h^-(z)}{2\bar{\mu}_h(z)}\right)\bar{\mu}_h(z)\right)^2}{\sum_z z^2 \bar{\mu}_h(z)} \\
&\geq \frac{1}{2}\frac{\left(\sum_z z \left(\bar{\mu}_h^+(z) - \bar{\mu}_h^-(z)\right)\right)^2}{\sum_z z^2 \bar{\mu}_h(z)} \\
&= \frac{1}{2}\frac{(\mathbb{E}^+[U_h] - \mathbb{E}^-[U_h])^2}{\mathrm{Var}[U_h]} \\
&\geq 4(1 - (2\theta^2)^{-1}) > 0,
\end{aligned}$$

where we used (6.3.21) on the second line, Cauchy-Schwarz on the third line (after rearranging), and Lemmas 6.3.10 and 6.3.11 on the last line.

■

Remark 6.3.12. *The proof above and a correlation inequality of [EKPS00, Theorem 1.4] give a lower bound on the probability of reconstruction of the majority estimator.*

Impossibility of reconstruction

The previous result was based on showing that majority voting, that is, $\hat{\sigma}_0^{\mathrm{Maj}}$, produces a good root-state estimator—up to $p = p_*$. Here we establish that this result is best possible. Majority is not in fact the best root-state estimator: in general its error probability can be higher than $\hat{\sigma}_0^{\mathrm{MAP}}$ as the latter also takes into account the *configuration of the states at level h*. However, perhaps surprisingly, it turns out that the critical threshold for $\hat{\sigma}_0^{\mathrm{Maj}}$ coincides with that of $\hat{\sigma}_0^{\mathrm{MAP}}$ in the CFN model.

To prove the second part of Theorem 6.3.9 we analyze the MAP estimator. Recall that $\mu_h(s_0|\mathbf{s}_h)$ is the conditional probability of the root state s_0 given the states $\mathbf{s}_h$ at level h. It will be more convenient to work with the following "root magnetization"

$$R_h := \mu_h(+1|\boldsymbol{\sigma}_h) - \mu_h(-1|\boldsymbol{\sigma}_h),$$

which, as a function of $\boldsymbol{\sigma}_h$, is a *random variable*. Note that $\mathbb{E}[R_h] = 0$ by symmetry. By Bayes' rule and the fact that $\mu_h(+1|\boldsymbol{\sigma}_h) + \mu_h(-1|\boldsymbol{\sigma}_h) = 1$, we have the following alternative formulas which will prove useful

$$R_h = \frac{1}{2\mu_h(\boldsymbol{\sigma}_h)}[\mu_h^+(\boldsymbol{\sigma}_h) - \mu_h^-(\boldsymbol{\sigma}_h)], \tag{6.3.22}$$

$$R_h = 2\mu_h(+1|\boldsymbol{\sigma}_h) - 1 = \frac{\mu_h^+(\boldsymbol{\sigma}_h)}{\mu_h(\boldsymbol{\sigma}_h)} - 1, \tag{6.3.23}$$

$$R_h = 1 - 2\mu_h(-1|\boldsymbol{\sigma}_h) = 1 - \frac{\mu_h^-(\boldsymbol{\sigma}_h)}{\mu_h(\boldsymbol{\sigma}_h)}. \tag{6.3.24}$$

It turns out to be enough to prove an upper bound on the variance of R_h.

Lemma 6.3.13 (Second moment bound)**.** *It holds that*

$$\|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}} \leq \sqrt{\mathbb{E}[R_h^2]}.$$

Proof. By (6.3.22),

$$\begin{aligned}
&\frac{1}{2}\sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} |\mu_h^+(\mathbf{s}_h) - \mu_h^-(\mathbf{s}_h)| \\
&= \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h(\mathbf{s}_h)\,|\mu_h(+1|\mathbf{s}_h) - \mu_h(-1|\mathbf{s}_h)| \\
&= \mathbb{E}|R_h| \\
&\leq \sqrt{\mathbb{E}[R_h^2]},
\end{aligned}$$

where we used Cauchy-Schwarz on the last line. ∎

Let $\bar{z}_h = \mathbb{E}[R_h^2]$. In view of Lemma 6.3.13, the proof of Theorem 6.3.9 (ii) will follow from establishing the limit

$$\lim_{h \to +\infty} \bar{z}_h = 0.$$

We apply the same kind of recursive argument we used for the analysis of majority (see in particular Lemma 6.3.11): we condition on the root to exploit conditional independence; we use the Markov transition matrix on the top edges.

We first derive a recursion for R_h itself—as a random variable. We proceed in two steps:

- *Step 1:* we break up the first h levels of the tree into two identical $(h-1)$-level trees with an additional edge at their respective root through conditional independence;

- *Step 2:* we account for that edge through the Markov transition matrix.

We will need some notation. Let $\dot{\boldsymbol{\sigma}}_h$ be the states at level h (from the root) below the left child of the root and let $\dot{\mu}_h$ be the distribution of $\dot{\boldsymbol{\sigma}}_h$ (and use a superscript $+$ to denote the conditional probability given the root is $+$, and so on). Define

$$\dot{Y}_h = \dot{\mu}_h(+1|\dot{\boldsymbol{\sigma}}_h) - \dot{\mu}_h(-1|\dot{\boldsymbol{\sigma}}_h),$$

where $\dot{\mu}_h(s_0|\dot{\mathbf{s}}_h)$ is the conditional probability that the root is s_0 given that $\dot{\boldsymbol{\sigma}}_h = \dot{\mathbf{s}}_h$. Similarly, denote with a double dot the same quantities with respect to the subtree below the right child of the root. Expressions similar to (6.3.22), (6.3.23) and (6.3.24) also hold.

Lemma 6.3.14 (Recursion: Step 1)**.** *It holds almost surely that*

$$R_h = \frac{\dot{Y}_h + \ddot{Y}_h}{1 + \dot{Y}_h \ddot{Y}_h}.$$

Proof. Using $\mu_h^+(\mathbf{s}_h) = \dot{\mu}_h^+(\dot{\mathbf{s}}_h)\ddot{\mu}_h^+(\ddot{\mathbf{s}}_h)$ by conditional independence, (6.3.22) applied to R_h, and (6.3.23) and (6.3.24) applied to $\dot{Y}_h$ and $\ddot{Y}_h$, we get

$$\begin{aligned}
R_h &= \frac{1}{2}\sum_{\gamma=+,-}\gamma\frac{\mu_h^\gamma(\boldsymbol{\sigma}_h)}{\mu_h(\boldsymbol{\sigma}_h)} \\
&= \frac{1}{2}\frac{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)}{\mu_h(\boldsymbol{\sigma}_h)}\sum_{\gamma=+,-}\gamma\frac{\dot{\mu}_h^\gamma(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h^\gamma(\ddot{\boldsymbol{\sigma}}_h)}{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)} \\
&= \frac{1}{2}\frac{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)}{\mu_h(\boldsymbol{\sigma}_h)}\sum_{\gamma=+,-}\gamma\left(1+\gamma\dot{Y}_h\right)\left(1+\gamma\ddot{Y}_h\right) \\
&= \frac{\dot{\mu}_h(\dot{\boldsymbol{\sigma}}_h)\ddot{\mu}_h(\ddot{\boldsymbol{\sigma}}_h)}{\mu_h(\boldsymbol{\sigma}_h)}(\dot{Y}_h + \ddot{Y}_h).
\end{aligned}$$

The factor in front can be computed as follows

$$\begin{aligned}\frac{\mu_h(\boldsymbol{\sigma}_h)}{\dot\mu_h(\dot{\boldsymbol{\sigma}}_h)\ddot\mu_h(\ddot{\boldsymbol{\sigma}}_h)} &= \sum_{\gamma=+,-}\frac{1}{2}\frac{\mu_h^{\gamma}(\boldsymbol{\sigma}_h)}{\dot\mu_h(\dot{\boldsymbol{\sigma}}_h)\ddot\mu_h(\ddot{\boldsymbol{\sigma}}_h)}\\ &= \sum_{\gamma=+,-}\frac{1}{2}\frac{\dot\mu_h^{\gamma}(\dot{\boldsymbol{\sigma}}_h)\ddot\mu_h^{\gamma}(\ddot{\boldsymbol{\sigma}}_h)}{\dot\mu_h(\dot{\boldsymbol{\sigma}}_h)\ddot\mu_h(\ddot{\boldsymbol{\sigma}}_h)}\\ &= \frac{1}{2}\sum_{\gamma=+,-}\left(1+\gamma\dot Y_h\right)\left(1+\gamma\ddot Y_h\right)\\ &= 1+\dot Y_h\ddot Y_h.\end{aligned}$$

That proves the claim. ■

For the second step of the recursion, we define

$$\dot D_h = \dot\nu_h(+1|\dot{\boldsymbol{\sigma}}_h) - \dot\nu_h(-1|\dot{\boldsymbol{\sigma}}_h),$$

where $\dot\nu_h(\dot s_0|\dot{\mathrm{s}}_h)$ is the conditional probability that the left child of the root is $\dot s_0$ given that the states at level h (from the root) below the left child are $\dot{\boldsymbol{\sigma}}_h = \dot{\mathrm{s}}_h$; and similarly for the right child of the root. Again expressions similar to (6.3.22), (6.3.23) and (6.3.24) hold. The following lemma is left as an exercise (see Exercise 6.18).

Lemma 6.3.15 (Recursion: Step 2)**.** *It holds almost surely that*

$$\dot Y_h = \theta \dot D_h.$$

We are now ready to prove the second half of our main theorem.

Proof of Theorem 6.3.9 (ii). Putting Lemmas 6.3.14 and 6.3.15 together, we get

$$R_h = \frac{\theta(\dot D_h + \ddot D_h)}{1+\theta^2\dot D_h\ddot D_h}. \tag{6.3.25}$$

We now take expectations. Recall that we seek to compute the second moment of

R_h. However an important simplification arises from the following observation

$$\begin{aligned}
\mathbb{E}^+[R_h] &= \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h^+(\mathbf{s}_h) R_h(\mathbf{s}_h) \\
&= \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h(\mathbf{s}_h) \frac{\mu_h^+(\mathbf{s}_h)}{\mu_h(\mathbf{s}_h)} R_h(\mathbf{s}_h) \\
&= \sum_{\mathbf{s}_h \in \{+1,-1\}^{2^h}} \mu_h(\mathbf{s}_h)(1 + R_h(\mathbf{s}_h)) R_h(\mathbf{s}_h) \\
&= \mathbb{E}[(1 + R_h) R_h] \\
&= \mathbb{E}[R_h^2],
\end{aligned}$$

where we used (6.3.23) on the third line and $\mathbb{E}[R_h] = 0$ on the fifth line. *So it suffices to compute the conditional first moment.*

Using the expansion

$$\frac{1}{1+r} = 1 - r + \frac{r^2}{1+r},$$

with $r = \theta^2 \dot{D}_h \ddot{D}_h$, we have by (6.3.25) that

$$\begin{aligned}
R_h &= \theta(\dot{D}_h + \ddot{D}_h) - \theta^3(\dot{D}_h + \ddot{D}_h)\dot{D}_h \ddot{D}_h + \theta^4 \dot{D}_h^2 \ddot{D}_h^2 R_h \\
&\leq \theta(\dot{D}_h + \ddot{D}_h) - \theta^3(\dot{D}_h + \ddot{D}_h)\dot{D}_h \ddot{D}_h + \theta^4 \dot{D}_h^2 \ddot{D}_h^2,
\end{aligned} \tag{6.3.26}$$

where we used $|R_h| \leq 1$.

We will need the conditional first and second moments of $\dot{D}_h$. For the first moment, note that by symmetry (more precisely, by the fact that R_{h-1} conditioned on $\sigma_0 = -1$ is equal in distribution to $-R_{h-1}$ conditioned on $\sigma_0 = +1$)

$$\begin{aligned}
\mathbb{E}^+[\dot{D}_h] &= (1-p)\mathbb{E}^+[R_{h-1}] + p\,\mathbb{E}^-[R_{h-1}] \\
&= (1-p)\mathbb{E}^+[R_{h-1}] + p\,\mathbb{E}^+[-R_{h-1}] \\
&= (1-2p)\mathbb{E}^+[R_{h-1}] \\
&= \theta\,\mathbb{E}^+[R_{h-1}].
\end{aligned}$$

Similarly, for the second moment, we have

$$\begin{aligned}
\mathbb{E}^+[\dot{D}_h^2] &= (1-p)\mathbb{E}^+[R_{h-1}^2] + p\,\mathbb{E}^-[R_{h-1}^2] \\
&= \mathbb{E}[R_{h-1}^2] \\
&= \mathbb{E}^+[R_{h-1}],
\end{aligned}$$

where we used that $\mathbb{E}^+[R_{h-1}^2] = \mathbb{E}^-[R_{h-1}^2]$ by symmetry so that $\mathbb{E}[R_{h-1}^2] = (1/2)\mathbb{E}^+[R_{h-1}^2] + (1/2)\mathbb{E}^-[R_{h-1}^2] = \mathbb{E}^+[R_{h-1}^2]$.

Taking expectations in (6.3.26), using conditional independence, and plugging in the formulas for $\mathbb{E}^+[\dot{D}_h]$ and $\mathbb{E}^+[\dot{D}_h^2]$ above, we obtain

$$\begin{aligned}\bar{z}_h &= \mathbb{E}^+[R_h] \\ &\leq \theta(\mathbb{E}^+[\dot{D}_h] + \mathbb{E}^+[\ddot{D}_h]) - \theta^3(\mathbb{E}^+[\dot{D}_h^2]\,\mathbb{E}^+[\ddot{D}_h] + \mathbb{E}^+[\ddot{D}_h^2]\,\mathbb{E}^+[\dot{D}_h]) \\ &\quad +\theta^4\mathbb{E}^+[\dot{D}_h^2]\,\mathbb{E}^+[\ddot{D}_h^2] \\ &= 2\theta^2\mathbb{E}^+[R_{h-1}] - 2\theta^4\mathbb{E}^+[R_{h-1}]^2 + \theta^4\mathbb{E}^+[R_{h-1}]^2 \\ &= 2\theta^2\bar{z}_{h-1} - \theta^4\bar{z}_{h-1}^2. \end{aligned} \tag{6.3.27}$$

We analyze this recursion next. At $h = 0$, we have $\bar{z}_0 = \mathbb{E}^+[R_0] = 1$.

- When $2\theta^2 < 1$, the sequence $\bar{z}_h$ decreases to 0 exponentially fast

$$\bar{z}_h \leq (2\theta^2)^h, \qquad h \geq 0.$$

- When $2\theta^2 = 1$ on the other hand, convergence to 0 occurs at a slower rate. We show by induction that

$$\bar{z}_h \leq \frac{4}{h}, \quad h \geq 0.$$

Note that $\bar{z}_1 \leq \bar{z}_0 - \theta^4\bar{z}_0^2 = 3/4 \leq 4$ since $\theta^4 = 1/4$, which proves the base case. Assuming the bound holds for $h - 1$, we have from (6.3.27) that

$$\begin{aligned}\bar{z}_h &\leq \bar{z}_{h-1} - \frac{1}{4}\bar{z}_{h-1}^2 \\ &\leq \frac{4}{h-1} - \frac{4}{(h-1)^2} \\ &= 4\frac{h-2}{(h-1)^2} \\ &\leq \frac{4}{h},\end{aligned}$$

where the last line follows from checking that $h(h-2) \leq (h-1)^2$.

Since

$$\lim_{h\to+\infty} \bar{z}_h = 0,$$

the claim follows from Lemma 6.3.13. ■

Remark 6.3.16. *While Theorem 6.3.9 part (i) can be generalized beyond the CFN model (see, e.g., [MP03]), part (ii) cannot. A striking construction of [Mos01] shows that, under more general models, certain root-state estimators taking into account the configuration of the states at level h can "beat" the Kesten-Stigum bound.*

6.4 ▷ *Finale: the phase transition of the Erdős-Rényi model*

A compelling way to view an Erdős-Rényi random graph—as its density varies—is the following coupling or "evolution." For each pair $\{i, j\}$, let $U_{\{i,j\}}$ be independent uniform random variables in $[0, 1]$ and set $\mathcal{G}(p) := ([n], \mathcal{E}(p))$ where $\{i, j\} \in \mathcal{E}(p)$ if and only if $U_{\{i,j\}} \leq p$. Then $\mathcal{G}(p)$ is distributed according to $\mathbb{G}_{n,p}$. As p varies from 0 to 1, we start with an empty graph and progressively add edges until the complete graph is obtained.

We showed in Section 2.3.2 that $\frac{\log n}{n}$ is a threshold function for connectivity. Before connectivity occurs in the evolution of the random graph, a quantity of interest is the size of the largest connected component. As we show in the current section, this quantity itself undergoes a remarkable phase transition: when $p = \frac{\lambda}{n}$ with $\lambda < 1$, the largest component has size $\Theta(\log n)$; as λ crosses 1, many components quickly merge to form a so-called "giant component" of size $\Theta(n)$.

This celebrated result is often referred to as "the" phase transition of the Erdős-Rényi graph model. Although the proof is quite long, it is well worth studying in details. It employs most tools we have seen up to this point: first and second moment methods, Chernoff-Cramér bounds, martingale techniques, coupling and stochastic domination, and branching processes. It is quintessential discrete probability.

6.4.1 Statement and proof sketch

Before stating the main theorems, we recall a basic result from Chapter 2.

- *(Poisson tail)* Let S_n be a sum of n i.i.d. $\mathrm{Poi}(\lambda)$ variables. Recall from (2.4.10) and (2.4.11) that for $a > \lambda$

$$-\frac{1}{n} \log \mathbb{P}[S_n \geq an] \geq a \log\left(\frac{a}{\lambda}\right) - a + \lambda =: I_\lambda^{\mathrm{Poi}}(a), \tag{6.4.1}$$

and similarly for $a < \lambda$

$$-\frac{1}{n} \log \mathbb{P}[S_n \leq an] \geq I_\lambda^{\mathrm{Poi}}(a). \tag{6.4.2}$$

To simplify the notation, we let

$$I_\lambda := I_\lambda^{\mathrm{Poi}}(1) = \lambda - 1 - \log \lambda \geq 0, \tag{6.4.3}$$

where the inequality follows from the convexity of I_λ and the fact that it attains its minimum at $\lambda = 1$ where it is 0.

We let $p = \frac{\lambda}{n}$ and denote by $\mathcal{C}_{\max}$ a largest connected component. In the subcritical case, that is, when $\lambda < 1$, we show that the largest connected component has logarithmic size in n.

Theorem 6.4.1 (Subcritical case: upper bound on the largest cluster)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda \in (0,1)$. For all $\kappa > 0$,*

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_{\max}| \geq (1+\kappa) I_\lambda^{-1} \log n\right] = o(1),$$

where I_λ is defined in (6.4.3)*.*

We also give a matching logarithmic lower bound on the size of $\mathcal{C}_{\max}$ in Theorem 6.4.11.

In the supercritical case, that is, when $\lambda > 1$, we prove the existence of a unique connected component of size linear in n, which is referred to as the *giant component*.

giant component

Theorem 6.4.2 (Supercritical regime: giant component)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda > 1$. For any $\gamma \in (1/2, 1)$ and $\delta < 2\gamma - 1$,*

$$\mathbb{P}_{n,p_n}[||\mathcal{C}_{\max}| - \zeta_\lambda n| \geq n^\gamma] = O(n^{-\delta}),$$

where ζ_λ be the unique solution in $(0,1)$ to the fixed point equation

$$1 - e^{-\lambda\zeta} = \zeta.$$

In fact, with probability $1 - O(n^{-\delta})$, there is a unique largest component and the second largest connected component has size $O(\log n)$.

See Figure 6.4 for an illustration.

At a high level, the proof goes as follows:

- *(Subcritical regime)* In the subcritical case, we use an exploration process and a domination argument to approximate the size of the connected components with the progeny of a branching process. The result then follows from the hitting-time theorem and the Poisson tail.

- *(Supercritical regime)* In the supercritical case, a similar argument gives a bound on the expected size of the giant component, which is related to the survival of the branching process. Chebyshev's inequality gives concentration. The hard part there is to bound the variance.

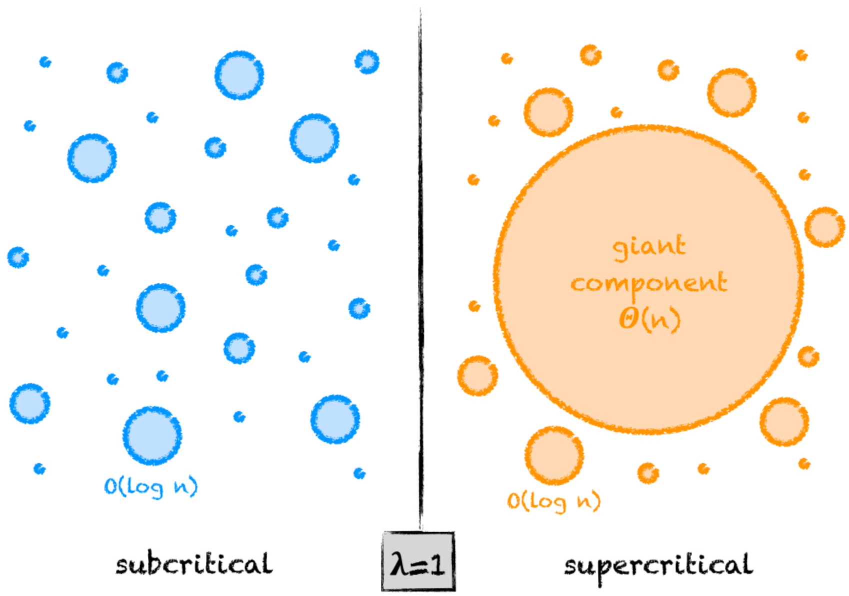

Figure 6.4: Illustration of the phase transition.

6.4.2 Bounding cluster size: domination by branching processes

For a vertex $v \in [n]$, let $\mathcal{C}_v$ be the connected component containing v, which we also refer to as the *cluster* of v. To analyze the size of $\mathcal{C}_v$, we use the exploration process introduced in Section 6.2.1 and show that it is dominated above and below by branching processes.

cluster

Exploration process

Recall that the exploration process started at v has 3 types of vertices: the active vertices $\mathcal{A}_t$, the explored vertices $\mathcal{E}_t$, and the neutral vertices $\mathcal{N}_t$. We start with $\mathcal{A}_0 := \{v\}$, $\mathcal{E}_0 := \emptyset$, and $\mathcal{N}_0$ contains all other vertices in G_n. We imagine revealing the edges of G_n as they are encountered in this process and we let $(\mathcal{F}_t)$ be the corresponding filtration. In words, starting with v, the cluster of v is progressively grown by adding to it at each time a vertex adjacent to one of the previously explored vertices and uncovering its remaining neighbors in G_n.

Let as before $A_t := |\mathcal{A}_t|$, $E_t := |\mathcal{E}_t|$, and $N_t := |\mathcal{N}_t|$, and

$$\tau_0 := \inf\{t \geq 0 : A_t = 0\} = |\mathcal{C}_v|,$$

where the rightmost equality is from Lemma 6.2.1. Recall that (E_t) is non-decreasing while (N_t) is non-increasing, and that the process is fixed for all $t > \tau_0$. Since $E_t = t$ for all $t \leq \tau_0$ (as exactly one vertex is explored at each time until the set of active vertices is empty) and $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t)$ forms a partition of $[n]$ for all t, we have

$$A_t + t + N_t = n, \qquad \forall t \leq \tau_0. \tag{6.4.4}$$

Hence, in tracking the size of the exploration process, we can work with A_t or N_t. Moreover at $t = \tau_0$ we have

$$|\mathcal{C}_v| = \tau_0 = n - N_{\tau_0}. \tag{6.4.5}$$

Similarly to the case of a Galton-Watson tree, the processes (A_t) and (N_t) admit a simple recursive form. Conditioning on $\mathcal{F}_{t-1}$:

- *(Active vertices)* If $A_{t-1} = 0$, the exploration process has finished its course and $A_t = 0$. Otherwise, (a) one active vertex becomes explored and (b) its neutral neighbors become active vertices. That is,

$$A_t = A_{t-1} + \mathbf{1}_{\{A_{t-1}>0\}}\big[\underbrace{-1}_{\text{(a)}} + \underbrace{X_t}_{\text{(b)}}\big], \tag{6.4.6}$$

  where X_t is binomial with parameters N_{t-1} and p_n. By (6.4.4), N_{t-1} can be written in terms of A_{t-1} as $N_{t-1} = n - (t-1) - A_{t-1}$. For the coupling arguments below, it will be useful to think of X_t as a sum of independent Bernoulli variables. That is, let $(I_{t,j} \,:\, t \geq 1, j \geq 1)$ be an array of independent, identically distributed $\{0,1\}$-variables with $\mathbb{P}[I_{1,1} = 1] = p_n$. We write

$$X_t = \sum_{i=1}^{N_{t-1}} I_{t,i}. \tag{6.4.7}$$

- *(Neutral vertices)* Similarly, if $A_{t-1} > 0$, that is, $N_{t-1} < n - (t-1)$, X_t neutral vertices become active. That is,

$$N_t = N_{t-1} - \mathbf{1}_{\{N_{t-1}<n-(t-1)\}}\, X_t. \tag{6.4.8}$$

Poisson branching process approximation

With these observations, we now relate the size of the cluster of v to the total progeny of a Poisson branching process with an appropriately chosen offspring mean. The intuition is simple: when $p_n = \lambda/n$, the number of neighbors of a vertex is well approximated by a Poisson distribution; therefore, exploration of the cluster of v is similar to that of the corresponding branching process. We will see that this holds long enough to prove accurate results about the subcritical regime (see Lemma 6.4.6). It will also be useful in the supercritical regime, but additional arguments will be required there (see Lemmas 6.4.7 and 6.4.8).

Lemma 6.4.3 (Cluster size: Poisson branching process approximation)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda > 0$ and let $\mathcal{C}_v$ be the connected component of $v \in [n]$. Let W_λ be the total progeny of a branching process with offspring distribution* $\mathrm{Poi}(\lambda)$. *Then, for $1 \leq k_n = o(\sqrt{n})$,*

$$\mathbb{P}[W_\lambda \geq k_n] - O\left(\frac{k_n^2}{n}\right) \leq \mathbb{P}_{n,p_n}[|\mathcal{C}_v| \geq k_n] \leq \mathbb{P}[W_\lambda \geq k_n].$$

From Example 6.2.7, we have an explicit formula for the distribution of W_λ.

Before proving the lemma, recall the following simple domination results from Chapter 4:

- *(Binomial domination)* We have

$$n \geq m \implies \mathrm{Bin}(n,p) \succeq \mathrm{Bin}(m,p). \tag{6.4.9}$$

  The binomial distribution is also dominated by the Poisson distribution in the following way:

$$\lambda \in (0,1) \quad \implies \quad \mathrm{Poi}(\lambda) \succeq \mathrm{Bin}\left(n-1, \frac{\lambda}{n}\right). \tag{6.4.10}$$

  For the proofs, see Examples 4.2.4 and 4.2.8.

We use these domination results to relate the size of a connected component to the progeny of a branching process.

Proof of Lemma 6.4.3. We start with the upper bound.

Upper bound: Because $N_{t-1} = n-(t-1)-A_{t-1} \leq n-1$, conditioned on $\mathcal{F}_{t-1}$, the following stochastic domination relations hold

$$\mathrm{Bin}\left(N_{t-1}, \frac{\lambda}{n}\right) \preceq \mathrm{Bin}\left(n-1, \frac{\lambda}{n}\right) \preceq \mathrm{Poi}(\lambda),$$

by (6.4.9) and (6.4.10). Observe that the center and rightmost distributions do not depend on N_{t-1}. Let $(X_t^{\succ})$ be a sequence of independent $\mathrm{Poi}(\lambda)$.

Using the coupling in Example 4.2.8, we can couple the processes $(I_{t,j})_j$ and $(X_t^{\succ})$ in such way that $X_t^{\succ} \geq \sum_{j=1}^{n-1} I_{t,j}$ almost surely for all t. Then by induction on t,

$$A_t \leq A_t^{\succ},$$

almost surely for all t, where we define (recalling (6.4.6))

$$A_t^{\succ} := A_{t-1}^{\succ} + \mathbf{1}_{\{A_{t-1}^{\succ}>0\}}\left[-1 + X_t^{\succ}\right], \tag{6.4.11}$$

with $A_0^{\succ} := 1$. In words, $(A_t^{\succ})$ is the size of the active set of a *Galton-Watson branching process with offspring distribution* $\mathrm{Poi}(\lambda)$, as defined in Section 6.2.1.

As a result, letting

$$W_\lambda = \tau_0^{\succ} := \inf\{t \geq 0 \,:\, A_t^{\succ} = 0\},$$

be the total progeny of this branching process, we immediately get

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| \geq k_n] = \mathbb{P}_{n,p_n}[\tau_0 \geq k_n] \leq \mathbb{P}[\tau_0^{\succ} \geq k_n] = \mathbb{P}[W_\lambda \geq k_n].$$

Lower bound: In the other direction, we proceed in two steps. We first show that, up to a certain time, the process is bounded from below by a branching process with binomial offspring distribution. In a second step, we show that this binomial branching process can be approximated by a Poisson branching process.

1. *(Domination from below)* Let $A_t^{\prec}$ be defined as (again recalling (6.4.6))

$$A_t^{\prec} := A_{t-1}^{\prec} + \mathbf{1}_{\{A_{t-1}^{\prec}>0\}}\left[-1 + X_t^{\prec}\right], \tag{6.4.12}$$

with $A_0^{\prec} := 1$, where

$$X_t^{\prec} := \sum_{i=1}^{n-k_n} I_{t,j}. \tag{6.4.13}$$

Note that we use the same $I_{t,j}$s as in the definition of X_t, that is, we couple the two processes. This time $(A_t^{\prec})$ is the size of the active set in the exploration process of a Galton-Watson branching process with offspring distribution $\mathrm{Bin}(n - k_n, p_n)$. Let

$$\tau_0^{\prec} := \inf\{t \geq 0 \,:\, A_t^{\prec} = 0\},$$

be the total progeny of this branching process. We prove the following relationship between τ_0 and $\tau_0^{\prec}$.

Lemma 6.4.4. *We have*

$$\mathbb{P}[\tau_0^{\prec} \geq k_n] \leq \mathbb{P}_{n,p_n}[\tau_0 \geq k_n].$$

Proof. We claim that A_t is bounded from below by $A_t^{\prec}$ *up to the stopping time*

$$\sigma_{n-k_n} := \inf\{t \geq 0 \,:\, N_t \leq n - k_n\},$$

which by convention is $+\infty$ if the event is not reached (i.e., if the cluster is "small"; see below). Indeed, $N_0 = n - 1$ and for all $t \leq \sigma_{n-k_n}$, $N_{t-1} > n - k_n$ by definition. Hence, by the coupling (6.4.7) and (6.4.13), $X_t \geq X_t^{\prec}$ for all $t \leq \sigma_{n-k_n}$ and as a result, by induction on t,

$$A_t \geq A_t^{\prec}, \qquad \forall t \leq \sigma_{n-k_n},$$

where we used the recursions (6.4.6) and (6.4.13).

Because the inequality between A_t and $A_t^{\prec}$ holds only up to time σ_{n-k_n}, we cannot compare directly τ_0 and $\tau_0^{\prec}$. However, we will use the following observation: the size of the cluster of v is at least the total number of active and explored vertices at any time t. In particular, when $\sigma_{n-k_n} < +\infty$,

$$\tau_0 = |\mathcal{C}_v| \geq A_{\sigma_{n-k_n}} + E_{\sigma_{n-k_n}} = n - N_{\sigma_{n-k_n}} \geq k_n.$$

On the other hand, when $\sigma_{n-k_n} = +\infty$, $N_t > n - k_n$ for all t—in particular for $t = \tau_0$—and therefore $|\mathcal{C}_v| = \tau_0 = n - N_{\tau_0} < k_n$ by (6.4.5). Moreover in that case, because $A_t \geq A_t^{\prec}$ for all $t \leq \sigma_{n-k_n} = +\infty$, it holds in addition that $\tau_0^{\prec} \leq \tau_0 < k_n$. To sum up, we have proved the implications

$$\tau_0^{\prec} \geq k_n \implies \sigma_{n-k_n} < +\infty \implies \tau_0 \geq k_n.$$

In particular, we have proved the lemma. ■

2. *(Poisson approximation)* Our next step is approximate the tail of $\tau_0^{\prec}$ by that of $\tau_0^{\succ}$.

Lemma 6.4.5. *We have*

$$\mathbb{P}[\tau_0^{\prec} \geq k_n] = \mathbb{P}[\tau_0^{\succ} \geq k_n] + O\left(\frac{k_n^2}{n}\right).$$

Proof. By Theorem 6.2.6,

$$\mathbb{P}[\tau_0^{\prec} = t] = \frac{1}{t}\mathbb{P}\left[\sum_{i=1}^{t} X_i^{\prec} = t - 1\right], \tag{6.4.14}$$

where the $X_i^{\prec}$s are independent $\mathrm{Bin}(n-k_n, p_n)$. Note further that, because the sum of independent binomials with the same success probability is binomial,

$$\sum_{i=1}^{t} X_i^{\prec} \sim \mathrm{Bin}(t(n-k_n), p_n).$$

Recall on the other hand that $(X_t^{\succ})$ is $\mathrm{Poi}(\lambda)$ and, because a sum of independent Poisson is Poisson (see Exercise 6.7), we have

$$\mathbb{P}[\tau_0^{\succ} = t] = \frac{1}{t}\mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} = t-1\right], \tag{6.4.15}$$

where

$$\sum_{i=1}^{t} X_i^{\succ} \sim \mathrm{Poi}(t\lambda).$$

We use the Poisson approximation result in Theorem 4.1.18 to compare the probabilities on the right-hand sides of (6.4.14) and (6.4.15). In fact, because the Poisson approximation is in terms of the total variation distance—which bounds any event—one might be tempted to apply it directly to the tails of $\tau_0^{\prec}$ and $\tau_0^{\succ}$ by summing over t. However note that the factor of $1/t$ in (6.4.14) and (6.4.15) prevents us from doing so.

Instead, we argue for each t separately and use that

$$\begin{aligned}\left|\mathbb{P}\left[\sum_{i=1}^{t} X_i^{\prec} = t-1\right] - \mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} = t-1\right]\right| \\ \leq \|\mathrm{Bin}(t(n-k_n), p_n) - \mathrm{Poi}(t\lambda)\|_{\mathrm{TV}},\end{aligned}$$

by the observations in the previous paragraph. Theorem 4.1.18 tells us that

$$\begin{aligned}\|\mathrm{Bin}(t(n-k_n), p_n) - \mathrm{Poi}\left(t(n-k_n)[-\log(1-p_n)]\right)\|_{\mathrm{TV}} \\ \leq \frac{1}{2}t(n-k_n)[-\log(1-p_n)]^2.\end{aligned}$$

We must adjust the mean of the Poisson distribution. To do so, we argue as in Example 4.1.12 to get

$$\begin{aligned}\|\mathrm{Poi}\left(t(n-k_n)[-\log(1-p_n)]\right) - \mathrm{Poi}\left(t\lambda\right)\|_{\mathrm{TV}} \\ \leq |t\lambda - t(n-k_n)(-\log(1-p_n))|\,.\end{aligned}$$

Finally, recalling that $p_n = \lambda/n$, combining the last three displays and using the triangle inequality for the total variation distance,

$$\begin{aligned}
&\left|\mathbb{P}\left[\sum_{i=1}^{t} X_i^{\prec} = t-1\right] - \mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} = t-1\right]\right| \\
&\qquad \leq \frac{1}{2} t(n-k_n)[-\log(1-p_n)]^2 + |t\lambda - t(n-k_n)(-\log(1-p_n))| \\
&\qquad \leq \frac{1}{2} tn \left(\frac{\lambda}{n} + O\left(\frac{\lambda^2}{n^2}\right)\right)^2 + \left|t\lambda - t(n-k_n)\left(\frac{\lambda}{n} + O\left(\frac{\lambda^2}{n^2}\right)\right)\right| \\
&\qquad = O\left(\frac{tk_n}{n}\right),
\end{aligned}$$

where we used that $k_n \geq 1$ and λ is fixed.

So, by (6.4.14) and (6.4.15), dividing by t and then summing over $t < k_n$ gives

$$\left|\mathbb{P}[\tau_0^{\prec} < k_n] - \mathbb{P}[\tau_0^{\succ} < k_n]\right| = O\left(\frac{k_n^2}{n}\right).$$

Rearranging proves the lemma. ∎

Putting together Lemmas 6.4.4 and 6.4.5 gives

$$\begin{aligned}
\mathbb{P}_{n,p_n}[|\mathcal{C}_v| \geq k_n] &= \mathbb{P}_{n,p_n}[\tau_0 \geq k_n] \\
&\geq \mathbb{P}[\tau_0^{\succ} \geq k_n] - O\left(\frac{k_n^2}{n}\right) \\
&= \mathbb{P}[W_\lambda \geq k_n] - O\left(\frac{k_n^2}{n}\right),
\end{aligned}$$

as claimed. ∎

Subcritical regime: largest cluster

We are now ready to analyze the subcritical regime, that is, the case $\lambda < 1$.

Lemma 6.4.6 (Subcritical regime: upper bound on cluster size)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda \in (0,1)$ and let $\mathcal{C}_v$ be the connected component of $v \in [n]$. For all $\kappa > 0$,*

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| \geq (1+\kappa) I_\lambda^{-1} \log n\right] = O(n^{-(1+\kappa)}).$$

Proof. We use the Poisson branching process approximation (Lemma 6.4.3). To apply the lemma we need to bound the tail of the progeny W_λ of a Poisson branching process. Using the notation of Lemma 6.4.3, by Theorem 6.2.6,

$$\mathbb{P}\left[W_\lambda \geq k_n\right] = \mathbb{P}\left[W_\lambda = +\infty\right] + \sum_{t \geq k_n} \frac{1}{t} \mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} = t - 1\right], \tag{6.4.16}$$

where the $X_i^{\succ}$s are i.i.d. $\mathrm{Poi}(\lambda)$. Both terms on the right-hand side depend on whether or not the mean λ is smaller or larger than 1. When $\lambda < 1$, the Poisson branching process goes extinct with probability 1 by the extinction theory (Theorem 6.1.6). Hence $\mathbb{P}[W_\lambda = +\infty] = 0$.

As to the second term, the sum of the $X_i^{\succ}$s is $\mathrm{Poi}(\lambda t)$. Using the Poisson tail (6.4.1) for $\lambda < 1$ and $k_n = \omega(1)$,

$$\begin{aligned}
\sum_{t \geq k_n} \frac{1}{t} \mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} = t - 1\right] &\leq \sum_{t \geq k_n} \mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} \geq t - 1\right] \\
&\leq \sum_{t \geq k_n} \exp\left(-t I_\lambda^{\mathrm{Poi}}\left(\frac{t-1}{t}\right)\right) \\
&\leq \sum_{t \geq k_n} \exp\left(-t(I_\lambda - O(t^{-1}))\right) \\
&\leq \sum_{t \geq k_n} C \exp\left(-t I_\lambda\right) \\
&= O\left(\exp\left(-I_\lambda k_n\right)\right),
\end{aligned} \tag{6.4.17}$$

for some constant $C > 0$.

Let $c = (1 + \kappa) I_\lambda^{-1}$ for $\kappa > 0$. By Lemma 6.4.3,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| \geq c \log n\right] \quad \leq \quad \mathbb{P}\left[W_\lambda \geq c \log n\right].$$

By (6.4.16) and (6.4.17),

$$\mathbb{P}\left[W_\lambda \geq c \log n\right] \quad = \quad O\left(\exp\left(-I_\lambda c \log n\right)\right), \tag{6.4.18}$$

which proves the claim. ■

As before, let $\mathcal{C}_{\max}$ be a largest connected component of G_n (choosing the component containing the lowest label if there is more than one such component). A union bound and the previous lemma immediately imply an upper bound on the size of $\mathcal{C}_{\max}$ in the subcritical case.

Proof of Theorem 6.4.1. Let again $c = (1+\kappa)I_\lambda^{-1}$ for $\kappa > 0$. By a union bound and symmetry,

$$\begin{aligned} \mathbb{P}_{n,p_n}\left[|\mathcal{C}_{\max}| \geq c\log n\right] &= \mathbb{P}_{n,p_n}\left[\exists v,\ |\mathcal{C}_v| > c\log n\right] \\ &\leq n\,\mathbb{P}_{n,p_n}\left[|\mathcal{C}_1| \geq c\log n\right]. \end{aligned} \tag{6.4.19}$$

By Lemma 6.4.6,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_{\max}| \geq c\log n\right] = O(n \cdot n^{-(1+\kappa)}) = O(n^{-\kappa}) \to 0,$$

as $n \to +\infty$. ∎

In fact we prove below that the largest component is indeed of size roughly $I_\lambda^{-1}\log n$. But first we turn to the supercritical regime.

Supercritical regime: two phases

Applying the Poisson branching process approximation in the supercritical regime gives the following.

Lemma 6.4.7 (Supercritical regime: extinction)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda > 1$. and let $\mathcal{C}_v$ be the connected component of $v \in [n]$. Let ζ_λ be the unique solution in $(0,1)$ to the fixed point equation*

$$1 - e^{-\lambda\zeta} = \zeta.$$

For any $\kappa > 0$,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| \geq (1+\kappa)I_\lambda^{-1}\log n\right] = \zeta_\lambda + O\left(\frac{\log^2 n}{n}\right).$$

Note the small—but critical difference—with Lemma 6.4.6: this time the branching process can survive. This happens with probability ζ_λ by extinction theory (Theorem 6.1.6). In that case, we will need further arguments to nail down the cluster size. Observe also that the result holds for a *fixed* vertex v—and therefore does not yet tell us about the *largest* cluster. We come back to the latter in the next subsection.

Proof of Lemma 6.4.7. We adapt the proof of Lemma 6.4.6, beginning with (6.4.16) which recall states

$$\mathbb{P}\left[W_\lambda \geq k_n\right] = \mathbb{P}\left[W_\lambda = +\infty\right] + \sum_{t \geq k_n} \frac{1}{t}\mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} = t-1\right],$$

where the $X_i^{\succ}$s are i.i.d. $\mathrm{Poi}(\lambda)$. When $\lambda > 1$, $\mathbb{P}[W_\lambda = +\infty] = \zeta_\lambda$, where $\zeta_\lambda > 0$ is the survival probability of the branching process by Example 6.1.10. As to the second term, using (6.4.2) for $\lambda > 1$,

$$\begin{aligned}\sum_{t \geq k_n} \frac{1}{t} \mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} = t-1\right] &\leq \sum_{t \geq k_n} \mathbb{P}\left[\sum_{i=1}^{t} X_i^{\succ} \leq t\right] \\ &\leq \sum_{t \geq k_n} \exp\left(-t I_\lambda\right) \\ &\leq C \exp\left(-I_\lambda k_n\right), \qquad (6.4.20)\end{aligned}$$

for a constant $C > 0$.

Now let $c = (1+\kappa) I_\lambda^{-1}$ for $\kappa > 0$. By Lemma 6.4.3,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| \geq c \log n\right] = \mathbb{P}\left[W_\lambda \geq c \log n\right] + O\left(\frac{\log^2 n}{n}\right). \qquad (6.4.21)$$

By (6.4.16) and (6.4.20),

$$\begin{aligned}\mathbb{P}\left[W_\lambda \geq c \log n\right] &= \zeta_\lambda + O\left(\exp\left(-c I_\lambda \log n\right)\right) \\ &= \zeta_\lambda + O(n^{-(1+\kappa)}). \qquad (6.4.22)\end{aligned}$$

Combining (6.4.21) and (6.4.22), for any $\kappa > 0$,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| \geq c \log n\right] = \zeta_\lambda + O\left(\frac{\log^2 n}{n}\right), \qquad (6.4.23)$$

as claimed. ■

Recall that the Poisson branching process approximation was based on the fact that the degree of a vertex is well approximated by a Poisson distribution. When the exploration process goes on for too long however (i.e., when k_n is large), this approximation is not as accurate because of a saturation effect: at each step of the exploration, we uncover edges *to the neutral vertices* (which then become active); and, because an Erdős-Rényi graph has a finite pool of vertices from which to draw these edges, as the number of neutral vertices decreases so does the expected number of uncovered edges. Instead we use the following lemma which explicitly accounts for the dwindling size of $\mathcal{N}_t$. Roughly speaking, we model the set of neutral vertices as a process that discards a fraction p_n of its current set at each time step (i.e., those neutral vertices with an edge to the current explored vertex).

Lemma 6.4.8. *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda > 0$ and let $\mathcal{C}_v$ be the connected component of $v \in [n]$. Let $Y_t \sim \mathrm{Bin}(n-1, 1-(1-p_n)^t)$. Then, for any t,*

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| = t] \leq \mathbb{P}[Y_t = t-1].$$

Proof. We work with neutral vertices. By (6.4.4) and Lemma 6.2.1, for any t,

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| = t] = \mathbb{P}_{n,p_n}[\tau_0 = t] \leq \mathbb{P}_{n,p_n}[N_t = n-t]. \tag{6.4.24}$$

Recall that $N_0 = n-1$ and

$$N_t = N_{t-1} - \mathbf{1}_{\{N_{t-1} < n-(t-1)\}} \sum_{i=1}^{N_{t-1}} I_{t,i}.$$

It is easier to consider the process *without the indicator* as it has a simple distribution. Define $N_0^0 := n-1$ and

$$N_t^0 := N_{t-1}^0 - \sum_{i=1}^{N_{t-1}^0} I_{t,i},$$

and observe that $N_t \geq N_t^0$ for all t, as the two processes agree up to time τ_0 at which point N_t stays fixed. The interpretation of N_t^0 is straightforward: starting with $n-1$ vertices, at each time each remaining vertex is discarded with probability p_n. Hence, the number of surviving vertices at time t has distribution

$$N_t^0 \sim \mathrm{Bin}(n-1, (1-p_n)^t),$$

by the independence of the steps. Arguing as in (6.4.24),

$$\begin{aligned}
\mathbb{P}_{n,p_n}[|\mathcal{C}_v| = t] &\leq \mathbb{P}_{n,p_n}[N_t^0 = n-t] \\
&= \mathbb{P}_{n,p_n}[(n-1) - N_t^0 = t-1] \\
&= \mathbb{P}[Y_t = t-1],
\end{aligned}$$

which concludes the proof. ■

The previous lemma gives the following additional bound on the cluster size in the supercritical regime. Together with Lemma 6.4.7 it shows that, when $|\mathcal{C}_v| > c \log n$, the cluster size is in fact linear in n with high probability. We will have more to say about the largest cluster in the next subsection.

Lemma 6.4.9 (Supercritical regime: saturation)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda > 1$ and let $\mathcal{C}_v$ be the connected component of $v \in [n]$. Let ζ_λ be the unique solution in $(0,1)$ to the fixed point equation*

$$1 - e^{-\lambda\zeta} = \zeta.$$

For any $\alpha < \zeta_\lambda$ and any $\delta > 0$, there exists $\kappa_{\delta,\alpha} > 0$ large enough so that

$$\mathbb{P}_{n,p_n}\left[(1+\kappa_{\delta,\alpha})I_\lambda^{-1}\log n \le |\mathcal{C}_v| \le \alpha n\right] = O(n^{-(1+\delta)}). \tag{6.4.25}$$

Proof. By Lemma 6.4.8,

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| = t] \le \mathbb{P}[Y_t = t-1] \le \mathbb{P}[Y_t \le t],$$

where $Y_t \sim \mathrm{Bin}(n-1, 1-(1-p_n)^t)$. Roughly, the right-hand side is negligible until the mean $\mu_t := (n-1)(1-(1-\lambda/n)^t)$ is of the order of t. Let ζ_λ be as above, and recall that it is a solution to

$$1 - e^{-\lambda\zeta} - \zeta = 0.$$

Note in particular that, when $t = \zeta_\lambda n$,

$$\mu_t = (n-1)(1-(1-\lambda/n)^{\zeta_\lambda n}) \approx n(1-e^{-\lambda\zeta_\lambda}) = \zeta_\lambda n = t.$$

Let $\alpha < \zeta_\lambda$.

For any $t \in [c\log n, \alpha n]$, by the Chernoff bound for Poisson trials (Theorem 2.4.7 (ii)(b)),

$$\mathbb{P}[Y_t \le t] \le \exp\left(-\frac{\mu_t}{2}\left(1-\frac{t}{\mu_t}\right)^2\right). \tag{6.4.26}$$

For $t/n \le \alpha < \zeta_\lambda$, using $1-x \le e^{-x}$ for $x \in (0,1)$ (see Exercise 1.16), there is $\gamma_\alpha > 1$ such that

$$\begin{aligned}
\mu_t &\ge (n-1)(1-e^{-\lambda(t/n)}) \\
&= t\left(\frac{n-1}{n}\right)\frac{1-e^{-\lambda(t/n)}}{t/n} \\
&\ge t\left(\frac{n-1}{n}\right)\frac{1-e^{-\lambda\alpha}}{\alpha} \\
&\ge \gamma_\alpha t,
\end{aligned}$$

for n large enough, where we used that $1-e^{-\lambda x}$ is increasing in x on the third line and that $1-e^{-\lambda x} - x > 0$ for $0 < x < \zeta_\lambda$ on the fourth line (as can be checked by computing the first and second derivatives). Plugging this back into (6.4.26), we get

$$\mathbb{P}[Y_t \le t] \le \exp\left(-t\left\{\frac{\gamma_\alpha}{2}\left(1-\frac{1}{\gamma_\alpha}\right)^2\right\}\right).$$

Therefore

$$\begin{aligned}\sum_{t=c\log n}^{\alpha n} \mathbb{P}_{n,p_n}[|\mathcal{C}_v| = t] &\le \sum_{t=c\log n}^{\alpha n} \mathbb{P}[Y_t \le t]\\ &\le \sum_{t=c\log n}^{+\infty} \exp\left(-t\left\{\frac{\gamma_\alpha}{2}\left(1-\frac{1}{\gamma_\alpha}\right)^2\right\}\right)\\ &= O\left(\exp\left(-c\log n\left\{\frac{\gamma_\alpha}{2}\left(1-\frac{1}{\gamma_\alpha}\right)^2\right\}\right)\right).\end{aligned}$$

Taking $\kappa > 0$ large enough proves (6.4.25). ■

6.4.3 Concentration of cluster size: second moment bounds

To characterize the size of the largest cluster in the supercritical case, we use Chebyshev's inequality. We also use a related second moment argument to give a lower bound on the largest cluster in the subcritical regime.

Supercritical regime: giant component

Assume $\lambda > 1$. Our goal is to characterize the size of the largest component. We do this by bounding what is *not* in it (i.e., intuitively those vertices whose exploration process goes extinct). For $\delta > 0$ and $\alpha < \zeta_\lambda$, let $\kappa_{\delta,\alpha}$ be as defined in Lemma 6.4.9. Set

$$\underline{k}_n := (1+\kappa_{\delta,\alpha})I_\lambda^{-1}\log n \qquad \text{and} \qquad \bar{k}_n := \alpha n.$$

We call a vertex v such that $|\mathcal{C}_v| \le \underline{k}_n$ a *small vertex*. *small vertex*

Let

$$S_k := \sum_{v\in[n]} \mathbf{1}_{\{|\mathcal{C}_v|\le k\}}.$$

It will also be useful to work with

$$B_k = n - S_k = \sum_{v\in[n]} \mathbf{1}_{\{|\mathcal{C}_v|>k\}}.$$

The quantity $S_{\underline{k}_n}$ is the number of small vertices. By Lemma 6.4.7, its expectation is

$$\mathbb{E}_{n,p_n}[S_{\underline{k}_n}] = n(1 - \mathbb{P}_{n,p_n}[|\mathcal{C}_v| > \underline{k}_n]) = (1-\zeta_\lambda)n + O\left(\log^2 n\right). \qquad (6.4.27)$$

Using Chebyshev's inequality (Theorem 2.1.2), we prove that $S_{\underline{k}_n}$ is concentrated.

Lemma 6.4.10 (Concentration of $S_{\underline{k}_n}$)**.** *For any* $\gamma \in (1/2, 1)$ *and* $\delta < 2\gamma - 1$*,*

$$\mathbb{P}_{n,p_n}[|S_{\underline{k}_n} - (1-\zeta_\lambda)n| \geq n^\gamma] = O(n^{-\delta}).$$

Lemma 6.4.10, which is proved below, leads to our main result in the supercritical case: the existence of the *giant component*, a unique cluster $\mathcal{C}_{\max}$ of size linear in n.

giant component

Proof of Theorem 6.4.2. Take $\alpha \in (\zeta_\lambda/2, \zeta_\lambda)$ and let $\underline{k}_n$, $\bar{k}_n$, and γ be as above. Let $\mathcal{B}_{1,n} := \{|B_{\underline{k}_n} - \zeta_\lambda n| \geq n^\gamma\}$. Because $\gamma < 1$, the event $\mathcal{B}_{1,n}^c$ implies that

$$\sum_{v\in[n]} \mathbf{1}_{\{|\mathcal{C}_v| > \underline{k}_n\}} = B_{\underline{k}_n} > \zeta_\lambda n - n^\gamma \geq 1,$$

for n large enough. That is, there is at least one "large" cluster of size $> \underline{k}_n$. In turn, that implies

$$|\mathcal{C}_{\max}| \leq B_{\underline{k}_n},$$

since there are at most $B_{\underline{k}_n}$ vertices in that large cluster.

Let $\mathcal{B}_{2,n} := \{\exists v, |\mathcal{C}_v| \in [\underline{k}_n, \bar{k}_n]\}$. If $\mathcal{B}_{2,n}^c$ holds, in addition to $\mathcal{B}_{1,n}^c$, then

$$|\mathcal{C}_{\max}| \leq B_{\underline{k}_n} = B_{\bar{k}_n},$$

since there is no cluster whose size falls in $[\underline{k}_n, \bar{k}_n]$. Moreover there is equality across the last display if there is a unique cluster of size greater than $\bar{k}_n$.

This is indeed the case under $\mathcal{B}_{1,n}^c \cap \mathcal{B}_{2,n}^c$: if there were two distinct clusters of size $\bar{k}_n$, then since $2\alpha > \zeta_\lambda$ we would have for n large enough

$$B_{\underline{k}_n} = B_{\bar{k}_n} > 2\bar{k}_n = 2\alpha n > \zeta_\lambda n + n^\gamma,$$

a contradiction. Hence we have proved that under $\mathcal{B}_{1,n}^c \cap \mathcal{B}_{2,n}^c$

$$|\mathcal{C}_{\max}| = B_{\underline{k}_n} = B_{\bar{k}_n}.$$

Take $\delta < 2\gamma - 1$. Applying Lemmas 6.4.9 and 6.4.10

$$\mathbb{P}[\mathcal{B}_{1,n} \cup \mathcal{B}_{2,n}] \leq O(n^{-\delta}) + n \cdot O(n^{-(1+\delta)}) = O(n^{-\delta}),$$

which concludes the proof. ■

It remains to prove Lemma 6.4.10.

Proof of Lemma 6.4.10. As mentioned above, we use Chebyshev's inequality. Hence our main task is to bound the variance of $S_{\underline{k}_n}$.

Our starting point is the following expression for the second moment

$$\begin{aligned}\mathbb{E}_{n,p_n}[S_k^2] &= \sum_{u,v\in[n]} \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \leq k, |\mathcal{C}_v| \leq k] \\ &= \sum_{u,v\in[n]} \big\{\mathbb{P}_{n,p_n}[|\mathcal{C}_u| \leq k, |\mathcal{C}_v| \leq k, u \leftrightarrow v] \\ &\qquad +\mathbb{P}_{n,p_n}[|\mathcal{C}_u| \leq k, |\mathcal{C}_v| \leq k, u \not\leftrightarrow v]\big\}, \quad (6.4.28)\end{aligned}$$

where $u \leftrightarrow v$ indicates that u and v are in the same connected component.

To bound the first term in (6.4.28), we note that $u \leftrightarrow v$ implies that $\mathcal{C}_u = \mathcal{C}_v$. Hence,

$$\begin{aligned}\sum_{u,v\in[n]} \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \leq k, |\mathcal{C}_v| \leq k, u \leftrightarrow v] &= \sum_{u,v\in[n]} \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \leq k, v \in \mathcal{C}_u] \\ &= \sum_{u,v\in[n]} \mathbb{E}_{n,p_n}[\mathbf{1}_{\{|\mathcal{C}_u|\leq k\}}\mathbf{1}_{\{v\in\mathcal{C}_u\}}] \\ &= \sum_{u\in[n]} \mathbb{E}_{n,p_n}\left[\mathbf{1}_{\{|\mathcal{C}_u|\leq k\}}\sum_{v\in[n]}\mathbf{1}_{\{v\in\mathcal{C}_u\}}\right] \\ &= \sum_{u\in[n]} \mathbb{E}_{n,p_n}[|\mathcal{C}_u|\mathbf{1}_{\{|\mathcal{C}_u|\leq k\}}] \\ &= n\,\mathbb{E}_{n,p_n}[|\mathcal{C}_1|\mathbf{1}_{\{|\mathcal{C}_1|\leq k\}}] \\ &\leq nk. \quad (6.4.29)\end{aligned}$$

To bound the second term in (6.4.28), we sum over the size of $\mathcal{C}_u$ and note that, conditioned on $\{|\mathcal{C}_u| = \ell, u \not\leftrightarrow v\}$, the size of $\mathcal{C}_v$ has the same distribution as the unconditional size of $\mathcal{C}_1$ in a $\mathbb{G}_{n-\ell,p_n}$ random graph, that is,

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| \leq k \,|\, |\mathcal{C}_u| = \ell, u \not\leftrightarrow v] = \mathbb{P}_{n-\ell,p_n}[|\mathcal{C}_1| \leq k].$$

Observe that the probability on the right-hand side is increasing in ℓ (as can be seen, e.g., by coupling; see below for a related argument). Hence

$$\begin{aligned}
&\sum_{u,v\in[n]}\sum_{\ell\le k}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|=\ell,|\mathcal{C}_v|\le k,u\not\leftrightarrow v]\\
&\quad=\sum_{u,v\in[n]}\sum_{\ell\le k}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|=\ell,u\not\leftrightarrow v]\,\mathbb{P}_{n,p_n}[|\mathcal{C}_v|\le k\,|\,|\mathcal{C}_u|=\ell,u\not\leftrightarrow v]\\
&\quad=\sum_{u,v\in[n]}\sum_{\ell\le k}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|=\ell,u\not\leftrightarrow v]\,\mathbb{P}_{n-\ell,p_n}[|\mathcal{C}_v|\le k]\\
&\quad\le\sum_{u,v\in[n]}\sum_{\ell\le k}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|=\ell]\,\mathbb{P}_{n-k,p_n}[|\mathcal{C}_v|\le k]\\
&\quad=\sum_{u,v\in[n]}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|\le k]\,\mathbb{P}_{n-k,p_n}[|\mathcal{C}_v|\le k].
\end{aligned}$$

To get a bound on the variance of S_k, we need to relate this last expression to $(\mathbb{E}_{n,p_n}[S_k])^2$, where we will use that

$$\mathbb{E}_{n,p_n}[S_k]=\mathbb{E}_{n,p_n}\left[\sum_{v\in[n]}\mathbf{1}_{\{|\mathcal{C}_v|\le k\}}\right]=\sum_{v\in[n]}\mathbb{P}_{n,p_n}[|\mathcal{C}_v|\le k]. \tag{6.4.30}$$

We define

$$\Delta_k:=\mathbb{P}_{n-k,p_n}[|\mathcal{C}_1|\le k]-\mathbb{P}_{n,p_n}[|\mathcal{C}_1|\le k].$$

Then, plugging this back above, we get

$$\begin{aligned}
&\sum_{u,v\in[n]}\sum_{\ell\le k}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|=\ell,|\mathcal{C}_v|\le k,u\not\leftrightarrow v]\\
&\quad\le\sum_{u,v\in[n]}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|\le k](\mathbb{P}_{n,p_n}[|\mathcal{C}_v|\le k]+\Delta_k)\\
&\quad\le(\mathbb{E}_{n,p_n}[S_k])^2+n^2|\Delta_k|,
\end{aligned}$$

by (6.4.30). It remains to bound Δ_k.

We use a coupling argument. Let $H\sim\mathbb{G}_{n-k,p_n}$ and construct $H'\sim\mathbb{G}_{n,p_n}$ in the following manner: let H' coincide with H on the first $n-k$ vertices then pick the rest the edges independently. Then clearly $\Delta_k\ge 0$ since the cluster of 1 in H' includes the cluster of 1 in H. In fact, Δ_k is the probability that under this coupling the cluster of 1 has at most k vertices in H but not in H'. That implies in particular that at least one of the vertices in the cluster of 1 in H is connected to a vertex in $\{n-k+1,\dots,n\}$. Hence by a union bound over those k^2 potential edges

$$\Delta_k\le k^2p_n,$$

and

$$\sum_{u,v\in[n]} \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \le k, |\mathcal{C}_v| \le k, u \leftrightarrow v] \le (\mathbb{E}_{n,p_n}[S_k])^2 + \lambda n k^2. \tag{6.4.31}$$

Combining (6.4.29) and (6.4.31), we get

$$\mathrm{Var}[S_k] \le 2\lambda n k^2.$$

The result follows from (6.4.27) and Chebyshev's inequality

$$\begin{aligned}
&\mathbb{P}[|S_{\underline{k}_n} - (1-\zeta_\lambda)n| \ge n^\gamma] \\
&\quad \le \mathbb{P}[|S_{\underline{k}_n} - \mathbb{E}_{n,p_n}[S_{\underline{k}_n}]| \ge n^\gamma - C\log^2 n] \\
&\quad \le \frac{2\lambda n \underline{k}_n^2}{(n^\gamma - C\log^2 n)^2} \\
&\quad \le \frac{2\lambda n (1+\kappa_{\delta,\alpha})^2 I_\lambda^{-2} \log^2 n}{C' n^{2\gamma}} \\
&\quad \le C'' n^{-\delta},
\end{aligned}$$

for constants $C, C', C'' > 0$ and n large enough, where we used that $2\gamma > 1$ and $\delta < 2\gamma - 1$. ■

Subcritical regime: second moment argument

A second moment argument also gives a lower bound on the size of the largest component in the subcritical case. We proved in Theorem 6.4.1 that, when $\lambda < 1$, the probability of observing a connected component of size larger than $I_\lambda^{-1} \log n$ is vanishingly small. In the other direction, we get:

Theorem 6.4.11 (Subcritical regime: lower bound on the largest cluster)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{\lambda}{n}$ with $\lambda \in (0,1)$. For all $\kappa \in (0,1)$,*

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_{\max}| \le (1-\kappa) I_\lambda^{-1} \log n\right] = o(1).$$

Proof. Recall that

$$B_k = \sum_{v\in[n]} \mathbf{1}_{\{|\mathcal{C}_v| > k\}}.$$

It suffices to prove that with probability $1 - o(1)$ we have $B_k > 0$ when $k = (1-\kappa) I_\lambda^{-1} \log n$. To apply the second moment method (Theorem 2.3.2), we need an upper bound on the second moment of B_k and a lower bound on its first moment. The following lemma is closely related to Lemma 6.4.10. Exercise 6.12 asks for a proof.

Lemma 6.4.12 (Second moment of X_k)**.** *Assume $\lambda < 1$. There is a constant $C > 0$ such that*

$$\mathbb{E}_{n,p_n}[B_k^2] \le (\mathbb{E}_{n,p_n}[B_k])^2 + Cnke^{-kI_\lambda}, \quad \forall k \ge 0.$$

Lemma 6.4.13 (First moment of X_k)**.** *Let $k_n = (1-\kappa)I_\lambda^{-1}\log n$. Then, for any $\beta \in (0,\kappa)$ we have that*

$$\mathbb{E}_{n,p_n}[B_{k_n}] = \Omega(n^\beta),$$

for n large enough.

Proof. By Lemma 6.4.3,

$$\begin{aligned}\mathbb{E}_{n,p_n}[B_{k_n}] &= n\,\mathbb{P}_{n,p_n}[|\mathcal{C}_1| > k_n] \\ &\ge n\,\mathbb{P}[W_\lambda > k_n] - O\left(\lceil k_n \rceil^2\right).\end{aligned} \tag{6.4.32}$$

Once again, we use the random-walk representation of the total progeny of a branching process (Theorem 6.2.6). In contrast to the proof of Lemma 6.4.6, we need a lower bound this time. For this purpose, we use the explicit expression for the law of the total progeny W_λ from Example 6.2.7

$$\mathbb{P}[W_\lambda > k_n] = \sum_{t > k_n} \frac{1}{t} e^{-\lambda t} \frac{(\lambda t)^{t-1}}{(t-1)!}.$$

Using Stirling's formula (see Appendix A) and (6.4.3), we note that

$$\begin{aligned}\frac{1}{t} e^{-\lambda t} \frac{(\lambda t)^{t-1}}{(t-1)!} &= e^{-\lambda t} \frac{(\lambda t)^{t-1}}{t!} \\ &= e^{-\lambda t} \frac{(\lambda t)^t}{\lambda t (t/e)^t \sqrt{2\pi t}(1+o(1))} \\ &= \frac{1-o(1)}{\lambda\sqrt{2\pi t^3}} \exp\left(-t\lambda + t\log\lambda + t\right) \\ &= \frac{1-o(1)}{\lambda\sqrt{2\pi t^3}} \exp\left(-tI_\lambda\right).\end{aligned}$$

Hence, for any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}[W_\lambda > k_n] &\ge \lambda^{-1} \sum_{t > k_n} \exp\left(-t(I_\lambda + \varepsilon)\right) \\ &= \Omega\left(\exp\left(-k_n(I_\lambda + \varepsilon)\right)\right),\end{aligned}$$

for n large enough. For any $\beta \in (0, \kappa)$, taking ε small enough we have

$$\begin{aligned} n\,\mathbb{P}[W_\lambda > k_n] &= \Omega\left(n \exp\left(-k_n(I_\lambda + \varepsilon)\right)\right) \\ &= \Omega\left(\exp\left(\{1 - (1-\kappa)I_\lambda^{-1}(I_\lambda + \varepsilon)\}\log n\right)\right) \\ &= \Omega(n^\beta). \end{aligned}$$

Plugging this back into (6.4.32) gives

$$\mathbb{E}_{n,p_n}[B_{k_n}] = \Omega(n^\beta),$$

which proves the claim. ■

We return to the proof of Theorem 6.4.11. Let again $k_n = (1-\kappa)I_\lambda^{-1}\log n$. By the second moment method and Lemmas 6.4.12 and 6.4.13,

$$\begin{aligned} \mathbb{P}_{n,p_n}[B_{k_n} > 0] &\geq \frac{(\mathbb{E}B_{k_n})^2}{\mathbb{E}[B_{k_n}^2]} \\ &\geq \left(1 + \frac{O(nk_n e^{-k_n I_\lambda})}{\Omega(n^{2\beta})}\right)^{-1} \\ &= \left(1 + \frac{O(nk_n e^{(\kappa-1)\log n})}{\Omega(n^{2\beta})}\right)^{-1} \\ &= \left(1 + \frac{O(k_n n^\kappa)}{\Omega(n^{2\beta})}\right)^{-1} \\ &\to 1, \end{aligned}$$

for β close enough to κ. That proves the claim. ■

6.4.4 Critical case via martingales

It remains to consider the critical case, that is, when $\lambda = 1$. As it turns out, the model goes through a "double jump": as λ crosses 1, the largest cluster size goes from order $\log n$ to order $n^{2/3}$ to order n. Here we use martingale methods to show the following.

Theorem 6.4.14 (Critical case: upper bound on the largest cluster)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{1}{n}$. For all $\kappa > 1$,*

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_{\max}| > \kappa n^{2/3}\right] \leq \frac{C}{\kappa^{3/2}},$$

for some constant $C > 0$.

Remark 6.4.15. *One can also derive a lower bound on the probability that* $|\mathcal{C}_{\max}| > \kappa n^{2/3}$ *for some* $\kappa > 0$ *[ER60]. Exercise 6.20 provides a sketch based on counting tree components; the combinatorial approach has the advantage of giving insights into the structure of the graph (see [Bol01] for more on this). See also [NP10] for a martingale proof of the lower bound as well as a better upper bound.*

The key technical bound is the following.

Lemma 6.4.16. *Let* $G_n \sim \mathbb{G}_{n,p_n}$ *where* $p_n = \frac{1}{n}$ *and let* $\mathcal{C}_v$ *be the connected component of* $v \in [n]$*. There are constants* $c, c' > 0$ *such that for all* $k \geq c$

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| > k] \leq \frac{c'}{\sqrt{k}}.$$

Before we establish the lemma, we prove the theorem assuming it.

Proof of Theorem 6.4.14. Recall that

$$B_k = \sum_{v \in [n]} \mathbf{1}_{\{|\mathcal{C}_v| > k\}}.$$

Take

$$k_n := \kappa n^{2/3}.$$

By Markov's inequality (Theorem 2.1.1) and Lemma 6.4.16,

$$\begin{aligned}
\mathbb{P}_{n,p_n}[|\mathcal{C}_{\max}| > k_n] &\leq \mathbb{P}_{n,p_n}[B_{k_n} > k_n] \\
&\leq \frac{\mathbb{E}_{n,p_n}[B_{k_n}]}{k_n} \\
&= \frac{n\,\mathbb{P}_{n,p_n}[|\mathcal{C}_v| > k_n]}{k_n} \\
&\leq \frac{nc'}{k_n^{3/2}} \\
&\leq \frac{C}{\kappa^{3/2}},
\end{aligned}$$

for some constant $C > 0$. ■

It remains to prove the lemma.

Proof of Lemma 6.4.16. Once again, we use the exploration process defined in Section 6.4.2 started at v. Let $(\mathcal{F}_t)$ be the corresponding filtration and let $A_t = |\mathcal{A}_t|$ be the size of the active set.

Domination by a martingale Recalling (6.4.6), we define

$$M_t := M_{t-1} + \left[-1 + \widetilde{X}_t\right], \tag{6.4.33}$$

with $M_0 := 1$ and $(\widetilde{X}_t)$ are i.i.d. $\mathrm{Bin}(n, 1/n)$. We couple (A_t) and (M_t) through the equation (6.4.7) by letting

$$\widetilde{X}_t = \sum_{i=1}^{n} I_{t,i}.$$

In particular $M_t \geq A_t$ for all t.

Furthermore, we have

$$\mathbb{E}[M_t \,|\, \mathcal{F}_{t-1}] = M_{t-1} - 1 + n\frac{1}{n} = M_{t-1}.$$

So (M_t) is a martingale. We define the stopping time

$$\tilde{\tau}_0 := \inf\{t \geq 0 \,:\, M_t = 0\}.$$

Recalling that

$$\tau_0 = \inf\{t \geq 0 \,:\, A_t = 0\} = |\mathcal{C}_v|,$$

by Lemma 6.2.1, we have $\tilde{\tau}_0 \geq \tau_0 = |\mathcal{C}_v|$ almost surely. So

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| > k\right] \leq \mathbb{P}[\tilde{\tau}_0 > k].$$

The tail of $\tilde{\tau}_0$ To bound the tail of $\tilde{\tau}_0$, we introduce a modified stopping time. For $h > 0$, let

$$\tau_h' := \inf\{t \geq 0 \,:\, M_t = 0 \text{ or } M_t \geq h\}.$$

We will use the inequality

$$\mathbb{P}[\tilde{\tau}_0 > k] = \mathbb{P}[M_t > 0, \forall t \leq k] \leq \mathbb{P}[\tau_h' > k] + \mathbb{P}[M_{\tau_h'} \geq h],$$

and we will choose h below to minimize the rightmost expression (or, more specifically, an upper bound on it). The rest of the analysis is similar to the gambler's ruin problem in Example 3.1.41, with some slight complications arising from the fact that the process is not nearest-neighbor.

We note that by the exponential tail of hitting times on finite state spaces (Lemma 3.1.25), the stopping time τ_h' is almost surely finite and, in fact, has a finite expectation. By two applications of Markov's inequality,

$$\mathbb{P}[M_{\tau_h'} \geq h] \leq \frac{\mathbb{E}[M_{\tau_h'}]}{h},$$

and

$$\mathbb{P}[\tau'_h > k] \leq \frac{\mathbb{E}\tau'_h}{k}.$$

We bound the expectations on the right-hand sides.

Bounding $\mathbb{E}M_{\tau'_h}$ and $\mathbb{E}\tau'_h$ To compute $\mathbb{E}M_{\tau'_h}$, we use the optional stopping theorem in the uniformly bounded case (Theorem 3.1.38 (ii)) to the stopped process $(M_{t\wedge\tau'_h})$ (which is also a martingale by Lemma 3.1.37) to get that

$$\mathbb{E}[M_{\tau'_h}] = \mathbb{E}[M_0] = 1.$$

We conclude that

$$\mathbb{P}[M_{\tau'_h} \geq h] \leq \frac{1}{h}. \tag{6.4.34}$$

To compute $\mathbb{E}\tau'_h$, we use a different martingale (adapted from Example 3.1.31), specifically

$$L_t := M_t^2 - \sigma^2 t,$$

where we let $\sigma^2 := n\frac{1}{n}\left(1 - \frac{1}{n}\right) = \left(1 - \frac{1}{n}\right)$, which is $\geq \frac{1}{2}$ when $n \geq 2$. To see that (L_t) is a martingale, note that by taking out what is known (Lemma B.6.13) and using the fact that (M_t) is itself a martingale

$$\begin{aligned}
\mathbb{E}[L_t \,|\, \mathcal{F}_{t-1}] &= \mathbb{E}[(M_{t-1} + (M_t - M_{t-1}))^2 - \sigma^2 t \,|\, \mathcal{F}_{t-1}] \\
&= \mathbb{E}[M_{t-1}^2 + 2M_{t-1}(M_t - M_{t-1}) + (M_t - M_{t-1})^2 - \sigma^2 t \,|\, \mathcal{F}_{t-1}] \\
&= M_{t-1}^2 + 2M_{t-1} \cdot 0 + \sigma^2 - \sigma^2 t \\
&= L_{t-1}.
\end{aligned}$$

By Lemma 3.1.37, the stopped process $(L_{t\wedge\tau'_h})$ is also a martingale; and it has bounded increments since

$$\begin{aligned}
|L_{(t+1)\wedge\tau'_h} - L_{t\wedge\tau'_h}| &\leq |M^2_{(t+1)\wedge\tau'_h} - M^2_{t\wedge\tau'_h}| + \sigma^2 \\
&\leq \left|(-1 + \widetilde{X}_{t+1})^2 + 2h|-1 + \widetilde{X}_{t+1}|\right| + \sigma^2 \\
&\leq n^2 + 2hn + 1.
\end{aligned}$$

We use the optional stopping theorem in the bounded increments case (Theorem 3.1.38 (iii)) on $(L_{t\wedge\tau'_h})$ to get

$$\mathbb{E}[M^2_{\tau'_h} - \sigma^2\tau'_h] = \mathbb{E}[M^2_{\tau'_h}] - \sigma^2\mathbb{E}\tau'_h = 1.$$

After rearranging (6.4.35)

$$\mathbb{E}\tau'_h \leq \frac{1}{\sigma^2}\mathbb{E}[M^2_{\tau'_h}] \leq 2\,\mathbb{E}[M^2_{\tau'_h}], \tag{6.4.35}$$

where we used the fact that $\sigma^2 \geq 1/2$.

To bound $\mathbb{E}[M^2_{\tau'_h}]$, we need to control by how much the process "overshoots" h. A stochastic domination argument gives the desired bound; Exercise 6.21 asks for a proof.

Lemma 6.4.17 (Overshoot bound)**.** *Let f be an increasing function and $W \sim \mathrm{Bin}(n, 1/n)$. Then*

$$\mathbb{E}[f(M_{\tau'_h} - h) \,|\, M_{\tau'_h} \geq h] \leq \mathbb{E}[f(W)].$$

The lemma implies that

$$\begin{aligned}\mathbb{E}[M^2_{\tau'_h} \,|\, M_{\tau'_h} \geq h] &= \mathbb{E}[(M_{\tau'_h} - h)^2 + 2(M_{\tau'_h} - h)h + h^2 \,|\, M_{\tau'_h} \geq h] \\ &\leq (\sigma^2 + 1) + 2h + h^2 \\ &\leq 4h^2.\end{aligned}$$

Plugging back into (6.4.35) gives

$$\mathbb{E}\tau'_h \leq 2\left\{\frac{1}{h}\mathbb{E}[M^2_{\tau'_h} \,|\, M_{\tau'_h} \geq h]\right\} \leq 8h,$$

where we used (6.4.34).

Putting everything together Finally take $h := \sqrt{\frac{k}{8}}$. Putting everything together

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| > k] \leq \mathbb{P}[\tilde{\tau}_0 > k] \leq \mathbb{P}[\tau'_h > k] + \mathbb{P}[M_{\tau'_h} \geq h] \leq \frac{8h}{k} + \frac{1}{h} = 2\sqrt{\frac{8}{k}}.$$

That concludes the proof. ∎

6.4.5 ▷ *Encore: random walk on the Erdős-Rényi graph*

So far in this section we have used techniques from all chapters of the book—with the exception of Chapter 5. Not to be outdone, we discuss one last result that will make use of spectral techniques. We venture a little further down the evolution of the Erdős-Rényi graph model to the connected regime. Specifically, recall from Section 2.3.2 that $G_n = (V_n, E_n) \sim \mathbb{G}_{n,p_n}$ is connected with probability $1 - o(1)$ when $np_n = \omega(\log n)$.

We show in that regime that lazy simple random walk (X_t) on G_n "mixes fast." Recall from Example 1.1.29 that, when the graph is connected, the corresponding transition matrix P is reversible with respect to the stationary distribution

$$\pi(v) := \frac{\delta(v)}{2|E_n|},$$

where $\delta(v)$ is the degree of v. For a fixed $\varepsilon > 0$, the mixing time (see Definition 1.1.35) is

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) = \inf\{t \geq 0 \,:\, d(t) \leq \varepsilon\},$$

where

$$d(t) = \sup_{x \in V_n} \|P^t(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}}.$$

By convention, we let $\mathrm{t}_{\mathrm{mix}}(\varepsilon) = +\infty$ if the graph is not connected. Our main result is the following.

Theorem 6.4.18 (Mixing on a connected Erdős-Rényi graph)**.** *Let $G_n \sim \mathbb{G}_{n,p_n}$ with $np_n = \omega(\log n)$. With probability $1 - o(1)$, the mixing time is $O(\log n)$.*

Edge expansion We use Cheeger's inequality (Theorem 5.3.5) which, recall, states that

$$\gamma \geq \frac{\Phi_*^2}{2},$$

where γ is the spectral gap of P (see Definition 5.2.11) and

$$\Phi_* = \min\left\{\Phi_{\mathrm{E}}(S; c, \pi) \,:\, S \subseteq V,\ 0 < \pi(S) \leq \frac{1}{2}\right\},$$

is the edge expansion constant (see Definition 5.3.2), with

$$\Phi_{\mathrm{E}}(S; c, \pi) = \frac{c(S, S^c)}{\pi(S)},$$

for a subset of vertices $S \subseteq V_n$. Here, for a pair of vertices x, y connected by an edge,

$$c(x, y) = \pi(x)P(x, y) = \frac{\delta(x)}{2|E_n|}\frac{1}{\delta(x)} = \frac{1}{2|E_n|}.$$

Hence

$$c(S, S^c) = \frac{|E(S, S^c)|}{2|E_n|},$$

where $E(S, S^c)$ is the set of edges between S and S^c. Similarly,

$$\pi(S) = \frac{\sum_{x \in S} \delta(x)}{2|E_n|}.$$

The numerator is referred to as the volume of S and we use the notation $\mathrm{vol}(S) = \sum_{x \in S} \delta(x)$. So

$$\frac{c(S, S^c)}{\pi(S)} = \frac{|E(S, S^c)|}{\mathrm{vol}(S)}. \tag{6.4.36}$$

Because the random walk is lazy, the spectral gap is equal to the absolute spectral gap (see Definition 5.2.11), and as a consequence the relaxation time (see Definition 5.2.12) is

$$\mathrm{t}_{\mathrm{rel}} = \gamma^{-1}.$$

Using Theorem 5.2.14, we get

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) \mathrm{t}_{\mathrm{rel}} \leq \log\left(\frac{1}{\varepsilon \pi_{\min}}\right) \frac{2}{\Phi_*^2}, \tag{6.4.37}$$

where

$$\pi_{\min} = \min_x \pi(x) = \min_x \frac{\delta(x)}{2|E_n|} = \frac{\min_x \delta(x)}{\sum_y \delta(y)}.$$

So our main task is to bound $\delta(x)$ and $|E(S, S^c)|$ with high probability. We do this next.

Bounding the degrees In fact, we have already done half the work. Indeed in Example 2.4.18 we studied the maximum degree of G_n

$$D_n = \max_{v \in V_n} \delta(v),$$

in the regime $np_n = \omega(\log n)$. We showed that for any $\zeta > 0$, as $n \to +\infty$,

$$\mathbb{P}\left[|D_n - np_n| \leq 2\sqrt{(1+\zeta) np_n \log n}\right] \to 1.$$

The proof of that result actually shows something stronger: all degrees satisfy the inequality simultaneously, that is,

$$\mathbb{P}\left[\forall v \in V_n,\ |\delta(v) - np_n| \leq 2\sqrt{(1+\zeta) np_n \log n}\right] = 1 - o(1). \tag{6.4.38}$$

We will use the fact that $2\sqrt{(1+\zeta) np_n \log n} = o(np_n)$ when $np_n = \omega(\log n)$. In essence, all degrees are roughly np_n. That implies the following claims.

Lemma 6.4.19 (Bounds on stationary distribution and volume)**.** *The following hold with probability* $1 - o(1)$.

(i) The smallest stationary probability satisfies

$$\pi_{\min} \geq \frac{1 - o(1)}{n}.$$

(ii) For any set of vertices $S \subseteq V_n$ *with* $|S| > 2n/3$*, we have*

$$\pi(S) > \frac{1}{2}.$$

(iii) For any set of vertices $S \subseteq V_n$ *with* $s := |S|$

$$\text{vol}(S) = snp_n(1 + o(1)).$$

Proof. We assume that the event in (6.4.38) holds.

For (i), that means

$$\pi_{\min} \geq \frac{np_n - 2\sqrt{(1+\zeta)np_n \log n}}{n(np_n + 2\sqrt{(1+\zeta)np_n \log n})} = \frac{1}{n}(1 - o(1)),$$

when $np_n = \omega(\log n)$.

For (ii), we get

$$\pi(S) = \frac{\sum_{x \in S} \delta(x)}{\sum_{x \in V_n} \delta(x)} \geq \frac{|S|(np_n - 2\sqrt{(1+\zeta)np_n \log n})}{n(np_n + 2\sqrt{(1+\zeta)np_n \log n})} > \frac{2}{3}(1 - o(1)).$$

Finally (iii) follows similarly. ■

Bounding the cut size An application of Bernstein's inequality (Theorem 2.4.17) gives the following bound.

Lemma 6.4.20 (Bound on the edge expansion)**.** *With probability* $1 - o(1)$,

$$\Phi_* = \Omega(1).$$

Proof. By the definition of Φ_* and Lemma 6.4.19 (ii), we can restrict ourselves to sets S of size at most $2n/3$. Let S be such a set with $s = |S|$. Then $|E(S, S^c)|$ is $\text{Bin}(s(n-s), p_n)$. By Bernstein's inequality with $c = 1$ and $\nu_i = p_n(1 - p_n)$,

$$\mathbb{P}_{n,p_n}[|E(S, S^c)| \leq s(n-s)p_n - \beta] \leq \exp\left(-\frac{\beta^2}{4s(n-s)p_n(1-p_n)}\right),$$

for $\beta \leq s(n-s)p_n(1-p_n)$. We take $\beta = \frac{1}{2}s(n-s)p_n$ and get

$$\mathbb{P}_{n,p_n}\left[|E(S,S^c)| \leq \frac{1}{2}s(n-s)p_n\right] \leq \exp\left(-\frac{s(n-s)p_n}{16(1-p_n)}\right).$$

By a union bound over all sets of size s and using the fact that $\binom{n}{s} \leq \left(\frac{ne}{s}\right)^s$ (see Appendix A), there is a constant $C > 0$ such that

$$\begin{aligned}\mathbb{P}_{n,p_n}&\left[\exists S,\ |S| = s,\ |E(S,S^c)| \leq \frac{1}{2}s(n-s)p_n\right] \\ &\leq \binom{n}{s}\exp\left(-\frac{s(n-s)p_n}{16(1-p_n)}\right) \\ &\leq \exp\left(-s\frac{np_n}{48} + s\log(ne/s)\right) \\ &\leq \exp\left(-Csnp_n\right),\end{aligned}$$

for n large enough, where we also used that $n - s \geq n/3$ and $np_n = \omega(\log n)$. Summing over s gives, for a constant $C' > 0$,

$$\mathbb{P}_{n,p_n}\left[\exists S,\ 1 \leq |S| \leq 2n/3,\ |E(S,S^c)| \leq \frac{1}{2}|S|(n-|S|)p_n\right] \leq C'\exp\left(-Cnp_n\right),$$

which goes to 0 as $n \to +\infty$.

Using (6.4.36) and Lemma 6.4.19 (iii), any set S such that $|E(S,S^c)| > \frac{1}{2}|S|(n-|S|)p_n$ has edge expansion

$$\Phi_{\mathrm{E}}(S;c,\pi) \geq \frac{\frac{1}{2}|S|(n-|S|)p_n}{|S|np_n(1+o(1))} \geq \frac{1}{6}(1-o(1)).$$

That proves the claim. ■

Proof of the theorem Finally, we are ready to prove the main result.

Proof of Theorem 6.4.18. Plugging Lemma 6.4.19 (i) and Lemma 6.4.20 into (6.4.37) gives

$$t_{\mathrm{mix}}(\varepsilon) \leq \log\left(\frac{1}{\varepsilon\pi_{\min}}\right)\frac{2}{\Phi_*^2} \leq C''\log(\varepsilon^{-1}n(1+o(1))) = O(\log n),$$

for some constant $C'' > 0$. ■

Remark 6.4.21. *A mixing time of $O(\log n)$ in fact holds for lazy simple random walk on $\mathbb{G}_{n,p_n}$ when $p_n = \frac{\lambda \log n}{n}$ with $\lambda > 1$ [CF07]. See also [Dur06, Section 6.5]. Mixing time on the giant component has also been studied. See, e.g., [FR08, BKW14, DKLP11].*

Exercises

Exercise 6.1 (Galton-Watson process: subcritical case)**.** We use Markov's inequality to analyze the subcritical case.

(i) Let (Z_t) be a Galton-Watson process with offspring distribution mean $m < 1$. Use Markov's inequality (Theorem 2.1.1) to prove that extinction occurs almost surely.

(ii) Prove the equivalent result in the multitype case, that is, prove (6.1.8).

Exercise 6.2 (Galton-Watson process: geometric offspring)**.** Let (Z_t) be a Galton-Watson branching process with geometric offspring distribution (started at 0), that is, $p_k = p(1-p)^k$ for all $k \geq 0$, for some $p \in (0,1)$. Let $q := 1-p$, let m be the mean of the offspring distribution, and let $W_t = m^{-t} Z_t$.

(i) Compute the probability generating function f of $\{p_k\}_{k\geq 0}$ and the extinction probability $\eta := \eta_p$ as a function of p.

(ii) If G is a 2×2 matrix, define

$$G(s) := \frac{G_{11}s + G_{12}}{G_{21}s + G_{22}}.$$

Show that $G(H(s)) = (GH)(s)$.

(iii) Assume $m \neq 1$. Use (ii) to derive

$$f_t(s) = \frac{pm^t(1-s) + qs - p}{qm^t(1-s) + qs - p}.$$

Deduce that when $m > 1$

$$\mathbb{E}[\exp(-\lambda W_\infty)] = \eta + (1-\eta)\frac{(1-\eta)}{\lambda + (1-\eta)}.$$

(iv) Assume $m = 1$. Show that

$$f_t(s) = \frac{t - (t-1)s}{t + 1 - ts},$$

and deduce that

$$\mathbb{E}[e^{-\lambda Z_t/t} \,|\, Z_t > 0] \to \frac{1}{1+\lambda}.$$

Exercise 6.3 (Supercritical branching process: infinite line of descent)**.** Let (Z_t) be a supercritical Galton-Watson branching process with offspring distribution $\{p_k\}_{k\geq 0}$. Let η be the extinction probability and define $\zeta := 1 - \eta$. Let Z_t^∞ be the number of individuals in the t-th generation with an infinite line of descent, i.e., whose descendant subtree is infinite. Denote by $\mathcal{S}$ the event of nonextinction of (Z_t). Define $p_0^\infty := 0$ and

$$p_k^\infty := \zeta^{-1} \sum_{j \geq k} \binom{j}{k} \eta^{j-k} \zeta^k p_j.$$

(i) Show that $\{p_k^\infty\}_{k\geq 0}$ is a probability distribution and compute its expectation.

(ii) Show that for any $k \geq 0$

$$\mathbb{P}[Z_1^\infty = k \,|\, \mathcal{S}] = p_k^\infty.$$

[Hint: Condition on Z_1.]

(iii) Show by induction on t that, conditioned on nonextinction, the process (Z_t^∞) has the same distribution as a Galton-Watson branching process with offspring distribution $\{p_k^\infty\}_{k\geq 0}$.

Exercise 6.4 (Multitype branching processes: a special case)**.** Extend Lemma 6.1.20 to the case $\mathbf{S}^{(\mathbf{u})} = \mathbf{0}$. [Hint: Show that $U_t = \mathbf{Z}_0\mathbf{u}$ for all t almost surely.]

Exercise 6.5 (Galton-Watson: Inverting history)**.** Let

$$H = (X_1, \ldots, X_{\tau_0}),$$

be the history (see Section 6.2) of the Galton-Watson process (Z_i). Write Z_i as a function of H, for all i.

Exercise 6.6 (Spitzer's lemma)**.** Prove Theorem 6.2.5.

Exercise 6.7 (Sum of Poisson)**.** Let Q_1 and Q_2 be independent Poisson random variables with respective means λ_1 and λ_2. Show by direct computation of the convolution that the sum $Q_1 + Q_2$ is Poisson with mean $\lambda_1 + \lambda_2$. [Hint: Recall that $\mathbb{P}[Q_1 = k] = e^{-\lambda_1}\lambda_1^k/k!$ for all $k \in \mathbb{Z}_+$.]

Exercise 6.8 (Percolation on bounded-degree graphs)**.** Let $G = (V, E)$ be a countable graph such that all vertices have degree bounded by $b + 1$ for $b \geq 2$. Let 0 be a distinguished vertex in G. For bond percolation on G, prove that

$$p_{\mathrm{c}}(G) \geq p_{\mathrm{c}}(\widehat{\mathbb{T}}_b),$$

by bounding the expected size of the cluster of 0. [Hint: Consider self-avoiding paths started at 0.]

Exercise 6.9 (Percolation on $\widehat{\mathbb{T}}_b$: critical exponent of $\theta(p)$)**.** Consider bond percolation on the rooted infinite b-ary tree $\widehat{\mathbb{T}}_b$ with $b > 2$. For $\varepsilon \in [0, 1 - \frac{1}{b}]$ and $u \in [0,1]$, define

$$h(\varepsilon, u) := u - \left(\left(1 - \tfrac{1}{b} - \varepsilon\right)(1-u) + \tfrac{1}{b} + \varepsilon\right)^b.$$

(i) Show that there is a constant $C > 0$ not depending on ε, u such that

$$\left| h(\varepsilon, u) - b\varepsilon u + \frac{b-1}{2b} u^2 \right| \leq C(u^3 \vee \varepsilon u^2).$$

(ii) Use (i) to prove that

$$\lim_{p \downarrow p_c(\widehat{\mathbb{T}}_b)} \frac{\theta(p)}{(p - p_c(\widehat{\mathbb{T}}_b))} = \frac{2b^2}{b-1}.$$

Exercise 6.10 (Percolation on $\widehat{\mathbb{T}}_2$: higher moments of $|\mathcal{C}_0|$)**.** Consider bond percolation on the rooted infinite binary tree $\widehat{\mathbb{T}}_2$. For density $p < \frac{1}{2}$, let Z_p be an integer-valued random variable with distribution

$$\mathbb{P}_p[Z_p = \ell] = \frac{\ell\, \mathbb{P}_p[|\mathcal{C}_0| = \ell]}{\mathbb{E}_p|\mathcal{C}_0|}, \qquad \forall \ell \geq 1.$$

(i) Using the explicit formula for $\mathbb{P}_p[|\mathcal{C}_0| = \ell]$ derived in Section 6.2.4, show that for all $0 < a < b < +\infty$

$$\mathbb{P}_p\left[\frac{Z_p}{(1/4)(\frac{1}{2} - p)^{-2}} \in [a,b]\right] \to C \int_a^b x^{-1/2} e^{-x} \mathrm{d}x,$$

as $p \uparrow \frac{1}{2}$, for some constant $C > 0$.

(ii) Show that for all $k \geq 2$ there is $C_k > 0$ such that

$$\lim_{p \uparrow p_c(\widehat{\mathbb{T}}_2)} \frac{\mathbb{E}_p|\mathcal{C}_0|^k}{(p_c(\widehat{\mathbb{T}}_2) - p)^{-1-2(k-1)}} = C_k.$$

(iii) What happens when $p \downarrow p_c(\widehat{\mathbb{T}}_2)$?

Exercise 6.11 (Branching process approximation: improved bound)**.** Let $p_n = \frac{\lambda}{n}$ with $\lambda > 0$. Let W_{n,p_n}, respectively W_λ, be the total progeny of a branching process with offspring distribution $\mathrm{Bin}(n, p_n)$, respectively $\mathrm{Poi}(\lambda)$.

(i) Show that

$$|\mathbb{P}[W_{n,p_n} \geq k] - \mathbb{P}[W_\lambda \geq k]| \\ \leq \max\{\mathbb{P}[W_{n,p_n} \geq k, W_\lambda < k], \mathbb{P}[W_{n,p_n} < k, W_\lambda \geq k]\}.$$

(ii) Couple the two processes step-by-step and use (i) to show that

$$|\mathbb{P}[W_{n,p_n} \geq k] - \mathbb{P}[W_\lambda \geq k]| \leq \frac{\lambda^2}{n} \sum_{i=1}^{k-1} \mathbb{P}[W_\lambda \geq i].$$

Exercise 6.12 (Subcritical Erdős-Rényi: second moment)**.** Prove Lemma 6.4.12.

Exercise 6.13 (Random binary search tree: property (BST))**.** Show that the (BST) property is preserved by the algorithm described at the beginning of Section 6.3.1.

Exercise 6.14 (Random binary search tree: limit)**.** Consider the equation (6.3.1).

(i) Show that there exists a unique solution greater than 1.

(ii) Prove that the expression on the left-hand side is strictly decreasing at that solution.

Exercise 6.15 (Random binary search tree: height is well-defined)**.** Let $\mathcal{T}$ be an infinite binary tree. Assign an independent $U[0,1]$ random variable Z_v to each vertex v in $\mathcal{T}$, set $S_\rho = n$ and then recursively from the root down

$$S_{v'} := \lfloor S_v Z_v \rfloor \quad \text{and} \quad S_{v''} := \lfloor S_v (1 - Z_v) \rfloor,$$

where v' and v'' are the left and right descendants of v in $\mathcal{T}$.

(i) Show that, for any v, it holds that $S_{v'} + S_{v''} = S_v - 1$ almost surely provided $S_v \geq 1$.

(ii) Show that, for any v, there is almost surely a descendant w of v (not necessarily immediate) such that $S_w = 1$.

(iii) Let

$$H_n = \max\{h : \exists v \in \mathcal{V}_h,\ S_v = 1\},$$

where $\mathcal{V}_h$ is the set of vertices of $\mathcal{T}$ at topological distance h from the root. Show that $H_n \leq n$.

Exercise 6.16 (Ising vs. CFN)**.** Let $\mathcal{T}_h$ be a rooted complete binary tree with h levels. Fix $0 < p < 1/2$. Assign to each vertex v a state $\sigma_v \in \{+1, -1\}$ at random according to the CFN model described in Section 6.3.2. Show that this distribution is equivalent to a ferromagnetic Ising model on $\mathcal{T}_h$ and determine the inverse temperature β in terms of p. [Hint: Write the distribution of the states under the CFN model as a product over the edges.]

Exercise 6.17 (Monotonicity of $\|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}}$)**.** Let μ_h^+, μ_h^- be as in Section 6.3.2. Show that

$$\|\mu_{h+1}^+ - \mu_{h+1}^-\|_{\mathrm{TV}} \leq \|\mu_h^+ - \mu_h^-\|_{\mathrm{TV}}.$$

[Hint: Use the Markovian nature of the process.]

Exercise 6.18 (Unsolvability: recursion)**.** Prove Lemma 6.3.15.

Exercise 6.19 (Cayley's formula)**.** Let (Z_t) be a Poisson branching process with offspring mean 1 started at $Z_0 = 1$ and let T be the corresponding Galton-Watson tree. Let W be the total of size of the progeny, that is, the number of vertices in T. Recall from Example 6.2.7 that

$$\mathbb{P}[W = n] = \frac{n^{n-1}e^{-n}}{n!}.$$

(i) Given $W = n$, label the vertices of T uniformly at random with the integers $1, \ldots, n$. Show that every rooted labeled tree on n vertices arises with probability $e^{-n}/n!$. [Hint: Label the vertices as you grow the tree and observe that a lot of terms cancel out or simplify.]

(ii) Derive Cayley's formula: the number of labeled trees on n vertices is n^{n-2}.

Exercise 6.20 (Critical regime: tree components)**.** Let $G_n \sim \mathbb{G}_{n,p_n}$ where $p_n = \frac{1}{n}$.

(i) Let $\gamma_{n,k}$ be the expected number of isolated tree components of size k in G_n. Justify the formula

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

[Hint: We did a related calculation in Section 2.3.2.]

(ii) Show that, if $k = \omega(1)$ and $k = o(n^{3/4})$,

$$\gamma_{n,k} \sim n\frac{k^{-5/2}}{\sqrt{2\pi}} \exp\left(-\frac{k^3}{6n^2}\right).$$

(iii) Conclude that for $0 < \delta < 1$ the expectation of U, the number of isolated tree components of size in $[(\delta n)^{2/3}, n^{2/3}]$, is $\Omega(\delta^{-1})$ as $\delta \to 0$.

(iv) For $1 \leq k_1 \leq k_2 \leq n - k_1$, let σ_{n,k_1,k_2} be the expected number of pairs of isolated tree components where the first one has size k_1 and the second one has size k_2. Justify the formula

$$\sigma_{n,k_1,k_2} = \binom{n}{k_1} k_1^{k_1-2} \left(\frac{1}{n}\right)^{k_1-1} \left(1 - \frac{1}{n}\right)^{k_1(n-k_1)+\binom{k_1}{2}-(k_1-1)} \times \binom{n-k_1}{k_2} k_2^{k_2-2} \left(\frac{1}{n}\right)^{k_2-1} \left(1 - \frac{1}{n}\right)^{k_2(n-(k_1+k_2))+\binom{k_2}{2}-(k_2-1)},$$

and show that

$$\sigma_{n,k_1,k_2} \leq \gamma_{n,k_1}\gamma_{n,k_2}.$$

[Hint: You may need to prove that, for $0 < a \leq 1 \leq b$, it holds that $1 - ab \leq (1-a)^b$.]

(v) Prove that $\mathrm{Var}[U] = O(\mathbb{E}[U])$. [Hint: Use (2.1.6), (iv), and (ii).]

Exercise 6.21 (Critical regime: overshoot bound)**.** The goal of this exercise is to prove Lemma 6.4.17. We use the notation of Section 6.4.4.

(i) Let $W, Z \sim \mathrm{Bin}(n, 1/n)$ and $0 \leq r \leq n$. Show that $W - r$ conditioned on $W \geq r$ is stochastically dominated by Z. [Hint: Use the representation of W as a sum of indicators. Thinking of the partial sums as a Markov chain, consider the first time it reaches r.]

(ii) Show that $M_{\tau'_h} - h$ conditioned on $M_{\tau'_h} \geq h$ is stochastically dominated by Z from (i). [Hint: By the tower property, it suffices to show that

$$\mathbb{P}[M_{\tau'_h} - h \geq z \,|\, \tau'_h = \ell, M_{\ell-1} = h - r, M_\ell \geq h] \leq \mathbb{P}[Z \geq z],$$

for the relevant ℓ, r, z.]

(iii) Use (ii) to prove Lemma 6.4.17.

Bibliographic remarks

Section 6.1 See [Dur10, Section 5.3.4] for a quick introduction to branching processes. A more detailed overview relating to its use in discrete probability can be found in [vdH17, Chapter 3]. A classical reference on branching processes is [AN04]. The Kesten-Stigum Theorem is due to Kesten and Stigum [KS66b]. Our proof of a weaker version with the second moment condition follows [Dur10, Example 5.4.3]. Section 6.1.4 is based loosely on [AN04, Chapter V]. A proof of Theorem 6.1.18 can be found in [Har63]. A good reference for the Perron-Frobenius Theorem (Theorem 6.1.17 as well as more general versions) is [HJ13, Chapter 8]. The central limit theorem for $\rho \geq \lambda^2$ referred to at the end of Section 6.1.4 is due to Kesten and Stigum [KS66a, KS67]. The critical percolation threshold for percolation on Galton-Watson trees is due to R. Lyons [Lyo90].

Section 6.2 The exploration process in Section 6.2.1 dates back to [ML86] and [Kar90]. The hitting-time theorem (Theorem 6.2.5) in the case $\ell = 1$ was first proved in [Ott49]. For alternative proofs, see for example [vdHK08] or [Wen75]. Spitzer's combinatorial lemma (Lemma 6.2.8) is from [Spi56]. See also [Fel71, Section XII.6]. The presentation in Section 6.2.4 follows [vdH10]. See also [Dur85].

Section 6.3 Section 6.3.1 follows [Dev98, Section 2.1] from the excellent volume [HMRAR98]. Section 6.3.2 is partly a simplified version of [BCMR06]. Further applications in phylogenetics, specifically to the sample complexity of phylogeny inference algorithms, can be found in, for example, [Mos04, Mos03, Roc10, DMR11, RS17]. The reconstruction problem also has applications in community detection [MNS15b]. See [Abb18] for a survey.

Section 6.4 The phase transtion of the Erdős-Rényi graph model was first studied in [ER60]. For much more, see for example [vdH17, Chapter 4], [JLR11, Chapter 5] and [Bol01, Chapter 6]. In particular a central limit theorem for the giant component, proved by several authors including Martin-Löf [ML98], Pittel [Pit90], and Barraez, Boucheron, and de la Vega [BBFdlV00], is established in [vdH17, Section 4.5]. Section 6.4.4 is based on [NP10]. See also [Per09, Sections 2 and 3]. Much more is known about the critical regime; see, e.g., [Ald97, Bol84, Lu90, LuPW94]. Section 6.4.5 is based partly on [Dur06, Section 6.5]. For a lot more on random walk on random graphs (not just Erdős-Rényi), see [Dur06, Chapter 6]. For more on the spectral properties of random graphs, see [CL06].