Chapter 5

Branching processes

Branching processes arise naturally in the study of stochastic processes on trees and locally tree-like graphs. After a review of the basic extinction theory of branching processes, we give a few classical examples of applications in discrete probability.

5.1 Background

To be written. See [Dur06, Section 5.3.4] and [vdH14, Section 3.3].

5.1.1 ▷ *Random walk on Galton-Watson trees*

To be written. See [LP, Theorem 3.5 and Corollary 5.10].*

5.2 Comparison to branching processes

We begin with an example whose connection to branching processes is clear: percolation on trees. Translating standard branching process results into their percolation counterpart immediately gives a more detailed picture of the behavior of the process than was derived in Section 2.2.3. We then tackle the phase transition of Erdös-Rényi graphs using a comparison to branching processes.

5.2.1 ▷ *Percolation on trees: critical exponents*

In this section, we use branching processes to study bond percolation on the infinite b-ary tree $\widehat{\mathbb{T}}_b$. The same techniques can be adapted to $\mathbb{T}_d$ with $d = b + 1$ in a straightforward manner.

*Requires: Section 2.2.3 and 3.1.1.

We denote the root by 0. 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 "children" 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 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 Lemma **??**, this leads immediately to a (second) proof of:

Claim 5.1.

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

The generating function of the offspring distribution is $\phi(s) := ((1-p)+ps)^b$. So, by Lemma **??** again, 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

$$s = ((1-p) + ps)^b. \tag{5.1}$$

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 5.2. *For $b = 2$,*

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

The expected size of the population at generation n is $(bp)^n$ so for $p \in [0, \frac{1}{b})$

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

For $p \in (\frac{1}{b}, 1)$, the total progeny is almost surely infinite, but it is of interest to compute the expected cluster size *conditioned on* $|\mathcal{C}_0| < +\infty$. We use the duality principle, Lemma **??**. For $0 \leq k \leq b$, let

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

where we used (5.1) and implicitly defined the dual density

$$\hat{p} := \frac{p\,\eta(p)}{(1-p) + p\,\eta(p)}. \tag{5.2}$$

In particular $\{\hat{p}_k\}$ is indeed a probability distribution. In fact it is binomial with parameters b and $\hat{p}$. The corresponding generating function is

$$\hat{\phi}(s) := ((1-\hat{p}) + \hat{p}s)^b = \eta(p)^{-1} \phi(s\,\eta(p)),$$

where the second expression can be seen directly from the definition of $\{\hat{p}_k\}$. Moreover,

$$\hat{\phi}'(s) = \eta(p)^{-1}\phi'(s\,\eta(p))\,\eta(p) = \phi'(s\,\eta(p)),$$

so $\hat{\phi}'(1^-) = \phi'(\eta(p)) < 1$ by the proof of Lemma **??**, confirming that percolation with density $\hat{p}$ is subcritical. Summarizing:

Claim 5.3. *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* (5.2).

Therefore:

Claim 5.4.

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

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 5.5. *For $b = 2$,*

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

In fact, the random walk representation of the process, Lemma **??**, gives an explicit formula for the distribution $|\mathcal{C}_0|$. Namely, because $|\mathcal{C}_0| \stackrel{\mathrm{d}}{=} \tau_0$ 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 and further

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

we have

$$\mathbb{P}_p[|\mathcal{C}_0| = \ell] = \frac{1}{\ell}\,\mathbb{P}\left[\sum_{i\leq\ell} X_\ell = \ell - 1\right] = \frac{1}{\ell}\binom{b\ell}{\ell-1}p^{\ell-1}(1-p)^{b\ell-(\ell-1)}, \quad (5.3)$$

where we used that a sum of independent binomials with the same p is still binomial. In particular, at criticality, using Stirling's formula 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 $\ell \to +\infty$.

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 5.2 and 5.5 imply for instance that, as $p \to p_c$,

critical exponent

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

and

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

In fact, as can be seen from Claim 5.4, the critical exponent of $\chi^{\mathrm{f}}(p)$ does not depend on b. The same holds for $\theta(p)$. See Exercise 5.5. Using (5.3), the higher moments of $|\mathcal{C}_0|$ can also be studied around criticality. See Exercise 5.6.

5.2.2 ▷ *Random binary search trees: height*

To be written. See [Dev98, Section 2.1].

5.2.3 ▷ *Erdös-Rényi graphs: the phase transition*

A compelling way to view Erdös-Rényi graphs as the 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 2.2.4 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 this 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 of Erdös and Rényi, which is often referred to as "the" phase transition of the Erdös-Rényi graph, is related to the phase transition in percolation. That should be clear from the similarities between the proofs, specifically the branching process approach to percolation on trees of Section 5.2.1.

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 bound, martingale techniques, coupling and stochastic domination, and branching processes. It is quintessential discrete probability.

For quick reference, we recall two results from previous chapters:

- *(Binomial domination)* We have

$$n \geq m \implies \mathrm{Bin}(n,p) \succeq \mathrm{Bin}(m,p). \tag{5.4}$$

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

$$\lambda \in (0,1) \quad \implies \quad \mathrm{Poi}(\lambda) \succeq \mathrm{Bin}\left(n-1, \frac{\lambda}{n}\right). \tag{5.5}$$

  For the proofs, see Examples 4.4 and 4.8.

- *(Poisson tail)* Let S_n be a sum of n i.i.d. $\mathrm{Poi}(\lambda)$ variables. Recall from (2.28) and (2.29) that for $a > \lambda$

$$-\frac{1}{n}\log \mathbb{P}[S_n \geq an] \geq a \log\left(\frac{a}{\lambda}\right) - a + \lambda =: I_\lambda^{\mathrm{Poi}}(a), \tag{5.6}$$

  and similarly for $a < \lambda$

$$-\frac{1}{n}\log \mathbb{P}[S_n \leq an] \geq I_\lambda^{\mathrm{Poi}}(a). \tag{5.7}$$

  To simplify the notation, we let

$$I_\lambda := I_\lambda^{\mathrm{Poi}}(1) = \lambda - 1 - \log \lambda \geq 0, \tag{5.8}$$

  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.

Exploration process For a vertex $v \in [n]$, let $\mathcal{C}_v$ be the connected component containing v, also referred to as the *cluster* of v. To analyze the size of $\mathcal{C}_v$, we introduce a natural procedure to explore $\mathcal{C}_v$ and show that it is dominated above and below by branching processes.

cluster

The exploration process started at v has 3 types of vertices:

active, explored, and neutral vertices

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

We start with $\mathcal{A}_0 := \{v\}$, $\mathcal{E}_0 := \emptyset$, and $\mathcal{N}_0$ contains all other vertices in G_n. At time t, if $\mathcal{A}_{t-1} = \emptyset$ 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 a random element, a_t, from $\mathcal{A}_{t-1}$ and set:

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

We imagine revealing the edges of G_n as they are encountered in the exploration 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 neighbors in G_n. 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 5.1.

Let $A_t := |\mathcal{A}_t|$, $E_t := |\mathcal{E}_t|$, and $N_t := |\mathcal{N}_t|$. Note that (E_t) is non-decreasing while (N_t) is non-increasing. Let

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

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. Moreover, for all t, $(\mathcal{A}_t, \mathcal{E}_t, \mathcal{N}_t)$ forms a partition of $[n]$ so

$$A_t + t + N_t = n, \qquad \forall t \leq \tau_0. \tag{5.9}$$

Hence, in tracking the size of the exploration process, we can work alternatively with A_t or N_t. Specifically, the size of the cluster of v can be characterized as follows.

Lemma 5.6.

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

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. ■

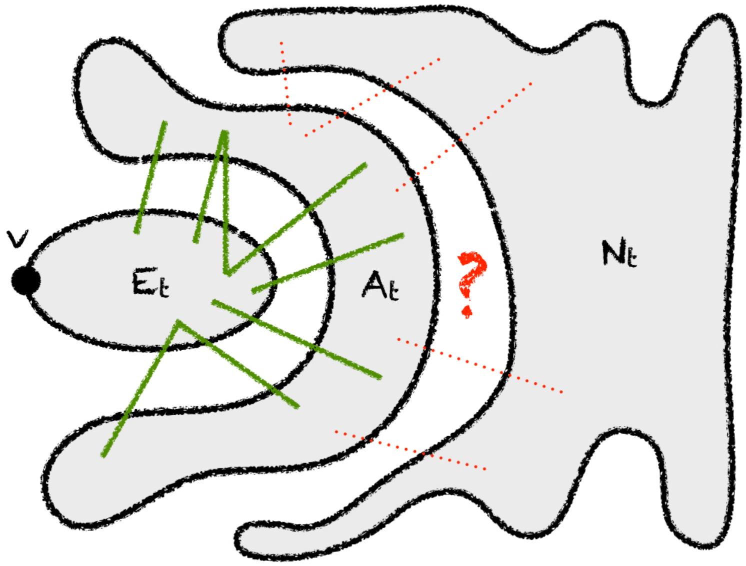

Figure 5.1: Exploration process for $\mathcal{C}_v$. The green edges are in $\mathcal{F}_t$. The red ones are not.

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 an explored vertex and (b) its neutral neighbors become active vertices. That is,

$$A_t = A_{t-1} + \mathbb{1}_{\{A_{t-1}>0\}} \big[\underbrace{-1}_{\text{(a)}} + \underbrace{Z_t}_{\text{(b)}} \big], \tag{5.10}$$

  where Z_t is binomial with parameters $N_{t-1} = n - (t-1) - A_{t-1}$ and p_n. For the coupling arguments below, it will be useful to think of Z_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 $\mathbb{P}[I_{11} = 1] = p_n$. We write

$$Z_t = \sum_{i=1}^{N_{t-1}} I_{t,i}. \tag{5.11}$$

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

$$N_t = N_{t-1} - \mathbb{1}_{\{N_{t-1}<n-(t-1)\}} Z_t. \tag{5.12}$$

Branching process arguments With these observations, we now relate the cluster size of v to the total progeny of a certain branching process. This is the key lemma.

Lemma 5.7 (Cluster size: 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 $k_n = o(\sqrt{n})$,*

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

Proof. 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 (5.4) and (5.5). Observe that the r.h.s. does not depend on N_{t-1}. Let $(Z_t^{\succ})$ be a sequence of independent $\mathrm{Poi}(\lambda)$. Using the coupling in Example 4.8, we can couple the processes $(I_{t,j})_j$ and $(Z_t^{\succ})$ in such way that $Z_t^{\succ} \geq \sum_{j=1}^{n-1} I_{t,j}$ a.s. for all t. Then by induction on t, for all t, $0 \leq A_t \leq A_t^{\succ}$ a.s. where we define

$$A_t^{\succ} := A_{t-1}^{\succ} + \mathbb{1}_{\{A_{t-1}^{\succ}>0\}}\big[-1+Z_t^{\succ}\big], \tag{5.13}$$

with $A_0^{\succ} := 1$. (In fact, this is a domination of Markov transition matrices, as defined in Definition 4.16.) In words, (A_t) is stochastically dominated by the exploration process of a *branching process with offspring distribution* $\mathrm{Poi}(\lambda)$. As a result, letting

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

be the total progeny of the 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

$$A_t^{\prec} := A_{t-1}^{\prec} + \mathbb{1}_{\{A_{t-1}^{\prec}>0\}}\big[-1 + Z_t^{\prec}\big], \tag{5.14}$$

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

$$Z_t^{\prec} := \sum_{i=1}^{n-k_n} I_{t,j}. \tag{5.15}$$

Note that $(A_t^{\prec})$ is the size of the active set in the exploration process of a 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 claim that A_t is bounded from below by $A_t^{\prec}$ up to time

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

Indeed, for all $t \leq \sigma_{n-k_n}$, $N_{t-1} > n - k_n$. Hence, by (5.11) and (5.15), $Z_t \geq Z_t^{\prec}$ for all $t \leq \sigma_{n-k_n}$ and as a result, by induction on t,

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

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, observe that 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 $\sigma_{n-k_n} < +\infty$,

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

On the other hand, when $\sigma_{n-k_n} = +\infty$, $N_t > n - k_n$ for all t—in particular for all $t \geq \tau_0$—and therefore $|\mathcal{C}_v| < k_n$. Moreover in that case, because $A_t \geq A_t^{\prec}$ for all t, 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,

$$\mathbb{P}[\tau_0^{\prec} \geq k_n] \leq \mathbb{P}_{n,p_n}[\tau_0 \geq k_n]. \tag{5.16}$$

2. *(Poisson approximation)* By Lemma **??**,

$$\mathbb{P}[\tau_0^{\prec} = t] = \frac{1}{t}\mathbb{P}\left[\sum_{i=1}^{t} Z_i^{\prec} = t-1\right], \tag{5.17}$$

where the $Z_i^{\prec}$s are independent $\mathrm{Bin}(n-k_n, p_n)$. Note that $\sum_{i=1}^t Z_i^{\prec} \sim \mathrm{Bin}(t(n-k_n), p_n)$. Recall the definition of $(Z_t^{\succ})$ from (5.13). By Lemmas **??** and **??** and the triangle inequality for total variation distance,

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

So by (5.17)

$$\begin{aligned}
\mathbb{P}[\tau_0^{\prec} \geq k_n] &= 1 - \mathbb{P}[\tau_0^{\prec} < k_n] \\
&= 1 - \mathbb{P}[\tau_0^{\succ} < k_n] + O\left(\frac{k_n^2}{n}\right) \\
&= \mathbb{P}[\tau_0^{\succ} \geq k_n] + O\left(\frac{k_n^2}{n}\right).
\end{aligned}$$

Plugging this approximation back into (5.16) gives

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

■

Remark 5.8. *In fact one can get a slightly better lower bound. See Exercise 5.7.*

When k_n is large, the branching process approximation above is not as accurate because of the saturation effect: an Erdös-Rényi graph has a finite pool of vertices from which to draw edges; as the number of neutral vertices decreases, so does the expected number of uncovered edges at each time. Instead we use the following lemma.

Lemma 5.9 (Cluster size: saturation)**.** *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 Lemma 5.6 and (5.9), for any t,

$$\mathbb{P}_{n,p_n}[|\mathcal{C}_v| = t] = \mathbb{P}_{n,p_n}[\tau_0 = t] \leq \mathbb{P}_{n,p_n}[N_t = n-t]. \tag{5.18}$$

Recall that $N_0 = n-1$ and

$$N_t = N_{t-1} - \mathbb{1}_{\{N_{t-1} < n-(t-1)\}} \sum_{i=1}^{N_{t-1}} I_{t,i}. \tag{5.19}$$

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

$$N_t^0 := N_{t-1}^0 - \sum_{i=1}^{N_{t-1}^0} I_{t,i}, \tag{5.20}$$

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 (5.18),

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

which concludes the proof. ■

Combining the previous lemmas we get:

Lemma 5.10 (Bound on the cluster size)**.** *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* $v \in [n]$*.*

- (Subcritical case) *Assume* $\lambda \in (0,1)$. *For all* $\kappa > 0$,

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

- (Supercritical case) *Assume* $\lambda > 1$. *Let* ζ_λ *be the unique solution in* $(0,1)$ *to the fixed point equation*

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

Note that ζ_λ *is the survival probability of a branching process with offspring distribution* $\mathrm{Poi}(\lambda)$. *For any* $\kappa > 0$,

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

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

$$\mathbb{P}_{n,p_n}\left[(1+\kappa_{\delta,\alpha})I_\lambda^{-1}\log n \le |\mathcal{C}_v| \le \alpha n\right] = O(n^{-(1+\delta)}). \tag{5.21}$$

Proof. In both cases we use Lemma 5.7. 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 5.7, by Lemma **??**

$$\mathbb{P}\left[W_\lambda > k_n\right] = \mathbb{P}\left[W_\lambda = +\infty\right] + \sum_{t > k_n} \frac{1}{t}\mathbb{P}\left[\sum_{i=1}^{t} Z_i^{\succ} = t-1\right], \tag{5.22}$$

where the $Z_i^{\succ}$s are i.i.d. $\mathrm{Poi}(\lambda)$. Both terms on the r.h.s. depend on whether or not the mean λ is smaller or larger than 1. We start with the first term. When $\lambda < 1$, the Poisson branching process goes extinct with probability 1. Hence $\mathbb{P}[W_\lambda = +\infty] = 0$. When $\lambda > 1$ on the other hand, $\mathbb{P}[W_\lambda = +\infty] = \zeta_\lambda$, where $\zeta_\lambda > 0$ is the survival probability of the branching process. As to the second term, the sum of the $Z_i^{\succ}$ is λt. When $\lambda < 1$, using (5.6),

$$\begin{aligned}
\sum_{t > k_n} \frac{1}{t}\mathbb{P}\left[\sum_{i=1}^{t} Z_i^{\succ} = t-1\right] &\le \sum_{t > k_n} \mathbb{P}\left[\sum_{i=1}^{t} Z_i^{\succ} \ge t-1\right] \\
&\le \sum_{t > k_n} \exp\left(-t I_\lambda^{\mathrm{Poi}}\left(\frac{t-1}{t}\right)\right) \\
&\le \sum_{t > k_n} \exp\left(-t(I_\lambda - O(t^{-1}))\right) \\
&\le \sum_{t > k_n} C' \exp\left(-t I_\lambda\right) \\
&\le C \exp\left(-I_\lambda k_n\right),
\end{aligned} \tag{5.23}$$

for some constants $C, C' > 0$, where we assume that $k_n = \omega(1)$. When $\lambda > 1$,

$$\begin{aligned}\sum_{t>k_n} \frac{1}{t}\mathbb{P}\left[\sum_{i=1}^{t} Z_i^{\succ} = t-1\right] &\leq \sum_{t>k_n} \mathbb{P}\left[\sum_{i=1}^{t} Z_i^{\succ} \leq t\right] \\ &\leq \sum_{t>k_n} \exp\left(-tI_\lambda\right) \\ &\leq C\exp\left(-I_\lambda k_n\right), \end{aligned} \qquad (5.24)$$

for a possibly different $C > 0$.

Subcritical case: Assume $0 < \lambda < 1$ and let $c = (1+\kappa)I_\lambda^{-1}$ for $\kappa > 0$. By Lemma 5.7,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_1| > c\log n\right] \leq \mathbb{P}\left[W_\lambda > c\log n\right].$$

By (5.22) and (5.23),

$$\mathbb{P}\left[W_\lambda > c\log n\right] = O\left(\exp\left(-I_\lambda c\log n\right)\right), \qquad (5.25)$$

which proves the claim.

Supercritical case: Now assume $\lambda > 1$ and again let $c = (1+\kappa)I_\lambda^{-1}$ for $\kappa > 0$. By Lemma 5.7,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| > c\log n\right] = \mathbb{P}\left[W_\lambda > c\log n\right] + O\left(\frac{\log^2 n}{n}\right), \qquad (5.26)$$

By (5.22) and (5.24),

$$\begin{aligned}\mathbb{P}\left[W_\lambda > c\log n\right] &= \zeta_\lambda + O\left(\exp\left(-cI_\lambda\log n\right)\right) \\ &= \zeta_\lambda + O(n^{-(1+\kappa)}).\end{aligned} \qquad (5.27)$$

Combining (5.26) and (5.27), for any $\kappa > 0$,

$$\mathbb{P}_{n,p_n}\left[|\mathcal{C}_v| > c\log n\right] = \zeta_\lambda + O\left(\frac{\log^2 n}{n}\right). \qquad (5.28)$$

Next, we show that in the supercritical case when $|\mathcal{C}_v| > c\log n$ the cluster size is in fact linear in n with high probability. By Lemma 5.9

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

where $Y_t \sim \mathrm{Bin}(n-1, 1-(1-p_n)^t)$. Roughly, the r.h.s. is negligible until the mean $\mu_t := (n-1)(1-(1-\lambda/n)^t)$ is of the order of t. Let ζ_λ be the unique solution in $(0,1)$ to the fixed point equation

$$f(\zeta) := 1 - e^{-\lambda\zeta} = \zeta.$$

The solution is unique because $f(0) = 0$, $f(1) < 1$, and the f is increasing, strictly concave and has derivative $\lambda > 1$ at 0. Note in particular that, when $t = \zeta_\lambda n$, $\mu_t \approx t$. Let $\alpha < \zeta_\lambda$. For any $t \in [c\log n, \alpha n]$, by a Chernoff bound for Poisson trials (Theorem 2.32),

$$\mathbb{P}[Y_t \le t] \le \exp\left(-\frac{\mu_t}{2}\left(1 - \frac{t}{\mu_t}\right)^2\right). \tag{5.29}$$

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

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

for n large enough, by the properties of f mentioned above. Plugging this back into (5.29), 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 (5.21).

■

Let $\mathcal{C}_{\max}$ be the largest connected component of G_n (choosing the component containing the lowest label if there is more than one such component). Our goal is to characterize the size of $\mathcal{C}_{\max}$. Let

$$X_k := \sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}_v| > k\}},$$

be the number of vertices in clusters of size at least k. There is a natural connection between X_k and $\mathcal{C}_{\max}$, namely,

$$|\mathcal{C}_{\max}| > k \iff X_k > 0 \iff X_k > k.$$

A first moment argument on X_k and the previous lemma immediately imply an upper bound on the size of $\mathcal{C}_{\max}$ in the subcritical case.

Theorem 5.11 (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* $\kappa > 0$,

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

Proof. Let $c = (1+\kappa) I_\lambda^{-1}$ for $\kappa > 0$. We use the first moment method on X_k. By symmetry and the first moment method (Corollary 2.5),

$$\begin{aligned} \mathbb{P}_{n,p_n}\left[|\mathcal{C}_{\max}| > c \log n\right] &= \mathbb{P}_{n,p_n}\left[X_{c \log n} > 0\right] \\ &\leq \mathbb{E}_{n,p_n}\left[X_{c \log n}\right] \\ &= n\, \mathbb{P}_{n,p_n}\left[|\mathcal{C}_1| > c \log n\right]. \end{aligned} \tag{5.30}$$

By Lemma 5.10,

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

as $n \to +\infty$. ■

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

Second moment arguments To characterize the size of the largest cluster in the supercritical case, we need a second moment argument. Assume $\lambda > 1$. For $\delta > 0$ and $\alpha < \zeta_\lambda$, let $\kappa_{\delta,\alpha}$ be as defined in Lemma 5.10. 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 $|\mathcal{C}_v| \leq \underline{k}_n$ a *small vertex*. Let

small vertex

$$Y_k := \sum_{v \in [n]} \mathbb{1}_{\{|\mathcal{C}_v| \leq k\}}.$$

Then $Y_{\underline{k}_n}$ is the number of small vertices. Observe that by definition $Y_k = n - X_k$. Hence by Lemma 5.10, the expectation of $Y_{\underline{k}_n}$ is

$$\mathbb{E}_{n,p_n}[Y_{\underline{k}_n}] = n(1 - \mathbb{P}_{n,p_n}[|\mathcal{C}_v| > \underline{k}_n]) = (1-\zeta_\lambda)n + O\left(\log^2 n\right). \qquad (5.31)$$

Using Chebyshev's inequality (Theorem 2.13), we prove that Y_{k_n} is close to its expectation:

Lemma 5.12 (Concentration of $Y_{\underline{k}_n}$)**.** *For any $\gamma \in (1/2, 1)$ and $\delta < 2\gamma - 1$,*

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

Lemma 5.12, which is proved below, leads to our main result in the supercritical case: the existence of a unique cluster of size linear in n which is referred to as the *giant component*.

giant component

Theorem 5.13 (Supercritical regime: giant component)**.** *For any $\gamma \in (1/2, 1)$ and $\delta < 2\gamma - 1$,*

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

In fact, with probability $1 - o(1)$, there is a unique largest component and the second largest cluster has size $\Omega(\log n)$.

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

$$|\mathcal{C}_{\max}| \leq X_{\underline{k}_n}.$$

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

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

There is equality in 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

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

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

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

Applying Lemmas 5.10 and 5.12 concludes the proof. ■

It remains to prove Lemma 5.12.

Proof of Lemma 5.12. The main task is to bound the variance of $Y_{\underline{k}_n}$. Note that

$$\begin{aligned}
\mathbb{E}_{n,p_n}[Y_k^2] &= \sum_{u,v\in[n]} \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \le k, |\mathcal{C}_v| \le k] \\
&= \sum_{u,v\in[n]} \big\{\mathbb{P}_{n,p_n}[|\mathcal{C}_u| \le k, |\mathcal{C}_v| \le k, u \Leftrightarrow v] \\
&\qquad + \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \le k, |\mathcal{C}_v| \le k, u \not\Leftrightarrow v]\big\}, \qquad (5.32)
\end{aligned}$$

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

To bound the first term in (5.32), we note that $u \Leftrightarrow v$ implies that $\mathcal{C}_u = \mathcal{C}_v$. Hence,

$$\begin{aligned}
\sum_{u,v\in[n]} \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \le k, |\mathcal{C}_v| \le k, u \Leftrightarrow v] &= \sum_{u,v\in[n]} \mathbb{P}_{n,p_n}[|\mathcal{C}_u| \le k, v \in \mathcal{C}_u] \\
&= \sum_{u,v\in[n]} \mathbb{E}_{n,p_n}[\mathbb{1}_{\{|\mathcal{C}_u|\le k\}} \mathbb{1}_{\{v\in\mathcal{C}_u\}}] \\
&= \sum_{u\in[n]} \mathbb{E}_{n,p_n}[|\mathcal{C}_u| \mathbb{1}_{\{|\mathcal{C}_u|\le k\}}] \\
&= n\, \mathbb{E}_{n,p_n}[|\mathcal{C}_1| \mathbb{1}_{\{|\mathcal{C}_1|\le k\}}] \\
&\le nk. \qquad (5.33)
\end{aligned}$$

To bound the second term in (5.32), 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,

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

Observe that this probability is increasing in ℓ. Hence

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

To get a bound on the variance of Y_k, we need to relate this last expression to $(\mathbb{E}_{n,p_n}[Y_k])^2$. For this purpose 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

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

and 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 edges

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

and

$$\sum_{u,v\in[n]}\mathbb{P}_{n,p_n}[|\mathcal{C}_u|\le k,|\mathcal{C}_v|\le k,u\Leftrightarrow v]\le(\mathbb{E}_{n,p_n}[Y_k])^2+\lambda k^2n. \tag{5.34}$$

Combining (5.33) and (5.34), we get

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

The result follows from Chebyshev's inequality (Theorem 2.13) and (5.31).

■

A similar second moment argument also gives a lower bound on the size of the largest component in the subcritical case. We proved in Theorem 5.11 that, when $\lambda < 1$, the probability of observing a connected component of size significantly larger than $I_\lambda^{-1} \log n$ is vanishingly small. In the other direction, we get:

Theorem 5.14 (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* $\kappa \in (0,1)$*,*

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

Proof. Recall that

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

It suffices to prove that with probability $1 - o(1)$ we have $X_k > 0$ when $k = (1-\kappa) I_\lambda^{-1} \log n$. To apply the second moment method, we need an upper bound on the second moment of X_k and a lower bound on its first moment. The following lemma is closely related to Lemma 5.12. Exercise 5.8 asks for a proof.

Lemma 5.15 (Second moment of X_k)**.** *Assume* $\lambda < 1$*. There is* $C > 0$ *such that*

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

Lemma 5.16 (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*

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

for n *large enough.*

Proof. By Lemma 5.7,

$$\begin{aligned}\mathbb{E}_{n,p_n}[X_{k_n}] &= n\,\mathbb{P}_{n,p_n}[|\mathcal{C}_1| > k_n] \\ &= n\,\mathbb{P}[W_\lambda > k_n] - O\left(k_n^2\right). \end{aligned} \tag{5.35}$$

Once again, we use the random-walk representation of the total progeny of a branching process (Lemma **??**). Using the notation of Lemma 5.7,

$$\begin{aligned}\mathbb{P}[W_\lambda > k_n] &= \sum_{t > k_n} \frac{1}{t} \mathbb{P}\left[\sum_{i=1}^t Z_i^\succ = t-1\right] \\ &= \sum_{t > k_n} \frac{1}{t} e^{-\lambda t} \frac{(\lambda t)^{t-1}}{(t-1)!}.\end{aligned}$$

Using Stirling's formula, we note that

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

For any $\varepsilon > 0$, for n large enough,

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

For any $\beta \in (0, \kappa)$, plugging the last line back into (5.35) and taking ε small enough gives

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

■

We return to the proof of Theorem 5.11. Let $k_n = (1-\kappa) I_\lambda^{-1} \log n$. By the second moment method (Corollary 2.19) and Lemmas 5.15 and 5.16,

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

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

Critical regime via martingales To be written. See [NP10].

5.3 Further applications

5.3.1 ▷ *Uniform random trees: local limit*

To be written. See [Gri81].

Exercises

Exercise 5.1 (Galton-Watson process: geometric offspring)**.** Let (Z_t) be a Galton-Watson branching process with geometric offspring distribution (started at 0), i.e., $p_k = p(1-p)^k$ 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 $M_t = m^{-t} Z_t$.

a) Compute the probability generating function f of $\{p_k\}_{k\geq 0}$ and the extinction probability $\eta := \eta_p$ as a function of p.

b) If G is a 2×2 matrix, define

$$G(s) := \frac{G_{11}s + G_{12}}{G_{21}s + G_{22}}.$$

Show that $G(H(s)) = (GH)(s)$.

c) Assume $m \neq 1$. Use b) to derive

$$f_t(s) = \frac{pm^t(1-s) + qs - p}{qm^t(1-s) + qs - p}.$$

Deduce that when $m > 1$

$$\mathbb{E}[\exp(-\lambda M_\infty)] = \eta + (1-\eta)\frac{(1-\eta)}{\lambda + (1-\eta)}.$$

d) 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 5.2 (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

$$p_k^\infty := \zeta^{-1} \sum_{j \geq k} \binom{j}{k} \eta^{j-k} \zeta^k p_j.$$

a) Show that $\{p_k^\infty\}_{k\geq 0}$ is a probability distribution and compute its expectation.

b) Show that for any $k \geq 0$

$$\mathbb{P}[Z_1^\infty = k \,|\, \mathcal{S}] = p_k^\infty.$$

[Hint: Condition on Z_1.]

c) 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 5.3 (Hitting-time theorem: nearest-neighbor walk)**.** Let $X_1, X_2, \ldots$ be i.i.d. random variables taking value $+1$ with probability p and -1 with probability $q := 1 - p$. Let $S_t := \sum_{i=1}^t X_i$ with $S_0 := 0$ and $M_t := \max\{S_i \,:\, 0 \leq i \leq t\}$.

a) For $r \geq 1$, use the reflection principle to show that

$$\mathbb{P}[M_t \geq r, S_t = b] = \begin{cases} \mathbb{P}[S_t = b], & b \geq r, \\ (q/p)^{r-b}\mathbb{P}[S_t = 2r - b], & b < r. \end{cases}$$

b) Use a) to give a proof of the hitting-time theorem in this special case: letting τ_b be the first time S_t hits $b > 0$, then show that for all $t \geq 1$

$$\mathbb{P}[\tau_b = t] = \frac{b}{t}\mathbb{P}[S_t = b].$$

[Hint: Consider the probability $\mathbb{P}[M_{t-1} = S_{t-1} = b - 1, S_t = b]$.]

Exercise 5.4 (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

$$p_c(G) \geq p_c(\widehat{\mathbb{T}}_b),$$

by bounding the expected size of the cluster of 0. [Hint: Consider self-avoiding paths started at 0.]

Exercise 5.5 (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

$$h(\varepsilon, u) := u - \left(\left(1 - \tfrac{1}{b} - \varepsilon\right)(1 - u) + \tfrac{1}{b} + \varepsilon\right)^b.$$

a) Show that there is a constant $C > 0$ not depending on ε, u such that

$$\left| h(\varepsilon, u) - b\varepsilon u + \frac{b-1}{2b} u^2 \right| \le C(u^3 \vee \varepsilon u^2).$$

b) Use a) 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 5.6 (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

$$\mathbb{P}_p[Z_p = \ell] = \frac{\ell\, \mathbb{P}_p[|\mathcal{C}_0| = \ell]}{\mathbb{E}_p|\mathcal{C}_0|}, \qquad \forall \ell \ge 1.$$

a) Using the explicit formula for $\mathbb{P}_p[|\mathcal{C}_0| = \ell]$ derived in Section 5.2.1, 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 $C > 0$.

b) Show that for all $k \ge 2$ there is $C_k > 0$ such that

$$\lim_{p \uparrow p_c(\widehat{\mathbb{T}}_2)} \frac{\mathbb{E}_p|\mathcal{C}_0|^k}{(p_c(\widehat{\mathbb{T}}_2) - p)^{-1-2(k-1)}} = C_k.$$

c) What happens when $p \downarrow p_c(\widehat{\mathbb{T}}_2)$?

Exercise 5.7 (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 $\mathrm{Poi}(\lambda)$.

a) Show that

$$|\mathbb{P}[W_{n,p_n} \geq k] - \mathbb{P}[W_\lambda \geq k]| \\ \leq \max\{\mathbb{P}[W_{n,p_n} \geq k, W_\lambda < k], \mathbb{P}[W_{n,p_n} < k, W_\lambda \geq k]\}.$$

b) Couple the two processes step-by-step and use a) 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 5.8 (Subcritical Erdös-Rényi: Second moment)**.** Prove Lemma 5.15.

Bibliographic remarks

Section 5.1 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 [vdH14, Chapter 3]. The classical reference on branching processes is [AN04]. The Kesten-Stigum theorem is due to Kesten and Stigum [KS66a, KS66b, KS67]. Our proof of a weaker version with an L^2 condition follows [Dur10, Example 5.4.3]. Spitzer's combinatorial lemma is due to Spitzer [Spi56]. The proof presented here follows [Fel71, Section XII.6]. The hitting-time theorem was first proved by Otter [Ott49]. Several proofs of a generalization can be found in [Wen75]. The critical percolation threshold for percolation on Galton-Watson trees is due to R. Lyons [Lyo90].

Section 5.2 The presentation in Section 5.2.1 follows [vdH10]. See also [Dur85].

For much more on the phase transition of Erdös-Rényi graphs, see e.g. [vdH14, 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 [vdH14, Section 4.5].

Section 5.3