PERCOLATION ON GALTON-WATSON TREES

Marcus Michelen

A DISSERTATION

in

Mathematics

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Robin Pemantle, Merriam Term Professor of Mathematics

Graduate Group Chairperson

Julia Hartmann, Professor of Mathematics

Dissertation Committee:
Ryan Hynd, Associate Professor of Mathematics
Robin Pemantle, Merriam Term Professor of Mathematics
Henry Towsner, Associate Professor of Mathematics

Acknowledgments

There are many people I'd like to thank, but first and foremost is my advisor, Robin Pemantle, for all of his help and support over the past five years. I'd also like to specially thank Josh Rosenberg and Julian Sahasrabudhe for countless mathematical conversations and their friendship, without which our collaborations would not have happened. Other Penn faculty I'd like to thank are Professors Ryan Hynd, Greta Panova and Henry Towsner as well as Steven Melczer for their advice and fantastic topics courses. There are many graduate students I'd like to thank: Debapratim Banerjee, Konstantinos Karatapanis, Somabha Mukherjee, and Mateo Wirth from the probability group; Michael Gerapetritis, The Gia Hoang, Yuhang Liu, Benjamin Pollak, Dominick Villano, Zhen Zeng from my cohort; and many other graduate students in both the math department and the broader Penn and Philadelphia communities.

I'd like to thank my parents along with the rest of my family; my partner Nate Biagiotti; and all of my friends. A last special thanks to Monica Pallanti, Paula Scarborough, Reshma Tanna and Robin Toney, without whom the Penn

math department would probably fall apart.

ABSTRACT

PERCOLATION ON GALTON-WATSON TREES

Marcus Michelen

Robin Pemantle, Advisor

We consider both Bernoulli and invasion percolation on Galton-Watson trees. In the former case, we show that the quenched survival function is smooth on the supercritical window and smooth from the right at criticality. We also study critical percolation conditioned to reach depth n, and construct the incipient infinite cluster by taking $n \to \infty$; quenched limit theorems are proven for the asymptotic size of the layers of the incipient infinite cluster. In the case of invasion percolation, we show that the law of the unique ray in the invasion cluster is absolutely continuous with respect to the limit uniform measure. All results are under assumptions for the offspring distribution of the underlying Galton-Watson tree.

Contents

Chapter 1

Introduction

Percolation is a catch-all term for processes that randomly thin out large graphs. The most classical instance of this is Bernoulli bond percolation: for a given number $p \in [0, 1]$, delete each edge of a graph independently with probability $1-p$. Already a large number of questions appear; a fundamental question asks if there a positive probability that there is an infinite connected component after deleting.

Percolation is interesting from many different angles. From an applied mathematics perspective, it is an oft-used model in material science for modeling how a liquid percolates through some medium. The local properties of the medium are modeled randomly and may be tuned in unison by the parameter p. For mathematicians, percolation is interesting primarily because it exhibits a phase transition. In the case where the underlying graph is $\mathbb{Z}^2$ with nearest neighbor edges, the Kesten-Harris theorem [Kes80, Har60] states that the probability there is an infinite cluster in p-percolation is positive if and only if $p > p_c := \frac{1}{2}$. This phase transition is quite dramatic: Menshikov [Men86, MMS86] showed that for $p < p_c$, the probability that the percolation cluster containing the origin has radius at least n decays exponentially in n. These results generalize to a wide variety of infinite graphs: there is a critical p_c above which there is a positive probability of an infinite cluster existing, and below which percolation clusters are quite small.

Many major open problems in probability theory concern the near-critical behavior in percolation as well as other processes that exhibit phase transitions [GBGL08, Chapter IV.25]. One central open question concerns whether or not there is percolation at criticality on $\mathbb{Z}^d$: namely, if we perform p_c percolation on $\mathbb{Z}^d$, is there an infinite cluster with positive probability? This has been shown to hold for $d \geq 11$ [HS94, FvdH17] by showing that critical percolation exhibits mean-field behavior in these dimensions.

The work of this thesis concerns percolation on Galton-Watson trees, a certain class of random trees. These graphs are geometrically simple—they have no cycles, by definition. However, the stochastic roughness introduced by working on a random graph eliminates much of the symmetry typically exploited in percolation problems. Furthermore, there is a relationship between percolation and random walk on trees, as discussed in depth in [LP17, Chapter 5], further motivating the study of percolation on trees.

This thesis consists of the content of three papers, each given their own chapter:

Chapters 2, 3 and 4 correspond to the works [MPR18], [Mic19], and [MPR17] respectively, where the first and last works are joint with Robin Pemantle and Josh Rosenberg. The intention is for each chapter to be self-contained, and thus each chapter retains its introduction. Many proofs are omitted for brevity, and thus a more detail-oriented reader should refer to the paper versions for completeness. In addition, a brief history of invasion percolation appears in Section 1.1.

1.1 A History of Invasion Percolation

Invasion percolation is a self-tuning algorithm that chooses a subgraph from a large graph at random. Its precise description has evolved slightly since its creation and has converged to various equivalent definitions; we take our definition from [DHS18]: Let $G = (V, E)$ be an infinite graph and assign independent uniform $[0, 1]$ variables—which are referred to as weights—to each edge. Initialize V_0 to be a single vertex and E_0 to be the empty set. At time n, let e_n be the edge in the boundary

$$\partial G_{n-1} = \{\{x, y\} : x \in V_{n-1}, \{x, y\} \notin E_{n-1}\}$$

of minimal weight. We then define

$$V_n = V_{n-1} \cup \{y\}, \quad E_n = E_{n-1} \cup \{e_n\}, \quad G_n = (V_n, E_n) .$$

The family of graphs G_n is increasing, and the *invasion percolation cluster* I is defined as the union over G_n. Importantly, there are no input parameters. In a sense that has been made rigorous in many settings, invasion percolation looks more and more like critical percolation as $n \to \infty$; this means that invasion percolation is an example of a system that exhibits self-organized criticality.

Material Science Origins

Invasion percolation was originally developed in the material science community to study how two different fluids interact in a porous medium. As a precursor, a 1978 work by de Gennes and Guyon [dGG78] models a porous medium as a graph and places directed edges in both directions at each edge. We imagine a source of water with pressure p is placed at a specified vertex $\mathbf{0}$, and the oil is placed at all other vertices. For each edge $\vec{ab}$, there is an associated quantity $\Phi_{\vec{ab}}$ denoting how difficult it is for water to cross from a to b thereby invading the oil via capillary action. The authors then assume that $\Phi_{\vec{ab}}$ are jointly independent random variables with $\Phi_{\vec{ab}}$ and $\Phi_{\vec{ba}}$ having the same distribution. Given neighboring vertices a and b with water at a and oil at b, water spreads to b provided $\Phi_{\vec{ab}} \leq p$, where we recall that p is the pressure of the water source.

Importantly, the presence of the external parameter p distinguishes this model from what is now known as invasion percolation. In fact, as [dGG78] note, this is the same as performing independent directed bond percolation and restricting the cluster only to sites reachable from $\mathbf{0}$ through open bonds. As the authors state,

> Nous nous préoccupons surtout de savoir si le comportement près du

> seuil, dans des expériences d'injection ou de succion, est régi par les mêmes exposants [critique] que ceux de la percolation.[1]

They note as well that similar percolation models were considered contemporaneously by [LSD77]. The model of [dGG78] was further tweaked and studied via Monte Carlo simulations in [LB80, LZS83, Len85]. Despite the differences between what is now known as invasion percolation, many authors cite [dGG78] or [LB80] as the beginnings of invasion percolation, including foundational mathematical works on this subject [CCN85, Zha95], despite the strong differences.

Invasion percolation as it is now known was more-or-less introduced in 1980 with [CKLW82] and was given the name invasion percolation in a subsequent work [WW83] bearing the name "Invasion percolation: a new form of percolation theory."

Most work in this era consists of simulations and observations concerning the results, although the work [CKLW82] contains quite a bit of mathematical content. In particular, they examine the case of a large $L \times W$ rectangle in $\mathbb{Z}^2$ with periodic boundary conditions. They impose a *trapping condition*: once the invasion cluster traps a component of its complement, it may not further invade the trapped component. They observe a power law for the fraction of vertices that end up in a trapped component.

They note that these exponents roughly match the analogue in the case of critical

[1] Rough translation: "Our main concern is whether near-critical behavior, in injection or suction experiments, is governed by the same [critical] exponents as percolation."

Bernoulli percolation and state that this matching of exponents "suggest[s] that [invasion percolation] is indeed at a critical point." Further, the authors heuristically state that eventually all added bonds have weight in the range $[0, p_c]$ although this is not quite right since no percolation occurs at criticality in $\mathbb{Z}^2$; in truth, for each $\varepsilon > 0$, all added bonds eventually have weight in the range $[0, p_c + \varepsilon)$.

Wilkinson—an author on [CKLW82]—continued this scaling-exponent approach with Willemsen in [WW83], and consider invasion percolation both with and without the trapping condition. They compare exponents of various quantities from their simulated data of invasion percolation with those of Bernoulli percolation. The authors concede that their work is "essentially descriptive" and that

> "It would be of interest to study [invasion percolation] in a more formal manner, ... even the simplest problem of growing a cluster from a point into an infinite lattice without the trapping rule appears intractable."

They conclude by stating "The development of a mathematical framework for discussing this structure poses a very interesting problem."

Enter Mathematics

Mathematical analysis begins in 1983 with a work of Nickel and Wilkinson [NW83] concerning invasion percolation on regular trees. Using a generating function approach, the authors show that the probability of adding a bond with weight larger than p_c at step n is on the order of $1/\sqrt{n}$ and that for any $\varepsilon > 0$, the probability of adding a bond of weight larger than $p_c + \varepsilon$ decays exponentially. Further, they find the scaling form for the number of invaded vertices in depth m at time n.

These results were generalized by Chayes, Chayes and Newman in 1985 [CCN85], written contemporaneously with [NW83]. The authors state their methods work for a many examples, but restrict their attention to $\mathbb{Z}^d$ for simplicity. They prove

- Let $Q_n(x)$ be the portion of bonds added up to time n with weight at most x. Then $Q_n(x)$ converges to the step function

$$Q(x) = \begin{cases} 1 & x < p_c \\ \\ 0 & x > p_c \end{cases}$$

  under since-verified hypotheses concerning Bernoulli percolation.

- For $y > p_c$, let $A_n(y)$ denote the event that the weight of the bond added at time n is at least y. Then

$$-\log(\mathbf{P}(A_n(y))) = \Theta(n^{(d-1)/d})$$

  as $n \to \infty$.

- In $d = 2$, the invaded region has zero volume fraction.

As Chayes, Chayes and Newman observe, the last point represents a step towards computing the fractal dimension of the invasion cluster, a problem resolved in 1995 by Zhang [Zha95].

A follow-up work [CCN87] by the same trio of authors began a productive feedback loop between invasion and Bernoulli percolation; they use invasion percolation to prove that for Bernoulli percolation on $\mathbb{Z}^d$

- $-\log(\mathbf{P}[x \text{ and } y \text{ belong to the same finite cluster}]) = \Theta(\|x - y\|)$.
- The probability that the root is contained in an infinite cluster is a smooth function of p on the open supercritical window.
- The probability that the cluster containing the origin has exactly n vertices is bounded above by $\exp\left(-c_p n^{(d-1)/d} / \log(n)\right)$.

While [CCN87] using invasion percolation for theoretical results, physicists began using invasion percolation to approximate critical percolation parameters with Monte Carlo simulations; an example of this is the work of [McC87] which studies the case of Voronoi percolation.

Certain aspects of [CCN85, CCN87] have been generalized to quasi-transitive and semi-transitive graphs in [HPS99]. In particular, [HPS99] uses invasion percolation examine the uniqueness of infinite clusters for coupled p-percolation on these graphs. In the process, they prove that for any $p > p_c$, invasion percolation adds only finitely many bonds with weight larger than p.

More work on $\mathbb{Z}^2$

From this point, the study of invasion percolation spread out in various different directions. In the case of $\mathbb{Z}^2$, much work has been done comparing invasion percolation to critical percolation. Járai [Jár03] showed that moments of the number of invaded vertices in the box $[-n,n]^2$ is on the same order as critical percolation conditioned to reach the boundary of this box. Similarly, Járai showed that the local structure of the invasion cluster when viewed sufficiently far from the origin is the same as that of the *incipient infinite cluster*, which is roughly critical percolation conditioned to percolate to infinity.

More involved features were studied: [vdBJV07] showed that the size of *the first pond*—the portion of the invasion cluster up until the edge of maximal total weight is added—has tails comparable to the radius of the critical percolation cluster, up to a logarithmic factor. Comparisons continued in [DSV09], which showed that certain k-point functions for the invasion cluster match analogues for critical percolation; however, in the same paper, it is shown that the laws of the incipient infinite cluster and invasion cluster are mutually singular. In a similar vein, [Sap11] showed the incipient infinite cluster doesn't stochastically dominate the invasion cluster. Further work on $\mathbb{Z}^2$ often concerns ponds and the weights connecting them [vdBJV07, DS11, DS12]. The connections between critical and invasion percolation are still being explored: as recently as 2018, Damron, Hanson and Sosoe [DHS18] examine so-called *arm events*—the existence of a family of disjoint paths with prescribed open and closed conditions connecting a given vertex to a large box—and study which open/closed conditions yield arm events whose probabilities roughly match in invasion percolation and critical percolation.

Other work on $\mathbb{Z}^2$ includes [DHS13] which shows that random walks on the invasion cluster are subdiffusive. Their analysis relies on Russo-Seymore-Welsh estimates, and note that their results extend to planar lattices with similar estimates. Recently, [GPS18] proved that the invasion cluster of the two-dimensional triangular lattice has a unique scaling limit.

More work on Trees

Invasion percolation on regular trees was explored in depth by Angel, Goodman, den Hollander, and Slade in [AGdHS08]. The authors examine the scaling behavior of the r-point function as well as the volume at and up to a given height. An important ingredient of their work is the representation of the invasion cluster as an infinite non-backtracking path called *the backbone* with subcritical percolation clusters added along the way. Interestingly, [AGdHS08] show that the law of the incipient cluster stochastically dominates that of the invasion cluster on regular trees, however the two are mutually singular. This differs quite a bit from the case of $\mathbb{Z}^2$ in which there is no stochastic dominance, as shown in [Sap11].

Angel and Goodman continued their work on regular trees in a 2013 work with Merle [AGM13] which identifies the scaling limit of both the invasion cluster and incipient infinite cluster on regular trees. The work [MPR17]—the basis for Chapter 4 of this thesis—generalizes certain facts of [AGdHS08] to Galton-Watson trees, thereby bringing the mathematical study of invasion percolation into the realm of random graphs. On almost-every Galton-Watson trees, the invasion cluster almost surely contains a single infinite path; [MPR17] studies the law of this path, and shows that this law is absolutely continuous with respect to the limit uniform measure.

The backbone decompositions central to [AGdHS08, AGM13] was taken to a locally-infinite limit setting in [ABGK12], in which each vertex has countably infinitely-many children, but the edge weights are no longer uniform.

Chapter 2

Super-Critical Percolation

The present chapter is based on excerpts from [MPR18], which is joint with Robin Pemantle and Josh Rosenberg. For brevity, proofs are omitted and abbreviated from certain sections.

2.1 Introduction

As earlier, let $\mathtt{GW}$ denote the measure on locally finite rooted trees induced by the Galton-Walton process for some fixed progeny distribution $\{p_n\}$ whose mean will be denoted μ. A random tree generated according to the measure $\mathtt{GW}$ will be denoted as $\mathbf{T}$. Throughout, we let Z denote a random variable with distribution $\{p_n\}$ and assume that $\mathbf{P}[Z = 0] = 0$; passing to the reduced tree as described in [AN72, Chapter 1.D.12], no generality is lost for any of the questions in the paper.

The growth rate and regularity properties of both random and deterministic trees can be analyzed by looking at the behavior of a number of different statistics. The Hausdorff dimension of the boundary and the escape speed of random walk are almost surely constant for a fixed Galton-Watson measure. Quantities that are random but almost surely well defined include the martingale limit $W := \lim Z_n/\mu^n$, the resistance to infinity when edges at level n carry resistance x^n for a fixed $x < \mu$, and the probability $\theta_{\mathbf{T}}(p)$ that $\mathbf{T}$ survives Bernoulli-p percolation, i.e., the probability there is a path of open edges from the root to infinity, where each edge is declared open with independent probability p. In this paper we seek to understand $\texttt{GW}$-almost sure regularity properties of the survival function $\theta_{\mathbf{T}}(\cdot)$ and to compute its derivatives at criticality.

The properties of the Bernoulli-p percolation survival function have been studied extensively in certain other cases, such as on the deterministic d-dimensional integer lattice, $\mathbb{Z}^d$. When $d = 2$, the Harris-Kesten Theorem [Har60, Kes80] states that the critical percolation parameter p_c is equal to $1/2$ and that critical percolation does not survive: $\theta_{\mathbb{Z}^2}(1/2) = 0$; more interesting is the nondifferentiability from the right of the survival function at criticality [KZ87]. When $d \geq 3$, less is known, despite the high volume of work on the subject. The precise value of the critical probability $p_c(d)$ is unknown for each $d \geq 3$; for $d \geq 19$, mean-field behavior has been shown to hold, implying that percolation does not occur at criticality [HS94]. This has recently been upgraded with computer assistance and shown to hold for $d \geq 11$ [FvdH17], while the cases of $3 \leq d \leq 10$ are still open. Lower bounds on the survival probability of $\mathbb{Z}^d$ in the supercritical region are an area of recent work [DCT16], but exact behavior near criticality is not known in general. On the question of regularity, the function $\theta_{\mathbb{Z}^d}(p)$ is smooth on $(p_c(d), 1]$ for each $d \geq 2$ [Gri99, Theorem 8.92].

There is less known about the behavior of $\theta_{\mathbf{T}}(\cdot)$ for random trees than is known on the integer lattice. We call the random function $\theta_{\mathbf{T}}(\cdot)$ the *quenched survival function* to distinguish it from the *annealed survival function* θ, where $\theta(x)$ is the probability of survival at percolation parameter x averaged over the GW distribution. For the regular d-ary tree, T_d, the classical theory of branching processes implies that the critical percolation parameter p_c is equal to $1/d$, that $\theta_{T_d}(1/d) = 0$ (that is, there is no percolation at criticality), and that for $p > p_c$, the quantity $\theta_{T_d}(p)$ is equal to the largest fixed point of $1 - (1 - px)^d$ in $[0, 1]$ (see, for instance, [AN72] for a treatment of this theory).

For Galton-Watson trees, a comparison of the quenched and annealed survival functions begins with the following classical result of Lyons, showing that p_c is the same in both cases.

Theorem 2.1.1 ([Lyo90])**.** *Let* $\mathbf{T}$ *be the family tree of a Galton-Watson process with mean* $\mathbf{E}[Z] =: \mu > 1$*, and let* $p_c(\mathbf{T}) = \sup\{p \in [0, 1] : \theta_{\mathbf{T}}(p) = 0\}$*. Then* $p_c(\mathbf{T}) = \frac{1}{\mu}$ *almost surely. Together with the fact that* $\theta_{\mathbf{T}}(1/\mu) = 0$*, this implies* $\theta_{\mathbf{T}}(p_c) = 0$ *almost surely.* $\square$

To dig deeper into this comparison, observe first that the annealed survival probability $\theta(x)$ is the unique fixed point on $[0,1)$ of the function $1-\phi(1-px)$ where $\phi(z) = \mathbf{E}[z^Z]$ is the probability generating function of the offspring distribution. In the next section we show that the annealed survival function $\theta(p)$ is smooth on $(p_c, 1)$ and, under moment conditions on the offspring distribution, the derivatives extend continuously to p_c. This motivates us to ask whether the same holds for the quenched survival function. Our main results show this to be the case, giving regularity properties of $\theta_{\mathbf{T}}(p)$ on the supercritical region.

Let r_j be the coefficients in the asymptotic expansion of the annealed function g at p_c. These are shown to exist in Proposition 2.2.6 below. In Theorem 2.3.1, under appropriate moment conditions, we will construct for each $j \geq 1$ a martingale $\{M_n^{(j)} : n \geq 1\}$ with an almost sure limit $M^{(j)}$, that is later proven to equal the jth coefficient in the aymptotic expansion of the quenched survival function g at p_c. Throughout the analysis, the expression W denotes the martingale limit $\lim Z_n/\mu^n$.

Theorem 2.1.2 (main results)**.**

(i) For `GW` *a.e. tree* $\mathbf{T}$*, the quantity* $\theta_{\mathbf{T}}(x)$ *is smooth as a function of* x *on* $(p_c, 1)$*.*

(ii) If $\mathbf{E}Z^{2k+1+\beta} < \infty$ *for some positive integer* k *and some* $\beta > 0$*, then we have the* k*-th order approximation*

$$\theta_{\mathbf{T}}(p_c + \varepsilon) = \sum_{j=1}^{k} M^{(j)} \varepsilon^j + o(\varepsilon^k)$$

for `GW` *a.e. tree* $\mathbf{T}$*, where* $M^{(j)}$ *is the quantity given explicitly in Theorem 2.3.1. Additionally,* $M^{(1)} = Wr_1$ *and* $\mathbf{E}[M^{(j)}] = r_j$*, where* W *is the martingale limit* *for* $\mathbf{T}$ *and* $j!r_j$ *are the derivatives of the annealed survival function, for which explicit expressions are given in Proposition 2.2.6.*

(iii) If $\mathbf{E}Z^{2k^2+3+\beta} < \infty$ *for some* $\beta > 0$*, then* $\texttt{GW}$*-almost surely* $\theta_{\mathbf{T}}(\cdot)$ *is of class* C^k *from the right at* p_c *and* $g^{(j)}(\mathbf{T}, p_c^+) = j!M^{(j)}$ *for all* $j \leq k$*; see the beginning of Section 2.2.1 for calculus definitions.*

Remark 2.1.3. Smoothness of $\theta_{\mathbf{T}}(\cdot)$ on $(p_c, 1)$ does not require any moment assumptions, in fact even when $\mathbf{E}Z = \infty$ one has $p_c = 0$ and smoothness of $\theta_{\mathbf{T}}(\cdot)$ on $(0, 1)$. The moment conditions relating to expansion at criticality given in (ii) are probably not best possible, but are necessary in the sense that if $\mathbf{E}Z^k = \infty$ for some k then not even the annealed survival function is smooth (see Proposition 2.2.4 below).

The proofs of the first two parts of Theorem 2.1.2 are independent of each other. Part (*ii*) is proved first, in Section 2.3. Part (*i*) is proved in Section 2.4.2 after some preliminary work in Section 2.4.1. Finally, part (*iii*) is proved in Section 2.4.3. The key to these results lies in a number of different expressions for the probability of a tree T surviving p-percolation and for the derivatives of this with respect to p. The first of these expressions is obtained via inclusion-exclusion. The second, Theorem 2.4.1 below, is a Russo-type formula [Rus81] expressing the derivative in terms of the *expected branching depth*

$$\frac{d}{dp}\theta_T(p) = \frac{1}{p}\mathbf{E}_T|B_p|$$

for $\texttt{GW}$-almost every T and every $p \in (p_c, 1)$, where $|B_p|$ is the depth of the deepest vertex B_p whose removal disconnects the root from infinity in p-percolation. The third generalizes this to a combinatorial construction suitable for computing higher moments.

A brief outline of the chapter is as follows. Section 2.2 contains definitions, preliminary results on the annealed survival function, and a calculus lemma. Section 2.3 writes the event of survival to depth n as a union over the events of survival of individual vertices, then obtains bounds via inclusion-exclusion. Let $X_n^{(j)}$ denote the expected number of cardinality j sets of surviving vertices at level n, and let $X_n^{(j,k)}$ denote the expected kth falling factorial of this quantity. These quantities diverge as $n \to \infty$ but inclusion-exclusion requires only that certain signed sums converge as $n \to \infty$. The Bonferroni inequalities give upper and lower bounds on $\theta_T(\cdot)$ for each n. Strategically choosing n as a function of ε and using a modified Strong Law argument allows us to ignore all information at height beyond n (Proposition 2.3.10). Each term in the Bonferroni inequalities is then individually Taylor expanded, yielding an expansion of $\theta_T(p_c + \varepsilon)$ with coefficients depending on n. Letting $\mathbf{T} \sim \mathtt{GW}$ and $n \to \infty$, the variables $X_n^{(j,k)}$ separate into a martingale part and a combinatorial part. The martingale part converges exponentially rapidly (Theorem 2.3.6). The martingale property for the cofficients themselves (Lemma 2.3.13) follows from some further analysis (Lemma 2.3.12) eliminating the combinatorial part when the correct signed sum is taken.

Section 2.4.1 proves the above formula for the derivative of θ (Theorem 2.4.1) via a Markov property for the coupled percolations as a function of the percolation parameter p. Section 2.4.2 begins with a well-known branching process description of the subtree of vertices with infinite lines of descent. It then goes on to describe higher order derivatives in terms of combinatorial gadgets denoted $\mathcal{D}$ which are moments of the numbers of edges in certain rooted subtrees of the percolation cluster and generalize the branching depth. We then prove an identity for differentiating these and apply it repeatedly to $\theta'(T,p) = p^{-1}\mathbf{E}B_p$, to write $(\partial/\partial p)^k\theta(T,p)$ as a finite sum $\sum_\alpha \mathcal{D}_\alpha$ of factorial moments of sets of surviving vertices. This suffices to prove smoothness of the quenched survival function on the supercritical region $p_c < p < 1$.

For continuity of the derivatives at p_c, an analytic trick is required. If a function possessing an order N asymptotic expansion at the left endpoint of an interval $[a,b]$ ceases to be of class C^k at the left endpoint for some k, then the $k+1$st derivative must blow up faster than $(x-a)^{-N/k}$ (Lemma 2.2.1). This is combined with bounds on how badly things can blow up at p_c (Proposition 2.4.11) to prove continuity from the right at p_c of higher order derivatives.

The paper ends by listing some questions left open, concerning sharp moment conditions and whether an asymptotic expansion ever exists without higher order derivatives converging at p_c.

2.2 Constructions, preliminary results, and annealed survival

2.2.1 Smoothness of real functions at the left endpoint

Conclusion (*iii*) glues together the conclusions of (*i*) and (*ii*) to show that the random function $\theta_{\mathbf{T}}(\cdot)$ is in fact smooth on the set $[p_c, 1)$. A useful fact is the following analytic Lemma:

Lemma 2.2.1. *Let $f : [a, b] \to \mathbb{R}$ be C^∞ on (a, b) with*

$$f(a + \varepsilon) = c_1\varepsilon + \cdots + c_k\varepsilon^k + \cdots + c_N\varepsilon^N + o(\varepsilon^N) \tag{2.2.1}$$

for some k, N with $1 \le k < N$, and assume

$$\lim_{\varepsilon\to 0} f^{(j)}(a + \varepsilon) = j!c_j \tag{2.2.2}$$

for all j such that $1 \le j < k$. If $f^{(k)}(a+\varepsilon) \not\to k!c_k$ as $\varepsilon \to 0^+$, then there must exist positive numbers $u_n \downarrow 0$ such that

$$\left|f^{(k+1)}(u_n)\right| = \omega\left(u_n^{-\frac{N}{k}}\right).$$

□

2.2.2 Galton-Watson trees

Since we will be working with probabilities on random trees, it will be useful to explicitly describe our probability space and notation. We begin with some notation we use for all trees, random or not. Let $\mathcal{U}$ be the canonical *Ulam-Harris* tree [ABF13]. The vertex set of $\mathcal{U}$ is the set $\mathbf{V} := \bigcup_{n=1}^{\infty} \mathbb{N}^n$, with the empty sequence $\mathbf{0} = \emptyset$ as the root. There is an edge from any sequence $\mathbf{a} = (a_1, \ldots, a_n)$ to any extension $\mathbf{a} \sqcup j := (a_1, \ldots, a_n, j)$. The depth of a vertex v is the graph distance between v and $\mathbf{0}$ and is denoted $|v|$. We work with trees T that are locally finite rooted subtrees of $\mathcal{U}$. The usual notations are in force: T_n denotes the set of vertices at depth n; $T(v)$ is the subtree of T at v, canonically identified with a rooted subtree of $\mathcal{U}$, in other words the vertex set of $T(v)$ is $\{w : v \sqcup w \in V(T)\}$ and the least common ancestor of v and w is denoted $v \wedge w$.

Turning now to Galton-Watson trees, let $\phi(z) := \sum_{n=1}^{\infty} p_n z^n$ be the offspring generating function for a supercritical branching process with no death, i.e., $\phi(0) = 0$. We recall,

$$
\begin{aligned}
\phi'(1) &= \mathbf{E}Z =: \mu \\
\phi''(1) &= \mathbf{E}[Z(Z-1)]
\end{aligned}
$$

where Z is a random variable with probability generating function ϕ. We will work on the canonical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ where $\Omega = (\mathbb{N} \times [0,1])^{\mathbf{V}}$ and $\mathcal{F}$ is the product Borel σ-field. We take $\mathbf{P}$ to be the probability measure making the coordinate functions $\omega_v = (\deg_v, U_v)$ i.i.d. with the law of (Z, U), where U is uniform on $[0,1]$ and independent of Z. The variables $\{\deg_v\}$, where $\deg_v$ is interpreted as the number of children of vertex v, will construct the Galton-Watson tree, while the variables $\{U_v\}$ will be used later for percolation. Let $\mathbf{T}$ be the random rooted subtree of $\mathcal{U}$ which is the connected component containing the root of the set of vertices that are either the root or are of the form $v \sqcup j$ such that $0 \leq j < \deg_v$. This is a Galton-Watson tree with offspring generating function ϕ. Let $\mathcal{T} := \sigma(\{\deg_v\})$ denote the σ-field generated by the tree $\mathbf{T}$. The $\mathbf{P}$-law of $\mathbf{T}$ on $\mathcal{T}$ is $\mathtt{GW}$.

As is usual for Galton-Watson branching processes, we denote $Z_n := |\mathbf{T}_n|$. Extend this by letting $Z_n(v)$ denote the number of offspring of v in generation $|v| + n$; similarly, extend the notation for the usual martingale $W_n := \mu^{-n} Z_n$ by letting $W_n(v) := \mu^{-n} Z_n(v)$. We know that $W_n(v) \to W(v)$ for all v, almost surely and in L^q if the offspring distribution has q moments. This is stated without proof for integer values of $q \geq 2$ in [Har63, p. 16] and [AN72, p. 33, Remark 3]; for a proof for all $q > 1$, see [BD74, Theorems 0 and 5]. Further extend this notation by letting $v^{(i)}$ denote the ith child of v, letting $Z_n^{(i)}(v)$ denote nth generation descendants of v whose ancestral line passes through $v^{(i)}$, and letting $W_n^{(i)}(v) := \mu^{-n} Z_n^{(i)}(v)$. Thus, for every v, $W(v) = \sum_i W^{(i)}(v)$. For convenience, we define $p_c := 1/\mu$, and recall that p_c is in fact $\mathtt{GW}$-a.s. the critical percolation parameter of T as per Theorem 2.1.1.

Bernoulli percolation

Next, we give the formal construction of Bernoulli percolation on random trees. For $0 < p < 1$, simultaneously define Bernoulli(p) percolations on rooted subtrees T of $\mathcal{U}$ by taking the percolation clusters to be the connected component containing $\mathbf{0}$ of the induced subtrees of T on all vertices v such that $U_v \leq p$. Let $\mathcal{F}_n$ be the

σ-field generated by the variables $\{U_v, \deg_v : |v| < n\}$. Because percolation is often imagined to take place on the edges rather than vertices, we let U_e be a synonym for U_v, where v is the farther of the two endpoints of e from the root. Write $v \leftrightarrow_{T,p} w$ if $U_e \leq p$ for all edges e on the geodesic from v to w in T. Informally, $v \leftrightarrow_{T,p} w$ iff v and w are both in T and are connected in the p-percolation. The event of successful p-percolation on a fixed tree T is denoted $H_T(p) := \{\mathbf{0} \leftrightarrow_{T,p} \infty\}$. The event of successful p-percolation on the random tree $\mathbf{T}$, is denoted $H_{\mathbf{T}}(p)$ or simply $H(p)$. Let $\theta_T(p) := \mathbf{P}[H_T(p)]$ denote the probability of p-percolation on the fixed tree T. Evaluating at $T = \mathbf{T}$ gives the random variable $\theta_{\mathbf{T}}(p)$ which is easily seen to equal the conditional expectation $\mathbf{P}(H(p) \,|\, \mathcal{T})$. Taking unconditional expectations we see that $\theta(p) = \mathbf{E}\theta_{\mathbf{T}}(p)$.

2.2.3 Smoothness of the annealed survival function θ

By Lyons' theorem, $\theta(p_c) = \mathbf{E}\theta_{\mathbf{T}}(p_c) = 0$. We now record some further properties of the annealed survival function θ.

Proposition 2.2.2. *The derivative from the right $K := \partial_+\theta(p_c)$ exists and is given by*

$$K = \frac{2}{p_c^3 \phi''(1)} \,. \tag{2.2.3}$$

where $1/\phi''(1)$ is interpreted as $\lim_{\xi \to 1^-} 1/\phi''(\xi)$.

Proof. Let $\phi_p(z) := \phi(1 - p + pz)$ be the offspring generating function for the Galton-Watson tree thinned by p-percolation for $p \in (p_c, 1)$. The fixed point of ϕ_p

is $1-\theta(p)$. In other words, $\theta(p)$ is the unique $s \in (0,1)$ for which $1-\phi_p(1-s) = s$, i.e. $1-\phi(1-ps) = s$. By Taylor's theorem with Mean-Value remainder, there exists a $\xi \in (1-p\theta(p), 1)$ so that

$$1-\phi(1-p\theta(p)) = p\theta(p)\phi'(1) - \frac{p^2\theta(p)^2}{2}\phi''(\xi) = \frac{p}{p_c}\theta(p) - \frac{p^2\theta(p)^2}{2}\phi''(\xi)\,.$$

Setting this equal to $\theta(p)$ and solving yields

$$\frac{\theta(p)}{p-p_c} = \frac{2}{p_c p^2 \phi''(\xi)}\,.$$

Taking $p \downarrow p_c$ and noting $\xi \to 1$ completes the proof. $\square$

Corollary 2.2.3. *(i) The function θ is analytic on $(p_c,1)$. (ii) If p_n decays exponentially then θ is analytic on $[p_c,1)$, meaning that for some $\varepsilon > 0$ there is an analytic function $\widetilde{\theta}$ on $(p_c-\varepsilon,1)$ such that $\theta(p) = \widetilde{\theta}(p)\mathbf{1}_{p>p_c}$.*

Proof. Recall that for $p \in (p_c,1)$, $\theta(p)$ is the unique positive s that satisfies $s = 1-\phi(1-ps)$. It follows that for all $p \in (p_c,1)$, $\theta(p)$ is the unique s satisfying

$$F(p,s) := s + \phi(1-ps) - 1 = 0\,.$$

Also note that since $\phi(1-ps)$ is analytic with respect to both variables for $(p,s) \in (p_c,1)\times(0,1)$, this means F is as well.

We aim to use the implicit function theorem to show that we can parameterize s as an analytic function of p on $(p_c,1)$; we thus must show $\frac{\partial F}{\partial s} \neq 0$ at all points $(p,\theta(p))$ for $p \in (p_c,1)$. Direct calculation gives

$$\frac{\partial F}{\partial s} = 1 - p\phi'(1-ps)\,.$$

Because ϕ is strictly convex on $(p_c, 1)$, we see that $\frac{\partial F}{\partial s}$ is positive for $p \in (p_c, 1)$ at the fixed point. Therefore, $\theta(p)$ is analytic on $(p_c, 1)$.

To prove (ii), observe that ϕ extends analytically to a segment $[0, 1+\varepsilon]$, which implies that $1-\phi(1-ps)$ is analytic on a real neighborhood of zero. Also $1-\phi(1-ps)$ vanishes at $s = 0$, therefore $\psi(p, s) := (1 - \phi(1 - ps))/s$ is analytic near zero and for $(p, s) \in (p_c, 1) \times (0, 1)$, the least positive value of s satisfying $\psi(p, s) = 1$ yields $\theta(p)$. Observe that

$$\frac{\partial \psi}{\partial p}(p_c, 0) = \lim_{s \to 0} \frac{s\phi'(1 - p_c s)}{s} = \phi'(1) = \mu \, .$$

By implicit differentiation,

$$\partial_+ \theta(p_c) = -\frac{\partial \psi / \partial s}{\partial \psi / \partial p}(p_c, 0)$$

which is equal to $1/K$ by Proposition 2.2.2. In particular, $(\partial \psi / \partial s)(p_c, 0) = -\mu/K$ is nonvanishing. Therefore, by the analytic implicit function theorem, solving $\psi(p, s) = 1$ for s defines an analytic function $\widetilde{\theta}$ taking a neighborhood of p_c to a neighborhood of zero, with $\widetilde{\theta}(p) > 0$ if and only if $p > p_c$. We have seen that $\widetilde{\theta}$ agrees with θ to the right of p_c, proving (ii). $\square$

In contrast to the above scenario in which Z has exponential moments and θ is analytic at p_c^+, the function θ fails to be smooth at p_c^+ when Z does not have all moments. The next two results quantify this: no kth moment implies $\theta \notin C^k$ from the right at p_c, and conversely, $\mathbf{E} Z^k < \infty$ implies $\theta \in C^j$ from the right at p_c for all $j < k/2$.

Proposition 2.2.4. *Assume $k \geq 2$, $\mathbf{E}[Z^k] < \infty$, and $\mathbf{E}[Z^{k+1}] = \infty$. Then $\theta^{(k+1)}(p)$ does not extend continuously to p_c from the right.*

In Section 2.2.4 we will prove the following partial converse.

Proposition 2.2.5. *For each $k \geq 1$, if $\mathbf{E}[Z^{2k+1}] < \infty$, then $\theta \in C^k$ from the right at p_c.*

2.2.4 Expansion of the annealed survival function θ at p_c^+

A good part of the quenched analysis requires only the expansion of the annealed survival function θ at p_c^+, not continuous derivatives. Proposition 2.2.6 below shows that $k+1$ moments are enough to give the order k expansion. Moreover, we give explicit expressions for the coefficients. We require the following combinatorial construction: let $\mathcal{C}_j(k)$ denote the set of compositions of k into j parts, i.e. ordered j-tuples of positive integers $(a_1, \ldots, a_j)$ with $a_1 + \cdots + a_j = k$; for a composition $a = (a_1, \ldots, a_j)$, define $\ell(a) = j$ to be the *length* of a, and $|a| = a_1 + \cdots + a_j$ to be the *weight* of a. Let $\mathcal{C}(\leq k)$ denote the set of compositions with weight at most k.

Proposition 2.2.6. *Suppose $\mathbf{E}[Z^{k+1}] < \infty$. Then there exist constants $r_1, \ldots, r_k$ such that $\theta(p_c+\varepsilon) = r_1\varepsilon + \cdots + r_k\varepsilon^k + o(\varepsilon^k)$. Moreover, the r_j's are defined recursively* *via*

$$r_1 = g'(p_c) = \frac{2}{p_c^3 \phi''(1)};$$

$$r_j = \frac{2}{p_c^2 \phi''(1)} \sum_{\substack{a \in \mathcal{C}(\leq j) \\ a \neq (j)}} r_{a_1} \cdots r_{a_{\ell(a)}} \binom{\ell(a)+1}{j-|a|} p_c^{|a|+\ell(a)+1-j} (-1)^{\ell(a)} \frac{\phi^{(\ell(a))+1}(1)}{(\ell(a)+1)!}. \quad (2.2.4)$$

Proof. To start, we utilize the identity $1 - \phi(1 - p\theta(p)) = \theta(p)$ for $p = p_c + \varepsilon$, and take a Taylor expansion:

$$\sum_{j=1}^{k+1} (p_c + \varepsilon)^j \theta(p_c + \varepsilon)^j (-1)^{j-1} \frac{\phi^{(j)}(1)}{j!} + o(((p_c + \varepsilon)\theta(p_c + \varepsilon))^{k+1}) = \theta(p_c + \varepsilon).$$

Divide both sides by $\theta(p_c + \varepsilon)$ and bound $\theta(p_c + \varepsilon) = O(\varepsilon)$ to get

$$\sum_{j=1}^{k+1} (p_c + \varepsilon)^j \theta(p_c + \varepsilon)^{j-1} (-1)^{j-1} \frac{\phi^{(j)}(1)}{j!} - 1 = o(\varepsilon^k). \quad (2.2.5)$$

Proceeding by induction, if we assume that the proposition holds for all $j < k$ for some $k \geq 2$, and we set

$$p_k(\varepsilon) := \frac{\theta(p_c + \varepsilon) - \sum_{j=1}^{k-1} r_j \varepsilon^j}{\varepsilon^k},$$

then (2.2.5) gives us

$$\begin{aligned} o(\varepsilon^k) &= \sum_{j=1}^{k+1} (p_c + \varepsilon)^j \theta(p_c + \varepsilon)^{j-1} (-1)^{j-1} \frac{\phi^{(j)}(1)}{j!} - 1 \\ &= \sum_{j=1}^{k+1} (p_c + \varepsilon)^j \left(\sum_{i=1}^{k-1} r_i \varepsilon^i + p_k(\varepsilon)\varepsilon^k \right)^{j-1} (-1)^{j-1} \frac{\phi^{(j)}(1)}{j!} - 1. \end{aligned} \quad (2.2.6)$$

Noting that the assumption that the proposition holds for $j = k - 1$ implies that $p_k(\varepsilon) = o(\varepsilon^{-1})$, we find that the expression on the right hand side in (2.2.6) is the sum of a polynomial in ε, the value $\frac{-p_c^2\phi''(1)}{2}p_k(\varepsilon)\varepsilon^k$, and an error term which is $o(\varepsilon^k)$. This implies that all terms of this polynomial that are of degree less than k must cancel, and that the sum of the term of order k and $\frac{-p_c^2\phi''(1)}{2}p_k(\varepsilon)\varepsilon^k$ must be $o(\varepsilon^k)$. This leaves only terms of degree greater than k. It follows that $p_k(\varepsilon)$ must be equal to $C + o(1)$, for some constant C.

To complete the induction step, it remains to show that $C = r_k$. To do so we must find the coefficient of ε^k in each term. We use the notation $[\varepsilon^j]$ to denote the coefficient of ε^j. For any j, we calculate

$$\begin{aligned}
[\varepsilon^k]\ (p_c+\varepsilon)^j \left(\sum_{i=1}^{k-1} r_i\varepsilon^i\right)^{j-1} &= \sum_{r=1}^{k}\left([\varepsilon^r]\ \left(\sum_{i=1}^{k-1} r_i\varepsilon^i\right)^{j-1}\right)\left([\varepsilon^{k-r}]\ (p_c+\varepsilon)^j\right) \\
&= \sum_{r=1}^{k}\left(\sum_{a\in\mathcal{C}_{j-1}(r)} r_{a_1}\cdots r_{a_{j-1}}\right)\binom{j}{k-r}p_c^{j+r-k}.
\end{aligned}$$

Putting this together with (2.2.6) we now obtain the desired equality $C = r_k$. Finally, noting that the base case $k = 1$ follows from Proposition 2.2.2, we see that the proposition now follows by induction. □

From here, Proposition 2.2.5 follows from Proposition 2.2.6 and Lemma 2.2.1 along with careful bookkeeping. A complete proof is contained in [MPR18].

2.3 Proof of part (ii) of Theorem 2.1.2: behavior at criticality

This section is concerned with the expansion of $\theta_{\mathbf{T}}(\cdot)$ at criticality. Section 2.3.1 defines the quantities that yield the expansion. Section 2.3.2 constructs some martingales and asymptotically identifies the expected number of k-subsets of T_n that survive critical percolation as a polynomial of degree $k-1$ whose leading term is a constant multiple of W (a consequence of Theorem 2.3.6, below). Section 2.3.3 finishes computing the ℓ-term Taylor expansion for $\theta_{\mathbf{T}}(\cdot)$ at criticality.

2.3.1 Explicit expansion

Throughout the paper we use $\{r_j\}$ to denote the coefficients of the expansion of θ when they exist, given by the explicit formula (2.2.4). For $m \geq 1$, the mth power of θ has a k-order expansion at p_c^+ whenever θ does. Generalizing the notation for r_j, we denote the coefficients of the expansion of θ^m at p_c^+ by $\{r_{m,j}\}$ where

$$\theta(p_c+\varepsilon)^m = \sum_{j=1}^{\ell} r_{m,j}\varepsilon^j + o(\varepsilon^\ell) \tag{2.3.1}$$

for any ℓ for which such an expansion exists.

We prove part (ii) of Theorem 2.1.2 by identifying the expansion. To do so, we need a notation for certain expectations. Fix a tree T. For $n \geq 0$, $j \geq 1$ and $v \in T$, define

$$X_n^{(j)}(v) := \sum_{\{v_1,\ldots,v_j\}\in\binom{T_n(v)}{j}} \mathbf{P}_T[v \leftrightarrow_{p_c} v_1, v_2, \ldots, v_j]$$

where $v \leftrightarrow_{p_c} v_1, v_2, \ldots, v_j$ is the event that v is connected to each of $v_1, \ldots, v_j$ under critical percolation. We omit the argument v when it is the root; thus $X_n^{(j)} := X_n^{(j)}(\mathbf{0})$. Note that

$$X_n^{(1)} = W_n, \qquad \text{and} \qquad X_n^{(2)} = \sum_{\{u,v\}\in\binom{T_n}{2}} p_c^{2n-|u\wedge v|} .$$

The former is the familiar martingale associated to a branching process, while the latter is related to the energy of the uniform measure on T_n.

Extend this definition further: for integers j and k, define

$$X_n^{(j,k)} := \sum_{\{v_i\}\in\binom{T_n}{j}} \binom{|T(v_1,\ldots,v_j)|}{k} p_c^{|T(v_1,\ldots,v_j)|}$$

where $T(v_1, \ldots, v_j)$ is the smallest rooted subtree of T containing each v_i and $|T(v_1, \ldots, v_j)|$ refers to the number of edges this subtree contains. Note that $X_n^{(j,0)} = X_n^{(j)}$.

Part (ii) of Theorem 2.1.2 follows immediately from the following expansion, which is the main work of this section.

Theorem 2.3.1. *Define*

$$M_n^{(i)} := M_n^{(i)}(T) := \mu^i \sum_{j=1}^{i} (-1)^{j+1} \sum_{d=j}^{i} p_c^d r_{j,d} X_n^{(j,i-d)} . \tag{2.3.2}$$

Suppose that $\mathbf{E}\left[Z^{(2\ell+1)(1+\beta)}\right] < \infty$ *for some integer* $\ell \geq 1$ *and real* $\beta > 0$. *(i) The quantities* $\{M_n^{(i)} : n \geq 1\}$ *are a* $\{\mathcal{T}_n\}$*-martingale with mean* r_i. *(ii) For* `GW`-

almost every tree T the limits $M^{(i)} := \lim_{n\to\infty} M_n^{(i)}$ exist. (iii) These limits are the coefficients in the expansion

$$\theta_T(p_c + \varepsilon) = \sum_{i=1}^{\ell} M^{(i)} \varepsilon^i + o(\varepsilon^\ell) \,. \tag{2.3.3}$$

Remark 2.3.2. The quantities $X_n^{(j,i)}$ do not themselves have limits as $n \to \infty$. In fact for fixed i and j the sum over d of $X_n^{(j,i-d)}$ is of order n^{i-1}. Therefore it is important to take the alternating outer sum before taking the limit.

2.3.2 Critical Survival of k-Sets

To prove Theorem 2.3.1 we need to work with centered variables. Centering at the unconditional expectation is not good enough because these mean zero differences are close to the nondegenerate random variable $n^{i-1}W$ and therefore not summable. Instead we subtract off a quantity that can be handled combinatorially, leaving a convergent martingale.

Throughout the rest of the paper, the notation Δ in front of a random variable with a subscript (and possibly superscripts as well) denotes the backward difference in the subscripted variable. Thus, for example,

$$\Delta X_n^{(j,i)} := X_n^{(j,i)} - X_{n-1}^{(j,i)} \,.$$

Let $X_n^{(j,i)} = Y_n^{(j,i)} + A_n^{(j,i)}$ denote the Doob decomposition of the process $\{X_n^{(j,i)} : n = 1, 2, 3, \ldots\}$ on the filtration $\{\mathcal{T}_n\}$. To recall what this means, ignoring superscripts for a moment, the Y and A processes are uniquely determined by requiring the Y process to be a martingale and the A process to be predictable, meaning that $A_n \in \mathcal{T}_{n-1}$ and $A_0 = 0$. The decomposition can be constructed inductively in n by letting $A_0 = 0, Y_0 = \mathbf{E}X_0$ and defining

$$\begin{aligned} \Delta A_n &:= \mathbf{E}\left(\Delta X_n \,|\, \mathcal{T}_{n-1}\right) ; \\ \Delta Y_n &:= \Delta X_n - \Delta A_n . \end{aligned}$$

We begin by identifying the predictable part.

Lemma 2.3.3. *Let $\mathcal{C}_i(j)$ denote the set of compositions of j of length i into strictly positive parts. Let $m_r := \mathbf{E}\binom{Z}{r}$ and define constants $c_{j,i}$ by*

$$c_{j,i} := p_c^j \sum_{\alpha \in \mathcal{C}_i(j)} m_{\alpha_1} m_{\alpha_2} \cdots m_{\alpha_i} .$$

Then for each $k \geq 0$,

$$\begin{aligned} \Delta A_{n+1}^{(j,k)} &= -X_n^{(j,k)} + \sum_{i=1}^{j} c_{j,i} \sum_{d=0}^{k} \binom{j}{k-d} X_n^{(i,d)} \\ &= \sum_{d=0}^{k-1} \binom{j}{k-d} X_n^{(j,d)} + \sum_{i=1}^{j-1} \sum_{d=0}^{k} c_{j,i} \binom{j}{k-d} X_n^{(i,d)} . \end{aligned} \tag{2.3.4}$$

Proof. For distinct vertices $v_1, \ldots, v_j$ in $\mathbf{T}_{n+1}$, their set of parents $u_1, \ldots, u_\ell$ form a subset of $\mathbf{T}_n$ with at most j elements. In order to sum over all j-sets of $\mathbf{T}_{n+1}$, one first sums over all sets of parents. For a fixed parent set $u_1, \ldots, u_\ell$ in $\mathbf{T}_{n-1}$, the total number of j-sets with parent set $\{u_1, \ldots, u_\ell\}$ is

$$\sum_{\alpha \in \mathcal{C}_\ell(j)} \binom{Z_1(u_1)}{\alpha_1} \cdots \binom{Z_1(u_\ell)}{\alpha_\ell} .$$

Furthermore, we have

$$\binom{|\mathbf{T}(v_1,\ldots,v_j)|}{k} = \binom{|\mathbf{T}(u_1,\ldots,u_\ell)|+j}{k} = \sum_{d=0}^{k}\binom{j}{k-d}\binom{|\mathbf{T}(u_1,\ldots,u_\ell)|}{d}.$$

This gives the expansion

$$\begin{aligned}
X_{n+1}^{(j,k)} &= \sum_{\{v_i\}\in\binom{\mathbf{T}_{n+1}}{j}} \binom{|\mathbf{T}(v_1,\ldots,v_j)|}{k} p_c^{|\mathbf{T}(v_1,\ldots,v_j)|} \\
&= \sum_{\ell=1}^{j} \sum_{\{u_i\}\in\binom{\mathbf{T}_n}{\ell}} \sum_{d=0}^{k} \binom{j}{k-d}\binom{\mathbf{T}(u_1,\ldots,u_\ell)}{d} p_c^{|\mathbf{T}(u_1,\ldots,u_\ell)|} \\
&\qquad \times \sum_{\alpha\in\mathcal{C}_\ell(j)} p_c^j \binom{Z_1(u_1)}{\alpha_1}\cdots\binom{Z_1(u_\ell)}{\alpha_\ell}.
\end{aligned}$$

Taking conditional expectations with respect to $\mathcal{T}_n$ completes the proof of the first identity, with the second following from rearrangement of terms. $\square$

The following corollary is immediate from Lemma 2.3.3 and the fact that $X_0^{(j,k)} = Y_0^{(j,k)}$.

Corollary 2.3.4. *For each j so that $\mathbf{E}[Z^j] < \infty$ and each k, the terms of the Y martingale are given by*

$$\begin{aligned}
Y_n^{(j,k)} &= Y_0^{(j,k)} + \sum_{m=1}^{n} \Delta Y_m^{(j,k)} \\
&= X_n^{(j,k)} - \sum_{m=0}^{n-1}\left[\sum_{d=0}^{k-1}\binom{j}{k-d}X_m^{(j,d)} + \sum_{i=1}^{j-1} c_{j,i}\sum_{d=0}^{k}\binom{j}{k-d}X_m^{(i,d)}\right].
\end{aligned}$$

$\square$

We want to show that these martingales converge both almost surely and in some appropriate L^p space; this will require us to take $L^{1+\beta}$ norms for some $\beta \in (0,1]$.

The following randomized version of the Marcinkiewicz-Zygmund inequality will be useful.

Lemma 2.3.5. *Let* $\{\xi_k\}_{k=1}^{\infty}$ *be i.i.d. with* $\mathbf{E}[\xi_1] = 0$ *and* $\mathbf{E}[|\xi_1|^{1+\beta}] < \infty$ *for some* $\beta \in (0, 1]$, *and let* N *be a random variable in* $\mathbb{N}$ *independent from all* $\{\xi_k\}$ *and with* $\mathbf{E}[N] < \infty$. *If we set* $S_n = \sum_{k=1}^{n} \xi_k$, *then there exists a constant* $c > 0$ *depending only on* β *so that*

$$\mathbf{E}[|S_N|^{1+\beta}] \le c\mathbf{E}[|\xi_1|^{1+\beta}]\mathbf{E}[N] \, .$$

In particular, if $\xi(v)$ *are associated to vertices* $v \in \mathbf{T}_s$, *and are mutually independent from* $\mathbf{T}_s$, *then*

$$\left\| p_c^s \sum_{v \in \mathbf{T}_s} \xi(v) \right\|_{L^{1+\beta}} \le c' p_c^{s\beta/(1+\beta)} \|\xi(v)\|_{L^{1+\beta}} \, .$$

Proof. Suppose first that N is identically equal to a constant n. The Marcinkiewicz-Zygmund inequality (e.g. [CT97, Theorem 10.3.2]) implies that there exists a constant $c > 0$ depending only on β such that

$$\mathbf{E}[|S_n|^{1+\beta}] \le c\mathbf{E}\left[\left(\sum_{k=1}^{n} |\xi_k|^2 \right)^{(1+\beta)/2} \right] .$$

Because $1+\beta \le 2$ and the ℓ^p norms descend, we have $\|(\xi_k)_{k=1}^n\|_{\ell^2} \le \|(\xi_k)_{k=1}^n\|_{\ell^{1+\beta}}$ deterministically; this completes the proof when N is constant. Writing

$$\mathbf{E}[|S_N|^{1+\beta}] = \mathbf{E}\left[\mathbf{E}[|S_N|^{1+\beta} \mid N]\right]$$

and applying the bound from the constant case completes the proof. ☐

We now show that the martingales $\{Y_n^{(j,k)} : n \ge 0\}$ converge.

Theorem 2.3.6. *Suppose* $\mathbf{E}[Z^{j(1+\beta)}] < \infty$ *for some* $\beta > 0$. *Then*

(a) $\|\Delta Y_{n+1}^{(j,k)}\|_{L^{1+\beta}} \leq Ce^{-cn}$ *where* C *and* c *are positive constants depending on* j, k, β *and the offspring distribution.*

(b) $Y_n^{(j,k)}$ *converges almost surely and in* $L^{1+\beta}$ *to a limit, which we denote* $Y^{(j,k)}$.

(c) *There exists a positive constant* $c'_{j,k}$ *depending only on* j, k *and the offspring distribution so that*

$X_n^{(j,k)} n^{-(j+k-1)} \to c'_{j,k} W$ *almost surely and in* $L^{1+\beta}$.

Proof.

Step 1: $(a) \implies (b)$. For any fixed j and k: the triangle inequality and (a) show that $\sup_n \|Y_n^{(j,k)}\|_{L^{1+\beta}} < \infty$, from which (b) follows from the L^p martingale convergence theorem. Next, we prove an identity representing $X_n^{(j,k)}$ as a multiple sum over values of $X^{(j',k')}$ with $(j',k') < (j,k)$ lexicographically.

Step 2: Some computation. For a set of vertices $\{v_1, \ldots, v_j\}$, let $v = v_1 \wedge v_2 \wedge \cdots \wedge v_j$ denote their most recent common ancestor. In order for $\mathbf{0} \leftrightarrow_{p_c} v_1, \ldots, v_j$ to hold, we must first have $\mathbf{0} \leftrightarrow_{p_c} v$. For the case of $j \geq 2$, looking at the smallest tree containing v and $\{v_i\}$, we must have that this tree branches into some number of children $a \in [2, j]$ immediately after v. We may thus sum over all possible v, first by height, setting $s = |v|$, then choosing how many children of v will be the ancestors of the $v_1, \ldots, v_j$. We then choose those children $\{u_r\}$, and choose how to distribute the $\{v_\ell\}$ among them. In order for critical percolation to reach each

$v_1, \ldots, v_j$, it must first reach v, then survive to each child of v that is an ancestor of some $\{v_\ell\}$ and then survive to the $\{v_\ell\}$ from there. Finally, in order to choose the k-element subset corresponding to $\binom{|\mathbf{T}(v_1,\ldots,v_j)|}{k}$, we may choose α_0 elements from the tree $T(u_1, \ldots, u_a)$ and α_ℓ elements from each subtree of u_ℓ. Putting this all together, we have the decomposition

$$\begin{aligned} X_n^{(j,k)} &= \sum_{s=0}^{n-1} p_c^s \sum_{v \in \mathbf{T}_s} \sum_{a=2}^{j} \sum_{u \in \binom{\mathbf{T}_1(v)}{a}} p_c^a \sum_{\beta \in \mathcal{C}_a(j)} \sum_{\alpha_0=0}^{k} \\ &\quad \times \sum_{\alpha \in \widetilde{\mathcal{C}}_a(k-\alpha_0)} \binom{s+a}{\alpha_0} X_{n-s-1}^{(\beta_1,\alpha_1)}(u_1) \cdots X_{n-s-1}^{(\beta_a,\alpha_a)}(u_a) \qquad (2.3.5) \\ &= \sum_{s=0}^{n-1} p_c^s \sum_{v \in \mathbf{T}_s} \Theta_{n-s-1}^{(j,k)}(v) \end{aligned}$$

where $\Theta_{n-s-1}^{(j,k)}(v)$ is defined as the inner quintuple sum in the previous line and $\widetilde{\mathcal{C}}_a(k)$ denotes the set of weak a-compositions of k; observe that the notation $\Theta_{n-s-1}^{(j,k)}(v)$ hides the dependence on $s = |v|$.

The difference $\Delta Y_n^{(j,k)}$ can now be computed as follows:

$$\begin{aligned} \Delta Y_n^{(j,k)} &= X_n^{(j,k)} - \sum_{i=1}^{j} \sum_{d=0}^{k} \binom{j}{k-d} c_{j,i} X_{n-1}^{(i,d)} \qquad (2.3.6) \\ &= \sum_{s=0}^{n-1} p_c^s \sum_{v \in \mathbf{T}_s} \Theta_{n-s-1}^{(j,k)}(v) - \sum_{s=0}^{n-2} p_c^s \sum_{v \in \mathbf{T}_s} \sum_{i=1}^{j} \sum_{d=0}^{k} \binom{j}{k-d} c_{j,i} \Theta_{n-s-2}^{(i,d)}(v) \\ &= \sum_{s=0}^{n-2} p_c^s \sum_{v \in \mathbf{T}_s} \left(\Theta_{n-s-1}^{(j,k)}(v) - \sum_{i=2}^{j} \sum_{d=0}^{k} \binom{j}{k-d} c_{j,i} \Theta_{n-s-2}^{(i,d)}(v) \right) \\ &\qquad + \left[p_c^{n-1} \sum_{v \in \mathbf{T}_{n-1}} \Theta_0^{(j,k)}(v) - c_{j,1} \sum_{d=0}^{k} \binom{j}{k-d} X_{n-1}^{(1,d)} \right] \\ &= \sum_{s=0}^{n-2} p_c^s \sum_{v \in \mathbf{T}_s} U_n^{(j,k)}(v) + V_n^{(j,k)} \end{aligned}$$

where

$$U_n^{(j,k)} = \left(\Theta_{n-s-1}^{(j,k)}(v) - \sum_{i=2}^{j}\sum_{d=0}^{k}\binom{j}{k-d}c_{j,i}\Theta_{n-s-2}^{(i,d)}(v)\right) ; \tag{2.3.7}$$

$$V_n^{(j,k)} = \left(p_c^{n-1}\sum_{v\in\mathbf{T}_{n-1}} p_c^j\binom{n+j-1}{k}\binom{Z_1(v)}{j}\right) - c_{j,1}\sum_{d=0}^{k}\binom{j}{k-d}X_{n-1}^{(1,d)} . \tag{2.3.8}$$

<u>Step 3: Proving (a) and (c) for $j=1$ and k arbitrary</u>. Specializing (2.3.6) to $j=1$ yields

$$\begin{aligned}\Delta Y_n^{(1,k)} &= \binom{n}{k}W_n - \binom{n-1}{k}W_{n-1} - \binom{n-1}{k-1}W_{n-1} \\ &= \binom{n-1}{k}(W_n - W_{n-1}) + \binom{n-1}{k-1}(W_n - W_{n-1}) .\end{aligned}$$

The quantity $W_n - W_{n-1}$ is the sum of independent contributions below each vertex in T_{n-1}; Lemma 2.3.5 shows this to be exponentially small in $L^{1+\beta}$ and proving (a), hence (b). Additionally, $Y_n^{(1,k)}n^{-k} \to W/k!$, thereby also showing (c) for $j=1$ and all k.

<u>Step 4: V is always small</u>. Using the identity $\binom{n+j-1}{k} = \sum_{d=0}^{k}\binom{n-1}{d}\binom{j}{k-d}$ and recalling that $X_{n-1}^{(1,d)} = \binom{n-1}{d}W_{n-1}$ shows that

$$V_n^{(j,k)} = \sum_{d=0}^{k}\binom{j}{k-d}\binom{n-1}{d}p_c^{n-1}\sum_{v\in\mathbf{T}_{n-1}} p_c^j\left[\binom{Z_1(v)}{j} - \mathbf{E}\binom{Z}{j}\right] .$$

Applying Lemma 2.3.5 shows that the innermost sum, when multiplied by p_c^{n-1}, has $L^{1+\beta}$ norm that is exponentially small in n. With k fixed and $d \le k$, the product with $\binom{n-1}{d}$ still yields an exponentially small variable, thus

$$||V_n^{(j,k)}||_{1+\beta} \le c_{j,k,\beta}e^{-\delta n} \tag{2.3.9}$$

for some $\delta = \delta(j, k, \beta) > 0$.

The remainder of the proof is an induction in two stages (Steps 5 and 6). In the first stage we fix $j > 1$, assume (a)–(c) for all (j', k') with $j' < j$, and prove (a) for (j, k) with k arbitrary. In the second stage, we prove (c) for (j, k) by induction on k, establishing (c) for $(j, 1)$ and then for arbitrary k by induction, assuming (a) for (j, k') where k' is arbitrary and (c) for (j, k') where $k' < k$.

Step 5: Prove (a) by induction on j. Fix $j \geq 2$ and assume for induction that (a) and (c) hold for all (j', k) with $j' < j$. The plan is this: The quantity $p_c^s \sum_{v \in \mathbf{T}_s} U_n^{(j,k)}(v)$ is W_n times the average of $U_n^{(j,k)}(v)$ over vertices $v \in \mathbf{T}_s$. Averaging many mean zero terms will produce something exponentially small in s. We will also show this quantity to be also exponentially small in $n - s$, whereby it follows that the outer sum over s is exponentially small, completing the proof.

Let us first see that $U_n^{(j,k)}(v)$ has mean zero. Expanding back the Θ terms gives

$$\begin{aligned} U_n^{(j,k)}(v) = \sum_{a=2}^{j} \sum_{u \in \binom{\mathbf{T}_1(v)}{a}} p_c^a \sum_{\alpha_0 = 0}^{k} \binom{s+a}{\alpha_0} & \\ \times \Bigg(\sum_{\beta \in \mathcal{C}_a(j)} \sum_{\alpha \in \widetilde{\mathcal{C}}_a(k - \alpha_0)} X_{n-s-1}^{(\beta_1, \alpha_1)}(u_1) \cdots X_{n-s-1}^{(\beta_a, \alpha_a)}(u_a) & \qquad (2.3.10) \\ - \sum_{i=2}^{j} \sum_{d=0}^{k} c_{j,i} \binom{j}{k-d} \sum_{\beta' \in \mathcal{C}_a(i)} \sum_{\alpha' \in \widetilde{\mathcal{C}}_a(d - \alpha_0)} X_{n-s-2}^{(\beta'_1, \alpha'_1)}(u_1) \cdots X_{n-s-2}^{(\beta'_a, \alpha'_a)}(u_a) \Bigg) . & \end{aligned}$$

Expanding the first product of X terms gives

$$\begin{aligned} & X_{n-s-1}^{(\beta_1, \alpha_1)}(u_1) \cdots X_{n-s-1}^{(\beta_a, \alpha_a)}(u_a) \\ & = \prod_{\ell=1}^{a} \left(\Delta Y_{n-s-1}^{(\beta_\ell, \alpha_\ell)}(u_\ell) + \sum_{\beta'_\ell = 1}^{\beta_\ell} c_{\beta_\ell, \beta'_\ell} \sum_{\alpha'_\ell = 0}^{\alpha_\ell} \binom{\beta_\ell}{\alpha_\ell - \alpha'_\ell} X_{n-s-2}^{(\beta'_\ell, \alpha'_\ell)}(u_\ell) \right) . \qquad (2.3.11) \end{aligned}$$

The vertices u_ℓ are all distinct children of v. Therefore, their subtrees are jointly independent, hence the pairs $(\Delta Y(u_\ell), X(u_\ell))$ are jointly independent. The product (2.3.11) expands to the sum of a-fold products of terms, each term in each product being either a ΔY or a weighted sum of X's, the a terms being jointly independent by the previous observation. Therefore, to see that the whole thing is mean zero, we need to check that the product of the a different sums of X terms in (2.3.11), summed over α and β to form the first half of the summand in (2.3.10), minus the subsequent sum over i, d, β' and α', has mean zero. In fact we will show that it vanishes entirely. For given compositions $\beta := (\beta_1, \ldots, \beta_a)$ and $\alpha := (\alpha_1, \ldots, \alpha_a)$, the product of the double sum of X terms inside the round brackets in (2.3.11) may be simplified:

$$\prod_{\ell=1}^{a} \left(\sum_{\beta'_\ell=1}^{\beta_\ell} c_{\beta_\ell,\beta'_\ell} \sum_{\alpha'_\ell=0}^{\alpha_\ell} \binom{\beta_\ell}{\alpha_\ell - \alpha'_\ell} X_{n-s-2}^{(\beta'_\ell,\alpha'_\ell)}(u_\ell) \right)$$
$$= \sum_{1 \preceq \beta' \preceq \beta} \sum_{0 \preceq \alpha' \preceq \alpha} \prod_{\ell=1}^{a} c_{\beta_\ell,\beta'_\ell} \binom{\beta_\ell}{\alpha_\ell - \alpha'_\ell} X_{n-s-2}^{(\beta'_\ell,\alpha'_\ell)}(u_\ell) \,.$$

Applying the identity

$$\sum_{\substack{\beta \in \mathcal{C}_a(j) \\ \beta \succeq \beta'}} \left(\prod_{\ell} c_{\beta_\ell,\beta'_\ell} \right) = c_{j,i} \,, \tag{2.3.12}$$

which follows by regrouping pieces of each composition in $\mathcal{C}_i(j)$ into smaller compositions each with β'_ℓ parts, then summing over α and β as in (2.3.10) and simplifying, using (2.3.12) in the last line, gives

$$\begin{aligned}
&\sum_{\beta\in\mathcal{C}_a(j)}\sum_{1\preceq\beta'\preceq\beta}\sum_{\alpha\in\widetilde{\mathcal{C}}_a(k-\alpha_0)}\sum_{0\preceq\alpha'\preceq\alpha}\prod_{\ell=1}^{a} c_{\beta_\ell,\beta'_\ell}\binom{\beta_\ell}{\alpha_\ell-\alpha'_\ell} X_{n-s-2}^{(\beta'_\ell,\alpha'_\ell)}(u_\ell)\\
&\qquad = \sum_{\beta\in\mathcal{C}_a(j)}\sum_{1\preceq\beta'\preceq\beta}\left(\prod_\ell c_{\beta_\ell,\beta'_\ell}\right)\sum_{d=0}^{k}\sum_{\alpha'\in\widetilde{\mathcal{C}}_a(d-\alpha_0)}\left(\prod_\ell X_{n-s-2}^{(\beta'_\ell,\alpha'_\ell)}(u_\ell)\right)\\
&\qquad\qquad \times \sum_{\substack{\alpha\in\widetilde{\mathcal{C}}_a(k-\alpha_0)\\ \alpha\geq\alpha'}}\prod_{\ell=1}^{a}\binom{\beta_\ell}{\alpha_\ell-\alpha'_\ell}\\
&\qquad = \sum_{\beta\in\mathcal{C}_a(j)}\sum_{1\preceq\beta'\preceq\beta}\left(\prod_\ell c_{\beta_\ell,\beta'_\ell}\right)\sum_{d=0}^{k}\sum_{\alpha'\in\widetilde{\mathcal{C}}_a(d-\alpha_0)}\left(\prod_\ell X_{n-s-2}^{(\beta'_\ell,\alpha'_\ell)}(u_\ell)\right)\binom{j}{k-d}\\
&\qquad = \sum_{i=2}^{j}\sum_{\beta'\in\mathcal{C}_a(i)}\sum_{\substack{\beta\in\mathcal{C}_a(j)\\ \beta\geq\beta'}}\left(\prod_\ell c_{\beta_\ell,\beta'_\ell}\right)\sum_{d=0}^{k}\sum_{\alpha'\in\widetilde{\mathcal{C}}_a(d-\alpha_0)}\left(\prod_\ell X_{n-s-2}^{(\beta'_\ell,\alpha'_\ell)}(u_\ell)\right)\binom{j}{k-d}\\
&\qquad = \sum_{i=2}^{j}\sum_{d=0}^{k} c_{j,i}\binom{j}{k-d}\sum_{\beta'\in\mathcal{C}_a(i)}\sum_{\alpha'\in\widetilde{\mathcal{C}}_a(d-\alpha_0)} X_{n-s-2}^{(\beta'_1,\alpha'_1)}(u_1)\cdots X_{n-s-2}^{(\beta'_a,\alpha'_a)}(u_a)\,.
\end{aligned}$$

This exactly cancels with the quadruple sum on the second line of (2.3.10), transforming (2.3.10) into

$$U_n^{(j,k)}(v) = \sum_{a=2}^{j}\sum_{u\in\binom{\mathbf{T}_1(v)}{a}} p_c^a\sum_{\alpha_0=0}^{k}\binom{s+a}{\alpha_0}\sum_{\beta\in\mathcal{C}_a(j)}\sum_{\alpha\in\widetilde{\mathcal{C}}_a(k-\alpha_0)}\prod_{\ell=1}^{a}(*)_\ell\,,$$

where $(*)_\ell = \Delta Y_{n-s-1}^{(\beta_\ell,\alpha_\ell)}(u_\ell)$ for at least one value of ℓ in $[1,a]$, and, when not equal to that, is equal to the last double sum inside the brackets in (2.3.11).

By the induction hypothesis, the ΔY terms have $(1+\beta)$ norm bounded above by something exponentially small:

$$\|\Delta Y_{n-s-1}^{(\beta_\ell,\alpha_\ell)}(u_\ell)\| = O\left(\exp\left[-\kappa_{\beta_\ell,\alpha_\ell}(n-s-1)\right]\right)\,. \tag{2.3.13}$$

We note that a and each β_ℓ and α_ℓ are all bounded above by j and that in each product $X^{(\beta_1,\alpha_1)}_{n-s-1}(u_1)\cdots X^{(\beta_a,\alpha_a)}_{n-s-1}(u_a)$, the terms are independent. The inductive hypothesis implies each factor $X_n^{(j,k)}$ has $L^{1+\beta}$ norm that is $O(n^{\lambda(j,k)})$.

Returning to (2.3.6), we may apply Lemma 2.3.5 to see that for each s, the quantity $p_c^s \sum_{v\in\mathbf{T}_s} U_n^{(j,k)}(v)$ is an average of $|\mathbf{T}_s|$ terms all having mean zero and $L^{1+\beta}$ bound exponentially small in $n-s$, and that averaging introduces another exponentially small factor, $\exp(-\nu s)$. Because the constants κ, λ and μ vary over a set of bounded cardinality, the product of these three upper bounds,

$$O\left(\exp(-\kappa(n-s))\cdot\exp(-\nu s)\cdot n^{\lambda(j,k)}\right)$$

decreases exponentially n.

<u>Step 6: Prove (c) by induction on (j,k)</u>. The final stage of the induction is to assume (a)–(c) for (j',k') lexicographically smaller than (j,k) and prove (c) for (j,k). We use the following easy fact.

Lemma 2.3.7. *If $a_n \to \infty$ and $a_n \sim b_n$ then the partials sums are also asymptotically equivalent: $\sum_{k=1}^n a_k \sim \sum_{k=1}^n b_k$.* □

We begin the inductive proof of with the case $k=0$. Rearranging the conclusion of Corollary 2.3.4, we see that

$$X_n^{(j,0)} = Y_n^{(j,0)} + \sum_{m=0}^{n-1}\sum_{i=1}^{j-1} X_m^{(i,0)}\,.$$

Using Lemma 2.3.7 the induction hypothesis, and the fact that $Y_n^{(j,0)} = O(1)$ simplifies this to

$$
\begin{aligned}
X_n^{(j,0)} &\sim \sum_{m=0}^{n-1} \left[\sum_{i=1}^{j-1} m^{i-1} c_i' W \right] \\
&\sim \sum_{m=0}^{n-1} c_{j-1}' m^{j-2} W \\
&\sim c_j' n^{j-1} W
\end{aligned}
$$

where $c_j' = \lim_{n\to\infty} \frac{c_{j-1}'}{n} \sum_{m=0}^{n-1} (m/n)^{j-2} = c_{j-1}'/(j-1)$.

The base case $k = 0$ being complete, we induct on k. The same reasoning, observing that the first inner sum is dominated by the $d = k-1$ term and the second by the $i = j-1$ and $d = k$ term, gives

$$
\begin{aligned}
X_n^{(j,k)} &= Y_n^{(j,k)} + \sum_{m=0}^{n-1} \left[\sum_{d=0}^{k-1} \binom{j}{k-d} X_m^{(j,d)} + \sum_{i=1}^{j-1} c_{j,i} \sum_{d=0}^{k} \binom{j}{k-d} X_m^{(i,d)} \right] \\
&\sim \sum_{m=0}^{n-1} \left(j X_m^{(j,k-1)} + c_{j,j-1} X_m^{(j-1,k)} \right) \\
&\sim \sum_{m=0}^{n-1} \left[j m^{j+k-2} W c_{j,k-1}' + c_{j,j-1} c_{j-1,k}' W m^{j+k-2} \right] \\
&\sim W \left(\frac{j c_{j,k-1}' + c_{j,j-1} c_{j-1,k}'}{j+k-1} \right) n^{j+k-1} .
\end{aligned}
$$

Setting $c_{j,k}' := \dfrac{j c_{j,k-1}' + c_{j,j-1} c_{j-1,k}'}{j+k-1}$ completes the almost-sure part of (c) by induction. The $L^{1+\beta}$ portion is similar, but we need one more easy fact.

Lemma 2.3.8. *If $a_n \to \infty$ and $b_n \to 0$ then $\sum_{k=1}^n a_n b_n = o\left(\sum_{k=1}^n a_k\right)$.* □

This allows us to calculate

$$
\begin{aligned}
&\left\| X_n^{(j,k)} n^{-(j+k-1)} - W c'_{j,k} \right\|_{L^{1+\beta}} \\
&\quad = \left\| n^{-(j+k-1)} \sum_{m=0}^{n-1} \left[\sum_{d=0}^{k-1} \binom{j}{k-d} X_m^{(j,d)} + \sum_{i=1}^{j-1} c_{j,i} \sum_{d=0}^{k} \binom{j}{k-d} X_m^{(i,d)} \right] \right. \\
&\qquad\qquad \left. - W c'_{j,k} - Y_n^{(j,k)} n^{-(j+k-1)} \right\|_{L^{1+\beta}} \\
&\quad \le o(1) + n^{-(j+k-1)} \left\| \sum_{m=0}^{n-1} \left(j X_m^{(j,k-1)} + c_{j,j-1} X_m^{(j-1,k)} \right) - n^{j+k-1} W c'_{j,k} \right\|_{L^{1+\beta}} \\
&\quad \le o(1) + n^{-(j+k-1)} \sum_{m=0}^{n-1} m^{j+k-2} \left(j \left\| \frac{X_m^{(j,k-1)}}{m^{j+k-2}} - W c'_{j,k-1} \right\|_{L^{1+\beta}} \right. \\
&\qquad\qquad \left. + c_{j,j-1} \left\| \frac{X_m^{(j-1,k)}}{m^{j+k-2}} - W c'_{j-1,k} \right\|_{L^{1+\beta}} \right) \\
&\quad = o(1)\,.
\end{aligned}
$$

This completes the induction, and the proof of Theorem 2.3.6. □

2.3.3 Expansion at Criticality

An easy inequality similar to classical Harris inequality [Har60] is as follows.

Lemma 2.3.9. *For finite sets of edges E_1, E_2, E_3, define A_j to be the event that all edges in E_j are open. Then*

$$\mathbf{P}[A_1 \cap A_2] \cdot \mathbf{P}[A_1 \cap A_3] \le \mathbf{P}[A_1] \cdot \mathbf{P}[A_1 \cap A_2 \cap A_3]\,.$$

Proof. Writing each term explicitly, this is equivalent to the inequality

$$p^{|E_1 \cup E_2| + |E_1 \cup E_3|} \le p^{|E_1| + |E_1 \cup E_2 \cup E_3|}\,.$$

Because $p \leq 1$, this is equivalent to

$$|E_1 \cup E_2| + |E_1 \cup E_3| \geq |E_1| + |E_1 \cup E_2 \cup E_3|,$$

which is easily proved for all triples E_1, E_2, E_3 by inclusion-exclusion. $\square$

Before finding the expansion at criticality, we show that focusing only on the first n levels of the tree and averaging over the remaining levels causes only a subpolynomial error in an appropriate sense.

Proposition 2.3.10. *Suppose* $\mathbf{E}[Z^{(2k-1)(1+\beta)}] < \infty$, *and set* $p = p_c + \varepsilon$. *Fix* $\delta > 0$ *and let* $n = n(\varepsilon) = \lceil \varepsilon^{-\delta} \rceil$. *Then for* δ *sufficiently small and each* $\ell > 0$,

$$\sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P}_\mathbf{T}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_k] \left(\theta_{\mathbf{T}(u_1)}(p) \cdots \theta_{\mathbf{T}(u_k)}(p) - \theta(p)^k\right) = o(\varepsilon^\ell) \qquad (2.3.14)$$

$\mathtt{GW}$-*almost surely as* $\varepsilon \to 0^+$.

Proof. For sufficiently small $\delta > 0$, we note that $(p_c+\varepsilon)^m \leq 2p_c^m$ for each $m \in [n, kn]$ and for ε sufficiently small. This will be of use throughout, and is responsible for the appearance of factors of 2 in the upper bounds.

Next, bound the variance of

$$\sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P}_\mathbf{T}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_k] \left[\theta_{\mathbf{T}(u_1)}(q) \cdots \theta_{\mathbf{T}(u_k)}(q) - \theta(q)^k\right]$$

for a fixed vertex, q. This expression has mean zero conditioned on $\mathcal{T}_n$. Its variance is equal to the expected value of its conditional variance given $\mathcal{T}_n$. We therefore square and take the expectation, where the second sum in the second and third lines are over pairs of disjoint k-tuples of points.

$$\mathbf{E}\left[\left(\sum_{\{u_i\}\in\binom{\mathbf{T}_n}{k}} \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_k]\left(\theta_{\mathbf{T}(u_1)}(q)\cdots\theta_{\mathbf{T}(u_k)}(q) - \theta(q)^k\right)\right)^2 \,\middle|\, \mathcal{T}_n\right]$$

$$= \frac{1}{(k!)^2}\sum_{r=1}^{k} r! \sum_{\{u_i\}_{i=1}^k, \{v_i\}_{i=r+1}^k \text{ dist.}} \binom{k}{r}^2$$

$$\times \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_k]\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_r, v_{r+1}, \ldots, v_k] C_r$$

$$\leq \frac{1}{(k!)^2}\sum_{r=1}^{k} r! \sum_{\{u_i\}_{i=1}^k, \{v_i\}_{i=r+1}^k \text{ dist.}} \binom{k}{r}^2$$

$$\times \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_r]\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_k, v_{r+1}, \ldots, v_k] C_r$$

$$\leq 4p_c^n \sum_{r=1}^{k} \binom{2k-r}{k}\binom{k}{r} C_r X_n^{(2k-r)}.$$

Here we have used the bounds $\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_r] \leq 2p_c^n$ and

$$\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p u_1, \ldots, v_k] \leq 2\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_{p_c} u_1, \ldots, v_k]$$

and we have defined

$$C_r := \mathbf{E}\Big[\left(\theta_{\mathbf{T}(u_1)}(q)\cdots\theta_{\mathbf{T}(u_k)}(q) - \theta(q)^k\right)$$
$$\times \left(\theta_{\mathbf{T}(u_1)}(q)\cdots\theta_{\mathbf{T}(u_r)}(q)\theta_{\mathbf{T}(v_{r+1})}(q)\cdots\theta_{\mathbf{T}(v_k)}(q) - \theta(q)^k\right)\Big].$$

Taking the expected value and using Theorem 2.3.6 along with Jensen's Inequality and induction gives that the variance is bounded above by $Cp_c^n n^{2k-2}$ for some constant C. This is exponentially small in n, so there exist constants $c_k, C_k > 0$ so that the variance is bounded above by $C_k e^{-c_k n}$.

Define $a = a(m,r) = \frac{1}{m} + \frac{r}{m^{\ell+2}}$ and $b = b(m,r) = \frac{1}{m} + \frac{r+1}{m^{\ell+2}}$. For each $\varepsilon \in (0,1)$ there exists a unique pair (m,r) such that $\varepsilon \in [1/m, 1/(m-1))$ and $\varepsilon \in [a,b)$.

Assume for now that $\lceil a^{-\delta} \rceil = \lceil b^{-\delta} \rceil$; the case in which the two differ is handled at the end of the proof. For all $\varepsilon \in [a, b)$ and $p = p_c + \varepsilon$, we have

$$\sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_p u_1, \ldots, u_k\right] \theta_{\mathbf{T}(u_1)}(p) \cdots \theta_{\mathbf{T}(u_k)}(p)$$

$$\leq \sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+b} u_1, \ldots, u_k\right] \theta_{\mathbf{T}(u_1)}(p_c + b) \cdots \theta_{\mathbf{T}(u_k)}(p_c + b)\,.$$

By Chebyshev's inequality, the conditional probability that the right-hand side is $b^{\ell+1}$ greater than its mean, given $\mathcal{T}_n$, is at most $C_k \cdot b^{-(2\ell+2)} e^{-c_k n}$. Because $n = \lceil b^{-\delta} \rceil$, this is finite when summed over all possible m and r, implying that all but finitely often

$$\sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+b} u_1, \ldots, u_k\right] \theta_{\mathbf{T}(u_1)}(p_c + b) \cdots \theta_{\mathbf{T}(u_k)}(p_c + b)$$

$$\leq \theta(p_c + b)^k \sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+b} u_1, \ldots, u_k\right] + b^{\ell+1}\,.$$

By a similar argument, we obtain the lower bound

$$\sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+b} u_1, \ldots, u_k\right] \theta_{\mathbf{T}(u_1)}(p_c + b) \cdots \theta_{\mathbf{T}(u_k)}(p_c + b)$$

$$\geq \theta(p_c + a)^k \sum_{\{u_i\} \in \binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+a} u_1, \ldots, u_k\right] - b^{\ell+1}\,.$$

Letting $(*)$ denote the absolute value of the left-hand-side of (2.3.14), we see that

$$
\begin{aligned}
(*) \le \theta(p_c+b)^k & \sum_{\{u_i\}\in\binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+b} u_1,\dots,u_k\right] \\
& - \theta(p_c+a)^k \sum_{\{u_i\}\in\binom{\mathbf{T}_n}{k}} \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+a} u_1,\dots,u_k\right] + 2b^{\ell+1} \\
\le 2(\theta(p_c+b)^k & - \theta(p_c+a)^k) X_n^{(k)} \\
& + \theta(p_c+b)^k \left(\mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+b} u_1,\dots,u_k\right] - \mathbf{P_T}\left[\mathbf{0} \leftrightarrow_{p_c+a} u_1,\dots,u_k\right]\right) + 2b^{\ell+1} \\
\le 2(\theta(p_c+b)^k & - \theta(p_c+a)^k) X_n^{(k)} + \theta(p_c+b)^k \frac{2\cdot n\cdot k(b-a)}{p_c} X_n^{(k)} + 2b^{\ell+1},
\end{aligned}
$$

where the last inequality is via the Mean Value Theorem.

Dividing by ε^ℓ and setting $C_k = 2k/p_c$, we have

$$
\begin{aligned}
2\frac{\theta(p_c+b)^k - \theta(p_c+a)^k}{\varepsilon^\ell} X_n^{(k)} + C_k \cdot n \cdot \theta(p_c+b)^k \frac{b-a}{\varepsilon^\ell} X_n^{(k)} + 2b(b/\varepsilon)^\ell \\
\le 2\frac{b-a}{\varepsilon^\ell} \cdot \frac{\theta(p_c+b)^k - \theta(p_c+a)^k}{b-a} X_n^{(k)} \\
+ C_k \cdot n \cdot \theta(p_c+b)^k \frac{b-a}{\varepsilon^\ell} X_n^{(k)} + 2b(b/a)^\ell \\
\le 2k\frac{b-a}{\varepsilon^\ell} \max_{x\in[p_c,1]} \theta'(x) X_n^{(k)} \\
+ C_k \cdot n \cdot \theta(p_c+b)^k \frac{b-a}{\varepsilon^\ell} X_n^{(k)} + 2b\left(\frac{b}{a}\right)^\ell
\end{aligned}
$$

again by the Mean Value Theorem.

By Theorem 2.3.6(c), $n^{-(k-1)} X_n^{(k)}$ converges as $n \to \infty$. By definition of b, a and n, $\frac{(b-a)n^k}{\varepsilon^\ell} \to 0$ as $\varepsilon \to 0$ for δ sufficiently small, thereby completing the proof except in the case when $\lceil a^{-\delta} \rceil \neq \lceil b^{-\delta} \rceil$.

When $\lceil a^{-\delta} \rceil$ and $\lceil b^{-\delta} \rceil$ differ, we can split the interval $[a,b)$ into subintervals $[a, c-\delta'), [c-\delta', c)$ and $[c,b)$, where $c \in (a,b)$ is the point where $\lceil x^{-\delta} \rceil$ drops.

Repeating the above argument for the first and third intervals, taking δ' sufficiently small, and exploiting continuity of the expression in (2.3.14) on $[a, c)$ provides us with desired asymptotic bounds for the middle interval, hence the proof is complete. $\square$

As a midway point in proving Theorem 2.3.1, we obtain an expansion for $\theta_{\mathbf{T}}(p_c + \varepsilon)$ that for a given ε is measurable with respect to $\mathcal{T}_{n(\varepsilon)}$, where $n(\varepsilon)$ grows like a small power of ε^{-1}.

Lemma 2.3.11. *Suppose* $\mathbf{E}[Z^{(2\ell+1)(1+\beta)}] < \infty$ *for some* $\ell \geq 1$ *and* $\beta > 0$*. Define* $n(\varepsilon) := \lceil \varepsilon^{-\delta} \rceil$*. Then for* $\delta > 0$ *sufficiently small, we have* $\mathtt{GW}$*-a.s. the following expansion as* $\varepsilon \to 0^+$*:*

$$\theta_{\mathbf{T}}(p_c + \varepsilon) = \sum_{i=1}^{\ell} \left(\sum_{j=1}^{i} (-1)^{j+1} \sum_{d=j}^{i} p_c^d r_{j,d} X_{n(\varepsilon)}^{(j,i-d)} \right) \mu^i \varepsilon^i + o(\varepsilon^\ell).$$

Proof. For each j and n, define

$$\widetilde{\mathsf{Bon}}_n^{(j)}(\varepsilon) := \sum_{\{v_i\} \in \binom{\mathbf{T}_n}{j}} \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p v_1, \dots, v_j] \theta_{\mathbf{T}(v_1)}(p) \cdots \theta_{\mathbf{T}(v_j)}(p)$$

$$\text{and } \mathsf{Bon}_n^{(j)}(\varepsilon) := \sum_{\{v_i\} \in \binom{\mathbf{T}_n}{j}} \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_p v_1, \dots, v_j] \theta(p)^j$$

where we write $p = p_c + \varepsilon$. Applying the Bonferroni inequalities to the event $\{\mathbf{0} \leftrightarrow_p \infty\} = \bigcup_{v \in T_n} \{\mathbf{0} \leftrightarrow_p v \leftrightarrow \infty\}$ yields

$$\sum_{i=1}^{2j} (-1)^{i+1} \cdot \widetilde{\mathsf{Bon}}_{n(\varepsilon)}^{(i)}(\varepsilon) \leq \theta_{\mathbf{T}}(p_c + \varepsilon) \leq \sum_{i=1}^{2j \pm 1} (-1)^{i+1} \cdot \widetilde{\mathsf{Bon}}_{n(\varepsilon)}^{(i)}(\varepsilon) \tag{2.3.15}$$

for each j, where the $\pm$ may be either a plus or minus.

For sufficiently small $\delta > 0$, Proposition 2.3.10 allows us to replace each $\widetilde{\mathsf{Bon}}^{(i)}_{n(\varepsilon)}(\varepsilon)$ with $\mathsf{Bon}^{(i)}_{n(\varepsilon)}(\varepsilon)$, introduce an $o(\varepsilon^\ell)$ error term, provided $\mathbf{E}[Z^{(2i-1)(1+\beta)}] < \infty$. Moreover, we note

$$\begin{aligned}\mathsf{Bon}^{(i)}_{n(\varepsilon)}(\varepsilon) &= \theta(p_c+\varepsilon)^i \sum_{\{v_r\}\in\binom{\mathbf{T}_{n(\varepsilon)}}{i}} \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow_{p_c+\varepsilon} v_1,\dots,v_i] \\ &\le C\theta(p_c+\varepsilon)^i X^{(i,0)}_{n(\varepsilon)} = o(\varepsilon^{i-1})\,.\end{aligned}$$

The constant C is introduced when we bound $(1+\varepsilon/p_c)^{|\mathbf{T}(v_1,\dots,v_i)|}$ from above by a constant C for δ sufficiently small; the limit follows from Theorem 2.3.6(c). For each j, apply (2.3.15) to show

$$\theta_{\mathbf{T}}(p_c+\varepsilon) = \sum_{j=1}^{\ell}(-1)^{j+1}\mathsf{Bon}^{(j)}_n(\varepsilon) + o(\varepsilon^\ell)\,. \tag{2.3.16}$$

Now expand

$$\begin{aligned}\mathsf{Bon}^{(j)}_n(\varepsilon) &= \theta(p_c+\varepsilon)^j \sum_{\{v_i\}\in\binom{\mathbf{T}_n}{j}} (p_c+\varepsilon)^{|\mathbf{T}(v_1,\dots,v_j)|} \\ &= \left(\sum_{i=j}^{\ell} r_{j,i}\varepsilon^i + o(\varepsilon^\ell)\right) \sum_{\{v_i\}\in\binom{\mathbf{T}_n}{j}} p_c^{|\mathbf{T}(v_1,\dots,v_j)|}(1+\varepsilon/p_c)^{|\mathbf{T}(v_1,\dots,v_j)|} \\ &= \left(\sum_{i=j}^{\ell} r_{j,i}\varepsilon^i + o(\varepsilon^\ell)\right) \sum_{\{v_i\}\in\binom{\mathbf{T}_n}{j}} p_c^{|\mathbf{T}(v_1,\dots,v_j)|} \\ &\qquad \times \left(\sum_{i=0}^{\ell}\binom{|\mathbf{T}(v_1,\dots,v_j)|}{i}\frac{\varepsilon^i}{p_c^i} + O(n^{\ell+1}\varepsilon^{\ell+1})\right) \\ &= \left(\sum_{i=j}^{\ell} r_{j,i}\varepsilon^i + o(\varepsilon^\ell)\right)\left(\sum_{i=0}^{\ell} X^{(j,i)}_n \frac{\varepsilon^i}{p_c^i} + o(\varepsilon^\ell)\right) \\ &= \sum_{i=j}^{\ell}\mu^i\varepsilon^i\left(\sum_{d=j}^{i} p_c^d r_{j,d} X^{(j,i-d)}_n\right) + o(\varepsilon^\ell)\,.\end{aligned} \tag{2.3.17}$$

Plugging this into (2.3.16) completes the Lemma. $\square$

We are almost ready to prove Theorem 2.3.1. We have dealt with the martingale part. What remains is to get rid of the predictable part. The following combinatorial identity is the key to making the predictable part disappear.

Lemma 2.3.12. *Fix $i \geq 1$ and suppose $\mathbf{E}[Z^{i+1}] < \infty$; then for each $a, b \leq i$ we have*

$$\sum_{d=1}^{i} \sum_{j=1}^{i} (-1)^{j-1} p_c^d r_{j,d} c_{j,a} \binom{j}{b-d} = (-1)^{a+1} p_c^b r_{a,b} \, .$$

Proof. Begin as in the proof of Proposition 2.2.6 with the identity

$$\Big[1 - \phi(1 - (p_c + \varepsilon)\theta(p_c + \varepsilon)) \Big]^a = \theta(p_c + \varepsilon)^a \, .$$

The idea is to take Taylor expansions of both sides and equate coefficients of ε^b; more technically, taking Taylor expansions of both sides up to terms of order $o(\varepsilon^i)$ yield two polynomials in ε of degree i whose difference is $o(\varepsilon^i)$ thereby showing the two polynomials are equal. The coefficient $[\varepsilon^b]\theta(p_c + \varepsilon)^a$ of ε^b on the right-hand side is $r_{a,b}$, by definition. On the left-hand-side, we write

$$\begin{aligned} &\Big[1 - \phi(1 - (p_c + \varepsilon)\theta(p_c + \varepsilon)) \Big]^a \\ &\qquad = \Big[\sum_{k=1}^{i} (-1)^{k+1} (1 + \varepsilon/p_c)^k \theta(p_c + \varepsilon)^k p_c^k \frac{\phi^{(k)}(1)}{k!} + o(\varepsilon^i) \Big]^a \\ &\qquad = (-1)^a \sum_{j=1}^{i} (-1)^j (1 + \varepsilon/p_c)^j \theta(p_c + \varepsilon)^j c_{j,a} + o(\varepsilon^i) \, . \end{aligned}$$

The coefficient of ε^b of $(1+\varepsilon/p_c)^j\theta(p_c+\varepsilon)^j$ is

$$\begin{aligned}[\varepsilon^b](1+\varepsilon/p_c)^j\theta(p_c+\varepsilon)^j &= \sum_{d=j}^{b}\left([\varepsilon^d]\theta(p_c+\varepsilon)^j\right)\left([\varepsilon^{b-d}](1+\varepsilon/p_c)^j\right) \\ &= \sum_{d=j}^{b} r_{j,d}\binom{j}{b-d}p_c^{-(b-d)}\,.\end{aligned}$$

Equating the coefficients of ε^b on both sides then gives

$$(-1)^a\sum_{j=1}^{i}(-1)^j c_{j,a}\sum_{d=j}^{b} r_{j,d}\binom{j}{b-d}p_c^{-(b-d)} = r_{a,b}\,.$$

Multiplying by $p_c^b(-1)^{a+1}$ on both sides completes the proof. $\square$

With Theorem 2.3.6 and Lemma 2.3.12 in place, the limits of $M_n^{(i)}$ fall out easily.

Lemma 2.3.13. *Suppose* $\mathbf{E}[Z^{i+1}]<\infty$ *for some* i *and let* $\beta>0$ *with* $\mathbf{E}[Z^{i(1+\beta)}]<\infty$. *Then*

(a) *The sequence* $(M_n^{(i)})_{n=1}^{\infty}$ *is a martingale with respect to the filtration* $(\mathcal{T}_n)_{n=1}^{\infty}$.

(b) *There exist positive constants* C,c *depending only on* i,β *and the progeny distribution so that* $\|M_{n+1}^{(i)}-M_n^{(i)}\|_{L^{1+\beta}}\le Ce^{-cn}$.

(c) *There exists a random variable* $M^{(i)}$ *so that* $M_n^{(i)}\to M^{(i)}$ *both almost surely and in* $L^{1+\beta}$.

Proof. Note first that (c) follows from (a) and (b) by the triangle inequality together with the L^p martingale convergence theorem. Parts (a) and (b) are proved simultaneously. Write

$$
\begin{aligned}
&\mu^{-i}\left(M_{n+1}^{(i)}-M_n^{(i)}\right)\\
&\quad=\sum_{j=1}^{i}(-1)^{j+1}\sum_{d=j}^{i}p_c^d r_{j,d}\left(X_{n+1}^{(j,i-d)}-X_n^{(j,i-d)}\right)\\
&\quad=\sum_{j=1}^{i}(-1)^{j+1}\sum_{d=j}^{i}p_c^d r_{j,d}\left(\Delta Y_{n+1}^{(j,i-d)}+\sum_{a=1}^{j}c_{j,a}\sum_{b=0}^{i-d}\binom{j}{i-d-b}X_n^{(a,b)}-X_n^{(j,i-d)}\right)\\
&\quad=\sum_{j=1}^{i}\sum_{d=j}^{i}(-1)^{j+1}p_c^d r_{j,d}\Delta Y_{n+1}^{(j,i-d)}\\
&\qquad+\sum_{j=1}^{i}\sum_{d=j}^{i}(-1)^{j+1}p_c^d r_{j,d}\left(\sum_{a=1}^{j}\sum_{b=0}^{i-d}c_{j,a}\binom{j}{i-d-b}X_n^{(a,b)}-X_n^{(j,i-d)}\right).
\end{aligned}
\tag{2.3.18}
$$

By Theorem 2.3.6, we have that $\Delta Y_{n+1}^{(j,i-d)}$ is exponentially small in $L^{1+\beta}$. This means that we simply need to handle the second sum in (2.3.18). We claim that it is identically equal to zero. This is equivalent to the claim that

$$
\sum_{j=1}^{i}\sum_{d=j}^{i}\sum_{a=1}^{j}\sum_{b=0}^{i-d}(-1)^{j+1}p_c^d r_{j,d}c_{j,a}\binom{j}{i-d-b}X_n^{(a,b)}=\sum_{a=1}^{i}\sum_{b=a}^{i}(-1)^{a+1}p_c^b r_{a,b}X_n^{(a,i-b)}.
\tag{2.3.19}
$$

To prove this, we rearrange the sums in the left-hand-side of (2.3.19). In order to handle the limits of each sum, we recall that $c_{j,a}=0$ for $j<a$ and $r_{j,d}=0$ for

$d < j$. Relabeling and swapping gives

$$\begin{aligned}
&\sum_{j=1}^{i}\sum_{d=j}^{i}\sum_{a=1}^{j}\sum_{b=0}^{i-d}(-1)^{j+1}p_c^d r_{j,d}c_{j,a}\binom{j}{i-d-b}X_n^{(a,b)} \\
&\qquad = \sum_{j=1}^{i}\sum_{d=j}^{i}\sum_{a=1}^{j}\sum_{b=d}^{i}(-1)^{j+1}p_c^d r_{j,d}c_{j,a}\binom{j}{b-d}X_n^{(a,i-b)} \\
&\qquad = \sum_{a=1}^{i}\sum_{b=a}^{i}X_n^{(a,i-b)}\left(\sum_{d=1}^{i}\sum_{j=1}^{i}(-1)^{j-1}p_c^d r_{j,d}c_{j,a}\binom{j}{b-d}\right).
\end{aligned}$$

Lemma 2.3.12 shows that the term in parentheses is equal to $(-1)^{a+1}p_c^b r_{a,b}$, thereby showing (2.3.19). $\square$

Proof of Theorem 2.3.1: Apply Lemma 2.3.11 to obtain some $\delta > 0$ sufficiently small so that

$$\theta_{\mathbf{T}}(p_c+\varepsilon) = \sum_{i=1}^{\ell} M_n^{(i)}\varepsilon^i + o(\varepsilon^\ell) \tag{2.3.20}$$

with $n = \lceil \varepsilon^{-\delta}\rceil$. The exponential convergence of $M_n^{(i)}$ from Lemma 2.3.13 together with Markov's inequality and Borel-Cantelli shows that

$$|M_n^{(i)} - M^{(i)}|n^N \to 0$$

almost surely for any fixed $N > 0$. Because $n = \lceil \varepsilon^{-\delta}\rceil$ implies $n^{-N} = o(\varepsilon^\ell)$ for N sufficiently large, (2.3.20) can be simplified to

$$\theta_{\mathbf{T}}(p_c+\varepsilon) = \sum_{i=1}^{\ell} M^{(i)}\varepsilon^i + o(\varepsilon^\ell)\,.$$

It remains only to show that $\mathbf{E}M^{(i)} = r_i$. Because $M_n^{(i)}$ converges in $L^{1+\beta}$, it also converges in L^1, implying $\mathbf{E}[M^{(i)}] = \mathbf{E}[M_1^{(i)}]$. Noting that $\mathbf{E}[X_1^{(j,k)}] = \binom{j}{k}c_{j,1}$, we use Lemma 2.3.12 with $a = 1$ and $b = i$ in the penultimate line to obtain

$$\begin{aligned} p_c^i \mathbf{E}[M_1^{(i)}] + \sum_{j=1}^{i} \sum_{d=j}^{i} (-1)^{j+1} p_c^d r_{j,d} \mathbf{E}[X_1^{(j,i-d)}] &= \sum_{j=1}^{i} \sum_{d=j}^{i} (-1)^{j+1} p_c^d r_{j,d} \binom{j}{i-d} c_{j,i} \\ &= (-1)^{1+1} p_c^i r_{1,i} \\ &= p_c^i r_i \, . \end{aligned}$$

□

2.4 Regularity on the Supercritical Region

In this section we prove Russo-type formulas expressing the derivatives of $\theta_T(p)$ as expectations of quantities measuring the number of pivotal bonds. The first and simplest of these is Theorem 2.4.1, expressing $\theta_T'(p)$ as the expected number of pivotal bonds multiplied by p^{-1}. In Section 2.4.2 we define some combinatorial gadgets to express more general expectations (Definitions 2.4.7) and show that these compute successive derivatives (Proposition 2.4.9). In Section 2.4.3, explicit estimates on these expectations are given in Proposition 2.4.11, which under suitable moment conditions lead to continuity of the first k derivatives at p_c^+, which is Theorem 2.4.10.

2.4.1 Smoothness on $(p_c, 1)$

To study regularity of $\theta_T(p)$, we obtain Russo-type formulas for the derivatives of $\theta_T(p)$ as expectations of quantities measuring the number of pivotal bonds. For brevity, we only sketch the proof that θ is continuously differentiable in $(p_c, 1)$, and give a birds-eye-view of the picture for higher derivatives.

Given T and p, let $T_p = T_p(\omega)$ denote the tree obtained from the p-percolation cluster at the root by removing all vertices v not connected to infinity in $T(v)$. Formally, $v \in T_p$ if and only if $\mathbf{0} \leftrightarrow_{T,p} v$ and $v \leftrightarrow_{T(v),p} \infty$. On the survival event $H_T(p)$ let B_p denote the first node at which T_p branches. The event $\{B_p = v\}$ is the intersection of three events $\mathsf{Open}(v), \mathsf{NoBranch}(v)$ and $\mathsf{Branch}(v)$ where:

- $\mathsf{Open}(v)$ is the event $\mathbf{0} \leftrightarrow_{T,p} v$ of the path from the root to v being open
- $\mathsf{NoBranch}(v)$ is the event that for each ancestor $w < v$, no child of w other than the one that is an ancestor of v is in T_p
- $\mathsf{Branch}(v)$ is the event that v has at least two children in T_p.

We call $|B_p|$ the *branching depth*. The main result of this subsection is the following.

Theorem 2.4.1. *The derivative of the quenched survival function is given by*

$$\theta'_T(p) = p^{-1}\mathbf{E}_T|B_p|\,,$$

which is finite and continuous on $(p_c, 1)$*.*

We provide a road-map to the proof of Theorem 2.4.1. From the classical theory of branching processes, we have:

Proposition 2.4.2. *For any* $p > p_c$ *define an offspring generating function*

$$\phi_p(z) := \frac{\phi(1 - p\theta(p)(1-z)) - \phi(1-p\theta(p))}{\theta(p)}\,. \tag{2.4.1}$$

Then the conditional distribution of $\mathbf{T}_p$ *given* $H(p)$ *is Galton-Watson with offspring generating function* ϕ_p*, which we will denote* $\mathtt{GW}_p$*.* $\square$

This gives us exponential moments for $|B_p|$.

Lemma 2.4.3 (annealed branching depth has exponential moments)**.** *Let*

$$A_p = A_p(\phi) := \phi_p'(0) \tag{2.4.2}$$

denote the probability under $\mathtt{GW}_p$ *that the root has precisely one child. Suppose* $r > 0$ *and* $p > p_c$ *satisfy* $(1+r)A_p < 1$*. Then* $\mathbf{E}(1+r)^{|B_p|} < \infty$*.*

Next we recast the p-indexed stochastic process $\{T_p : p \in [0,1]\}$ as a Markov chain. Define a filtration $\{\mathcal{G}_p : 0 \le p \le 1\}$ by $\mathcal{G}_p = \sigma(\mathcal{T}, \{U_e \vee p\})$. Clearly if $p > p'$ then $\mathcal{G}_p \subseteq \mathcal{G}_{p'}$, thus $\{\mathcal{G}_p\}$ is a filtration when p decreases from 1 to 0. Informally, $\mathcal{G}_p$ knows the tree, knows whether each edge e is open at "time" p, and if not, "remembers" the time $U(e)$ when e closed. The key is to note that $\{T_p\}$ is in fact Markovian:

Lemma 2.4.4. *Fix any tree* T*. The edge processes* $\{\mathbf{1}_{U(e)\le p}\}$ *are independent left-continuous two-state continuous time Markov chains. They have initial state 1 when* $p = 1$ *and terminal state 0 when* $p = 0$*, and they jump from 1 to 0 at rate* p^{-1}*. The process* $\{T_p\}$ *is a function of these and is also Markovian on* $\{\mathcal{G}_p\}$*.*

Next we define the quantity β as $\beta := \inf\{p : T_p \text{ is infinite}\}$. Thus $\theta_T(p) = \mathbf{P}_T(\beta \le p)$ and $\theta_T'(p)$ is the density, if it exists, of the $\mathbf{P}_T$-law of β. Before establishing Theorem 2.4.1, we will need one additional lemma.

Lemma 2.4.5. *With probability* 1*, at* $p = \beta$ *the root of* T_p *is connected to infinity,* $|B_p| < \infty$ *(i.e.* T_p *does branch somewhere for* $p = \beta$*), and there is a vertex* $v \le B_p$ *with* $U_v = \beta$*. Consequently, the event* $H(p)$ *is, up to measure* 0*, a disjoint union of the events* $\{B_\beta = v\} \cap \{\beta \le p\}$*.*

Proof of Theorem 2.4.1: By Lemma 2.4.5 $H_T(p)$ is equal to the union of the disjoint events $\{B_\beta = v\} \cap H_T(p)$. On $\{B_\beta = v\}$ the indicator $\mathbf{1}_{H(p)}$ jumps to zero precisely when $\mathsf{Open}(v)$ does so, which occurs at rate $p^{-1}|v|$. Because all jumps have the same sign, it now follows that

$$\frac{d}{dp}\,\theta_T(p) = \frac{1}{p}\sum_{v\in T}|v|\mathbf{P}(B_p = v) = \frac{1}{p}\mathbf{E}|B_p|\,,$$

which may be $+\infty$. Summing by parts, we also have

$$\frac{d}{dp}\,\theta_T(p) = \frac{1}{p}\sum_{v\neq \mathbf{0}}\mathbf{P}(B_p \ge v) \tag{2.4.3}$$

where $B_p \ge v$ denotes $B_p = w$ for some descendant w of v.

To see that this is finite and continuous on $(p_c, 1)$, consider any $p' > p_c$ and $r > 0$ with $(1+r)p'A_{p'} < 1$. For any $p \in (p', (1+r)p')$ we have

$$\begin{aligned}\mathbf{P}_T(B_p \ge v) &= \mathbf{P}_T(\mathsf{Open}(v))\mathbf{P}(\mathsf{NoBranch}(v,p))\theta_{T(v)}(p)\\ &\le (1+r)^{|v|}(p')^{|v|}\mathbf{P}_T(\mathsf{NoBranch}(v,p'))\,.\end{aligned}$$

Taking the expectation of the expression on the right and multiplying by $\theta(p')$

we observe that

$$\theta(p')\mathbf{E}\left[\sum_{v\in\mathbf{T}}(1+r)^{|v|}(p')^{|v|}\mathbf{P_T}(\mathsf{NoBranch}(v,p'))\right]$$
$$=\mathbf{E}\left[\sum_{v\in\mathbf{T}}(1+r)^{|v|}(p')^{|v|}\mathbf{P_T}(\mathsf{NoBranch}(v,p'))\theta_{\mathbf{T}(v)}(p')\right]$$
$$=\mathbf{E}\left[\sum_{n=0}^{\infty}(1+r)^{n}\mathbf{P}(B_{p'}\geq n)\right]<\infty$$

where the last inequality follows from Lemma 2.4.3. This now implies that for `GW`-almost every T, the right-hand-side of (2.4.3) converges uniformly for $p \in (p', (1+r)p')$, thus implying continuity on this interval. Covering $(p_c, 1)$ by countably many intervals of the form $(p', (1+r)p')$, the theorem follows by countable additivity. $\square$

2.4.2 Smoothness of θ on the Supercritical Region

Building on the results from the previous subsection, we establish the main result concerning the behavior of the quenched survival function in the supercritical region.

Theorem 2.4.6. *For* `GW`*-a.e.* T*,* $\theta_T(p) \in C^\infty((p_c, 1))$.

In order to prove this result, we define quantities generalizing the quantity $\mathbf{E}_T|B_p|$ and show that the derivative of a function in this class remains in the class. We present the central definitions and key ideas, although the longer proofs are omitted.

Throughout the remainder of this chapter, our trees are rooted and *ordered*, meaning that the children of each vertex have a specified order (usually referred to as left-to-right rather than smallest-to-greatest) and isomorphisms between trees are understood to preserve the root and the orderings.

Definition 2.4.7.

(*i*) **Collapsed trees.** Say the tree $\mathcal{V}$ is a collapsed tree if no vertex of $\mathcal{V}$ except possibly the root has precisely one child.

(*ii*) **Initial subtree.** The tree $\widetilde{T}$ is said to be an initial subtree of T if it has the same root, and if for every vertex $v \in \widetilde{T} \subseteq T$, either all children of v are in $\widetilde{T}$ or no children of v are in $\widetilde{T}$, with the added proviso that if v has only one child in T then it must also be in $\widetilde{T}$.

(*iii*) **The collapsing map Φ.** For any ordered tree T, let $\Phi(T)$ denote the isomorphism class of ordered trees obtained by collapsing to a single edge any path along which each vertex except possibly the last has only one child in T (see figure below).

(*iv*) **Notations $T(\mathcal{V})$ and $\mathcal{V} \preceq T$.** It follows from the above definitions that any collapsed tree $\mathcal{V}$ is isomorphic to $\Phi(\widetilde{T})$ for at most one initial tree $\widetilde{T} \subseteq T$. If there is one, we say that $\mathcal{V} \preceq T$ and denote this subtree by $T(\mathcal{V})$. We will normally use this for $T = T_p$. For example, when $\mathcal{V}$ is the tree with one edge then $\mathcal{V} \preceq T_p$ if and only if T_p has precisely one child of the root, in which case $T_p(\mathcal{V})$ is the path from the root to B_p.

(*v*) **The embedding map ι.** If e is an edge of $\mathcal{V}$ and $\mathcal{V} \preceq T$, let $\iota(e)$ denote the path in $T(\mathcal{V})$ that collapses to the edge carried to e in the above isomorphism. For a vertex $v \in \mathcal{V}$ let $\iota(v)$ denote the last vertex in the path $\iota(e)$ where e is the edge between e and its parent; if v is the root of $\mathcal{V}$ then by convention $\iota(v)$ is the root of T.

(vi) **Edge weights.** If $\mathcal{V} \preceq T$ and $e \in E(\mathcal{V})$, define $d(e) = d_{T,\mathcal{V}}(e)$ to be the length of the path $\iota(e)$.

(vii) **Monomials.** A monomial in (the edge weights of) a collapsed tree $\mathcal{V}$ is a set of nonnegative integers $\{F(e) : e \in E(\mathcal{V})\}$ indexed by the edges of $\mathcal{V}$, identified with the product

$$\langle T, \mathcal{V}, F\rangle := \begin{cases} \prod_{e\in E(\mathcal{V})} d(e)^{F(e)} & \text{if } \mathcal{V} \preceq T\,, \\ 0 & \text{otherwise.} \end{cases} . \tag{2.4.4}$$

A monomial F is only defined in reference to a weighted collapsed tree $\mathcal{V}$.

Definition 2.4.8 (monomial expectations)**.** Given T, p, a positive real number r, a finite collapsed tree $\mathcal{V}$, and a monomial F, define functions $\mathcal{R} = \mathcal{R}(T, r, \mathcal{V}, p)$ and $\mathcal{D} = \mathcal{D}(T, F, \mathcal{V}, p)$ by

$$\mathcal{R} \;:=\; \mathbf{E}_T\left[(1+r)^{|E(T_p(\mathcal{V}))|}\mathbf{1}_{\mathcal{V}\preceq T_p}\right] \tag{2.4.5}$$

$$\mathcal{D} \;:=\; \mathbf{E}_T\left[\langle T_p, \mathcal{V}, F\rangle\right] . \tag{2.4.6}$$

For example, if $\mathcal{V}_1$ is the tree with a single edge e and $F_1(e) = 1$, then $\langle T_p, \mathcal{V}_1, F_1\rangle = |B_p|$ and the conclusion of Theorem 2.4.1 is that for $p > p_c$,

$$\frac{d}{dp}\,\theta_T(p) = \frac{1}{p}\mathbf{E}_T|B_p| = \frac{1}{p}\mathcal{D}(T, F_1, \mathcal{V}_1, p)\,. \tag{2.4.7}$$

The main result of this section, from which Theorem 2.4.6 follows without too much further work, is the following representation.

Proposition 2.4.9. *Let $\mathcal{V}$ be a collapsed tree and let F be a monomial in the variables $d_{T,\mathcal{V}}(e)$. Then there exists a collection of collapsed trees $\mathcal{V}_1, \ldots, \mathcal{V}_m$ for some $m \geq 1$ and monomials $F_1, \ldots, F_m$, such that*

$$\frac{d}{dp}\mathbf{E}_T\langle T_p, \mathcal{V}, F\rangle = \frac{1}{p}\sum_{i=1}^{m} \mathbf{E}_T\langle T_p, \mathcal{V}_i, F_i\rangle \tag{2.4.8}$$

on $(p_c, 1)$ and is finite and continuous on $(p_c, 1)$ for GW*-a.e. tree T. Furthermore, each monomial F_i on the right-hand side of* (2.4.8) *satisfies* $\deg(F_i) = 1 + \deg(F)$ *and each of the edge sets $E(\mathcal{V}_i)$ satisfies $|E(\mathcal{V}_i)| \leq 2 + |E(\mathcal{V})|$.*

From here, Theorem 2.4.6 follows from (2.4.7) and Proposition 2.4.9 by induction.

2.4.3 Continuity of the derivatives at p_c

We now address the part of Theorem 2.1.2 concerning the behavior of the derivatives of g near criticality. We restate this result here as the following Theorem.

Theorem 2.4.10. *If $\mathbf{E}[Z^{(2k^2+3)(1+\beta)}] < \infty$ for some $\beta > 0$, then*

$$\lim_{p \to p_c^+} \theta_{\mathbf{T}}^{(j)}(p) = j! M^{(j)}$$

for every $j \leq k$ GW*-a.s. where $M^{(j)}$ are as in Theorem 2.3.1.*

To prove Theorem 2.4.10 we need to bound how badly the monomial expectations $\mathcal{D}(T, F, \mathcal{V}, p_c + \varepsilon)$ can blow up as $\varepsilon \downarrow 0$, then use Lemma 2.2.1 to see that they can't blow up at all.

Proposition 2.4.11. *Let $\mathcal{V}$ be a collapsed tree with ℓ leaves and $\mathcal{E}$ edges and let F be a monomial in the edges of $\mathcal{V}$. Suppose that the offspring distribution has at least m moments, where $m \geq \max_e F(e)$ and also $m \geq 3$. Then*

$$\mathcal{D}(T, F, \mathcal{V}, p_c + \varepsilon) = O\left(\varepsilon^{\lambda}\right)$$

for any $\lambda < 2\ell - \mathcal{E} - \deg(F)$ and $\mathtt{GW}$-almost every T.

Chapter 3

Critical Percolation and the Incipient Infinite Cluster

This chapter is based on [Mic19], and contains the entire paper verbatim.

3.1 Introduction

We consider percolation on a locally finite rooted tree T: each edge is open with probability $p \in (0,1)$, independently of all others. Let $\mathbf{0}$ denote the root of T and $\mathcal{C}_p$ be the open p-percolation cluster of the root. We may consider the *survival probability* $\theta_T(p) := \mathbf{P}[|\mathcal{C}_p| = +\infty]$ and note that θ_T is an increasing function of p. There thus exists a *critical percolation parameter* $p_c \in [0,1]$ so that $\theta_T(p) = 0$ for all $p \in [0, p_c)$ and $\theta_T(p) > 0$ for $p \in (p_c, 1]$. If T is a regular tree where each non-root vertex has degree $d+1$—i.e. each vertex has d children—then the classical theory of branching processes shows that $p_c = \frac{1}{d}$ and $\theta_T(p_c) = 0$ (see, for instance, [AN72]). Since critical percolation does not occur, we may consider the *incipient infinite cluster* (IIC), in which we condition on critical percolation reaching depth M of T and take M to infinity.

The IIC for regular trees was first constructed and considered by Kesten in [Kes86b]. In that work, along with [BK06], the primary focus was on simple random walk on the IIC for regular trees. Our focus is on three elementary quantities for random T: the probability that critical percolation reaches depth n; the number of vertices of $\mathcal{C}_p$ at depth n conditioned on percolation reaching depth n; and the number of vertices in the IIC at depth n. For regular trees, these questions were answered in the study of critical branching processes. In fact, these classical results apply to *annealed* critical percolation on Galton-Watson trees. If we generate a Galton-Watson tree T with progeny distribution $Z \geq 1$ with $\mathbf{E}[Z] > 1$, we may perform $p_c = 1/\mathbf{E}[Z]$ percolation at the same time as we generate T; this is known at the *annealed* process—in which we generate T and percolate simultaneously—and is equivalent to generating a Galton-Watson tree with offspring distribution $\widetilde{Z} := \text{Bin}(Z, p_c)$. Since $\mathbf{E}[\widetilde{Z}] = 1$, this is a critical branching process and thus the classical theory can be used:

Theorem 3.1.1 ([KNS66])**.** *Suppose* $\mathbf{E}[Z^2] < \infty$*, and set* Y_n *to be the set of vertices at depth* n *of* T *connected to the root in* $p_c = 1/\mathbf{E}[Z]$ *percolation. Then* (*a*) *The annealed probability of surviving to depth n satisfies*

$$n \cdot \mathbf{P}[|Y_n| > 0] \to \frac{2}{\mathrm{Var}\,[\widetilde{Z}]} = \frac{2\mathbf{E}[Z]^2}{\mathbf{E}[Z(Z-1)]}\,.$$

(*b*) *The annealed conditional distribution of $|Y_n|/n$ given $|Y_n| > 0$ converges in distribution to an exponential law with mean $\frac{\mathbf{E}[Z(Z-1)]}{2\mathbf{E}[Z]^2}$ as $n \to \infty$.*

Under the additional assumption of $\mathbf{E}[Z^3] < \infty$, parts (*a*) and (*b*) are due to Kolmogorov [Kol38] and Yaglom [Yag47] respectively; as such, they are commonly referred to as Kolmogorov's estimate and Yaglom's limit law. For a modern treatment of these classical results, see [LPP95] or [LP17, Section 12.4]. Although less widely known, Theorem 3.1.1 quickly gives a limit law for the size of the annealed IIC.

Corollary 3.1.2. *If $\mathbf{E}[Z^2] < \infty$, let C_n denote the number of vertices at depth n in the annealed incipient infinite cluster. Then C_n/n converges in distribution to the random variable with density $\lambda^2 x e^{-\lambda x}$ with $\lambda := \frac{2\mathbf{E}[Z]^2}{\mathbf{E}[Z(Z-1)]}$ on $[0,\infty)$. In other words,*

$$\lim_{n\to\infty}\left(\lim_{M\to\infty}\mathbf{P}[|Y_n|/n \in (a,b)\,|\,|Y_M| > 0]\right) = \int_a^b \lambda^2 x e^{-\lambda x}\,dx$$

for each $a < b$.

This can be easily proven from Theorem 3.1.1 using an argument similar to the proof of Theorem 3.3.11, and thus the details are omitted.

Our goal is to upgrade Theorem 3.1.1 and Corollary 3.1.2 to hold for the *quenched* process; that is, rather than generate T and perform percolation at the same time as in the annealed case, we generate T and then perform percolation on each resulting T. Before stating the quenched results, we recall some notation and facts from the theory of branching processes. If we allow $\mathbf{P}[Z=0]>0$ and condition on the resulting tree being infinite, we may pass to the reduced tree as in [LP17, Chapter 5.7] in which we remove all vertices that have finitely many descendants; this results in a new Galton-Watson process with some offspring distribution $\tilde{Z} \geq 1$. We therefore assume without loss of generality that $Z \geq 1$. For a Galton-Watson tree T, let Z_n denote the number of vertices at distance of n from the root; then the process $W_n = Z_n/(\mathbf{E}[Z])^n$ converges almost-surely to some random variable W.

A first quenched result is that of [Lyo90], which states that for a.e. supercritical Galton-Watson tree with progeny distribution Z, we have that the critical percolation probability is $p_c = 1/\mathbf{E}[Z]$; furthermore, for almost every Galton-Watson tree $\mathbf{T}$, $\theta_{\mathbf{T}}(p) = 0$ for $p \in [0, p_c]$ and $\theta_{\mathbf{T}}(p) > 0$ for $p \in (p_c, 1]$. For a fixed tree T, let $\mathbf{P}_T[\cdot]$ be the probability measure induced by performing p_c percolation on T. When T is random, this is a random variable and we may ask about the almost sure behavior of certain probabilities. Our main results are summarized in the following theorem:

Theorem 3.1.3. *Let $\mathbf{T}$ be a Galton-Watson tree with progeny distribution $Z \geq 1$ with $\mathbf{E}[Z] > 1$. Suppose $\mathbf{E}[Z^p] < \infty$ for each $p \geq 1$. Set $\lambda := \frac{2\mathbf{E}[Z]^2}{\mathbf{E}[Z(Z-1)]}$ and let Y_n be the set of vertices in depth n of $\mathbf{T}$ connected to the root in $p_c = 1/\mathbf{E}[Z]$ percolation. Then for a.e. $\mathbf{T}$ we have*

(a) $n \cdot \mathbf{P}_{\mathbf{T}}[|Y_n| > 0] \to W\lambda$ *a.s.*

(*b*) *The conditioned variable* $(|Y_n|/n \,|\, |Y_n| > 0)$ *converges in distribution to an exponential random variable with mean* λ^{-1} *a.s.*

(*c*) *Let* $\mathbf{C}_n$ *denote the number of vertices in the quenched IIC of* $\mathbf{T}$ *at depth* n*. Then* $\mathbf{C}_n/n$ *converges in distribution to the random variable with density* $\lambda^2 x e^{-\lambda x}$ *a.s.*

Note that, surprisingly, the limit laws of parts (*b*) and (*c*) of Theorem 3.1.3 do not depend at all on $\mathbf{T}$ itself but just on the distribution of Z. This is in sharp contrast to the case of near-critical and supercritical percolation on Galton-Watson trees, in which the behavior is dependent on the tree itself [MPR18]. One possible justification for this lack of dependence on W, for instance, is that conditioning on $|Y_n| > 0$ forces certain structure of the percolation cluster near the root; since W is mostly determined by the levels of $\mathbf{T}$ near the root, the behavior when conditioned on $|Y_n| > 0$ for large n does not depend on W. Part (*a*) of Proposition 3.3.8 corroborates this heuristic explanation.

The three parts of Theorem 3.1.3 are Theorems 3.3.3, 3.3.5 and 3.3.11 respectively. The proof of part (*a*) utilizes its annealed analogue, Theorem 3.1.1(*a*), along with a law of large numbers argument. Part (*b*) is proven by the method of moments building on the work of [MPR18]. Part (*c*) follows from there with a similar law of large numbers argument combined with two short facts about the structure of the percolation cluster conditioned on $|Y_n| > 0$ (this is Proposition 3.3.8).

Remark 3.1.4. Theorem 3.1.3 assumes that $\mathbf{E}[Z^p] < \infty$ for each $p \geq 1$, and we suspect that this condition is an artifact of the proof. Since we use the method of moments, it is natural that we require all moments of the underlying distribution to be finite. We suspect that less rigid conditions are sufficient, but this would require a different proof strategy than the method of moments, perhaps utilizing a stronger anti-concentration statement in the vein of Proposition 3.3.8.

3.2 Set-up and Notation

We begin with some notation and a brief description of the probability space on which we will work. Let Z be a random variable taking values in $\{1, 2, \ldots, \}$ with $\mu := \mathbf{E}[Z] > 1$ and $\mathbf{P}[Z = 0] = 0$. Define its *probability generating function* to be $\phi(z) := \sum \mathbf{P}[Z = k]z^k$. Let $\mathbf{T}$ be a random locally finite rooted tree with law equal to that of a Galton-Watson tree with progeny distribution Z and let $(\Omega_1, \mathcal{T}, \mathtt{GW})$ be the probability space on which it is defined. Since we will perform percolation on these trees, we also use variables $\{U_i\}_{i=1}^{\infty}$ where the U_i are i.i.d. random variables uniform on $[0, 1]$; let $(\Omega_2, \mathcal{F}_2, \mathbf{P}_2)$ be the corresponding probability space. Our canonical probability space will be $(\Omega, \mathcal{F}, \mathbf{P})$ with $\Omega := \Omega_1 \times \Omega_2$, $\mathcal{F} := \mathcal{T} \otimes \mathcal{F}_2$ and $\mathbf{P} := \mathtt{GW} \times \mathbf{P}_2$. We interpret an element $\omega = (T, \omega_2) \in \Omega$ as the tree T with edge weights given by the U_i random variables. To obtain p percolation, we restrict to the subtree of edges with weight at most p. Since we are concerned with quenched probabilities, we define the measure $\mathbf{P_T}[\cdot] := \mathbf{P}[\cdot \,|\, \mathbf{T}] = \mathbf{P}[\cdot \,|\, \mathcal{T}]$. Since this is a random variable, our goal is to prove theorems $\mathtt{GW}$-a.s.

We employ the usual notation for a rooted tree T, Galton-Watson or otherwise: $\mathbf{0}$ denotes the root; T_n is the set of vertices at depth n; and $Z_n := |T_n|$. In the case of a Galton-Watson tree $\mathbf{T}$, we define $W_n := Z_n/\mu^n$ and recall that $W_n \to W$ almost surely. Furthermore, if $\mathbf{E}[Z^p] < \infty$ for some $p \in [1, \infty)$, we in fact have $W_n \to W$ in L^p [BD74, Theorems 0 and 5]. In the Galton-Watson case, define $\mathcal{T}_n := \sigma(\mathbf{T}_n)$; then $(\mathcal{T}_n)_{n=0}^{\infty}$ is a filtration that increases to $\mathcal{T}$. For a vertex v of T, define $T(v)$ to be the descendant tree of v and extend our notation to include $T_n(v), Z_n(v), W_n(v)$ and $W(v)$. For vertices v and w, write $v \leq w$ if v is an ancestor of w.

For percolation, recall that the critical percolation probability for $\mathtt{GW}$-a.e. $\mathbf{T}$ is $p_c := 1/\mu$ and that percolation does not occur at criticality [Lyo90]. For vertices v and w with $v \leq w$, let $\{v \leftrightarrow w\}$ denote the event that there is an open path from v to w in p_c percolation; let $\{v \leftrightarrow (u, w)\}$ be the event that v is connected to both u and w in p_c percolation; for a subset S of $\mathbf{T}$, let $\{v \leftrightarrow S\}$ denote the event that v is connected to some element of S in p_c percolation; lastly, let Y_n be the set of vertices in $\mathbf{T}_n$ that are connected to $\mathbf{0}$ in p_c percolation.

3.3 Quenched Results

3.3.1 Moments

For $k \geq j$, let $\mathcal{C}_j(k)$ denote the set of j-compositions of k, i.e. ordered j-tuples of positive integers that sum to k. Define

$$c_{k,j} := p_c^k \sum_{a \in \mathcal{C}_j(k)} m_{a_1} m_{a_2} \cdots m_{a_j}$$

where $m_r := \mathbf{E}[\binom{Z}{r}]$. We use the following result from [MPR18]:

Theorem 3.3.1 ([MPR18])**.** *Define*

$$M_n^{(k)} := \mathbf{E}_{\mathbf{T}}\left[\binom{|Y_n|}{k}\right] - \sum_{i=1}^{k-1} c_{k,i} \sum_{j=0}^{n-1} \mathbf{E}_{\mathbf{T}}\left[\binom{|Y_j|}{i}\right].$$

If $\mathbf{E}[Z^{2k}] < \infty$, *then* $M_n^{(k)}$ *is a martingale with respect to the filtration* $(\mathcal{T}_n)$, *and there exist constants* C_k *and* c_k *so that*

$$\|M_{n+1}^{(k)} - M_n^{(k)}\|_{L^2} \leq C_k e^{-c_k n}.$$

While Theorem 3.3.1 is not stated precisely this way in [MPR18], the martingale property follows from [MPR18, Lemma 4.1], while the L^2 bound on the increments is given in [MPR18, Theorem 4.4]. This gives us the leading term of each $\mathbf{E}_{\mathbf{T}}\left[|Y_n|^k\right]$.

Proposition 3.3.2. *For each* k,

$$\mathbf{E}_{\mathbf{T}}\left[|Y_n|^k\right] n^{-(k-1)} \to k! \left(\frac{p_c^2 \phi''(1)}{2}\right)^{k-1} W$$

almost surely and in L^2.

Proof. By Theorem 3.3.1, $M_n^{(k)}$ is a martingale with uniformly bounded L^2 norm for each k. By the L^p martingale convergence theorem, $M_n^{(k)}$ converges in L^2 and almost surely. We now proceed by induction on k. For $k = 1$, $\mathbf{E_T}[|Y_n|] = W_n$ which converges to W. Suppose that the proposition holds for all $j < k$. Then by convergence of $M_n^{(k)}$,

$$\mathbf{E_T}\left[\binom{|Y_n|}{k}\right] n^{-(k-1)} = \sum_{i=1}^{k-1} c_{k,i} n^{-(k-1)} \sum_{j=0}^{n-1} \mathbf{E_T}\left[\binom{|Y_j|}{i}\right] + o(1)$$

where the $o(1)$ term is both in L^2 and almost surely. By induction, the leading term is the contribution from $i = k-1$. Noting that $c_{k,k-1} = (k-1)p_c^2 \frac{\phi''(1)}{2}$ and the fact that $\sum_{j=0}^{n-1} j^d \sim \frac{1}{d+1} n^{d+1}$ completes the proof. □

3.3.2 Survival Probabilities

Throughout, define $\lambda := \frac{2}{p_c^2 \phi''(1)}$. Our first task is to find a quenched analogue of Kolmogorov's estimate:

Theorem 3.3.3. *If* $\mathbf{E}[Z^4] < \infty$*, then*

$$n \cdot \mathbf{P_T}[|Y_n| > 0] \to W\lambda$$

almost surely.

The proof utilizes the Bonferroni inequalities. In order to control the second-order term, the variance of a sum of pairs is calculated, thereby introducing the requirement of $\mathbf{E}[Z^4] < \infty$. We begin first by proving upper and lower bounds:

Lemma 3.3.4. *For each n,*

$$\frac{n \cdot \mathbf{E_T}[|Y_n|]^2}{\mathbf{E_T}[|Y_n|^2]} \leq n \cdot \mathbf{P_T}[|Y_n| > 0] \leq \frac{2\overline{W}}{1-p_c}$$

where, $\overline{W} = \sup_n W_n$.

Proof. The lower bound is the Paley-Zygmund inequality. For the upper bound, we use [LP17, Theorem 5.24]:

$$\mathbf{P_T}[|Y_n| > 0] \leq \frac{2}{\mathscr{R}(\mathbf{0} \leftrightarrow \mathbf{T}_n)}$$

where $\mathscr{R}(\mathbf{0} \leftrightarrow \mathbf{T}_n)$ is the equivalent resistance between the root and $\mathbf{T}_n$ when all of $\mathbf{T}_n$ is shorted to a single vertex and each edge branching from depth $k-1$ to k has resistance $\frac{1-p_c}{p_c^k}$. Shorting together all vertices at depth k for each k gives the lower bound

$$\mathscr{R}(\mathbf{0} \leftrightarrow \mathbf{T}_n) \geq \sum_{k=1}^{n} \frac{1-p_c}{Z_k p_c^k} = \sum_{k=1}^{n} \frac{1-p_c}{W_k} \geq (1-p_c)\frac{n}{\overline{W}}\,.$$

$\square$

Proof of Theorem 3.3.3: For each fixed $m < n$, the Bonferroni inequalities imply

$$\left| n\mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_n] - n \sum_{v \in \mathbf{T}_m} \mathbf{P_T}[\mathbf{0} \leftrightarrow v \leftrightarrow \mathbf{T}_n] \right| \leq n \sum_{u,v \in \binom{\mathbf{T}_m}{2}} \mathbf{P_T}[\mathbf{0} \leftrightarrow (u,v) \leftrightarrow \mathbf{T}_n]\,. \tag{3.3.1}$$

If we can show that the right-hand side of (3.3.1) converges a.s. to zero for some choice of $m = m(n)$, then the survival probability is sufficiently close to a sum of i.i.d. random variables. The random variables $\mathbf{P_T}[\mathbf{0} \leftrightarrow v \leftrightarrow \mathbf{T}_n]$ are i.i.d. with mean $p_c^m \mathbf{P}[\mathbf{0} \leftrightarrow \mathbf{T}_{n-m}]$, implying that the sum is close to $W_m \mathbf{P}[\mathbf{0} \leftrightarrow \mathbf{T}_{n-m}]$.

Applying the annealed result Theorem 3.1.1 would then complete the proof after noting that $W_m \to W$ almost surely provided $m \to \infty$. The remainder of the proof follows this sketch.

Set $m = \lceil n^{1/4} \rceil$; we then bound the second moment

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{u,v \in \binom{\mathbf{T}_m}{2}} \mathbf{P_T}[\mathbf{0} \leftrightarrow (u,v) \leftrightarrow \mathbf{T}_n]\right)^2\right] \\
&\quad = \mathbf{E}\left[\left(\sum_{u,v \in \binom{\mathbf{T}_m}{2}} \mathbf{P_T}[\mathbf{0} \leftrightarrow (u,v)]\mathbf{P_T}[u \leftrightarrow \mathbf{T}_n]\mathbf{P_T}[v \leftrightarrow \mathbf{T}_n]\right)^2\right] \\
&\quad = \mathbf{E}\left[\mathbf{E}\left[\left(\sum_{u,v \in \binom{\mathbf{T}_m}{2}} \mathbf{P_T}[\mathbf{0} \leftrightarrow (u,v)]\mathbf{P_T}[u \leftrightarrow \mathbf{T}_n]\mathbf{P_T}[v \leftrightarrow \mathbf{T}_n]\right)^2 \,\middle|\, \mathcal{T}_m\right]\right] \\
&\quad = \mathbf{E}\left[\mathbf{E}\left[\left(\sum_{u,v \in \binom{\mathbf{T}_m}{2}} \mathbf{P_T}[\mathbf{0} \leftrightarrow (u,v)]\mathbf{P_T}[u \leftrightarrow \mathbf{T}_n]\mathbf{P_T}[v \leftrightarrow \mathbf{T}_n]\right)^2 \,\middle|\, \mathcal{T}_m\right]^{(1/2)\cdot 2}\right] \\
&\quad \le \mathbf{E}\left[\left(\sum_{u,v \in \binom{\mathbf{T}_m}{2}} \mathbf{P_T}[\mathbf{0} \leftrightarrow (u,v)] \left\| \mathbf{P_T}[u \leftrightarrow \mathbf{T}_n]\mathbf{P_T}[v \leftrightarrow \mathbf{T}_n] \right\|_{L^2}\right)^2\right] \\
&\quad \le \left(\frac{2}{1-p_c}\right)^4 \mathbf{E}[\overline{W}^2]^2 \cdot (n-m)^{-4} \mathbf{E}\left[\binom{|Y_m|}{2}^2\right] \\
&\quad \le Cm^2 n^{-4} \,.
\end{aligned}$$

Multiplying by n, the second moment of the right-hand side of (3.3.1) is bounded above by $Cm^2n^{-2} = O(n^{-3/2})$ which is summable in n. By Chebyshev's Inequality together with the Borel-Cantelli Lemma, the right-hand side of (3.3.1) converges to zero almost surely. This implies

$$
\begin{aligned}
n\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow \mathbf{T}_n] &= n \sum_{v \in \mathbf{T}_m} \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow v \leftrightarrow \mathbf{T}_n] + o(1) \\
&= \sum_{v \in \mathbf{T}_m} \frac{n\mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_n]}{\mu^m} + o(1)\,. \qquad (3.3.2)
\end{aligned}
$$

We want to show that the right-hand side of (3.3.2) converges to $W\lambda$, so we first calculate

$$
\begin{aligned}
&\mathrm{Var}\left[\sum_{v \in \mathbf{T}_m} \frac{n\mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_n] - n\mathbf{P}[\mathbf{0} \leftrightarrow \mathbf{T}_{n-m}]}{\mu^m}\right] \\
&\qquad = \mathbf{E}\left[\mathrm{Var}\left[\sum_{v \in \mathbf{T}_m} \frac{n\mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_n] - n\mathbf{P}[\mathbf{0} \leftrightarrow \mathbf{T}_{n-m}]}{\mu^m} \,\middle|\, \mathcal{T}_m\right]\right] \\
&\qquad = \mathbf{E}\left[\frac{1}{\mu^{2m}} \sum_{v \in \mathbf{T}_m} \mathrm{Var}\left[n\mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_n]\right]\right] \\
&\qquad \le \frac{C}{\mu^m}
\end{aligned}
$$

where the last inequality is via Lemma 3.3.4. Since this is summable in n, Chebyshev's Inequality and the Borel-Cantelli Lemma again imply

$$
\sum_{v \in \mathbf{T}_m} \frac{n\mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_n]}{\mu^m} = \sum_{v \in \mathbf{T}_m} \frac{n\mathbf{P}[\mathbf{0} \leftrightarrow \mathbf{T}_{n-m}]}{\mu^m} + o(1) = W_m(n \cdot \mathbf{P}[\mathbf{0} \leftrightarrow \mathbf{T}_{n-m}]) + o(1)\,.
$$

Taking $n \to \infty$ and utilizing Theorem 3.1.1 together with (3.3.2) completes the proof. $\square$

3.3.3 Conditioned Survival

Theorem 3.3.5. *Suppose $\mathbf{E}[Z^p] < \infty$ for all $p \ge 1$. Then the conditional variable $(|Y_n|/n \,|\, |Y_n| > 0)$ converges in distribution to an exponential random variable with mean λ^{-1} for $\mathtt{GW}$-almost every $\mathbf{T}$.*

By conditional random variable $(|Y_n|/n \,|\, |Y_n| > 0)$, we mean the random variable with law $\mathbf{P_T}[|Y_n|/n \in \cdot \,|\, |Y_n| > 0]$.

Proof. The proof is via the method of moments. In particular, since the moment generating function of an exponential random variable has a positive radius of convergence, its distribution is uniquely determined by its moments. Thus, any sequence of random variables with each moment converging to the moment of an exponential random variable must converge in distribution to that exponential random variable [Bil95, Theorems 30.1 and 30.2].

Let X_n be a random variable with distribution $(|Y_n|/n \,|\, |Y_n| > 0)$. It is sufficient to show $\mathbf{E_T}[X_n^k] \to k!\lambda^{-k}$ $\mathtt{GW}$-a.s. since $k!\lambda^{-k}$ is the kth moment of an exponential random variable. Proposition 3.3.2 and Theorem 3.3.3 imply

$$\begin{aligned}
\mathbf{E_T}[X_n^k] &= \frac{\mathbf{E_T}[|Y_n|^k]}{n^k \mathbf{P_T}[|Y_n| > 0]} \\
&= \frac{\mathbf{E_T}[|Y_n|^k}{n^{k-1}} \cdot \frac{1}{n \cdot \mathbf{P_T}[|Y_n| > 0]} \\
&\to k! W \lambda^{-(k-1)} \cdot \frac{1}{\lambda W} \\
&= k!\lambda^{-k} \,.
\end{aligned}$$

□

More can be said about the structure of the open percolation cluster of the root conditioned on $\mathbf{0} \leftrightarrow \mathbf{T}_n$, but we require two general, more or less standard lemmas first.

Lemma 3.3.6. *For any events A and B with $\mathbf{P}[B] \neq 0$,*

$$|\mathbf{P}[A \mid B] - \mathbf{P}[A]| \leq \mathbf{P}[B^c] .$$

Proof. Expand

$$\mathbf{P}[A] = \mathbf{P}[A \mid B](1 - \mathbf{P}[B^c]) + \mathbf{P}[A \mid B^c]\mathbf{P}[B^c]$$

and solve

$$\mathbf{P}[A] - \mathbf{P}[A \mid B] = (\mathbf{P}[A \mid B^c] - \mathbf{P}[A \mid B])\mathbf{P}[B^c] .$$

Taking absolute values and bounding $|\mathbf{P}[A \mid B^c] - \mathbf{P}[A \mid B]| \leq 1$ completes the proof.

$\square$

Lemma 3.3.7. *Let X_k be i.i.d. centered random variables with $\mathbf{E}[|X_1|^p] < \infty$ for some $p \in [2, \infty)$. Then there exists a constant C_p so that*

$$\mathbf{P}\left[\left|\sum_{k=1}^{n} \frac{X_k}{n}\right| > t\right] \leq C_p t^{-p} n^{-p/2} + 2\exp\left(-\frac{nt^2}{\mathrm{Var}\,[X_1]}\right)$$

for all $t > 0$.

Proof. This is a straightforward application of [Che09, Theorem 2.1] which states that for independent random variables M_i with $\mathbf{E}[M_i] = 0$ and $\mathbf{E}[|M_i|^p] < \infty$ for some $p > 2$ we have

$$\mathbf{P}\left[\left|\sum_{i=1}^{n} M_i\right| \geq t\right] \leq C_p t^{-p} \max\left(r_{n,p}(t), (r_{n,2}(t))^{p/2}\right) + \exp\left(-\frac{t^2}{16 b_n}\right)$$

where $r_{n,u}(t) = \sum_{i=1}^{n} \mathbf{E}(|M_i|^u \mathbf{1}_{|M_i| \geq 3b_n/t})$, $b_n = \sum_{i=1}^{n} \mathbf{E}[M_i^2]$ and C_p is a positive constant. Setting $M_i = X_i/n$ completes the proof. $\square$

For a fixed tree and $m < n$, define $B_m(n)$ to be the event that $\mathbf{0} \leftrightarrow \mathbf{T}_n$ through precisely one vertex at depth m.

Proposition 3.3.8. *Suppose $\mathbf{E}[Z^p] < \infty$ for all $p \geq 1$. There exists an $N = N(\mathbf{T})$ with $N < \infty$ almost surely so that for all $n \geq N$, we have*

(*a*) $\mathbf{P}_{\mathbf{T}}[B_m(n)^c \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] < Cn^{-1/4}$ *for* $m = m(n) := \lceil \frac{\log n}{4 \log \mu} \rceil$

(*b*) $\max_{v \in \mathbf{T}_n} \mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] = O(n^{-1/8})$

for some constant $C > 0$.

Proof. Note first that for the choice of m as in part (a), we have $\frac{1}{2\mu} W n^{1/4} \leq Z_m \leq 2\mu W n^{1/4}$ for sufficiently large n.

(a) Using Theorem 3.3.3 and Lemma 3.3.4, we bound

$$\begin{aligned}
\mathbf{P}_{\mathbf{T}}[B_m(n)^c \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] &\leq \frac{\left(\sum_{v \in T_m} \mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_n]\right)^2}{\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow \mathbf{T}_n]} \\
&\leq \left(\frac{2}{1-p_c}\right)^2 \left(\frac{\sum_{v \in \mathbf{T}_m} \overline{W}(v)}{Z_m}\right)^2 \frac{Z_m^2}{(n-m)^2 \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow \mathbf{T}_n]} \\
&\leq C \left(\frac{\sum_{v \in \mathbf{T}_m} \overline{W}(v)}{Z_m}\right)^2 W n^{-1/2} \qquad (3.3.3)
\end{aligned}$$

for n sufficiently large, and some choice of $C > 0$ depending on the distribution of Z. Applying Lemma 3.3.7 for $p = 9$ gives

$$\mathbf{P}\left[\left|\frac{\sum_{v \in \mathbf{T}_m} \overline{W}(v)}{Z_m} - \mathbf{E}[\overline{W}]\right| > n^{1/8}\right] \leq C_9 n^{-9/8} + 2\exp\left(-n^{1/4}/\mathrm{Var}\,[\overline{W}]\right)$$

where we use the trivial bound of $1 \leq Z_m$. Since this is summable in n, the Borel-Cantelli Lemma implies that this event only occurs finitely often. In particular, this means that for sufficiently large n

$$\mathbf{P_T}[B_m(n)^c \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] \le CWn^{-1/4} \tag{3.3.4}$$

for some constant $C > 0$ depending only on the distribution of Z.

(b) Applying Lemma 3.3.6 to the measure $\mathbf{P_T}[\cdot \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n]$ and recalling $B_m(n) \subseteq \mathbf{0} \leftrightarrow \mathbf{T}_n$,

$$\Big|\mathbf{P_T}[v \in Y_n \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] - \mathbf{P_T}[v \in Y_n \,|\, B_m(n)]\Big| \le \mathbf{P_T}\left[B_m(n)^c \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n\right]$$

which is $O(n^{-1/4})$ by part (a). It is thus sufficient to bound $\mathbf{P_T}[v \in Y_n \,|\, B_m(n)]$. For a vertex $v \in \mathbf{T}_n$ and $m < n$, let $P_m(v)$ be the ancestor of v in $\mathbf{T}_m$. We then have

$$\mathbf{P_T}[v \in Y_n \,|\, B_m(n)] \le \mathbf{P_T}[\mathbf{0} \leftrightarrow P_m(v) \leftrightarrow \mathbf{T}_n \,|\, B_m(n)]\,.$$

Conditioned on $B_m(n)$, there exists a unique vertex $w \in \mathbf{T}_m$ so that $\mathbf{0} \leftrightarrow w \leftrightarrow \mathbf{T}_n$; this vertex w is chosen with probability bounded above by

$$\begin{aligned}
&\mathbf{P_T}[\mathbf{0} \leftrightarrow w \leftrightarrow \mathbf{T}_n \,|\, B_m(n)] \\
&\qquad \le \frac{\mathbf{P_T}[\mathbf{0} \leftrightarrow w \leftrightarrow \mathbf{T}_n]}{\sum_{u \in \mathbf{T}_m} \mathbf{P_T}[\mathbf{0} \leftrightarrow u \leftrightarrow \mathbf{T}_n] - \sum_{(u_1,u_2) \in \binom{\mathbf{T}_m}{2}} \mathbf{P_T}[\mathbf{0} \leftrightarrow (u_1,u_2) \leftrightarrow \mathbf{T}_n]} \\
&\qquad \le \frac{\mathbf{P_T}[w \leftrightarrow \mathbf{T}_n]}{\sum_{u \in \mathbf{T}_m} \mathbf{P_T}[u \leftrightarrow \mathbf{T}_n] - \left(\sum_{u \in \mathbf{T}_m} \mathbf{P_T}[u \leftrightarrow \mathbf{T}_n]\right)^2} \\
&\qquad \le \frac{c(n-m)^{-1}\overline{W}(w)}{(1+o(1))\sum_{u \in \mathbf{T}_m} \mathbf{P_T}[u \leftrightarrow \mathbf{T}_n]}
\end{aligned} \tag{3.3.5}$$

where the latter inequality is by applying the bound of Lemma 3.3.4 to the numerator and arguing as in (3.3.3) to almost-surely bound the denominator. In particular, the $o(1)$ term is uniform in w.

We want to take the maximum over all possible $w \in \mathbf{T}_m$, and note that for any $\alpha > 0$,

$$
\begin{aligned}
\mathbf{P}\left[\max_{w\in\mathbf{T}_m} \overline{W}(w) > n^\alpha\right] &= \mathbf{E}\left[\mathbf{P}\left[\max_{w\in\mathbf{T}_m} \overline{W}(w) > n^\alpha \,\middle|\, \mathcal{T}_m\right]\right] \\
&\le \mathbf{E}[Z_m]\mathbf{P}[\overline{W} > n^\alpha] \\
&\le \mu^m \cdot \frac{\mathbf{E}[\overline{W}^{2/\alpha}]}{n^2} \\
&= O(n^{-7/4})
\end{aligned}
$$

which is summable, implying that for any fixed $\alpha > 0$, we eventually have $\max_{w\in\mathbf{T}_m} \overline{W}(w) \le n^\alpha$. It merely remains to bound the denominator of (3.3.5).

Note that by Proposition 3.3.2, the lower bound given in Lemma 3.3.4 converges almost surely to $\frac{W\lambda}{2}$ as $n \to \infty$. In particular, this means that if we set

$$
p_n := \mathbf{P}\left[\frac{W\lambda}{4} \le n\mathbf{P_T}[|Y_n| > 0]\right],
$$

then $p_n \to 1$. By Hoeffding's inequality together with Borel-Cantelli, the number of vertices $u \in \mathbf{T}_m$ for which we have

$$
\frac{W(u)\lambda}{4} \le (n-m)\mathbf{P_T}[u \leftrightarrow \mathbf{T}_n]
$$

is almost surely at least $1/2$ of $\mathbf{T}_m$ for n sufficiently large. This gives

$$
(n-m)\sum_{u\in\mathbf{T}_m} \mathbf{P_T}[u \leftrightarrow \mathbf{T}_n] \ge \frac{\lambda}{4}\sum_{u\in\mathbf{T}_m} W(u)\mathbf{1}_{W(u)\lambda/4 \le (n-m)\mathbf{P_T}[u\leftrightarrow\mathbf{T}_n]} = \Omega(Z_m)\,.
$$

Recalling that $Z_m = \Theta(Wn^{-1/4})$ and plugging the above into (3.3.5) completes the proof. $\square$

3.3.4 Incipient Infinite Cluster

As in [Kes86a], we sketch a proof of the construction of the IIC. For an infinite tree T, define $T[n]$ to be the finite subtree of T obtained by restricting to vertices of depth at most n.

Lemma 3.3.9. *Suppose* $\mathbf{E}[Z^4] < \infty$*; for a subtree* t *of* $\mathbf{T}[n]$*, we have*

$$\lim_{M\to\infty} \mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_M] = \frac{\sum_{v\in t_n} W(v)}{W} \mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t]$$

almost surely for each tree t*.*

The random measure $\mu_{\mathbf{T}}$ *on subtrees of* $\mathbf{T}$ *with marginals*

$$\mu_{\mathbf{T}}\big|_{\mathcal{T}_n}[t] := \frac{\sum_{v\in t_n} W(v)}{W} \mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t]$$

has a unique extension to a probability measure on rooted infinite trees GW *almost surely. The IIC is thus the random subtree of* $\mathbf{T}$ *with law* $\mu_{\mathbf{T}}$*.*

Proof. Since each $\mathbf{T}$ has countably many vertices, Theorem 3.3.3 assures that $n\mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_{n+|v|}] = \lambda W(v)$ for each vertex v of $\mathbf{T}$ a.s. When all of these limits hold, we then have

$$\begin{aligned}
\mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_M] &= \frac{\mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t, \mathbf{0} \leftrightarrow \mathbf{T}_M]}{\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow \mathbf{T}_M]} \\
&= \mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t] \left(\frac{\sum_{v\in t_n} \mathbf{P}_{\mathbf{T}}[v \leftrightarrow \mathbf{T}_M] + O(|t_n|^2 M^{-2})}{\mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow \mathbf{T}_M]} \right) \\
&\xrightarrow{M\to\infty} \mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t] \frac{\sum_{v\in t_n} W(v)}{W}
\end{aligned}$$

for each t. To show that the measure $\mu_{\mathbf{T}}$ can be extended, we note that its marginals are consistent, as can be seen via the recurrence $W(v) = p_c \sum_w W(w)$ where the sum is over all children of v. Applying the Kolmogorov extension theorem [Dur10, Theorem 2.1.14] completes the proof. $\square$

It is easy to show that the law of the IIC can in fact be generated by conditioning on $p > p_c$ percolation to survive and then taking $p \to p_c^+$:

Corollary 3.3.10. *For a subtree t of $\mathbf{T}[n]$, we have*

$$\lim_{p \to p_c^+} \mathbf{P}_{\mathbf{T}}[\mathcal{C}_p[n] = t \mid |\mathcal{C}_p| = \infty] = \frac{\sum_{v \in t_n} W(v)}{W} \mathbf{P}_{\mathbf{T}}[\mathcal{C}_{p_c}[n] = t]$$

almost surely.

Proof. As shown in [MPR18], we have

$$\lim_{p \to p_c} \frac{\mathbf{P}_{\mathbf{T}}[|\mathcal{C}_p| = \infty]}{p - p_c} = KW$$

almost-surely for some constant K depending only on the offspring distribution. The Corollary follows from Bayes' theorem in the same manner as Lemma 3.3.9. $\square$

In light of Lemma 3.3.9, it is natural to guess that the number of vertices in the IIC at depth n will asymptotically be the size-biased version of $(|Y_n| \mid \mathbf{0} \leftrightarrow \mathbf{T}_n)$: the sum $\sum_{v \in t_n} W(v)$ will be relatively close to $|t_n| W$, therefore biasing each choice of t by a factor of $|t_n|$. In order to make this argument rigorous, we will invoke Proposition 3.3.8 which shows that no single vertex has high probably of surviving conditionally. Throughout, we use the notation $n(a,b) = (na, nb)$ for $a < b$ and $\mathbf{C}$ to denote the IIC.

Theorem 3.3.11. *Suppose $\mathbf{E}[Z^p] < \infty$ for each $1 \le p < \infty$. Then for each $0 \le a < b$,*

$$\lim_{n\to\infty} \mathbf{P_T}[\mathbf{C}_n \in n(a,b)] = \int_a^b \lambda^2 x e^{-\lambda x}\, dx$$

almost surely. In fact, $\mathbf{C}_n/n$ converges in distribution to the random variable with density $\lambda^2 x e^{-\lambda x}$ for $\mathtt{GW}$-almost every $\mathbf{T}$.

Proof. To see that convergence in distribution follows from the almost sure limit, apply the almost sure limit to each interval (a,b) with $a,b \in \mathbb{Q}$; since there are only countably many such intervals, there exists a set of full $\mathtt{GW}$ measure on which these limits simultaneously exist for each rational interval, thereby implying convergence in distribution [Dur10, Theorem 3.2.5].

We have

$$\mathbf{P_T}[\mathbf{C}_n \in n(a,b)] = \lim_{M\to\infty} \mathbf{P_T}[Y_n \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_{n+M}]\,.$$

For a fixed n, write

$$\begin{aligned}
&\mathbf{P_T}[|Y_n| \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_{n+M}] \\
&\qquad = \frac{\mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_{n+M} \,|\, |Y_n| \in n(a,b)] \cdot \mathbf{P_T}[|Y_n| \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] \cdot \mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_n]}{\mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_{n+M}]}\,.
\end{aligned} \tag{3.3.6}$$

We then calculate

$$\begin{aligned}
&\mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_{n+M} \,|\, |Y_n| \in n(a,b)] \\
&\qquad = \sum_S \mathbf{P_T}[Y_n = S \,|\, |Y_n| \in n(a,b)] \mathbf{P_T}[S \leftrightarrow \mathbf{T}_{n+M}] \\
&\qquad = \sum_S \mathbf{P_T}[Y_n = S \,|\, |Y_n| \in n(a,b)] \sum_{v \in S} \mathbf{P_T}[v \leftrightarrow \mathbf{T}_{n+M}] + O(M^{-2}) \\
&\qquad = \sum_{v \in \mathbf{T}_n} \mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b)] \mathbf{P_T}[v \leftrightarrow \mathbf{T}_{n+M}] + O(M^{-2}) \,.
\end{aligned}$$

For a fixed n, we take $M \to \infty$ and utilize Theorem 3.3.3 to get

$$\lim_{M \to \infty} \frac{\mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_{n+M} \,|\, |Y_n| \in n(a,b)]}{\mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_{n+M}]} = \frac{1}{W} \sum_{v \in \mathbf{T}_n} \mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b)] \cdot W(v) \,. \tag{3.3.7}$$

We plug this into (3.3.6) to get the limit

$$\begin{aligned}
&\lim_{M \to \infty} \mathbf{P_T}[|Y_n| \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_{n+M}] \\
&\qquad\qquad = \left(\sum_{v \in \mathbf{T}_n} \frac{\mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b)]}{n} \cdot W(v) \right) \\
&\qquad\qquad\quad \times \left(\mathbf{P_T}[|Y_n| \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] \right) \left(\frac{n \cdot \mathbf{P_T}[\mathbf{0} \leftrightarrow \mathbf{T}_n]}{W} \right) \,.
\end{aligned}$$

Theorems 3.3.3 and 3.3.5 show that the latter two factors above have almost sure limits $\int_a^b \lambda e^{-\lambda x}\,dx$ and λ as $n \to \infty$, leaving only the first term. We note that

$$\begin{aligned}
\mathbf{E}\left[\sum_{v\in\mathbf{T}_n} \frac{\mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)]}{n} \cdot W(v) \,\middle|\, \mathcal{T}_n\right] &= \sum_{v\in\mathbf{T}_n} \frac{\mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)]}{n} \\
&= \mathbf{E}_{\mathbf{T}}\left[\frac{|Y_n|}{n} \,\middle|\, |Y_n| \in n(a,b)\right] \\
&= \frac{\mathbf{E}_{\mathbf{T}}\left[\frac{|Y_n|}{n} \cdot \mathbf{1}_{|Y_n|/n\in(a,b)} \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n\right]}{\mathbf{P}_{\mathbf{T}}\left[\frac{|Y_n|}{n} \in (a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n\right]} \\
&\to \frac{\int_a^b \lambda x e^{-\lambda x}\,dx}{\int_a^b \lambda e^{-\lambda x}\,dx}
\end{aligned}$$

where the limit is by the continuous mapping theorem [Dur10, Theorem 3.2.4] and Theorem 3.3.5. It's thus sufficient to show that

$$\left|\sum_{v\in\mathbf{T}_n} \frac{\mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)]}{n} \cdot (W(v) - 1)\right| \xrightarrow{n\to\infty} 0 \tag{3.3.8}$$

almost surely.

Our strategy is to use a conditional version of the Borel-Cantelli Lemma together with Chebyshev's inequality. We bound the conditional variance

$$\begin{aligned}
&\mathrm{Var}\left[\sum_{v\in\mathbf{T}_n} \frac{\mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)]}{n} \cdot (W(v) - 1) \,\middle|\, \mathcal{T}_n\right] \\
&\qquad = \mathrm{Var}\,(W) \sum_{v\in\mathbf{T}_n} \frac{\mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)]^2}{n^2} \\
&\qquad \le \mathrm{Var}\,(W) \max_{v\in\mathbf{T}_n} \mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)] \sum_{v\in\mathbf{T}_n} \frac{\mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)]}{n^2} \\
&\qquad \le \mathrm{Var}\,(W) \max_{v\in\mathbf{T}_n} \mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)] \cdot \frac{\mathbf{E}[Y_n \,|\, |Y_n| \in n(a,b)]}{n^2} \\
&\qquad \le \mathrm{Var}\,(W) \cdot \frac{b}{n} \cdot \max_{v\in\mathbf{T}_n} \mathbf{P}_{\mathbf{T}}[v \in Y_n \,|\, |Y_n| \in n(a,b)]\,.
\end{aligned} \tag{3.3.9}$$

We want to show that this is summable, and thus look to bound the max term. Applying Lemma 3.3.6 to the measure $\mathbf{P_T}[\cdot \,|\, |Y_n| \in n(a,b)]$ gives

$$\begin{aligned}
\Big|\mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b)] - \mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b), B_m(n)]\Big| \\
&\leq \mathbf{P_T}[B_m(n)^c \,|\, |Y_m| \in n(a,b)] \\
&\leq \frac{\mathbf{P_T}[B_m(n)^c \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n]}{\mathbf{P_T}[|Y_n| \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n]} \\
&= O(n^{-1/4}) \qquad (3.3.10)
\end{aligned}$$

by Proposition 3.3.8 and Theorem 3.3.5. Similarly,

$$\begin{aligned}
\mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b), B_m(n)] &= \frac{\mathbf{P_T}[v \in Y_n, |Y_n| \in n(a,b), B_m(n)]}{\mathbf{P_T}[|Y_n| \in n(a,b), B_m(n)]} \\
&\leq \frac{\mathbf{P_T}[v \in Y_n, B_m(n)]}{\mathbf{P_T}[|Y_n| \in n(a,b), B_m(n)]} \\
&= \frac{\mathbf{P_T}[v \in Y_n \,|\, B_m(n)]}{\mathbf{P_T}[|Y_n| \in n(a,b) \,|\, B_m(n)]}\,. \qquad (3.3.11)
\end{aligned}$$

Using Lemma 3.3.6 once again expands the denominator

$$\begin{aligned}
\Big|\mathbf{P_T}[|Y_n| \in n(a,b) \,|\, B_m(n)] - \mathbf{P_T}[|Y_n| \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n]\Big| &\leq \mathbf{P_T}[B_m(n)^c \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] \\
&\leq Cn^{-1/4}
\end{aligned}$$

by Proposition 3.3.8. Plugging into (3.3.11) gives the upper bound

$$\mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b), B_m(n)] \leq \frac{\mathbf{P_T}[v \in Y_n \,|\, B_m(n)]}{\mathbf{P_T}[|Y_n| \in n(a,b) \,|\, \mathbf{0} \leftrightarrow \mathbf{T}_n] - Cn^{-1/4}}\,. \qquad (3.3.12)$$

Combining (3.3.10), (3.3.12) and Proposition 3.3.8 bounds

$$\max_{v \in \mathbf{T}_n} \mathbf{P_T}[v \in Y_n \,|\, |Y_n| \in n(a,b)] = O(n^{-1/8})\,.$$

Thus, by (3.3.9), the conditional variance is almost surely summable. For any fixed $\delta > 0$, Chebyshev's inequality then implies

$$\mathbf{P}\left[\left|\sum_{v\in\mathbf{T}_n}\frac{\mathbf{P}_{\mathbf{T}}[v \in Y_n \mid |Y_n| \in (a,b)]}{n}\cdot(W(v)-1)\right| > \delta \,\middle|\, \mathcal{T}_n\right]$$

is summable almost surely. Applying a conditional Borel-Cantelli Lemma (e.g. [Che78]) shows that (3.3.8) holds almost surely. $\square$

Chapter 4

Invasion Percolation

This final chapter is based on excerpts from [MPR17], which is joint with Robin Pemantle and Josh Rosenberg. For brevity, proofs are omitted and abbreviated from certain sections.

4.1 Introduction

Given an infinite rooted tree, how might one sample, nearly uniformly, from the set of paths from the root to infinity? One motive for this question is that nearly uniform sampling leads to good estimates on the growth rate [JS89]. One might be trying to estimate the size of a search tree, or, in the case of [RS00], to determine the growth rate of the number of self-avoiding paths.

A number of methods have been studied. One is to do a random walk on the tree, with a "homesickness" parameter determining how much steps back toward the root are favored [LPP96]. The parameter needs to be tuned near criticality: too much homesickness and the walk gets stuck near the root; too little homesickness and the walk goes to infinity without taking the time to ensure that the path is well randomized. Randall and Sinclair [RS00] solve this by estimating the critical parameter as the walk progresses, re-tuning the homesickness to lie above this by an amount decreasing at an appropriate rate.

Another approach is to use percolation. One conditions the percolation cluster to survive to level N; as the percolation parameter decreases to criticality and N is taken to infinity, the law of this cluster approaches the law of the *incipient infinite cluster* (IIC). For many graphs—e.g. regular or Galton-Watson trees—the IIC almost surely contains a unique infinite path, thereby giving a mechanism for sampling such a path. In practice, the same considerations arise as with homesick random walks: tuning the percolation parameter too low yields too little likelihood of survival and too great a time cost to rejection sampling; too great a percolation parameter results in too many surviving paths and a selection problem which leads to poor randomization.

Invasion percolation was introduced as a model for how viscous fluid creeps through an environment in [WW83]. For a more complete history, see Section 1.1. Each site is given an independent $U[0,1]$ random variable, representing how great the percolation probability would have to be before the site would be open. The cluster then grows by adding, at each time step, the site with the least U value among sites neighboring the cluster but not in the cluster. On trees, it is not hard to see that the lim sup of U-values of bonds chosen is equal to the critical percolation parameter. In other words, instead of running percolation at p_c and conditioning to survive, one allows slightly supercritical bonds but less and less as the cluster grows. As is the case for the IIC, the invasion cluster almost surely contains only one infinite path in the case of regular or Galton-Watson trees, and thus gives a different mechanism for sampling paths. Unlike the IIC and homesick random walk, invasion percolation requires no tuning to criticality and is an instance of self-organized criticality.

The invasion cluster has some properties in common with the IIC but not all. For example, results of Kesten [Kes86a] and Zhang [Zha95] show that the growth exponents of the two are equal on the two-dimensional lattice; however the measures of the two clusters are mutually singular on the lattice [DSV09] as well as on a regular tree [AGdHS08]. Our focus is the comparison of the laws induced on paths by both the IIC and invasion percolation.

On a Galton-Watson tree T, there is a natural measure on paths, the limit-uniform measure μ_T, which although it does not restrict precisely to the uniform measure on each generation, approximates this as closely as possible. There is not, however, a fast algorithm for sampling from it. Rules such as "split equally at each node" lead to rapid sampling but the wrong entropy; in other words, the Radon-Nikodym derivative with respect μ_T on generation N will be exponential in N. It is not hard to show that on almost every Galton-Watson tree (assuming a $Z \log Z$ moment for the offspring distribution), the unique path in the IIC has law μ_T. Since sampling from the IIC is problematic, it is therefore natural to ask how close the law ν_T of the path chosen by the invasion cluster is to μ_T. It is easy to see that the two laws are typically not equal. In particular, this shows that the IIC does not stochastically dominate the invasion measure on Galton-Watson trees.

The best comparison one might hope for is that ν_T be absolutely continuous with respect to μ_T, perhaps even with Radon-Nikodym derivative in L^p. Our main result is as follows.

Theorem 4.1.1. *Suppose the offspring distribution Z has at least p moments and $\mathbf{P}[Z = 0] = 0$; set $p_1 := \mathbf{P}[Z = 1]$, let $\mu := \mathbf{E}[Z]$, and denote $q := \frac{\log \mu}{\log(1/p_1)}$. If*

$$2p^2q^2 + (3p^2 + 5p)q + (-p^2 + 11p - 4) < 0,$$

then $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$ almost surely.

The condition in Theorem 4.1.1 is a trade-off between p_1 and p. In the case of $p = \infty$, the condition becomes $p_1 < 1/\mu^{\frac{3+\sqrt{17}}{2}}$. In the case of $p_1 = 0$, the condition is $p > \frac{11+\sqrt{105}}{2}$.

A summary of the argument behind Theorem 4.1.1 is as follows. Let X_n be the KL-distance between the way that μ_T and ν_T split at the nth step γ_n of a path chosen from ν_T. A sufficient condition for absolute continuity is that $\sum_{n=1}^{\infty} \mathbf{E}X_n < \infty$. A precise statement is given in Lemma 4.2.7 below. A more detailed outline of the argument is given at the end of this section.

The reason we have a hope of estimating X_n is that there is a *backbone* decomposition for invasion percolation. Define the backbone to be the almost surely unique non-backtracking path $\gamma = (\mathbf{0}, \gamma_1, \gamma_2, \ldots)$ from the root to infinity. For any vertex v define the *pivot value* at v, denoted $\beta(v)$, to be the least p such that there is a path from v to infinity in the subtree at v with all U variables (not including the one at v) at most p. On a regular tree, invasion percolation was studied in [AGdHS08, AGM13]. For the purposes of studying ν_T, the regular tree is a degenerate case, because μ_T and ν_T are equal to each other and to the equally splitting measure. Further, on regular trees, the incipient infinite cluster stochastically dominates the invasion cluster; this fails to hold in the Galton-Watson case due to the fact that $\mu_T \neq \nu_T$. Despite these differences, the results on backbones and pivots in the regular case extend in a useful way to the Galton-Watson setting. In particular, [AGdHS08] prove that the process $\{\beta(\gamma_n) - p_c\}_{n \geq 0}$ converges to the *Poisson lower envelope* process when properly scaled; we prove similar results, and combine Theorem 4.6.2 and Corollaries 4.6.3 and 4.6.4 into the following:

Theorem 4.1.2. *Define* $h_n := \beta(\gamma_n) - p_c$*. Then*

(i) Let $\{U_j\}_{j=0}^{\infty}$ *be IID random variables each uniformly distributed on* $(0, 1)$ *and define* $M_n = \min\{U_0, \ldots, U_n\}$*. Then for each* $\varepsilon > 0$*, the process* $\{h_n\}$ *may be coupled with* $\{M_n\}$ *so that with probability* 1*,* h_n *satisfies* $(1 - \varepsilon)p_c M_n \leq h_n \leq (1 + \varepsilon)p_c M_n$ *for all sufficiently large* n*.*

(ii) For any $\varepsilon > 0$ as $k \to \infty$,

$$(kh_{\lceil kt \rceil}/p_c)_{t \geq \varepsilon} \stackrel{*}{\Longrightarrow} (L(t))_{t \geq \varepsilon}$$

where $\stackrel{*}{\Longrightarrow}$ denotes convergence in distribution of càdlàg paths in the Skorohod space $D[0, \infty)$ and $L(t)$ denotes the Poisson lower envelope process, defined in [AGdHS08] and Section 4.6.

(iii) The sequence $n \cdot h_n$ converges in distribution to $p_c \cdot \exp(1)$, where $\exp(1)$ is an exponential random variable with mean 1.

Conditioning on T, the way the invasion measure splits at v depends on the whole tree. However, if one also conditions on the pivot at v, then the way the invasion measure splits at v becomes independent of everything outside of the subtree at v. A similar statement is true if one conditions on the pivot of v being less than or equal to a certain value; these are the Markov properties of Propositions 4.4.4 and 4.4.6. The limiting behavior of these values is given in Theorem 4.4.8 and Section 4.6. Further, Lemma 4.5.1 shows that this conditioned splitting measure is close to a ratio of survival probabilities under supercritical Bernoulli percolation. The problem is thus reduced to proving estimates of the survival probabilities of Galton-Watson trees under supercritical Bernoulli percolation as in Section 4.3.

The remainder of the chapter is organized as follows. Section 4.2 sets up the notation and gives some preliminary results. Some care is required to set up the probability space so that we can easily speak of the random measures μ_T and

ν_T, which are conditional on the Galton-Watson tree. Section 4.2 culminates in Lemma 4.2.7 and Corollary 4.2.8. Section 4.3 estimates near-critical survival probabilities for Galton-Watson trees. Section 4.4 proves two Markov properties for the subtree from γ_n together with $\beta(\gamma_n)$. The remainder of the section extends the work of [AGdHS08] by proving a limit law for $\beta(\gamma_n)$ which then implies an upper bound on the rate at which $\beta(\gamma_n) \downarrow p_c$. In particular, Corollary 4.6.3 shows convergence to the Poisson lower envelope process, as in [AGdHS08]. Section 4.5 completes the proof of Theorem 4.1.1 by comparing the conditional invasion measure to the ratio of survival probabilities and utilizing the estimates on survival probabilities from Section 4.3.

Outline of Proof of Theorem 4.1.1

1. **Absolute continuity follows from summability of KL-divergence of splits**

   Let X_n be the KL-distance between the way that μ_T and ν_T split at the nth step γ_n of a path chosen from ν_T. A sufficient condition for absolute continuity is that $\sum_{n=1}^{\infty} X_n < \infty$. A precise statement is given in Lemma 4.2.7. In fact, we may replace the KL-distance with a process that differs from X_n for only finitely many n (Corollary 4.2.8).

2. **Shifting to the γ_n is the same as conditioning on the pivot being at most a certain value**

We show that shifting to the γ_n is the same as examining a fresh Galton-Watson tree with the pivot of the root conditioned to be at most a certain random variable that we call the *dual pivot*, β_n^*. This is the content of the Markov property given in Proposition 4.4.4.

3. **Understanding how β_n^* behaves for large n**

   As $n \to \infty$, the variables β_n^* approach p_c. We in fact will have that the convergence is quite rapid, as shown by Theorem 4.4.8. The variables β_n^* are closely related to the pivots β_n whose rate of decay is given in Theorem 4.1.2; the process $\{\beta_n^*\}_n$ is difficult to study by itself, although the pair (β_n, β_n^*) is Markov with transition kernel given explicitly in Proposition 4.4.7.

4. **Conditioned on the pivot of the root being at most p, the split of the invasion measure is close to the ratio of survival probabilities**

   With Steps 2 and 3 in mind, we examine the split of the invasion measure conditioned on the root having pivot at most p. Lemma 4.5.1 shows that this splitting measure may be closely approximated by splitting according to the probability that the subtree survives p-percolation.

5. **The ratio of survival probabilities is close to the split of the limit-uniform measure**

   The last remaining step is to show that if p is close to p_c, the ratios of the probabilities of surviving p-percolation closely approximate the splits of the limit uniform measure (Proposition 4.5.2). In order to show this, much work needs to be done to approximate the near-critical survival probabilities of Galton-Watson trees. This is the content and focus of Section 4.3.

4.2 Construction and preliminary results

4.2.1 Galton-Watson trees

We begin with some notation we use for all trees, random or not. Let $\mathcal{U}$ be the canonical *Ulam-Harris* tree [ABF13]. The vertex set of $\mathcal{U}$ is the set $\mathbf{V} := \bigcup_{n=1}^{\infty} \mathbb{N}^n$, with the empty sequence $\mathbf{0} := \emptyset$ as the root. There is an edge from any sequence $\mathbf{a} = (a_1, \ldots, a_n)$ to any extension $\mathbf{a} \sqcup j := (a_1, \ldots, a_n, j)$. The depth of a vertex v is the graph distance between v and $\mathbf{0}$ and is denoted $|v|$. We work with trees T that are locally finite rooted subtrees of $\mathcal{U}$. The usual notations are in force: T_n denotes the set of vertices at depth n; $T(v)$ is the subtree of T at v, canonically identified with a rooted subtree of $\mathcal{U}$, in other words the vertex set of $T(v)$ is $\{w : v \sqcup w \in V(T)\}$; ∂T denotes the set of infinite non-backtracking paths from the root; if $\gamma \in \partial T$ then γ_n $(n \geq 0)$ denotes the nth vertex in γ; the last common ancestor of v and w is denoted $v \wedge w$ and the last common vertex of γ and γ' is denoted $\gamma \wedge \gamma'$; $\overleftarrow{v}$ denotes the parent of v . Let μ_T^n denote the uniform measure on the nth generation of T. In some cases, for example for almost every Galton-Watson tree, the limit $\mu_T := \lim_{n\to\infty} \mu_T^n$ exists and is called the *limit-uniform* *measure* [LP17, Chapter 17.6].

Turning now to Galton-Watson trees, let $\phi(z) := \sum_{n=1}^{\infty} p_n z^n$ be the ordinary generating function for a supercritical branching process with no death, i.e., $\phi(0) = 0$. We recall,

$$\phi'(1) = \mathbf{E}Z =: \mu$$

$$\phi''(1) = \mathbf{E}[Z(Z-1)]$$

where Z is a random variable with probability generating function ϕ. Throughout, we assume $\mathbf{E}[Z^2] < \infty$; in particular, this also means that $\phi''(1) < \infty$. Moreover, since our focus is on ∂T, the assumption of $\phi(0) = 0$ can be made without loss of generality by considering the reduced tree, as in [AN72, Chapter I.12].

We will work on the canonical probability space $(\Omega, \mathcal{F}, \mathbf{P})$ where $\Omega = (\mathbb{N} \times [0,1])^{\mathbf{V}}$, $\mathcal{F}$ is the product Borel σ-field, and $\mathbf{P}$ is the probability measure making the coordinate functions $\omega_v = (\deg_v, U_v)$ IID with the law of (Z, U), where U is uniform on $[0,1]$ and independent of Z. The variables $\{\deg_v\}$—where $\deg_v$ is interpreted as the number of children of vertex v—will construct the Galton-Watson tree, while the variables $\{U_v\}$ will be used later for percolation. Let $\mathbf{T}$ be the random rooted subtree of $\mathcal{U}$ which is the connected component containing the root of the set of vertices that are either the root or are of the form $v \sqcup j$ such that $0 \le j < \deg_v$. This is a Galton-Watson tree with ordinary generating function ϕ.

As is usual for Galton-Watson branching processes, we denote $Z_n := |\mathbf{T}_n|$. Extend this by letting $Z_n(v)$ denote the number of offspring of v in generation $|v| + n$;

similarly, extend the notation for the usual martingale $W_n := \mu^{-n} Z_n$ by letting $W_n(v) := \mu^{-n} Z_n(v)$. We know that $W_n(v) \to W(v)$ for all v, almost surely and in L^p if the offspring distribution has p moments. This is stated without proof for integer values of $p \geq 2$ in [Har63, p. 16] and [AN72, p. 33, Remark 3]; for a proof for all $p > 1$, see [BD74, Theorems 0 and 5]. Further extend this notation by letting $v^{(i)}$ denote the ith child of v, letting $Z_n^{(i)}(v)$ denote nth generation descendants of v whose ancestral line passes through $v^{(i)}$, and letting $W_n^{(i)}(v) := \mu^{-n} Z_n^{(i)}(v)$. Thus, for every v, $W(v) = \sum_i W^{(i)}(v)$. For convenience, define $p_c := 1/\mu$ and recall that p_c is almost surely the critical percolation parameter for $\mathbf{T}$ [Lyo90].

4.2.2 Bernoulli and Invasion Percolation

In this subsection we give the formal construction of percolation on random trees, and for invasion percolation. Our approach is to define a simultaneous coupling of invasion percolations on all subtrees T of $\mathcal{U}$ via the U variables, then specialize to the random tree $\mathbf{T}$. Let $\mathcal{T} := \sigma(\{\deg_v : v \in \mathbf{V}\})$ denote the σ-field generated by the tree variables. Because $\mathcal{T}$ is independent from the U variables, this means we have constructed a process whose law, conditional on $\mathcal{T}$, is invasion percolation on $\mathbf{T}$. We use the notation $\mathbf{E}_{\mathbf{T}}$ to denote $\mathbf{E}[\cdot \mid \mathcal{T}]$; similarly $\mathbf{P}_{\mathbf{T}}[\cdot] := \mathbf{P}[\cdot \mid \mathcal{T}]$. Moreover, we use $\mathtt{GW} := \mathbf{P}|_{\mathcal{T}}$ to denote the Galton-Watson measure on trees.

We begin with a similar construction for ordinary percolation. For $0 < p < 1$, simultaneously define Bernoulli(p) percolations on rooted subtrees T of $\mathcal{U}$ by taking the percolation clusters to be the connected component containing $\mathbf{0}$ of the induced subtrees of T on all vertices v such that $U_v \le p$; note that we always include the root $\mathbf{0}$, and thus the uniform variables U_v may equivalently be thought of as being edge-weights connecting the parent of v to v. Let $\mathcal{F}_n$ be the σ-field generated by the variables $\{U_v, \deg_v : |v| < n\}$. Let $p_c = 1/\mu = 1/\phi'(1)$ denote the critical probability for percolation. Write $v \leftrightarrow_{T,p} w$ if $U_u \le p$ for all u on the geodesic from v to w in T. Informally, $v \leftrightarrow_{T,p} w$ iff v and w are both in T and are connected in the p-percolation. The event of successful p percolation on T is $H_T(p) := \{\mathbf{0} \leftrightarrow_{T,p} \infty\}$ and the event of successful p percolation on the random tree $\mathbf{T}$, is denoted $H_{\mathbf{T}}(p)$ or simply $H(p)$. Let $\theta(T,p) := \mathbf{P}[H_T(p)]$ denote the probability of p percolation on T. The conditional probability $\mathbf{P}_{\mathbf{T}}[H(p)]$ is measurable with respect to $\mathbf{T}$ and we may define $\theta_{\mathbf{T}}(p) := \mathbf{P}_{\mathbf{T}}[H(p)]$. Furthermore, we may define $\theta(p) = \mathbf{P}[H(p)] = \mathbf{E}\theta_{\mathbf{T}}(p)$. Since $p_c = 1/\mu$ is the critical percolation parameter for a.e. $\mathbf{T}$, note that $\theta_{\mathbf{T}}(p) = 0$ for all $p \in [0, p_c]$.

Define invasion percolation on an arbitrary tree T as follows. Start with $I_0^T = \mathbf{0}$ where we recall that $\mathbf{0}$ is the root of T. Inductively define I_{n+1}^T to consist of I_n^T along with the vertex of minimal weight U_v adjacent to I_n^T. The invasion percolation cluster is defined as $I^T := \bigcup_n I_n^T$. Note that I^T is measurable with respect to the U variables. Let $I := I^{\mathbf{T}}$ denote the invasion cluster of the random tree $\mathbf{T}$. By independence of the U variables and $\mathcal{T}$, the conditional distribution of I given $\mathbf{T}$ agrees with that of invasion percolation.

Proposition 4.2.1. *For any $p > p_c$, I almost surely reaches some vertex v such that $v \leftrightarrow_p \infty$ in $\mathbf{T}(v)$.*

Proof. We consider the following coupling that generates I at the same time as $\mathbf{T}$: begin with the root, and generate children according to Z, giving each new edge a $(0,1)$ weight uniformly and independently. We denote this height 1 weighted tree as L_1. The sequence of weighted trees $\{L_n\}$ is now defined inductively as follows: for each $n \geq 1$, L_{n+1} is obtained by assigning Z children (using an independent copy of Z) to the boundary vertex of L_n with the smallest corresponding edge weight, and then giving each of the new edges a $(0,1)$ weight uniformly and independently.

For each $n \geq 1$, define $\mathcal{F}_n$ to be the Borel σ-field inside of $\mathcal{F}$ that is generated by L_n. Next, we define the increasing sequence of stopping times $N_1, N_2, \ldots$ in the following way: set N_1 equal to the minimum value of n such that all edges connected to boundary vertices of L_n have weight at least p, and if no such value exists, set $N_1 = \infty$. For $j \geq 1$, set N_{j+1} equal to infinity if either $N_j = \infty$, or there is no $n > N_j$ such that all edges connected to boundary vertices of L_n have weight at least p, and otherwise set N_{j+1} equal to the minimum $n > N_j$ for which this last condition is satisfied. Observing that $\{N_1 = \infty\}$ is simply the event that all edges of the invasion cluster I have weight less than p, we see that $\mathbf{P}(N_1 = \infty) = \theta(p)$. In addition, since no edges in L_{N_j} are considered until time N_{j+1}, we also find that for every j, k with $1 \leq j \leq k < \infty$, the random variable $N_{j+1} - N_j$ is independent of $\mathcal{F}_k$ with respect to the probability measure $\mathbf{P}(\cdot | N_j = k)$. Finally, noting that $(N_{j+1} - N_j | N_j = k)$

has the same distribution as N_1, we find that $\mathbf{P}(N_{j+1} = \infty | N_j < \infty) = \theta(p)$.

Now define $A_p \in \mathcal{F}$ to be the event that I eventually invades a vertex with corresponding edge weight less than p. Since having a j for which $N_j = \infty$ implies A_p, we can now conclude from the above observations that

$$\mathbf{P}(A_p) = \mathbf{E}[\mathbf{P}(A_p|\mathbf{T})] \geq \sum_{j=0}^{\infty} \theta(p)\big(1 - \theta(p)\big)^j = 1 \implies \mathbf{P}_{\mathbf{T}}(A_p) = 1 \ \ \texttt{GW}\text{-a.s.},$$

thus completing the proof. □

Corollary 4.2.2. *For any $p > p_c$, the number of edges in I with weight greater than p is almost surely finite.*

This was proven for a large class of graphs by Häggström, Peres and Schonmann [HPS99], but this class doesn't cover the case of Galton-Watson trees conditioned on survival; they exploit quite a bit of symmetry that does not occur in the Galton-Watson case.

Proof. Let x be the first invaded vertex with an infinite subtree below with weights less than p. Then after x is invaded, no edges of weight larger than p will be invaded. □

Corollary 4.2.3. *There is almost surely only one infinite non-backtracking path from $\mathbf{0}$ in I. Equivalently, $\mathbf{T}$ is almost surely in the set of trees T such that I^T contains almost surely a unique infinite non-backtracking path from $\mathbf{0}$.*

Proof. Suppose that there are two distinct paths to infinity in I; by Corollary 4.2.2, there exist maximal weights M_1 and M_2 along these paths after they split,

$\mathbf{P}$-almost surely. If $M_1 > M_2$, the second infinite path would be invaded before the edge containing M_1. Similarly, we cannot have $M_2 > M_1$. Finally, $M_1 = M_2$ has $\mathbf{P}$-probability 0, completing the proof. $\square$

This proof is stated for invasion percolation on regular trees in [AGdHS08], but is identical for Galton-Watson trees once Corollary 4.2.2 is in place; the unique path guaranteed by Corollary 4.2.3 is typically called the backbone of I, and we continue this convention. Note that a regular tree is simply a Galton-Watson tree with Z almost-surely constant.

Definition 4.2.4 (the invasion path γ)**.** Let $\gamma^T := (\mathbf{0}, \gamma_1^T, \gamma_2^T, \ldots)$ be the random sequence whose nth element is the unique v with $|v| = n$ such that $v \leftrightarrow \infty$ via a downward path in the invasion cluster I^T. Let ν_T denote the law of γ^T given T. Let $\nu_{\mathbf{T}}$ denote the random measure on the random space $(\mathbf{T}, \partial\mathbf{T})$ induced by the $\gamma^{\mathbf{T}}$. In other words, for measurable $A \subseteq \partial\mathcal{U}$, $\nu_{\mathbf{T}}(A, \omega) = \mathbf{P}[\gamma^T \in A]$ evaluated at $T = \mathbf{T}(\omega)$. By Corollary 4.2.3, this is a well defined probability measure for almost every ω.

4.2.3 Preliminary comparison of limit-uniform and invasion measures

Our main goal is to see whether $\nu_{\mathbf{T}}$ is almost surely absolutely continuous with respect to $\mu_{\mathbf{T}}$. We give the summability criterion that establishes a sufficient condition for absolute continuity in terms of the KL-divergence of the two measures along a ray chosen from $\nu_{\mathbf{T}}$.

Definition 4.2.5 (the splits p and q at children of u, and their difference, X)**.** Let v be a vertex of $\mathbf{T}$ and let u be the parent of v. Define

$$p(v) := \mu_{\mathbf{T}}(v)/\mu_{\mathbf{T}}(u)$$

$$q(v) := \nu_{\mathbf{T}}(v)/\nu_{\mathbf{T}}(u)$$

$$X(u) := \sum_w q(w) \log[q(w)/p(w)]$$

where the sum is over all children w of u and $\nu_{\mathbf{T}}(v) = \nu_{\mathbf{T}}(\{\gamma : v \in \gamma\})$ and $\mu_{\mathbf{T}}(v)$ is defined similarly. The quantity X is known as KL-divergence. The KL-divergence $\mathcal{K}(\rho, \rho')$ is defined between any two probability measures ρ and ρ' on a finite set $\{1, \ldots, k\}$ by the formula

$$\mathcal{K}(\rho, \rho') := \sum_{i=1}^{k} \rho'(i) \log \frac{\rho'(i)}{\rho(i)} .$$

It is a measure of the difference between the two distributions. It is always non-negative but not symmetric. The following inequality shows that $\mathcal{K}$ behaves like quadratic distance away from $\rho = 0$.

Proposition 4.2.6. *Let ρ and ρ' be probability measures on the set $\{1, \ldots, k\}$ and denote $\varepsilon_i := \rho'(i)/\rho(i) - 1$. Then*

$$\mathcal{K}(\rho, \rho') \leq \sum_{i=1}^{k} \rho(i) \varepsilon_i^2 . \tag{4.2.1}$$

Proof. Define the function R on $(-1, \infty)$ by

$$R(x) := \frac{(1+x)\log(1+x) - x}{x^2}$$

if $x \neq 0$ and $R(0) := 1/2$. This makes R continuous, positive, and decreasing from 1 to 0 on $(-1, \infty)$. When $\varepsilon = \rho'/\rho - 1$, we may compute

$$\frac{\rho' \log(\rho'/\rho) - (\rho' - \rho)}{\rho} = \frac{(1+\varepsilon)\rho \log(1+\varepsilon) - \varepsilon\rho}{\rho} = \varepsilon^2 R(\varepsilon)\,.$$

Because $\sum_{i=1}^{k} \rho(i) = \sum_{i=1}^{k} \rho'(i) = 1$, we see that

$$\mathcal{K}(\rho, \rho') = \sum_{i=1}^{k} (\rho'(i) - \rho(i)) + \rho(i)\varepsilon_i^2 R(\varepsilon_i) = \sum_{i=1}^{k} \rho(i)\varepsilon_i^2 R(\varepsilon_i)$$

and the result follows from $0 < R(\varepsilon_i) < 1$. □

Applying Proposition 4.2.6 to $\rho' = q$ and $\rho = p$ gives

$$X(u) \leq \sum_{w} p(w)\varepsilon(w)^2 \tag{4.2.2}$$

where $\varepsilon(w) = \frac{q(w)}{p(w)} - 1$.

Lemma 4.2.7. *Let T be a fixed tree on which ν_T and μ_T are well defined on the Borel σ-field $\mathcal{B}$ on ∂T. If*

$$\sum_{n=1}^{\infty} \sum_{|v|=n} X(v)\nu_T(v) < \infty \tag{4.2.3}$$

then $\nu_T \ll \mu_T$.

Proof. On the measure space $(\partial T, \mathcal{B})$, define a filtration $\{\mathcal{G}_n\}$ by letting $\mathcal{G}_n$ denote the σ-field generated by the sets $\{\gamma : \gamma_n = v\}$. The Borel σ-field $\mathcal{B}$ is the increasing limit $\sigma(\bigcup_n \mathcal{G}_n)$. Let

$$M_n := \left.\frac{d\nu_T}{d\mu_T}\right|_{\mathcal{G}_n} .$$

In other words, $M_n(\gamma) = \nu_T(\gamma_n)/\mu_T(\gamma_n)$. Let $\overline{M} := \limsup_{n\to\infty} M_n$. The Radon-Nikodym martingale theorem [Dur10, Theorem 5.3.3] says that $\{M_n\}$ is a martingale with respect to $(\partial T, \mathcal{B}, \mu_T, \{\mathcal{G}_n\})$ and that $\nu_T \ll \mu_T$ is equivalent to $\nu_T(\{\overline{M} = \infty\}) = 0$. This is equivalent to $\nu_T(\{\underline{M} = 0\}) = 0$ where $\underline{M} = 1/\overline{M} = \liminf_n 1/M_n$. The sequence $\{1/M_n\}$ is a ν_T-martingale, therefore $\{\log(1/M_n)\}$ is a ν_T-supermartingale and to conclude that it ν_T-a.s. does not go to negative infinity, it suffices to show that its expectation is bounded from below.

We compute the conditional expected increment of $\log(1/M_n)$. Letting γ denote the ray $(\gamma_1, \gamma_2, \ldots)$,

$$\log \frac{1}{M_{n+1}(\gamma)} - \log \frac{1}{M_n(\gamma)} = \log \frac{\nu_T(\gamma_n)}{\mu_T(\gamma_n)} - \log \frac{\nu_T(\gamma_{n+1})}{\mu_T(\gamma_{n+1})} = -\log \frac{q(\gamma_{n+1})}{p(\gamma_{n+1})}\,.$$

Conditioning on $\mathcal{G}_n$, if $\gamma_n = u$, then the ν_T-probability of $\gamma_{n+1} = v$ is $q(v)$, whence

$$\mathbf{E}_{\nu_T}\left[\log \frac{1}{M_{n+1}} - \log \frac{1}{M_n} \,\middle|\, \mathcal{G}_n\right] = \sum_{v \text{ child of } u} -q(v) \log \frac{q(v)}{p(v)} = -X(u)\,.$$

Taking the unconditional expectation,

$$\mathbf{E}_{\nu_T}\left[\log \frac{1}{M_{n+1}} - \log \frac{1}{M_n}\right] = -\sum_{|v|=n} \nu_T(v) X(v)$$

and summing over n shows that (4.2.3) implies that $\log(1/M_n)$ has expectation bounded from below, establishing $\nu_T \ll \mu_T$. $\square$

Corollary 4.2.8. *Recall that γ denotes the invasion path on $\mathbf{T}$ and let X_n denote $X(\gamma_n)$.*

(i) If $\sum_{n=1}^{\infty} \mathbf{E} X_n < \infty$ then $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$ `GW`-almost surely.

(ii) Define the filtration $\{\mathcal{G}'_n\}$ on $(\Omega, \mathcal{F})$ by letting $\mathcal{G}'_n$ be the σ-field generated by $\mathcal{T}$ together with $\gamma_1, \ldots, \gamma_n$. Let $Y(v)$ be non-negative random variables such that $Y(v) \in \mathcal{G}'_{|v|}$ and

$$\mathbf{P}[X(\gamma_n) \neq Y(\gamma_n) \text{ infinitely often}] = 0\,.$$

Then $\sum_{n=1}^{\infty} \mathbf{E} Y_n < \infty$ implies that `GW`*-almost surely, $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$.*

Proof.

(i) Writing $\mathbf{E} X_n = \mathbf{E}\left[\mathbf{E}_{\mathbf{T}} X_n\right]$ we see that the hypothesis of (i), namely $\sum_{n=1}^{\infty} \mathbf{E} X_n < \infty$, implies $\mathbf{E} \sum_{n=1}^{\infty} \mathbf{E}_{\mathbf{T}} X_n < \infty$. This implies $\sum_{n=1}^{\infty} \mathbf{E}_{\mathbf{T}} X_n < \infty$ almost surely. A version of $\mathbf{E}_{\mathbf{T}} X_n$ is $\sum_{|v| \in \mathbf{T}_n} X(v) \nu_{\mathbf{T}}(v)$, whence (4.2.3) holds for `GW`-almost every $\mathbf{T}$, implying almost sure absolute continuity of $\mu_{\mathbf{T}}$ with respect to $\nu_{\mathbf{T}}$.

(ii) The argument used to prove Lemma 4.2.7 may be adapted as follows. Let $M_n := \left.\frac{d\nu_{\mathbf{T}}}{d\mu_{\mathbf{T}}}\right|_{\mathcal{G}'_n}$, a version of which is the function taking the value $\frac{\nu_{\mathbf{T}}(v)}{\mu_{\mathbf{T}}(v)}$ on $\{\gamma_n = v\}$. Again $\{M_n\}$ is a martingale and $\log(1/M_n)$ is a supermartingale which we need to show converges almost surely. The sequence

$$S_M := \sum_{n=1}^{M} \left(\log \frac{1}{M_{n+1}} - \log \frac{1}{M_n}\right) \mathbf{1}_{X(\gamma_n) = Y(\gamma_n)}$$

is a convergent supermartingale because its expected increments are either 0 or $-Y(\gamma_n)$; convergence of the unconditional expectations $\mathbf{E} Y(\gamma_n)$ implies almost sure convergence of the expected increments, implying almost sure convergence of the supermartingale $\{S_M\}$. The hypotheses of (ii) imply that the increments of S_M differ from the increments of $\log(1/M_n)$ finitely often almost surely, implying convergence of the supermartingale $\log(1/M_n)$ and hence $\nu_{\mathbf{T}} \ll \mu_{\mathbf{T}}$. $\square$

4.3 Survival function conditioned on the tree

This section is concerned with estimating the quenched survival function $\theta_{\mathbf{T}}(p)$. The ultimate goal will be to examine the behavior of $\theta_{\mathbf{T}}(p)$ for small $p - p_c$, as estimates on $\theta_{\mathbf{T}}(\cdot)$ will be central to step 5 of the outline. Before studying the random function $\theta_{\mathbf{T}}(p)$, we record some basic properties of the annealed function $\theta(p) = \mathbf{E}[\theta_{\mathbf{T}}(p)]$. For a more complete analysis of the function $\theta_{\mathbf{T}}(\cdot)$, see [MPR18].

4.3.1 Properties of the annealed function $\theta(p)$

We restate the necessary parts of Propositions 2.2.2 and 2.2.5.

Proposition 4.3.1. *The derivative from the right $K := \theta'(p_c)$ exists and is given by*

$$K := \frac{2}{p_c^3 \phi''(1)} . \tag{4.3.1}$$

Furthermore as, $p \downarrow p_c$, $\theta'(p) \to K$.

4.3.2 Preliminary estimates of $\theta_{\mathbf{T}}(p)$

We now move to estimating $\theta_{\mathbf{T}}(p)$, a random variable measurable with respect to $\mathcal{T}$. We first prove an upper bound on θ which gives a uniform bound on the L^q

norm of θ. Additionally, we show that conditioning on only the first n levels gives a random variable exponentially close to θ. Estimating this averaged random variable is a key element in the proof of Theorem 4.1.1, and is the content of Section 4.3.3.

The following result from [LP17] will be useful for obtaining an a.s. upper bound on $\theta_{\mathbf{T}}(p_c + \varepsilon)$.

Theorem 4.3.2 ([LP17, Theorem 5.24])**.** *For p-percolation, we have*

$$\frac{1}{\mathscr{R}(\mathbf{0} \leftrightarrow \infty) + 1} \leq \mathbf{P}_{\mathbf{T}}[\mathbf{0} \leftrightarrow \infty] \leq \frac{2}{\mathscr{R}(\mathbf{0} \leftrightarrow \infty) + 1}$$

where $\mathscr{R}(\mathbf{0} \leftrightarrow \infty)$ denotes the effective resistance from $\mathbf{0}$ to infinity when an edge connecting $\overleftarrow{u}$ to u is given resistance

$$r(e) = \frac{1-p}{p^{|u|}}.$$

□

From this, we deduce:

Proposition 4.3.3. *For any $\varepsilon \in (0, 1-p_c)$ and* GW*-almost surely,*

$$\theta_{\mathbf{T}}(p_c + \varepsilon) < \frac{2\varepsilon \overline{W}}{(1 - p_c - \varepsilon)p_c} \tag{4.3.2}$$

where $\overline{W} := \sup_n W_n(\mathbf{T})$ is almost surely finite because $\lim_{n\to\infty} W_n$ exists almost surely.

Proof. To get an upper bound on g, we need a lower bound on the resistance. For each height n, short together all nodes at this height. For every $p = p_c + \varepsilon$ this gives a lower bound of

$$
\begin{aligned}
\mathscr{R}(\mathbf{0} \leftrightarrow \infty) &\geq \sum_{n=1}^{\infty} \frac{1-p_c-\varepsilon}{Z_n(p_c+\varepsilon)^n} \\
&= (1-p_c-\varepsilon) \sum_{n=1}^{\infty} \frac{p_c^n}{W_n(p_c+\varepsilon)^n} \\
&\geq \frac{(1-p_c-\varepsilon)}{\overline{W}} \sum_{n=1}^{\infty} \frac{p_c^n}{(p_c+\varepsilon)^n} \\
&= \frac{(1-p_c-\varepsilon)p_c}{\overline{W}\varepsilon}.
\end{aligned}
$$

Using Theorem 4.3.2, we get

$$
\theta_{\mathbf{T}}(p_c+\varepsilon) \leq \frac{2}{1+\frac{(1-p_c-\varepsilon)p_c}{\overline{W}\varepsilon}} \leq \frac{2\varepsilon\overline{W}}{(1-p_c-\varepsilon)p_c}. \tag{4.3.3}
$$

$\square$

Proposition 4.3.4 (uniform L^q bound)**.** *Suppose the offspring distribution has a finite $q>1$ moment. Then for any $\delta>0$, there is a constant c_q such that for all $\varepsilon \in (0, 1-p_c-\delta)$,*

$$
\mathbf{E}\theta_{\mathbf{T}}(p_c+\varepsilon)^q \leq c_q \varepsilon^q
$$

where $c_q = c_q(\delta) > 0$.

Proof. First recall that if the offspring has a finite q-moment, then $M_q := \mathbf{E}W^q$ is finite as well. By the L^q maximal inequality (e.g., [Dur10, Theorem 5.4.3]), we have that

$$
\mathbf{E}\left[\left(\sup_{1\leq k\leq n} W_k\right)^q\right] \leq \left(\frac{q}{q-1}\right)^q \mathbf{E}W_n^q \leq \left(\frac{q}{q-1}\right)^q M_q
$$

because $\{W_n^q\}$ is a submartingale.

Note that $\left(\sup_{1\le k\le n} W_k\right)^q \uparrow \overline{W}^q$ as $n\to\infty$. By monotone convergence, this implies $\mathbf{E}[\overline{W}^q] \le (q/(q-1))^q M_q$. In particular, for any $\varepsilon < 1-p_c-\delta$, this together with Proposition 4.3.3 implies

$$\mathbf{E}[\theta_{\mathbf{T}}(p_c+\varepsilon)^q] \le \left(\frac{2\varepsilon}{(1-p_c-\varepsilon)p_c}\right)^q \mathbf{E}[\overline{W}^q] \le \left(\frac{2\varepsilon}{\delta p_c}\right)^q \left(\frac{q}{q-1}\right)^q M_q\,,$$

proving the proposition with $c_q = \left(\frac{2q}{(q-1)\delta p_c}\right)^q M_q$. $\square$

Let $\mathcal{T}_n$ denote $\sigma(\deg_v : |v|\le n)$. Because $\mathcal{T}_n \uparrow \mathcal{T}$ and θ is bounded, we know that $\mathbf{E}[\theta_{\mathbf{T}}(p)\,|\,\mathcal{T}_n] \to \theta_{\mathbf{T}}(p)$ almost surely and in L^1. It turns out that $\theta_{n,\mathbf{T}}(p) := \mathbf{E}[\theta_{\mathbf{T}}(p)\,|\,\mathcal{T}_n]$ is much easier to estimate than θ itself. Our strategy is to show this convergence is exponentially rapid, transferring the work from estimation of θ to estimation of θ_n.

Lemma 4.3.5. *For any $\delta > 0$, there are constants $c_i > 0$ such that for all $p \in (p_c, \sqrt{p_c}-\delta)$*

$$\left|\,\theta_{\mathbf{T}}(p) - \theta_{n,\mathbf{T}}(p)\,\right| \le c_1 e^{-c_2 n}$$

with probability at least $1-e^{-c_3 n}$.

Proof. Define a random set $S = S(n,p)$ to be the set of vertices $v \in \mathbf{T}_n$ such that $\mathbf{0} \leftrightarrow_p v$. Let $\pi_{\mathbf{T}}$ denote the law of the random variable S, an atomic probability measure on the subsets of the random set $\mathbf{T}_n$. Using

$$\theta_{\mathbf{T}}(p) = \mathbf{P}[H(p)\,|\,\mathcal{T}] = \mathbf{E}\left[\mathbf{P}[H(p)\,|\,\mathcal{F}'_n]\,|\,\mathcal{T}\right]$$

where $\mathcal{F}'_n$ be the σ-field generated by $\mathcal{F}_n$ and $\mathcal{T}$, we obtain the explicit representation

$$\theta_{\mathbf{T}}(p) = \sum_S \pi_{\mathbf{T}}(S) \left[1 - \left(\prod_{v \in S} (1 - \theta_{\mathbf{T}(v)}(p)) \right) \right] . \tag{4.3.4}$$

Order the vertices in $\mathbf{T}_n$ arbitrarily and define the revealed martingale $\{M_k\}$ by

$$M_k := \mathbf{E} \left[\theta_{\mathbf{T}}(p) \,|\, \mathcal{T}_n, \{\mathbf{T}(v_j) : j \leq k\} \right] \tag{4.3.5}$$

as k ranges from 0 to $|\mathbf{T}_n|$. By definition, $M_0 = \theta_{n,\mathbf{T}}$. Also, $M_{|\mathbf{T}_n|} = \theta_{\mathbf{T}}(p)$ because from $\mathcal{T}_n$ together with $\{\mathbf{T}(v) : v \in \mathbf{T}_n\}$ one can reconstruct $\mathbf{T}$. Arguing as in (4.3.4), we obtain the explicit representation

$$M_k = \sum_S \pi_{\mathbf{T}}(S) \left[1 - \prod_{v \in S_{\leq k}} (1 - \theta_{\mathbf{T}(v)}(p))(1 - \theta(p))^{|S_{>k}|} \right] \tag{4.3.6}$$

where for a given set $S \subset \mathbf{T}_n$, $S_{\leq k}$ denotes the vertices in S indexed $\leq k$ and $S_{>k}$ denotes the set indexed $> k$.

We claim the increments of $\{M_k\}$ are bounded by p^n. Indeed, (4.3.6) implies

$$|M_k - M_{k-1}| \leq \sum_{S \ni v_k} \pi_{\mathbf{T}}(S) |\theta_{\mathbf{T}(v_k)}(p) - \theta(p)| \leq \sum_{S \ni v_k} \pi_{\mathbf{T}}(S) = \mathbf{P}[\mathbf{0} \leftrightarrow_p v_k] = p^n.$$

Azuma's inequality [Azu67] implies that for any $c_1, c_2 > 0$, the bounded increments translate to an upper bound

$$\mathbf{P} \left[|\theta_{\mathbf{T}}(p) - \theta_{n,\mathbf{T}}(p)| > c_1 e^{-c_2 n} \,|\, \mathcal{T}_n \right] \leq \exp \left(- \frac{c_1^2 e^{-2c_2 n}}{2 |\mathbf{T}_n| p^{2n}} \right) . \tag{4.3.7}$$

Recall that for any $\gamma > 0$,

$$\mathbf{P}[|\mathbf{T}_n| \geq (\mu(1+\gamma))^n] = \mathbf{P}[W_n \geq (1 + \gamma^n)] \leq (1+\gamma)^{-n}$$

by Markov's inequality. Since $\mu p^2 < 1$ uniformly for $p \in [p_c, \sqrt{p_c} - \delta]$, we therefore may pick c_2 so that $e^{-2c_2 n}$ is exponentially larger in n than $|\mathbf{T}_n| p^{2n}$ with exponentially high probability. Conditioning on this event and applying (4.3.7) completes the proof. $\square$

4.3.3 Bounds on the difference between $\theta_{\mathbf{T}}(p)$ and $W\theta(p)$

For the purposes of proving Theorem 4.1.1, we will show that $\theta_{\mathbf{T}}(p_c + \varepsilon)$ is close to $W\theta(p_c + \varepsilon)$ for small $\varepsilon > 0$. For a fixed vertex v in a tree $\mathbf{T}$ define $E(v, \varepsilon)$ by

$$\theta_{\mathbf{T}(v)}(p_c + \varepsilon) = \theta(p_c + \varepsilon)\left(W(v) + E(v, \varepsilon)\right).$$

Proposition 4.3.6. *Suppose the offspring distribution of Z has $p \geq 2$ moments. Then for any δ, ℓ for which both $0 < \delta < 1$ and $0 < \ell < \frac{1}{2}$, there exist constants $C_i > 0$ so that for all ε sufficiently small*

$$|E(\mathbf{0}, \varepsilon)| \leq C_1 W \varepsilon^{1-\delta} + C_2 \varepsilon^{1-2\ell} \sum_{j=1}^{\lceil \varepsilon^{-\delta} \rceil - 1} W_j \tag{4.3.8}$$

with probability at least $1 - C_3 \varepsilon^{p\ell - \delta}$.

Proof. Let c_1, c_2, c_3 be the constants from Lemma 4.3.5, and fix $\delta > 0$. Then for $m = \lceil \varepsilon^{-\delta} \rceil$, we have

$$|\theta_{m,\mathbf{T}}(p_c + \varepsilon) - \theta_{\mathbf{T}}(p_c + \varepsilon)| < c_1 e^{-c_2/\varepsilon^{\delta}} \tag{4.3.9}$$

with probability at least $1 - e^{-c_3/\varepsilon^{\delta}}$, which implies that (4.3.9) holds for the root and all children of the root with probability at least $1 - (\mu + 1)e^{-c_3/\varepsilon^{\delta}}$. Utilizing

(4.3.9) and the fact that $\theta(p_c + \varepsilon) = \Theta(\varepsilon)$ as $\varepsilon \to 0^+$ (while also making sure to select $c_3 < c_2$) gives

$$\frac{1}{\theta(p_c+\varepsilon)} \left|\theta_{m,\mathbf{T}}(p_c+\varepsilon) - \theta_{\mathbf{T}}(p_c+\varepsilon)\right| < c_1 \frac{1}{\theta(p_c+\varepsilon)} e^{-c_2/\varepsilon^\delta} = O\left(e^{-c_3/\varepsilon^\delta}\right). \quad (4.3.10)$$

By [Dub71], there exist positive constants C_1' and c_2' so that

$$\mathbf{P}[W \leq a] \leq C_1' a^{c_2'}.$$

This implies that $C_1 e^{-c_3/\varepsilon^\delta} \leq W\varepsilon^{1-\delta}$ with probability at least $1 - Ce^{-c/\varepsilon^\delta}$ for some new constants. Thus, to show equation (4.3.8), it is sufficient to examine $\theta_{m,\mathbf{T}}(p_c + \varepsilon)$.

The Bonferroni inequalities imply that

$$\mathsf{Bon}_m^{(1)}(\mathbf{0},\varepsilon) - \mathsf{Bon}_m^{(2)}(\mathbf{0},\varepsilon) \leq \theta_{m,\mathbf{T}}(p_c+\varepsilon) \leq \mathsf{Bon}_m^{(1)}(\mathbf{0},\varepsilon)$$

where

$$\mathsf{Bon}_m^{(1)}(\mathbf{0},\varepsilon) := \left(1 + \frac{\varepsilon}{p_c}\right)^m W_m \theta(p_c+\varepsilon)$$

$$\text{and } \mathsf{Bon}_m^{(2)}(\mathbf{0},\varepsilon) := \theta(p_c+\varepsilon)^2 \sum_{\substack{u,w\in\mathbf{T}_m \\ u\neq w}} (p_c+\varepsilon)^{2m-|u\wedge w|}.$$

To bound $\theta_{m,\mathbf{T}}(p_c+\varepsilon) - \theta(p_c+\varepsilon)W$, we first bound $\frac{\mathsf{Bon}_m^{(1)}(\mathbf{0},\varepsilon)}{\theta(p_c+\varepsilon)} - W$. Write

$$\frac{\mathsf{Bon}_m^{(1)}(\mathbf{0},\varepsilon)}{\theta(p_c+\varepsilon)} - W = W\left(\left(1+\frac{\varepsilon}{p_c}\right)^m - 1\right) + [W_m - W](1+\varepsilon/p_c)^m.$$

Note first that $|(1+\varepsilon/p_c)^m - 1| \leq Cm\varepsilon/p_c$ for some $C > 0$. Recalling that $m = \lceil \varepsilon^{-\delta} \rceil$ gives a bound of $C\varepsilon^{1-\delta}$. Additionally, we have $(1+\varepsilon/p_c)^m \leq 2$ for ε sufficiently small. We now look towards $|W_m - W|$.

By [AN72, Chapter I.13], we have that

$$\mathrm{Var}\left[W_m - W \,|\, W_m\right] = \frac{W_m}{\mu^m}\left(\frac{\mathrm{Var}\left[Z\right]}{\mu^2-\mu}\right).$$

By the law of total variance, this implies that

$$\mathrm{Var}\left[W_m - W\right] = \frac{1}{\mu^m}\frac{\mathrm{Var}\left[Z\right]}{\mu^2-\mu} =: \frac{C_Z}{\mu^m}.$$

Chebyshev's inequality then gives

$$\mathbf{P}[|W_m - W| > \mu^{-m/3}] \leq C_Z \mu^{-m/3}.$$

Since $\mu^{-m/3} \leq \mu^{-\varepsilon^{-\delta}/3} \leq C_2 e^{-c_1/\varepsilon^{c_2}}$ for some positive constants C_2 and c_1, c_2, we have that

$$\frac{\left|\mathsf{Bon}_m^{(1)}(\mathbf{0},\varepsilon) - \theta(p_c+\varepsilon)W\right|}{\theta(p_c+\varepsilon)} \leq C_1 W \varepsilon^{1-\delta} + C_2 e^{-c_1/\varepsilon^{c_2}} \tag{4.3.11}$$

with probability at least $1 - C_Z\mu^{-m/3} = 1 - C_3 e^{-c_3/\varepsilon^{c_4}}$.

By computing the lower probabilities of W again, recall that there exist constants C_1' and c_2' so that

$$\mathbf{P}[W \leq a] \leq C_1' a^{c_2'}.$$

This implies that $C_2 e^{-c_1/\varepsilon^{c_2}} < C_1 W \varepsilon^{1-\delta}$ with probability at least $1 - Ce^{-c_2' c_1/\varepsilon^{c_2}}$. Relabeling constants, this means that for sufficiently small ε, we can upgrade (4.3.11) to

$$\frac{\left|\mathsf{Bon}_m^{(1)}(\mathbf{0},\varepsilon) - \theta(p_c+\varepsilon)W\right|}{\theta(p_c+\varepsilon)} \leq C_1 W \varepsilon^{1-\delta} \tag{4.3.12}$$

with probability at least $1 - e^{-c_1/\varepsilon^{c_2}}$.

The last piece is to bound $\mathsf{Bon}^{(2)}_m(\mathbf{0},\varepsilon)/\theta(p_c+\varepsilon)$. By Fubini's theorem,

$$\begin{aligned}\frac{\mathsf{Bon}^{(2)}_m(\mathbf{0},\varepsilon)}{\theta(p_c+\varepsilon)} &= \theta(p_c+\varepsilon)\sum_{\substack{u,w\in\mathbf{T}_m\\ u\neq w}}(p_c+\varepsilon)^{2m-|u\wedge w|}\\ &\leq 2\theta(p_c+\varepsilon)\sum_{j=0}^{m-1}p_c^{2m-j}\sum_{u,w:|u\wedge w|=j}1\\ &\leq 2\theta(p_c+\varepsilon)\sum_{j=0}^{m-1}p_c^j\sum_{v\in\mathbf{T}_j}\sum_{1\leq i<k}W^{(i)}_{m-j-1}(v)W^{(k)}_{m-j-1}(v)\\ &\leq \theta(p_c+\varepsilon)\sum_{j=0}^{m-1}p_c^j\sum_{v\in\mathbf{T}_j}W_{m-j}(v)^2\end{aligned}$$

where the second inequality is from the bound $(1+\frac{\varepsilon}{p_c})^{2m}\leq 2$ for sufficiently small ε.

Note that for each j the innermost sum is a sum of IID random variables. We utilize the Fuk-Nagaev inequality from [FN71] which states

$$\mathbf{P}\left[\sum_{u\in\mathbf{T}_j}[W_{m-j}(u)^2-\mathbf{E}W^2_{m-j}]>t\,\middle|\,Z_j\right]\leq C_p t^{-p/2}Z_j^{p/4}+\exp\left(-C\frac{t^2}{Z_j}\right).$$

Applying this bound for $t=\mathbf{E}W^2_{m-j}Z_j\varepsilon^{-2\ell}$ gives

$$\begin{aligned}\mathbf{P}\left[\sum_{u\in\mathbf{T}_j}[W_{m-j}(u)^2-\mathbf{E}W^2_{m-j}]>(\mathbf{E}W^2_{m-j})Z_j\varepsilon^{-2\ell}\,\middle|\,Z_j\right]&\\ \leq C'_p\varepsilon^{p\ell}(Z_j)^{-p/4}+\exp\left(-C'Z_j/\varepsilon^{4\ell}\right)&\\ \leq C''_p\varepsilon^{p\ell}&\end{aligned}$$

for some choice of $C''_p>C'_p$. By applying this bound and a union bound, we get

$$\begin{aligned}\frac{\mathsf{Bon}^{(2)}_m(\mathbf{0},\varepsilon)}{\theta(p_c+\varepsilon)} &\leq \theta(p_c+\varepsilon)(1+\varepsilon^{-2\ell})\sum_{j=0}^{m-1}\left(\mathbf{E}W^2_{m-j}\right)Z_jp_c^j\\ &\leq C\theta(p_c+\varepsilon)\varepsilon^{-2\ell}\sum_{j=0}^{m-1}W_j\end{aligned}$$

with probability at least $1 - C_p'' m\varepsilon^{p\ell}$ for some new choice of C. This means that

$$\mathbf{P}\left[\frac{\mathsf{Bon}_m^{(2)}(\mathbf{0},\varepsilon)}{\theta(p_c+\varepsilon)} > C\theta(p_c+\varepsilon)\varepsilon^{-2\ell}\sum_{j=0}^{m-1} W_j\right] \le mC_p''\varepsilon^{p\ell}.$$

Recalling that $\theta(p_c+\varepsilon) = \Theta(\varepsilon)$ now gives

$$\frac{\mathsf{Bon}_m^{(2)}(\mathbf{0},\varepsilon)}{\theta(p_c+\varepsilon)} \le C_2\varepsilon^{1-2\ell}\sum_{j=0}^{m-1} W_j$$

with probability at least $1 - C\varepsilon^{p\ell-\delta}$ for some new C. Along with equations (4.3.9) and (4.3.12), this now implies the proposition. □

From here, we extract the estimate that will be used to prove Theorem 4.1.1:

Corollary 4.3.7. *Suppose the offspring distribution of Z has $p > 1$ moments and $p_1 := \mathbf{P}[Z = 1]$. Let δ, ℓ, d be positive constants such that*

$$\alpha = 1 - 3\ell - (1+d)\delta \tag{4.3.13}$$

is greater than $\frac{1}{2}$. Then there exists a constant $C > 0$ such that for all $\varepsilon > 0$ sufficiently small

$$|E(v,\varepsilon)| \le CW(v)\varepsilon^{\alpha} \tag{4.3.14}$$

for the root and its children with probability at least $1 - C\varepsilon^{\delta'}$ for
$\delta' = \min\left\{p\ell - \delta, \frac{\log(1/p_1)}{\log(\mu)} d\delta\right\}$.

Proof. The first term in equation (4.3.8) is always eventually smaller than $W(v)\varepsilon^{\alpha}$ since the exponent on ε is larger. The final term in equation (4.3.8) can now be dealt with separately.

By [BD74, Theorems 0 and 5], if Z is in L^p, then $W_k \xrightarrow{L^p} W$, implying $\mathbf{E}[|W_k - W|^p] \leq C$ for some $C > 0$. Therefore,

$$\mathbf{P}[|W_k - W| > \varepsilon^{-\ell}] \leq C\varepsilon^{p\ell}.$$

For $m = \lceil \varepsilon^{-\delta} \rceil$, condition on Z_1, apply a union bound, and take expectation to see that

$$\sum_{k=1}^{m} W_k \leq m(\varepsilon^{-\ell} + W)$$

for the root and all of its children with probability at least $1 - C(1+\mu)\varepsilon^{p\ell-\delta}$. Applying this to the latter term in equation (4.3.8) gives

$$|E(v,\varepsilon)| \leq C_1 W(v)\varepsilon^{1-2\ell-\delta} + C_2\varepsilon^{1-3\ell-\delta}$$

with probability at least $1 - C\varepsilon^{p\ell-\delta}$.

In the case where $p_1 = 0$, the lower tails on W provided by [Dub71] show that for any $r_1, r_2 > 0$ we have $\mathbf{P}[W(v) < \varepsilon^{r_1}] = o(\varepsilon^{r_2})$, thereby showing $W(v) < \varepsilon^{r_1}$ with probability less than ε^{r_2} for ε sufficiently small. Setting $r_1 = d\delta$ and $r_2 = p\ell - \delta$ completes the proof when $p_1 = 0$.

When $p_1 > 0$, there exists a constant C so that for all $a \in (0,1)$

$$\mathbf{P}[W < a] \leq Ca^{\log(1/p_1)/\log(\mu)}.$$

This implies that for α as in (4.3.13),

$$\mathbf{P}[W(v) < \varepsilon^{1-3\ell-\delta-\alpha}] = O\left(\varepsilon^{\frac{\log(1/p_1)}{\log(\mu)}d\delta}\right). \tag{4.3.15}$$

Performing a union bound for the root and all of its children again completes the proof. $\square$

4.4 Pivot Sequence on the Backbone

Define the shift function $\sigma : \Omega \to \Omega$ by

$$(\sigma(\omega))_v := \omega_{\gamma_1 \sqcup v} \,. \tag{4.4.1}$$

Informally, σ shifts the values of random variables at nodes $\gamma_1 \sqcup v$ in $T(\gamma_1)$ back to node v; these values populate the whole Ulam tree; values of variables not in $T(\gamma_1)$ are discarded; this is a tree-indexed version of the shift for an ordinary Markov chain. The n-fold shift σ^n shifts n steps down the backbone.

The main purpose of this section will be to understand the shift function θ, and thereby understand the behavior of the pivots. While this section contains many intermediate results—a fair number of which may be of independent interest—only a handful will be directly of use in the proof of Theorem 4.1.1: the pair of Propositions 4.4.4(i) and 4.4.5 demonstrating that shifting down the backbone is the same as conditioning on the pivot being at most a certain value (this is step 2 in the outline in the introduction); also of use will be Theorem 4.4.8, which accomplishes step 3 of the outline by showing that $\beta_n^* - p_c$ approaches 0 rapidly.

Before showing the necessary Markov properties, a fair bit of notation is necessary. We begin with the definition of the dual pivots β_n^*; these variables will be central to the proof of Theorem 4.1.1, primarily due to their appearance in Proposition 4.4.4.

Definition 4.4.1 (dual trees and pivots)**.** Recall that $T(v)$ denotes the subtree from v, moved to the root. Let $T^*(v)$ denote the rooted subtree induced on all vertices $w \notin T(v)$, and let $\beta^*_{v,w}$ represent the pivot of the vertex w on $T^*(v)$, that is, the least x such that w is connected to infinity by a path with weights $\leq x$ that avoids going through v. The dual pivot β^*_v is defined to be $\min_{w<v} \beta^*_{v,w}$. In keeping with the notation for pivots, we denote $\beta^*_n := \beta^*_{\gamma_n}$.

Definition 4.4.2. We define the following σ-fields.

(i) For fixed $v \neq \mathbf{0}$, define $\mathcal{C}_v$ to be the σ-field generated by $\deg_w$ and U_w for all $w \neq v$ in $T(v)$ along with $\deg_v$. Define $\mathcal{B}^*_v$ to be the σ-field generated by all the other data: U_w and $\deg_w$ for all $w \in \mathcal{T}^*(v)$, along with U_v.

(ii) For $n \geq 1$, let $\mathcal{B}^*_n$ denote the σ-field containing γ_n and all sets of the form $\{\gamma_n = v\} \cap B$ where $B \in \mathcal{B}^*_v$. Informally, $\mathcal{B}^*_n$ is generated by γ_n and $\mathcal{B}^*_{\gamma_n}$.

(iii) Let $\mathcal{C}_n$ be the σ-field generated by $\sigma^n\omega$; in other words it contains $\deg(\gamma_n)$ and all pairs $(\deg_{\gamma_n \sqcup x}, U_{\gamma_n \sqcup x})$. It is not important, but this definition does not allow $\mathcal{C}_n$ to know the identity of γ_n.

It is elementary that $\{\mathcal{B}^*_n\}$ is a filtration, that $\mathcal{B}^*_n \cap \mathcal{C}_n$ is trivial, and that $\mathcal{B}^*_n \vee \mathcal{C}_n = \mathcal{F}$.

Definition 4.4.3. We define the following conditioned measures.

(i) For $x \in (p_c, 1)$, let $\mathbf{Q}_x := (\mathbf{P} \,|\, \beta_0 \leq x)$ denote the conditional law given $\mathbf{0} \leftrightarrow_x \infty$, in other words, $\mathbf{Q}_x[A] = \frac{\theta_A(x)}{\theta(x)}$ where

$$\theta_A(x) := \mathbf{P}[A \cap \{\beta_0 \leq x\}] .$$

(ii) Let $\mathcal{L}$ denote the law of β_0, the pivot at the root. By [Dur10, Theorem 5.1.9], one may define regular conditional distributions $\mathbf{P}_x := (\mathbf{P} \mid \beta_0 = x)$. These satisfy $\mathbf{P}_x[\beta_0 = x] = 1$ and $\int \mathbf{P}_x \, d\mathcal{L}(x) = \mathbf{P}$. Also, $\mathbf{Q}_y = (1/\theta(y)) \int \mathbf{P}_x \, d\mathcal{L}|_{[0,y]}(x)$.

A common null set for all the conditioned measures is the set where either the invasion ray is not well defined or $\beta(v) = \beta^*_v$ for some v. Equalities below are always interpreted as holding modulo this null set.

Proposition 4.4.4 (Markov property for dual pivots)**.**

(i) For any $A \in \mathcal{F}$,

$$\mathbf{P}[\sigma^n \omega \in A \mid \mathcal{B}^*_n] = \mathbf{Q}_{\beta^*_n}[A] \, .$$

(ii) More generally, if $0 < y \le 1$ then for any $A \in \mathcal{F}$,

$$\mathbf{Q}_y[\sigma^n \omega \in A \mid \mathcal{B}^*_n] = \mathbf{Q}_{\beta^*_n \wedge y}[A] \, .$$

(iii) Under $\mathbf{P}$, the sequence $\{\beta^*_n\}$ is a time homogeneous Markov chain adapted to $\mathcal{B}^*_n$ with transition kernel $p(x, S) = \mathbf{Q}_x[\beta^*_1 \wedge x \in S]$ and initial distribution δ_1.

It is immediate that $\mathbf{Q}_x \ll \mathbf{P}$ for all x. The following more quantitative statement will be useful, especially when used in conjunction with Proposition 4.4.4(i).

Proposition 4.4.5. *Let $q > 1$ and suppose that the offspring distribution has a finite q-moment. Then there exists a constant C_q such that for all $A \in \mathcal{T}$ and for all $\delta > 0$ and all $x \in (p_c, 1)$,*

$$\mathbf{P}[A] \le \delta \ \ \textit{implies} \ \ \mathbf{Q}_x[A] \le C_q \delta^{1-1/q} \, .$$

Proof. On $\mathcal{T}$, the density of $\mathbf{Q}_x$ with respect to $\mathbf{P}$ is given by

$$\frac{d\mathbf{Q}_x}{d\mathbf{P}}(T) = \frac{\theta_T(x)}{\theta(x)}\,.$$

Combining Proposition 4.3.1, which implies $\theta(x) \sim K(x-p_c)$, with Proposition 4.3.4, which shows $\int \theta_T(p_c+\varepsilon)^q \, d\mathtt{GW}(T) \leq c_q \varepsilon^q$ provided $p_c + \epsilon$ is bounded away from 1, we see that

$$\int \left| \frac{d\mathbf{Q}_x}{d\mathbf{P}}(T) \right|^q \, d\mathtt{GW}(T) \leq c'_q$$

for some constant c'_q and all $x \in (p_c, 1)$. Applying Hölder's inequality with $1/p + 1/q = 1$ then gives

$$\mathbf{Q}_x[A] = \int \mathbf{1}_A \frac{d\mathbf{Q}_x}{d\mathbf{P}} \, d\mathbf{P} \leq \left[\int \mathbf{1}_A \, d\mathbf{P} \right]^{1/p} \left[\int \left(\frac{d\mathbf{Q}_x}{d\mathbf{P}} \right)^q \, d\mathbf{P} \right]^{1/q} \leq C_q \delta^{1-1/q}$$

when $C_q = (c'_q)^{1/q}$. □

The measures $\mathbf{P}_x$ are in some sense more difficult to compute with than $\mathbf{Q}_x$ because of the conditioning on measure zero sets. Relations such as the Markov property, however, are conceptually somewhat simpler. The following statement of the Markov property generalizes what was proved in [AGdHS08, Theorem 1.2 and Proposition 3.1], with $\mathcal{B}_n^+$ representing the σ-field generated by $\mathcal{B}_n^*$ together with β_n. Note, however, that the only role Propositions 4.4.6 and 4.4.7 play in the proof of Theorem 4.1.1 is that they are utilized to prove Theorem 4.4.8. The proposition below is also of independent interest, and will be crucial for studying the forward maximal weight process in Section 4.6.

Proposition 4.4.6 (Markov property for pivots)**.** *For any $A \in \mathcal{F}$,*

$$\mathbf{P}\left[\sigma^n \omega \in A \,|\, \mathcal{B}_n^+\right] = \mathbf{P}_{\beta_n}[A]$$

on $(\Omega, \mathcal{F}, \mathbf{P}_x)$.

In fact, the pair $\{\beta_n, \beta_n^*\}$ is Markov:

Proposition 4.4.7. *The sequence $\{\beta_n, \beta_n^*\}$ is a time-homogeneous Markov chain adapted to $\{\mathcal{B}_n^+\}$ with initial distribution $\mathcal{L} \times \delta_1$. Further, if we define $h_n^* := \beta_n^* - p_c$ and $f(x) := \phi'(1-(p_c+x)\theta(p_c+x))$, then $\{h_n, h_n^*\}$ has transition probabilities given by $p(\{a,b\}, \cdot) = \nu_a \times \tilde{\nu}_{a,b}$ where*

$$\begin{aligned} \frac{d\nu_a}{dx} &= \frac{f(a)\theta'(p_c+x)}{\theta'(p_c+a)} \mathbf{1}_{x<a} + C_a \delta_a \\ \textit{and} \quad \frac{d\tilde{\nu}_{a,b}}{dx} &= -\frac{f'(x)}{f(a)} \mathbf{1}_{a<x<b} + \tilde{C}_{a,b}\delta_b \end{aligned}$$

with $C_a = f(a)(p_c + a)$ and $\tilde{C}_{a,b} = \frac{f(b)}{f(a)}$.

The decay rate of $\beta_n^* - p_c = h_n^*$ follows from analyzing this Markov chain; the following Theorem accomplishes Step 3 of the outline.

Theorem 4.4.8. *There exists $C > 0$ such that for any $t \in (1/2, 1)$, $\mathbf{P}[h_n^* > n^{-t}]$ is $O(e^{-Cn^{1-t}})$.*

4.5 Proof of Theorem 4.1.1

This section is devoted to the proof of Theorem 4.1.1. For a non-root vertex $v \in \mathbf{T}_{n+1}$ with $|v| = n+1$ and $p > p_c$, define

$$\tilde{q}(v,p) := \mathbf{Q}_p[\sigma^{-1}_{\overleftarrow{v}}\{v = \gamma_1\}]. \tag{4.5.1}$$

In words, $\tilde{q}(v,p)$ considers the tree rooted at $\overleftarrow{v}$ and finds the probability that v is in the backbone conditioned on the root having pivot at most p. We then have $q(v) = \mathbf{E}^{(n)}_{\mathbf{T}}[\tilde{q}(v,\beta^*_n)]$, where β^*_n is as defined in Definition 4.4.1 and $\mathbf{E}^{(n)}_{\mathbf{T}} := \mathbf{E}[\cdot|\mathcal{T},\gamma_n]$.

4.5.1 Comparing $\tilde{q}$ and the ratio of survival functions

The goal of this section is to accomplish step 4 of the outline. This takes the form of

Lemma 4.5.1. *Let $\{w_k\}_{k=1}^d$ be an enumeration of the children of v. Then for any $p > p_c$ and j,*

$$\left|\tilde{q}(w_j,p) - \frac{\theta_{\mathbf{T}(w_j)}(p)}{\sum_{k=1}^d \theta_{T(w_k)}(p)}\right| \le \frac{\theta_{T(v)}(p)}{1-\theta_{T(v)}(p)} \cdot \frac{\theta_{T(w_j)}(p)}{\sum_{k=1}^d \theta_{T(w_k)}(p)}. \tag{4.5.2}$$

Proof. Define

$$A_j = \tilde{q}(w_j,p) - \frac{\theta_{T(w_j)}(p)}{\sum_{k=1}^d \theta_{T(w_k)}(p)}$$

and write

$$\begin{aligned}\tilde{q}(w_j,p) &= \frac{\mathbf{P}_{\mathbf{T}}[U_{w_j} \vee \beta(w_j) \text{ is smallest } \mid \beta(v) \le p]}{\sum_{i=1}^d \mathbf{P}_{\mathbf{T}}[U_{w_i} \vee \beta(w_i) \text{ is smallest } \mid \beta(v) \le p]} \\ &= \frac{\mathbf{P}_{\mathbf{T}}[U_{w_j} \vee \beta(w_j) \text{ smallest and } U_{w_j} \vee \beta(w_j) \le p]}{\sum_{i=1}^d \mathbf{P}_{\mathbf{T}}[U_{w_i} \vee \beta(w_i) \text{ smallest and } U_{w_i} \vee \beta(w_i) \le p]}.\end{aligned}$$

For each j, we observe that

$$p \cdot \theta_{T(w_j)}(p)(1 - B_j) \le \mathbf{P_T}[U_{w_j} \vee \beta(w_j) \text{ smallest and } U_{w_j} \vee \beta(w_j) \le p]$$
$$\le p \cdot \theta_{T(w_j)}(p)$$

where $1 - B_j = \prod_{1 \le i \ne j \le d}(1 - p\theta_{T(w_i)}(p))$. The upper bound is the probability that $U_{w_j} \vee \beta(w_j) \le p$, while the lower bound is the probability that $U_{w_j} \vee \beta(w_j) \le p$, and that this does not hold for any of the siblings of w_j.

This gives the bounds

$$\frac{\theta_{T(w_j)}(p)(1 - B_j)}{\sum_{k=1}^d \theta_{T(w_k)}(p)} \le \tilde{q}(w_j, p) \le \frac{\theta_{T(w_j)}(p)}{\sum_{k=1}^d \theta_{T(w_k)}(p)(1 - B_k)}. \tag{4.5.3}$$

Sandwich bounds on the difference with survival ratios follow:

$$\frac{-B_j \theta_{T(w_j)}(p)}{\sum_{k=1}^d \theta_{T(w_k)}(p)} \le A_j \le \frac{\sum_{k=1}^d [\theta_{T(w_k)}(p)\theta_{T(w_j)}(p)B_k]}{(\sum_{k=1}^d \theta_{T(w_k)}(p))(\sum_{k=1}^d [\theta_{T(w_k)}(p)(1 - B_k)])}. \tag{4.5.4}$$

Finally, the simple bound of

$$B_k \le 1 - \prod_{i=1}^d (1 - p\theta_{T(w_i)}(p)) = \theta_{T(v)}(p)$$

allows us to rewrite equation (4.5.4) as

$$-\frac{\theta_{T(v)}(p)\theta_{T(w_j)}(p)}{\sum_{k=1}^d \theta_{T(w_k)}(p)} \le A_j \le \frac{\theta_{T(v)}(p)}{1 - \theta_{T(v)}(p)} \frac{\theta_{T(w_j)}(p)}{\sum_{k=1}^d \theta_{T(w_k)}(p)}. \tag{4.5.5}$$

$\square$

4.5.2 Completing the Argument

The main ingredients for showing that p and q are close are in place: Corollary 4.3.7 bounds the fluctuations of $\theta_{\mathbf{T}}(\cdot)$, which will allow us to complete step 5 of the outline; Lemma 4.5.1 shows that $\tilde{q}$ is close to the ratio of survival probabilities for a fixed p (step 4); and Propositions 4.4.4(i), 4.4.5 and Theorem 4.4.8 will allow us to translate bounds for a fixed p into a bound for the random variable β_n^* (steps 3 and 2 respectively). We now put these pieces together for one final bound:

Proposition 4.5.2. *Letting* $q := \frac{\log(\mu)}{\log(1/p_1)}$, *if*

$$2p^2q^2 + (3p^2 + 5p)q + (-p^2 + 11p - 4) < 0, \tag{4.5.6}$$

then there exists $M > 0$ *and* $t \in (1/2, 1)$ *such that, with probability* 1, *the set*

$$\bigcup_{n=1}^{\infty} \left\{ v \in \mathbf{T}_{n+1} : \left| \frac{q(v)}{p(v)} - 1 \right| > 3Mn^{-t}, \overleftarrow{v} = \gamma_n \right\}$$

is finite.

Proof of Theorem 4.1.1: As guaranteed by Proposition 4.5.6, let $M > 0$ and $t \in (1/2, 1)$ so that with probability 1, the set

$$\bigcup_{n=1}^{\infty} \left\{ v \in \mathbf{T}_{n+1} : \left| \frac{q(v)}{p(v)} - 1 \right| > 3Mn^{-t}, \overleftarrow{v} = \gamma_n \right\}$$

is finite. Define the event

$$A_n := \left\{ \left| \frac{q(v)}{p(v)} - 1 \right| \leq 3Mn^{-t} \text{ for all } v \in \mathbf{T}_{n+1} \text{ with } \overleftarrow{v} = \gamma_n \right\}.$$

Define $Y_n := X_n \mathbf{1}_{A_n}$; by the choice of M, t, $Y_n = X_n$ all but finitely often almost surely. Therefore, by Corollary 4.2.8, it is sufficient to show that $\sum \mathbf{E} Y_n < \infty$. By definition of A_n, Proposition 4.2.6 gives an upper bound of $Y_n \leq 9M^2 n^{-2t}$. Taking expectation and recalling $t \in (1/2, 1)$ completes the proof. □

4.6 The Forward Maximal Weight Process

This section will be devoted to describing the limiting behaviour of the process $\{\beta_n - p_c\}$. We begin by showing that $\{\beta_n\}$ is a time-homogeneous Markov chain and computing the transition probabilities.

Lemma 4.6.1.

(i) The sequence $\{\beta_n := \beta(\gamma_n)\}$ is a time-homogeneous Markov chain adapted to $\{\mathcal{B}_n^+\}$ with initial distribution $\mathcal{L}$.

(ii) Reparametrizing by letting $h_n := \beta_n - p_c$, a formula for the transition kernel of the chain $\{h_n\}$ is given in terms of the OGF ϕ by $p(a, \cdot) = \mu_a$ where

$$\frac{d\mu_a}{dx} = C_a\delta_a + \frac{\phi'\left(1 - (p_c + a)\theta(p_c + a)\right)\theta'(p_c + x)}{\theta'(p_c + a)}\mathbf{1}_{(0,a)}(x)$$

and

$$C_a = 1 - \frac{\phi'\left(1 - (p_c + a)\theta(p_c + a)\right)\theta(p_c + a)}{\theta'(p_c + a)}.$$

Theorem 4.6.2. *Let $U_0, U_1, \ldots$ be a sequence of IID random variables each uniformly distributed on $(0, 1)$, and let $M_n = \min\{U_0, U_1, \ldots, U_n\}$. For each C_1, C_2 such that $0 < C_1 < p_c < C_2$, the process $\{h_n\}$ can be coupled with the process $\{M_n\}$ so that, with probability 1, h_n eventually (meaning for all sufficiently large n) satisfies $C_1 \cdot M_n \le h_n \le C_2 \cdot M_n$.*

This coupling is enough to prove convergence on the level of paths. Let $\mathcal{P}$ be an intensity 1 Poisson point process on the upper-half-plane; define the *Poisson lower* *envelope process* by

$$L(t) := \min\{y > 0 : (x, y) \in \mathcal{P} \text{ for some } x \in [0, t]\}.$$

Then we have

Corollary 4.6.3. *For any $\varepsilon > 0$ as $k \to \infty$,*

$$(kh_{\lceil kt \rceil}/p_c)_{t \geq \varepsilon} \overset{*}{\Longrightarrow} (L(t))_{t \geq \varepsilon} \tag{4.6.1}$$

where $\overset{*}{\Longrightarrow}$ denotes convergence in distribution of càdlàg paths in the Skorohod space $D[\varepsilon, \infty)$.

Corollary 4.6.4. *The sequence $n \cdot h_n$ converges in distribution to $p_c \cdot \exp(1)$, where $\exp(1)$ is an exponential random variable with mean 1.*

Bibliography

[ABF13] L. Addario-Berry and K. Ford. Poisson-Dirichlet branching random walks. *Ann. Appl. Prob.*, 23:283–307, 2013.

[ABGK12] Louigi Addario-Berry, Simon Griffiths, and Ross J. Kang. Invasion percolation on the Poisson-weighted infinite tree. *Ann. Appl. Probab.*, 22(3):931–970, 2012.

[AGdHS08] O. Angel, J. Goodman, F. den Hollander, and G. Slade. Invasion percolation on regular trees. *Ann. Probab.*, 36:420–466, 2008.

[AGM13] Omer Angel, Jesse Goodman, and Mathieu Merle. Scaling limit of the invasion percolation cluster on a regular tree. *Ann. Probab.*, 41(1):229–261, 2013.

[AN72] K. Athreya and P. Ney. *Branching Processes.* Springer-Verlag, New York, 1972.

[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tôhoku Math. J. (2)*, 19:357–367, 1967.

[BD74] N. Bingham and R. Doney. Asymptotic properties of supercritical branching processes I: the Galton-Watson process. *Adv. Appl. Prob.*, 6:711–731, 1974.

[Bil95] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[BK06] Martin T Barlow and Takashi Kumagai. Random walk on the incipient infinite cluster on trees. *Illinois Journal of Mathematics*, 50(1-4):33–65, 2006.

[CCN85] J. T. Chayes, L. Chayes, and C. M. Newman. The stochastic geometry of invasion percolation. *Comm. Math. Phys.*, 101(3):383–407, 1985.

[CCN87] J. T. Chayes, L. Chayes, and C. M. Newman. Bernoulli percolation above threshold: an invasion percolation analysis. *Ann. Probab.*, 15(4):1272–1287, 1987.

[Che78] Louis H.Y. Chen. A short note on the conditional Borel-Cantelli lemma. *The Annals of Probability*, 6(4):699–700, 1978.

[Che09] Christophe Chesneau. A tail bound for sums of independent random variables and application to the Pareto distribution. *Appl. Math. E-Notes*, 9:300–306, 2009.

[CKLW82] Richard Chandler, Joel Koplik, Kenneth Lerman, and Jorge F Willemsen. Capillary displacement and percolation in porous media. *Journal of Fluid Mechanics*, 119:249–267, 1982.

[CT97] Yuan Shih Chow and Henry Teicher. *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, third edition, 1997. Independence, interchangeability, martingales.

[DCT16] Hugo Duminil-Copin and Vincent Tassion. A new proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model. *Comm. Math. Phys.*, 343(2):725–745, 2016.

[dGG78] P.G. de Gennes and E. Guyon. Lois générales pout l'injection d'un fluide dans un milieu poreux aléatoire. *J. de mécanique*, 39, 1978.

[DHS13] Michael Damron, Jack Hanson, and Philippe Sosoe. Subdiffusivity of random walk on the 2D invasion percolation cluster. *Stochastic Process. Appl.*, 123(9):3588–3621, 2013.

[DHS18] Michael Damron, Jack Hanson, and Philippe Sosoe. Arm events in two-dimensional invasion percolation. *J. Stat. Phys.*, 173(5):1321–1352, 2018.

[DS11] Michael Damron and Artëm Sapozhnikov. Outlets of 2D invasion percolation and multiple-armed incipient infinite clusters. *Probab. Theory Related Fields*, 150(1-2):257–294, 2011.

[DS12] Michael Damron and Artëm Sapozhnikov. Limit theorems for 2D invasion percolation. *Ann. Probab.*, 40(3):893–920, 2012.

[DSV09] M. Damron, A. Sapozhnikov, and B. Vágvölgyi. Relations between invasion percolation and critical percolation in two dimensions. *Ann. Probab.*, 37:2297–2331, 2009.

[Dub71] Serge Dubuc. Problèmes relatifs à l'itération de fonctions suggérés par les processus en cascade. *Ann. Inst. Fourier (Grenoble)*, 21(1):171–251, 1971.

[Dur10] R. Durrett. *Probability: Theory and Examples.* Duxbury Press, New York, NY, fourth edition, 2010.

[FN71] D. H. Fuk and S. V. Nagaev. Probabilistic inequalities for sums of independent random variables. *Teor. Verojatnost. i Primenen.*, 16:660–675, 1971.

[FvdH17] Robert Fitzner and Remco van der Hofstad. Mean-field behavior for nearest-neighbor percolation in $d > 10$. *Electron. J. Probab.*, 22:Paper No. 43, 65, 2017.

[GBGL08] Timothy Gowers, June Barrow-Green, and Imre Leader, editors. *The Princeton companion to mathematics.* Princeton University Press, Princeton, NJ, 2008.

[GPS18] Christophe Garban, Gábor Pete, and Oded Schramm. The scaling limits of the minimal spanning tree and invasion percolation in the plane. *Ann. Probab.*, 46(6):3501–3557, 2018.

[Gri99] G. Grimmett. *Percolation*, volume 321 of *Grundlehren der mathematischen Wissenschaften*. Springer, New York, second edition, 1999.

[Har60] T. Harris. A lower bound for the percolation probability in a certain percolation process. *Proc. Camb. Phil. Soc.*, 56:13–20, 1960.

[Har63] T. Harris. *The theory of branching processes*. Springer, Berlin, 1963.

[HPS99] O. Häggström, Y. Peres, and R. Schonmann. Percolation on transitive graphs as a coalescent process: relentless merging followed by simultaneous uniqueness. In *Perplexing problems in probability*, volume 44, pages 69–90. Birkhäuser, Boston, 1999.

[HS94] Takashi Hara and Gordon Slade. Mean-field behaviour and the lace expansion. In *Probability and phase transition (Cambridge, 1993)*, volume 420 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 87–122. Kluwer Acad. Publ., Dordrecht, 1994.

[Jár03] A. Járai. Invasion percolation and the incipient infinite cluster in 2d. *Comm. Math.Phys.*, 236:311–334, 2003.

[JS89] M. Jerrum and A. Sinclair. Approximate counting, uniform generation and rapidly mixing markov chains. *Inform. and Comput.*, 82:93–133, 1989.

[Kes80] Harry Kesten. The critical probability of bond percolation on the square lattice equals 1/2. *Communications in mathematical physics*, 74(1):41–59, 1980.

[Kes86a] H. Kesten. The incipient infinite cluster in two-dimensional percolation. *Prob Theory Rel. Fields*, 73:369–394, 1986.

[Kes86b] Harry Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist*, 22(4):425–487, 1986.

[KNS66] H. Kesten, P. Ney, and F. Spitzer. The Galton-Watson process with mean one and finite variance. *Teor. Verojatnost. i Primenen.*, 11:579–611, 1966.

[Kol38] A.N. Kolmorogov. On the solution of a problem in biology. *Izv. NII Mat. Mekh. Tomsk. Univ*, 2:7–12, 1938.

[KZ87] Harry Kesten and Yu Zhang. Strict inequalities for some critical exponents in two-dimensional percolation. *J. Statist. Phys.*, 46(5-6):1031–1055, 1987.

[LB80] R. Lenormand and S. Bories. Description d'un mécanisme de connexion de liaison destiné à l'étude du drainage avec piégeage en milieu poreux. *C. R. Acad. Sci.*, 291:279, 1980.

[Len85] Roland Lenormand. Capillary and viscous fingering in an etched network. In *Physics of Finely Divided Matter*, pages 289–294. Springer, 1985.

[LP17] R. Lyons and Y. Peres. *Probability on trees and networks.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2017.

[LPP95] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23(3):1125–1138, 1995.

[LPP96] R. Lyons, R. Pemantle, and Y. Peres. Biased random walks on galton-watson trees. *Prob. Th. Rel. Fields*, 106:249–264, 1996.

[LSD77] RG Larson, LE Scriven, and HT Davis. Percolation theory of residual phases in porous media. *Nature*, 268(5619):409, 1977.

[Lyo90] R. Lyons. Random walks and percolation on trees. *Ann. Probab.*, 18:931–958, 1990.

[LZS83] Roland Lenormand, Cesar Zarcone, and A Sarr. Mechanisms of the displacement of one fluid by another in a network of capillary ducts. *Journal of Fluid Mechanics*, 135:337–353, 1983.

[McC87] J. F. McCarthy. Invasion percolation on a random lattice. *J. Phys. A*, 20(11):3465–3469, 1987.

[Men86] M. V. Menshikov. Coincidence of critical points in percolation problems. *Dokl. Akad. Nauk SSSR*, 288(6):1308–1311, 1986.

[Mic19] Marcus Michelen. Critical percolation and the incipient infinite cluster on Galton-Watson trees. *Electronic Communications in Probability*, 24, 2019.

[MMS86] M. V. Menshikov, S. A. Molchanov, and A. F. Sidorenko. Percolation theory and some applications. In *Probability theory. Mathematical statistics. Theoretical cybernetics, Vol. 24 (Russian)*, Itogi Nauki i Tekhniki, pages 53–110, i. Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform., Moscow, 1986. Translated in J. Soviet Math. **42** (1988), no. 4, 1766–1810.

[MPR17] Marcus Michelen, Robin Pemantle, and Josh Rosenberg. Invasion percolation on Galton-Watson trees. *arXiv preprint arXiv:1711.10672*, 2017.

[MPR18] Marcus Michelen, Robin Pemantle, and Josh Rosenberg. Quenched survival of Bernoulli percolation on Galton-Watson trees. *arXiv preprint arXiv:1805.03693*, 2018.

[NW83] Bernie Nickel and David Wilkinson. Invasion percolation on the Cayley tree: exact solution of a modified percolation model. *Phys. Rev. Lett.*, 51(2):71–74, 1983.

[RS00] D. Randall and A. Sinclair. Self-testing algorithms for self-avoiding walks. *Journal of Mathematical Physics*, 41:1570–1584, 2000.

[Rus81] Lucio Russo. On the critical percolation probabilities. *Z. Wahrsch. Verw. Gebiete*, 56(2):229–237, 1981.

[Sap11] Artëm Sapozhnikov. The incipient infinite cluster does not stochastically dominate the invasion percolation cluster in two dimensions. *Electron. Commun. Probab.*, 16:775–780, 2011.

[vdBJV07] Jacob van den Berg, Antal A. Járai, and Bálint Vágvölgyi. The size of a pond in 2D invasion percolation. *Electron. Comm. Probab.*, 12:411–420, 2007.

[WW83] David Wilkinson and Jorge F. Willemsen. Invasion percolation: a new form of percolation theory. *J. Phys. A*, 16(14):3365–3376, 1983.

[Yag47] A. M. Yaglom. Certain limit theorems of the theory of branching random processes. *Doklady Akad. Nauk SSSR (N.S.)*, 56:795–798, 1947.

[Zha95] Y. Zhang. The fractal volume of the two-dimensional invasion percolation cluster. *Comm. Math.Phys.*, 167:237–254, 1995.