Probab. Theory Relat. Fields 116, 57–88 (2000) © Springer-Verlag 2000

Irene Hueter · Steven P. Lalley

Anisotropic branching random walks on homogeneous trees

Received: 30 June 1998 / Revised version: 10 March 1999

Abstract. Symmetric branching random walk on a homogeneous tree exhibits a *weak survival phase*: For parameter values in a certain interval, the population survives forever with positive probability, but, with probability one, eventually vacates every finite subset of the tree. In this phase, particle trails must converge to the *geometric boundary* Ω of the tree. The random subset Λ of the boundary consisting of all ends of the tree in which the population survives, called the *limit set* of the process, is shown to have Hausdorff dimension no larger than one half the Hausdorff dimension of the entire geometric boundary. Moreover, there is *strict inequality* at the phase separation point between weak and strong survival *except* when the branching random walk is *isotropic*. It is further shown that in all cases there is a distinguished probability measure μ supported by Ω such that the Hausdorff dimension of $\Lambda \cap \Omega_\mu$, where Ω_μ is the set of μ-*generic points* of Ω, converges to one half the Hausdorff dimension of Ω_μ at the phase separation point. Exact formulas are obtained for the Hausdorff dimensions of Λ and $\Lambda \cap \Omega_\mu$, and it is shown that the log Hausdorff dimension of Λ has critical exponent $1/2$ at the phase separation point.

1. Introduction

1.1. Background: weak survival/strong survival transition

Certain stochastic growth processes (e.g., branching random walks, contact and percolation processes) in spaces with hyperbolic geometries (e.g., the homogeneous tree of degree $d \geq 3$, the Poincaré plane, Fuchsian groups) exhibit a phase not present in the corresponding processes in spaces with Euclidean geometry. This is the *weak survival phase*, in which the "population" survives forever with positive probability, but with probability 1 eventually vacates every compact subset of the ambient space. Instances of processes for which this phase is known to exist include branching Brownian motion in the Poincaré plane [9], the isotropic contact

I. Hueter: University of Florida, Department of Mathematics, Gainesville, FL 32611, USA.
e-mail: `hueter@math.ufl.edu`

S.P. Lalley: University of Chicago, Department of Statistics, Chicago, IL, USA.
e-mail: `lalley@stat.purdue.edu`

Second author supported by NSF Grant DMS-9626590.

Mathematics Subject Classification (1991): Primary 60J15, 60J80; Secondary 60K35

Key words and phrases: Anisotropic branching random walk – Hausdorff dimension – Homogeneous tree – Weak survival

process on a homogeneous tree [16, 13, 17], isotropic branching random walk on a homogeneous tree [14], and site percolation[1] on a co-compact Fuchsian group [8].

The transition from weak survival to strong survival is only partially understood. Interesting questions center on the *limit set* Λ, defined to be the random subset of the *natural boundary* Ω consisting of those points of the boundary to which particle trajectories converge. (Spaces with hyperbolic geometry have nontrivial natural boundaries: for a tree, the natural boundary is the space of ends; for the Poincaré plane the boundary is the circle at infinity). It is known for hyperbolic branching Brownian motion [9], isotropic branching random walk on a homogeneous tree [14], and for the isotropic contact process on a homogeneous tree [10] that in the weak survival phase the Hausdorff dimension $\delta_H(\Lambda)$ of the limit set cannot exceed 1/2 the Hausdorff dimension $\delta_H(\Omega)$ of the natural boundary. For hyperbolic branching Brownian motion and isotropic branching random walk, explicit formulas can be given for $\delta_H(\Lambda)$: these show that at the critical point (i.e., the point of transition between weak and strong survival) the Hausdorff dimension of the limit set *equals* 1/2 the Hausdorff dimension of the natural boundary, and furthermore that $\delta_H(\Lambda)$ exhibits a square-root singularity, i.e., that as a function of the *growth parameter* λ (the fission rate for branching processes, the infection rate for contact processes),

$$\frac{1}{2}\delta_H(\Omega) - \delta_H(\Lambda) \sim C\sqrt{\lambda_c - \lambda} \tag{1}$$

for a suitable positive constant C (here λ_c is the critical value of λ).

1.2. *Anisotropic branching random walk*

The purpose of this paper is to investigate the extent to which *isotropy* affects the nature of the transition from weak to strong survival. We shall restrict our attention to anisotropic, symmetric, nearest neighbor branching random walk on a homogeneous tree of even degree $2d$ (with $d > 1$), a process for which exact calculations can be performed. The methods developed apply also, with minor modifications, to anisotropic, symmetric, nearest neighbor branching random walks on homogeneous trees of odd degrees $2d-1$; for ease of exposition, we shall discuss only the case of even degree.

1.2.1. *The tree as a Cayley graph*

The definition of anisotropic branching random walk relies on the representation of the homogeneous tree $\mathscr{T} = \mathscr{T}_{2d}$ of even degree as the Cayley graph of the free group $\mathscr{G} = \mathscr{G}_d$ on d generators. There is a similar representation of the homogeneous tree $\mathscr{T}_d$ of any degree $d \geq 3$ as the Cayley graph of the free product $\mathbb{Z}_2 \star \mathbb{Z}_2 \star \cdots \star \mathbb{Z}_2$ of d copies of the two element group $\mathbb{Z}_2$; this representation can be used to define a somewhat different class of branching random walks, to which our techniques may also be applied.

[1] For percolation, the analogue of the weak survival phase is the phase in which infinitely many infinite connected clusters co-exist.

Let $\mathscr{A}_+ = \{a_1, a_2, \dots, a_d\}$ be a set of symbols of cardinality d, define $\mathscr{A}_-$ to be the set of formal inverses $\{a_1^{-1}, a_2^{-1}, \dots, a_d^{-1}\}$ of the symbols in $\mathscr{A}_+$, and set $\mathscr{A} = \mathscr{A}_+ \cup \mathscr{A}_-$. The free group $\mathscr{G}$ with generators $\mathscr{A}_+$ is the set of finite reduced words from the alphabet $\mathscr{A}$ (a word is *reduced* if no letter $a \in \mathscr{A}$ is adjacent to its inverse); multiplication consists of concatenation followed by reduction, and the group identity 1 is the empty word. There is a natural bijection between $\mathscr{G}$ and the set of vertices of $\mathscr{T}$, in which $g, h \in \mathscr{G}$ are mapped to adjacent vertices of $\mathscr{T}$ if and only if $gh^{-1} \in \mathscr{A}$. Thus, vertices are uniquely "represented" by finite reduced words from $\mathscr{A}$. Henceforth, we shall not be careful to distinguish between vertices of $\mathscr{T}$ and the words (or group elements) to which they correspond, and we shall refer to $\mathscr{G}$ as the vertex set of $\mathscr{T}$. For any vertex w, denote by $|w|$ the length of its representative word; note that $|w|$ is also the distance from the root vertex 1 to vertex w in the graph $\mathscr{T}$.

The bijection between $\mathscr{G}$ and $\mathscr{T}$ extends in a canonical way to a bijection between the natural boundary of $\mathscr{T}$ and the set Ω of semi-infinite reduced words from the alphabet $\mathscr{A}$. If $\omega = x_1x_2 \dots \in \Omega$ then ω corresponds to the *end*[2] of $\mathscr{T}$ represented by the semi-infinite geodesic that passes through the vertices $1, x_1, x_1x_2, \dots$ in succession. For each real number $\alpha \in (0, 1)$ there is a natural metric d_α on Ω, defined by

$$d_\alpha(\omega, \omega') = \alpha^{N(\omega,\omega')} \tag{2}$$

where $N(\omega, \omega')$ is the maximum integer n such that the sequences ω and ω' agree in entries $1, 2, \dots, n$. The corresponding topology on Ω is, for any choice of α, the topology of coordinatewise convergence.

For any vertex w of $\mathscr{T}$, define $\mathscr{T}(w)$ to be the set of vertices v such that the geodesic segment from 1 to v contains w, equivalently, such that the unique word representing w is a prefix of the word representing v. Similarly, define $\Omega(w)$ to be the set of infinite reduced words $\omega = x_1x_2 \dots$ such that for some finite n, w is represented by the word $x_1x_2 \dots x_n$. Note that for every integer $n \geq 1$, the set $\{\Omega(w) : |w| = n\}$ is a finite open cover of Ω.

1.2.2. Random walk on the free group

Let $\mathscr{P} = \{p_g\}$ be a probability distribution on the group $\mathscr{G}$, and let $\xi_1, \xi_2, \dots$ be independent, identically distributed $\mathscr{G}$-valued random variables with common distribution $\mathscr{P}$. A *random walk* with step distribution $\mathscr{P}$ and initial position $x \in \mathscr{G}$ is a sequence of $\mathscr{G}$-valued random variables $Y_n, n \geq 0$ such that $Y_0 = x$ and for $n \geq 1$,

$$Y_n = x\xi_1\xi_2 \dots \xi_n \quad . \tag{3}$$

[2] An *end* of $\mathscr{T}$ is an equivalence class of semi-infinite geodesics, where two geodesics are equivalent if and only if the sets of vertices through which they pass differ in only finitely many vertices. A *geodesic* in $\mathscr{T}$ is a finite or semi-infinite sequence of distinct vertices $v_1, v_2, \dots$ such that for every $n \geq 1$, the vertices v_n and v_{n+1} are nearest neighbors.

If the initial position is not specified, it is assumed to be the root 1. We shall restrict attention to distributions $\mathscr{P}$ that are *nearest neighbor*, *symmetric*, *nondegenerate*, and *aperiodic*; in particular, we assume that

$$p_a > 0 \quad \text{if and only if } a \in \mathscr{A} \cup \{1\} \ , \tag{4}$$

$$p_a \ = \ p_{a^{-1}} \quad \text{for each } a \in \mathscr{A}_+ \ . \tag{5}$$

If all the probabilities $p_a, a \in \mathscr{A}$, are identical, the distribution $\mathscr{P}$ (and the corresponding random walk) is called *isotropic*; otherwise, $\mathscr{P}$ is *anisotropic*. Because $d \geq 2$, any random walk with nondegenerate nearest neighbor distribution is necessarily transient; in fact [4], if $p_1 > 0$ then there exist constants $C_x > 0$, $x \in \mathscr{G}$, and $1 < R < \infty$ such that for each fixed x,

$$P\{Y_n = x\} \sim \frac{C_x}{R^n n^{3/2}} \tag{6}$$

as $n \to \infty$. The exponential rate R is called the *spectral radius* of the random walk; it will also prove to be the critical fission rate for the branching random walk defined below. It is the positivity of $R - 1$ that accounts for the existence of a weak survival phase in the branching random walk (see [2] for more along these lines).

A crucial role in the results below will be played by a collection of generating functions $F_x(z)$ indexed by the vertices $x \neq 1$ of $\mathscr{G}$. These are defined as follows: for each $x \neq 1$, let T_x be the time of first passage to x by the random walk Y_n ($T_x = \infty$ on the event that Y_n never visits x); for $|z| \leq 1$ define

$$F_x(z) = E z^{T_x} I\{T_x < \infty\}; \tag{7}$$

and for $|z| > 1$ define $F_x(z)$ to be the direct analytic continuation of the function defined by (7). It will be seen below (section 3) that all of the functions F_x have radius of convergence R, and that $F_x(R) < \infty$. Furthermore, the functions F_x are *algebraic*: for each x, $w = F_x(z)$ satisfies a polynomial equation $Q_x(w, z) = 0$, and there is an algorithm for producing the polynomial Q_x. In the special case $d = 2$, the polynomials $Q_a(w, z), a \in \mathscr{A}$, are of degree 4, and so explicit expressions (involving iterated radicals) can be given for the functions F_a (see section 3 below).

1.2.3. Branching random walk on $\mathscr{G}$

This is the process of primary interest here. It evolves in discrete time $n = 0, 1, 2, \ldots$. At time $n = 0$ there is a single particle, located at a vertex $x \in \mathscr{G}$. (If the initial position x is not specified, it is assumed to be 1.) The change in the state of the population between times n and $n + 1$ takes place in two stages: (1) particle reproduction and (2) particle dispersal. Particles live for one unit of time. At the end of its lifetime, a particle ζ first fissions, creating a random number $N_\zeta \geq 1$ of offspring particles, all located at the same vertex as was ζ before the fission; for each particle ζ the random variable N_ζ has a geometric distribution with mean $\lambda > 1$, i.e.,

$$P\{N_\zeta \geq k\} = (1 - \lambda^{-1})^{k-1}, \qquad k = 1, 2, \ldots \ .$$

After the fissions, all particles move randomly to new vertices according to the step distribution $\mathscr{P}$. All random choices described in this evolution are assumed to be mutually independent. The random walk $\{Y_n\}_{n\geq 0}$ on $\mathscr{G}$ with step distribution $\mathscr{P}$ will be called the *marginal* random walk associated with the branching random walk.

The state of the population at any time n is determined by the number Z_n of particles in existence (before the next fission stage), and the unordered set $\{X_1^{(n)}, X_2^{(n)}, \dots, X_{Z_n}^{(n)}\}$ of particle locations. Observe that Z_n is a supercritical Galton–Watson process, so $\lim Z_n/\lambda^n$ exists and (since the geometric distribution has finite variance) is strictly positive. Conditional on the value of Z_n, the distribution of $X_1^{(n)}$ is the same as that of Y_n; consequently, for any vertex x the expected number of particles located at x at time n is $EZ_n P\{Y_n = x\}$, which by (6) is asymptotic to $C_x(\lambda/R)^n/n^{3/2}$. It follows that the process survives weakly if $\lambda \leq R$, and survives strongly if $\lambda > R$.

1.3. Main results: the limit set Λ

Henceforth we shall consider only the weak survival phase

$$1 < \lambda \leq R \ . \tag{8}$$

In the weak survival phase, the population eventually vacates every finite subset of $\mathscr{G}$. It follows that the population has a well-defined set Λ of accumulation points in Ω: Λ is the set of sequences $\omega = x_1x_2\dots \in \Omega$ such that for every (finite) prefix $x = x_1x_2\dots x_m$ of ω, eventually there are particles located at vertices in $\mathscr{T}(x)$. It is easily seen that Λ is a closed, and therefore compact, subset of Ω.

Theorem 1. *With probability* 1, *the Hausdorff dimension* $\delta(\lambda)$ *of* Λ (*relative to the metric* d_α) *is given by*

$$\delta(\lambda) = -\frac{\log\theta(\lambda)}{\log\alpha} \tag{9}$$

where $\theta(\lambda)$ *is the unique positive number such that*

$$\sum_{a\in\mathscr{A}} \frac{F_a(\lambda)}{\theta(\lambda) + F_a(\lambda)} = 1 \ . \tag{10}$$

The function $\delta(\lambda)$ *is strictly increasing in* λ, *and has critical exponent* $\frac{1}{2}$ *at the phase separation point* $\lambda = R$; *in particular, there is a constant* $C > 0$ *such that as* $\lambda \to R-$,

$$\delta(R) - \delta(\lambda) \sim C\sqrt{R-\lambda} \ . \tag{11}$$

Furthermore,

$$\delta(\lambda) \leq \frac{1}{2}\delta_H(\Omega) \ , \tag{12}$$

with equality holding if and only if the underlying random walk is isotropic and $\lambda = R$.

It is natural to ask what happens to Λ at the critical value $\lambda = R$. For this purpose, we partition the natural boundary Ω into *measure classes* Ω_μ. Recall that Ω is the set of infinite reduced words from the alphabet $\mathscr{A}$. Let $\sigma : \Omega \to \Omega$ be the one-sided shift operator on Ω, i.e.,

$$\sigma(x_1 x_2 \ldots) = x_2 x_3 \ldots \quad .$$

For any ergodic, σ-invariant probability measure μ on Ω, define Ω_μ to be the subset of Ω consisting of all $\omega \in \Omega$ such that for every continuous real-valued function $f : \Omega \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(\sigma^i \omega) = \int_\Omega f \, d\mu \quad . \tag{13}$$

Birkhoff's ergodic theorem implies that $\mu(\Omega_\mu) = 1$ (since the space of continuous functions on Ω is separable in the sup norm topology). Moreover, if μ and ν are distinct ergodic probability measures then $\Omega_\mu \cap \Omega_\nu = \emptyset$.

For any ergodic, $\sigma-$invariant probability measure μ on Ω denote by $h(\mu)$ the Kolmogorov-Sinai entropy of the measure-preserving system (Ω, μ, σ). Define the function $\varphi_\lambda : \Omega \to \mathbb{R}$ by $\varphi_\lambda(x_1 x_2 \ldots) = \log F_{x_1}(\lambda)$.

Theorem 2. *Fix $\lambda \le R$, and let μ be any ergodic, σ-invariant probability measure on Ω. If $h(\mu) < -\int \varphi_\lambda \, d\mu$ then with probability one, $\Lambda \cap \Omega_\mu = \emptyset$. If $h(\mu) \ge -\int \varphi_\lambda \, d\mu$ then with probability one the set $\Lambda \cap \Omega_\mu$ has Hausdorff dimension $\delta(\lambda; \mu)$ (relative to the metric d_α) given by*

$$\delta(\lambda; \mu) = -\frac{h(\mu) + \int_\Omega \varphi_\lambda \, d\mu}{\log \alpha} \quad . \tag{14}$$

This Hausdorff dimension satisfies the inequality

$$\delta(\lambda; \mu) \le \frac{1}{2} \delta_H(\Omega_\mu) \quad , \tag{15}$$

and equality holds in (15) *for one and only one ergodic probability measure μ_* and only when $\lambda = R$.*

The identification of the Hausdorff dimension (14) will be deduced from a more general result of Lalley and Sellke [11] concerning labelled Galton–Watson trees (see section 6.2 below). The more interesting part of Theorem 2 is the final statement, that there exists a probability measure μ_* on Ω for which equality holds in (15). This shows that the convergence of $\delta_H(\Lambda)$ to one-half the Hausdorff dimension of the natural boundary in the isotropic case does indeed have an analogue in the anisotropic case – in particular, the transition from weak to strong survival occurs precisely when, for some Ω_μ, the set Λ fills a subset of half the Hausdorff dimension of Ω_μ. The distinguished probability measure μ_* is defined as follows:

Define the *backscattering matrix* $M_2(z)$ to be the $2d \times 2d$ matrix, indexed by elements of $\mathscr{A}$, whose entries are given by

$$
\begin{aligned}
(M_2(z))_{ij} &= F_j(z)^2 \quad \text{if } j \neq i^{-1}, \\
&= 0 \qquad\qquad \text{if } j = i^{-1} \;.
\end{aligned}
\tag{16}
$$

Note that when $0 < z \leq R$ this is an irreducible nonnegative matrix, so the Perron-Frobenius theorem applies. In Corollary 2 below it will be shown that the lead (Perron-Frobenius) eigenvalue of $M_2(R)$ is 1. Consequently, if v is the (positive) right eigenvector, then

$$
p_2(i, j) = \frac{(M_2(R))_{ij} v_j}{v_i} \tag{17}
$$

are the entries of an irreducible stochastic matrix $\mathbb{P}_2$. The probability measure μ_* is the unique probability measure on Ω such that the induced coordinate process is the stationary Markov chain with transition probability matrix $\mathbb{P}_2$.

Remarks

(A) Analogous results can be proved for branching random walks whose marginal random walk has a step distribution that is symmetric, nondegenerate, aperiodic, and of *finite support* (but not nearest neighbor). The only new phenomenon is that the distinguished measure μ_* on the boundary may no longer be *Markov*; instead it may be a more general *Gibbs state*. See [8] for a proof of the local limit theorem and related saddlepoint approximations in the finite-range case, in which these more general Gibbs states arise.

(B) It is likely that there is an analogue of Theorem 2 for symmetric, anisotropic contact processes on homogeneous trees. However, entirely new techniques would be needed for the proof, as our proof revolves around the "superposition" property of branching random walk, and the highly algebraic Proposition 3 below. Moreover, even the existence of a weak survival phase has not yet been established for the anisotropic contact process.

(C) That $\delta_H(\Lambda)$ is almost surely constant is relatively straightforward. Let δ^* be the supremum of all real numbers δ such that $P\{\delta_H(\Lambda) \geq \delta\} > 0$. At each time $n \geq 0$ the population has at least one particle; thus, with auxiliary randomization, it is possible to choose one particle ζ_n at random from the offspring of ζ_{n-1} in such a way that the choice is independent of the future evolution of the branching random walk. Each particle ζ_n begets its own branching random walk (started at time n and at the location of ζ_n), which has its own limit set Λ_n. Clearly, $\Lambda_n \subset \Lambda$; but $\delta_H(\Lambda_n)$ has the same distribution as $\delta_H(\Lambda)$, since the branching random walk descendant from particle ζ_n has the same distribution as the original branching random walk, apart from the location shift, which does not affect the Hausdorff dimension of the limit set. Fix $\varepsilon > 0$, and for each n let H_n be the event that Λ_n has Hausdorff dimension at least $\delta^* - \varepsilon$. Then the indicator functions of $H_1, H_2, \ldots$ constitute a stationary sequence, which by Kolmogorov's 0-1 Law is ergodic. Thus, by Birkhoff's ergodic theorem, at least one of the events H_n must occur, with probability 1. Hence, with probability one $\delta_H(\Lambda) \geq \sup_n \delta_H(\Lambda_n) \geq \delta^* - \varepsilon$. Since $\varepsilon > 0$ is arbitrary, it follows that $\delta_H(\Lambda) = \delta^*$ almost surely.

A similar argument shows that for every ergodic probability measure μ on Ω, the Hausdorff dimension of $\Lambda \cap \Omega_\mu$ is constant.

2. The superposition property

The key probabilistic tool needed for proving Theorems 1 and 2 is the *superposition property* of the branching random walk, an extension of the strong Markov property. (In the theory of superprocesses the analogous property, first used by Dynkin, is sometimes called the "special Markov property".) The superposition property may be described informally as follows. Begin with a modification of the branching random walk in which a certain subset K of the vertex set acts as a "sticky" barrier – any particle entering the set K is "frozen", i.e., undergoes no subsequent fission and makes no further movements. Observe that arrivals of particles in K may take place at different times, and that some particles may move and reproduce forever without entering K. Label the particles of this modified branching random walk "blue". Now to each frozen particle ζ at a vertex $x \in K$ attach an independent branching random walk with initial particle located at x (and with starting time T_ζ, the time at which ζ first entered K). Label the particles of these attached branching random walks "red".

Superposition property. *The particle system consisting of all blue and red particles is a version of the* (*original*) *branching random walk.*

We refrain from giving a formal proof, which follows the usual line of proof of the strong Markov property. Note, however, that the construction can be reversed. For each particle ζ of a branching random walk, define the *trail* of ζ to be the path in $\mathscr{G} \times \mathbb{Z}_+$ followed by ζ and its ancestors. At any time t, let ζ be *blue* if its trail has not entered K by time t; otherwise, let ζ be *red*. Then the red and blue processes have the same joint distribution as in the construction above. The Superposition Property is equivalent to the assertion that each connected red cluster (in $\mathscr{G} \times \mathbb{Z}_+$) is a branching random walk, conditionally independent of the rest of the process.

Observe that the Superposition Property applies also to the attached branching random walks. Thus, versions of the branching random walk may be constructed in stages, first letting particles run until frozen at vertices in K_1; then, vertex by vertex $x \in K_1$, "unfreezing" frozen particles at x and letting particles of their descendant branching random walks run until frozen at vertices in $K_2(x)$; etc. Such multi-stage constructions will allow us to embed various Galton–Watson chains in the branching random walk – see section 6 below.

Henceforth, we will denote by $Z_n^K(H)$ the number of (blue) particles located in a set $H \subset \mathscr{G}$ at time n for the modified branching random walk with "freezing" at K. For any vertex x set $Z_n^K(x) = Z_n^K(\{x\})$, and set $Z_+^K(H) = \lim_{n\to\infty} Z_n^K(H)$. Observe that when $K = \emptyset$, the modified branching random walk is identical to the original branching random walk; in this case, we will write $Z_n(H) = Z_n^\emptyset(H)$.

3. Generating functions

The generating functions $F_x(z)$ defined by (7) above play a central role in the determination of $\delta(\lambda)$ and its detailed behavior at the transition between the weak survival and strong survival regimes. In this section we collect the necessary information about these generating functions.

3.1. Connection with the branching random walk

First we show how the generating functions $F_x(\lambda)$ arise naturally in connection with the branching random walk with fission rate λ, and deduce a fundamental fact about mean occupation numbers. Fix a vertex $x \neq 1$, and consider the modified branching random walk in which particles are "frozen" upon reaching x. Recall that $Z_+^x(x)$ is the number of particles that are ultimately frozen at x.

Lemma 1. $EZ_+^x(x) = F_x(\lambda)\ \forall\, x \neq 1$.

Proof. $Z_+^x(x) = \sum_{n=1}^{\infty} \xi_n$, where ξ_n is the number of particles of the branching random walk that are first frozen at x at time n. The trail of any particle counted in ξ_n must reach x for the *first* time at n; consequently, ξ_n is the number of particles of the *unmodified* branching random walk whose trails reach x for the first time at n. But the trail of a randomly chosen particle (from the Z_n particles in existence at time n) has the same distribution as the first n steps of the marginal random walk Y; consequently, the probability that the trail of this randomly chosen particle first reaches x at time n is $P\{T_x = n\}$. Since the expected total number of particles of the unmodified branching random walk in existence at time n is λ^n, it follows that $E\xi_n = \lambda^n P\{T_x = n\}$. Summing over n gives the advertised identity. □

Corollary 1. *For each $\lambda \in (0, R]$ and each vertex $x \in \mathscr{G} - \{1\}$,*

$$F_x(\lambda) < 1 \ . \tag{18}$$

Proof. Consider the total occupation time $\sum_{n=0}^{\infty} Z_n(x)$ of the vertex x by particles of the branching random walk. The particles counted in $\sum_{n=0}^{\infty} Z_n(x)$ may be partitioned into two groups: (1) those whose trails are visiting x for the first time, and (2) those that descend from particles which visited x earlier. By the Superposition Property, each particle in the first group gives rise to its own offspring branching random walk, identical in law (except for a shift in time and space) to the original branching random walk; hence, the expected occupation time of x by particles of this offspring branching random walk equals the expected occupation time $G(\lambda)$ of 1 by particles of the original branching random walk. It follows by Lemma 1 that

$$E \sum_{n=0}^{\infty} Z_n(x) = F_x(\lambda) G(\lambda) \ , \tag{19}$$

where $G(\lambda) = \sum_{n \geq 0} P\{Y_n = 1\} \lambda^n$. Observe that, by (6), $G(\lambda) < \infty$ for all $\lambda \leq R$, so the left side of (19) is finite.

Consider a particle ζ whose trail reaches x for the first time at some $n \geq 1$. This particle gives rise to an offspring branching random walk started at x. Particles of this offspring branching random walk may eventually return to the root vertex 1; by the symmetry of the branching random walk and the Superposition Principle, the expected occupation time of the root vertex 1 by such particles is $E \sum_{n=0}^{\infty} Z_n(x)$. Summing over all "first arrivals" at x and taking expectations, we obtain, by (19) and Lemma 1,

$$G(\lambda) = E \sum_{n=0}^{\infty} Z_n(1) > F_x(\lambda)^2 G(\lambda)$$

(the inequality is strict because not every particle that returns to the root vertex has a trail that passes through x). Since $G(\lambda) > 0$, the inequality (18) follows. □

3.2. Basic algebraic relations

The functions F_x are interrelated by a system of algebraic equations that derive from the Markov property and the tree-structure of the state space. Recall first that each element $x \neq 1$ of $\mathscr{G}$ has a unique representation $x = i_1 i_2 \cdots i_k$ as a word in the generators. In order that $Y_T = x$ at some finite time $T = T_x$, it is necessary and sufficient that the path $\{Y_n\}_{n \leq T}$ pass through each of the intermediate vertices $i_1 i_2 \cdots i_l$, $1 \leq l \leq k$, on the geodesic from 1 to x. The elapsed times between first visits to neighboring intermediate vertices on this geodesic are conditionally independent, by the strong Markov property, and the jth elapsed time has conditional distribution identical to the distribution of T_i, where $i \in \mathscr{A}$ is the jth letter in the word representation of x. Thus,

$$F_x = \prod_{j=1}^{k} F_{i_j} \quad . \tag{20}$$

The generating functions F_i, $i \in \mathscr{A}$, satisfy a system of $2d$ algebraic equations gotten by conditioning on the first step Y_1 of the random walk. Any $i \in \mathscr{A}$ may be first visited either on the first step or on a subsequent step; in the latter case, the first step must be to some $j \in \{1\} \cup \mathscr{A} - \{i\}$, after which the random walk must revisit 1 and then subsequently visit i. The elapsed time before the first revisit to 1 after the initial step to j has the same distribution as T_j (if $j \neq 1$), by symmetry, and the subsequent elapsed time until the first visit to i has the same (conditional) distribution as T_i. Thus,

$$F_i(z) = p_i z + p_1 z F_i(z) + \sum_{j \in \mathscr{A}} p_j z F_j(z) F_i(z) - p_i z F_i(z)^2 \quad . \tag{21}$$

Note that since $F_i = F_{i^{-1}}$ and $p_i = p_{i^{-1}}$, the d equations indexed by $i \in \mathscr{A}_-$ are redundant. Consequently, when referring to the system (21) we will sometimes include only those equations indexed by $i \in \mathscr{A}_+$. With this convention, (21) is a system of d polynomial equations in $d + 1$ variables.

It is evident from equations (20) and (21) that all of the functions F_x, $x \in \mathcal{G}$, have the same radius of convergence R. It is not difficult to see that R is also the spectral radius of the random walk Y_n (see [7]). In fact, the Markov property implies that the Green's function $G(z) = \sum_{n\geq 0} P\{Y_n = 1\}z^n$, whose radius of convergence is by definition the spectral radius of the random walk Y_n, is related to the functions $F_i(z)$ by

$$G(z) = 1/(1 - p_1 z - \sum_{i\in\mathcal{A}} p_i z F_i(z)) \ . \tag{22}$$

Since $G(z)$ is finite at its radius of convergence $z = R$ (see, e.g., [4] or [7]), relation (22) implies that

$$p_1 R + \sum_{i\in\mathcal{A}} p_i R F_i(R) < 1 \ . \tag{23}$$

Moreover, since the radius of convergence R is a singularity of $G(z)$, it follows that at least one, and therefore, by (21), all of the functions F_i must have singularities at $z = R$. It is known [4], [7] that as $z \uparrow R$, $G(R) - G(z) \sim C\sqrt{R-z}$ for a positive constant C (in fact, this accounts for the Local Limit Theorem (6) – see [4]). Consequently, by (22) and (21), there exist positive constants C_i such that as $z \uparrow R$,

$$F_i(R) - F_i(z) \sim C_i\sqrt{R-z} \ . \tag{24}$$

The algebraic equations (20), (21) and (22) imply that each of the functions $F_x(z)$ and $G(z)$ is an *algebraic function*, i.e., that it satisfies a polynomial equation with coefficients in the ring $\mathbb{C}[z]$. This polynomial may be obtained by a straightforward elimination algorithm: to obtain F_i, eliminate the other variables F_j one at a time by taking *resultants* (see, e.g., [12]). For example, when $d = 2$ (in which case there are two distinct generating functions, $F_a(z)$ and $F_b(z)$) the polynomial equation satisfied by $w = F_i$, $i = a, b$, is

$$\begin{aligned} &w^4 z^3(-6p_i^3 + 3p_i^2) + w^3 z^2(8p_i^2 - 4p_i) + w^2 z^3(4p_i^3 - 10p_i^2 + 6p_i - 1) \\ &\quad + w^2 z((1 - 2p_i) + z^3(2p_i^3 - p_i^2)) = 0 \end{aligned}$$

which may be solved by radicals (the correct branch may be identified using the fact that $F_i(z)$ is a probability generating function). Since equation (10) is also polynomial, it follows that $\theta(\lambda)$ is an algebraic function. When $d = 2$, (10) may be rewritten as

$$3F_a(\lambda)F_b(\lambda) + \theta(\lambda)(F_a(\lambda) + F_b(\lambda)) - \theta(\lambda)^2 = 0 \ ,$$

and thus an explicit algebraic expression for $\theta(\lambda)$ can be obtained.

3.3. Consequences

The next result shows that the critical exponent for the Hausdorff dimension at the phase separation point is $1/2$ (see equation (11)).

Proposition 1. *There exists a constant $C > 0$ such that as $z \uparrow R$,*

$$\theta(R) - \theta(\lambda) \sim C\sqrt{R - \lambda} \ . \tag{25}$$

Proof. Recall from (10) that, for $0 < \lambda \leq R$, $\theta(\lambda)$ is the unique positive solution of the equation

$$\sum_{i \in \mathscr{A}} Q(F_i(\lambda), \theta(\lambda)) = 1 \ , \tag{26}$$

where $Q(x, y) = x/(x + y)$. The function $Q(x, y)$ is continuously differentiable in x and y, and its partial derivatives

$$\frac{\partial Q}{\partial x} = \frac{y}{(x+y)^2}$$
$$\frac{\partial Q}{\partial y} = \frac{-x}{(x+y)^2}$$

are strictly positive and strictly negative, respectively, at $(x, y) = (F_i(\lambda), \theta(\lambda))$, for every $\lambda \in (0, R]$. Consequently, application of Taylor's formula in equation (26) at $\lambda = R$ gives

$$\sum_{i \in \mathscr{A}} a_i (F_i(R) - F_i(\lambda)) + \sum_{i \in \mathscr{A}} b_i (\theta(R) - \theta(\lambda))$$
$$+ o\left(|\theta(R) - \theta(\lambda)| + \sum_{i \in \mathscr{A}} |F_i(R) - F_i(\lambda)| \right) = 0 \ ,$$

where $a_i > 0$ and $b_i < 0$. The desired relation (25) follows, in view of (24). □

The proof that strict inequality must obtain in (12) of Theorem 1 except in the case of an *isotropic* branching random walk will rest on the following.

Proposition 2. *$F_i(R) = F_j(R)$ for all $i, j \in \mathscr{A}$ if and only if the branching random walk is isotropic, i.e, if and only if the one-step transition probabilities for the marginal random walk satisfy $p_i = p_j$ for all $i, j \in \mathscr{A}$.*

Proof. Suppose that $F_i(R) = f$ for all $i \in \mathscr{A}$. Then by the fundamental relation (21), for each i,

$$f = p_i R + p_1 R f + (1 - p_1 - p_i) R f^2 \quad \forall i \ .$$

Hence, for any two distinct indices i, j, the quantities on the right-hand side must be equal, and so the difference must be zero, i.e.

$$(p_i - p_j)R - (p_i - p_j)Rf^2 = 0 \ .$$

If for some pair $i, j \in \mathscr{A}$ it were the case that $p_i \neq p_j$ then it would follow that $f^2 = 1$, and hence that $f = 1$. But this would contradict inequality (23). □

The following proposition is the principal fact about the generating functions needed for the proof of Theorem 2.

Proposition 3.

$$\sum_{i \in \mathscr{A}} \frac{F_i(R)^2}{1 + F_i(R)^2} = 1 \ . \tag{27}$$

Proof. The functions $F_i(z)$ are analytic inside the circle of convergence $|z| < R$, and since F_i is represented by a power series with nonnegative coefficients, $z = R$ must be a singularity. Hence, by the (complex) Implicit Function Theorem (see, e.g., [6]), the Jacobian matrix of the $d \times d$ system of equations (21) must be *singular* at $z = R$. The Jacobian matrix may be written as $I - J(z)$, where $J(z)$ has entries

$$\begin{aligned} J(z)_{ij} &= p_1 z + 2 \sum_{k \in \mathscr{A}_+} p_k z F_k(z) \qquad \text{if } j = i, \\ &= 2 p_j z F_i(z) \qquad\qquad\qquad \text{if } j \neq i \ . \end{aligned} \tag{28}$$

Since these entries are nonnegative when $0 < z \leq R$, the spectrum of $J(z)$ is contained in the closed disk with radius $\beta(z)$, where $\beta(z)$ is defined to be the lead eigenvalue of $J(z)$. Moreover, since the entries of $J(z)$ are nondecreasing and analytic in z for $0 < z < R$, so is $\beta(z)$; and since all of the functions $F_i(z)$ are continuous and finite at $z = R$, so is $\beta(z)$. Near $z = 0$ the entries of $J(z)$ are small, since each entry is divisible by z, and so for $z > 0$ sufficiently small, $\beta(z) < 1$. Thus, the *smallest* $z > 0$ where $I - J(z)$ is singular must be the smallest z where $\beta(z) = 1$.

Now consider the eigenvalue equation $J(R)u = u$. Since 1 is the Perron-Frobenius eigenvalue, the vector u must have all entries nonnegative, and at least some strictly positive. Thus, u may be normalized so that $\sum_{j \in \mathscr{A}_+} R p_j u_j = 1$; with this normalization the eigenvalue equation $J(R)u = u$ may, by equation (28), be rewritten as

$$\begin{aligned} u_i &= (p_1 R + 2 \sum_{j \in \mathscr{A}_+} p_j R F_j(R)) u_i + 2 \sum_{j \in \mathscr{A}_+ - \{i\}} p_j R F_i(R) u_j \\ &= (p_1 R + 2 \sum_{j \in \mathscr{A}_+} p_j R F_j(R)) u_i + 2 \sum_{j \in \mathscr{A}_+} p_j R F_i(R) u_j - 2 p_i R F_i(R) u_i \\ &= 2 F_i(R) + \{p_1 R + 2 \sum_{j \in \mathscr{A}_+} p_j R F_j(R) - 2 p_i R F_i(R)\} u_i \ . \end{aligned}$$

Multiplying each side by $F_i(R)$ and then making the substitution

$$F_i(R) - p_i R - p_i R F_i(R)^2 = F_i(R)\{p_1 R + 2 \sum_{j \in \mathscr{A}_+} p_j R F_j(R) - 2 p_i R F_i(R)\}$$

(note that this follows from the fundamental relations (21)) gives

$$\begin{aligned} u_i F_i(R) &= 2F_i(R)^2 + (F_i(R) - p_i R)u_i - p_i R F_i(R)^2 u_i \\ &\Longleftrightarrow u_i p_i R\{1 + F_i(R)^2\} = 2F_i(R)^2 \\ &\Longleftrightarrow u_i p_i R = 2F_i(R)^2/\{1 + F_i(R)^2\} \ . \end{aligned}$$

Substituting the last expression for $u_i p_i R$ in the normalization relation

$$\sum_{i \in \mathscr{A}_+} u_i p_i R = 1$$

now gives

$$\sum_{i \in \mathscr{A}_+} \frac{2F_i(R)^2}{1 + F_i(R)^2} = 1 \ ,$$

which is equivalent to the desired equation (27). □

4. Backscatter and first-passage matrices

For any positive numbers ρ, λ, define $M_\rho(\lambda)$ to be the $2d \times 2d$ matrix with entries

$$\begin{aligned} \left(M_\rho(\lambda)\right)_{ij} &= F_j(\lambda)^\rho \quad \text{if } j \neq i^{-1}, \\ &= 0 \qquad\qquad \text{if } j = i^{-1} \ . \end{aligned} \tag{29}$$

Note that $M_2(\lambda)$ is the *backscatter matrix* defined by (16). Henceforth the matrix $M_1(\lambda)$ will be called the *first-passage matrix*. Both matrices will play important roles in the proofs of Theorems 1 and 2. For $\rho > 0$ and $0 < \lambda \leq R$ the matrix $M_\rho(\lambda)$ is aperiodic and irreducible, with nonnegative entries, and thus is subject to the conclusions of the Perron-Frobenius theorem. In particular, $M_\rho(\lambda)$ has a largest positive eigenvalue $\theta(\rho; \lambda)$. Since the positive entries $F_j(\lambda)^\rho$ increase strictly with $\lambda \in (0, R]$, the lead eigenvalue $\theta(\rho; \lambda)$ also increases strictly with λ.

Proposition 4. *The lead eigenvalue of $M_\rho(\lambda)$ is the unique positive solution $\theta(\rho; \lambda)$ of the equation*

$$\sum_{a \in \mathscr{A}} \frac{F_a(\lambda)^\rho}{\theta(\rho; \lambda) + F_a(\lambda)^\rho} = 1 \ . \tag{30}$$

For each $\rho > 0$, the function $\lambda \to \theta(\rho; \lambda)$ is strictly increasing in λ for $\lambda \leq R$.

Proof. Consider the eigenvalue equation $M_\rho(\lambda)u = \theta u$. If $\theta = \theta(\rho;\lambda)$ is the lead eigenvalue, then the vector u must have nonnegative entries, not all zero. Thus, u may be normalized so that $\sum_i F_i(\lambda)^\rho u_i = 1$. With this normalization, the eigenvalue equation may be rewritten as

$$\begin{aligned} \theta u_j &= \sum_i F_i(\lambda)^\rho u_i - F_j(\lambda)^\rho u_j \qquad \Longleftrightarrow \\ u_j &= 1/(\theta + F_j(\lambda)^\rho) \ . \end{aligned}$$

Hence, the normalization relation $\sum_i F_i(\lambda)^\rho u_i = 1$ is equivalent to equation (30).

For each $a \in \mathscr{A}$ the generating function $F_a(\lambda)$ is a strictly increasing function of λ. Since the matrix $M_\rho(\lambda)$ is irreducible and aperiodic, it follows by a routine argument that the lead eigenvalue $\theta(\rho;\lambda)$is strictly increasing in λ. □

Corollary 2. $\theta(2; R) = 1$.

Proof. This follows immediately from Proposition 3 and equation (30). □

5. Upper bounds

5.1. Upper bound for $\delta(\lambda)$

For each integer $m \geq 0$, define $\mathscr{G}_m$ to be the set of all vertices $x \in \mathscr{G}$ at distance m from the root 1 (i.e., all x such that $|x| = m$) and define $\mathscr{H}_m$ to be the set of all vertices $x \in \mathscr{G}_m$ that are visited by particles of the branching random walk. Then for each $m \geq 1$,

$$\Lambda \subset \bigcup_{x \in \mathscr{H}_m} \Omega(x) \ , \tag{31}$$

because a particle trajectory can approach a point $\omega \in \Omega(x)$ only if it eventually stays in $\mathscr{T}(x)$, and because the marginal random walk is nearest neighbor, this can occur only if the trajectory touches x. The inclusion (31) provides a simple (possibly crude) covering of the random set Λ by sets $\Omega(x)$, all of diameter α^m. Consequently, the Hausdorff dimension of Λ can be bounded above by estimating the cardinalities of the sets $\mathscr{H}_m$. Recall (Proposition 4) that $\theta(\lambda) = \theta(1;\lambda)$ is the lead eigenvalue of the first-passage matrix $M_1(\lambda)$ defined by (29).

Lemma 2. $\limsup_{m\to\infty} |\mathscr{H}_m|^{1/m} \leq \theta(\lambda)$.

Proof. Fix $x \in \mathscr{G}_m$, and consider the modified branching random walk with absorption (freezing) at x. Recall that $Z_+^x(x)$ is the number of distinct particles that are ultimately frozen at x. Then $P\{x \in \mathscr{H}_m\} \leq EZ_+^x(x)$.But $EZ_+^x(x) = F_x(\lambda)$, by Lemma 1, and by equation (20), $F_x(\lambda) = \prod_{j=1}^{m} F_{i_j}(\lambda)$, where $x = i_1 i_2 \cdots i_m$ is the word representation of x. Consequently,

$$\begin{aligned} E|\mathscr{H}_m| &\le \sum_{x \in \mathscr{G}_m} E Z_+^x(x) \\ &= \sum_{x \in \mathscr{G}_m} F_x(\lambda) \\ &= \sum_{i_1 i_2 \cdots i_m} \prod_{j=1}^{m} F_{i_j}(\lambda) \\ &= \mathbf{1}^t M_1(\lambda)^m \mathbf{1} \quad , \end{aligned}$$

where $\mathbf{1}$ is the $2d$-vector with all entries 1. Since $\theta(\lambda)$ is the lead eigenvalue of $M_1(\lambda)$ and $\mathbf{1}$ and $\mathbf{1}^t$ are dominated by multiples of the corresponding right and left eigenvectors,

$$\mathbf{1}^t M_1(\lambda)^m \mathbf{1} \le C\theta(\lambda)^m$$

for a suitable positive constant C. The desired result now follows routinely from the Chebyshev-Markov inequality and the Borel-Cantelli lemma. □

Corollary 3. *With probability* 1, $\delta(\lambda) \le -\log\theta(\lambda)/\log\alpha$.

This follows immediately from Lemma 2, and proves half of the relation (9).

5.2. Upper bound for $\delta(\lambda;\mu)$

Let μ be an ergodic, σ-invariant probability measure on the space Ω of semi-infinite reduced words from $\mathscr{A}$ (the natural boundary of $\mathscr{T}$). Recall that $\delta(\lambda;\mu)$ is the Hausdorff dimension (in the metric d_α) of $\Lambda \cap \Omega_\mu$, where Ω_μ is the subset of Ω consisting of all sequences ω that are "generic" for μ in the sense (13). Recall that $\varphi_\lambda : \Omega \to \mathbb{R}$ is the function defined by $\varphi_\lambda(x_1 x_2 \cdots) = \log F_{x_1}(\lambda)$. Since φ_λ is continuous on Ω, relation (13) holds with $f = \varphi_\lambda$.

Lemma 3. *For every* $\varepsilon > 0$ *there exist sets* $\Gamma_m = \Gamma_m(\mu) \subset \mathscr{G}_m$ *of vertices at distance* m *from the root* 1 *such that*

$$\lim_{m\to\infty} m^{-1} \log|\Gamma_m| \le h(\mu) + \varepsilon \quad , \tag{32}$$

$$\lim_{m\to\infty} \sup_{x_1 x_2 \ldots x_m \in \Gamma_m} |m^{-1} \sum_{j=1}^{m} \log F_{x_j}(\lambda) - \int_\Omega \varphi_\lambda \, d\mu| \le \varepsilon \quad , \tag{33}$$

and such that for every sequence $x_1 x_2 \cdots \in \Omega$,

$$x_1 x_2 \cdots \in \Omega_\mu \Rightarrow x_1 x_2 \cdots x_m \in \Gamma_m(\mu) \text{ infinitely often} \quad . \tag{34}$$

Proof. This is a routine consequence of the Shannon-McMillan theorem and the ergodic theorem. See the proof of Lemma 3 in [11] for details. □

The following proposition proves half of the equality (14).

Proposition 5. *If $h(\mu) < -\int \varphi_\lambda \, d\mu$ then with probability one, $\Lambda \cap \Omega_\mu = \emptyset$. If $h(\mu) \geq -\int \varphi_\lambda \, d\mu$ then with probability one*

$$\delta(\lambda; \mu) \leq -\frac{h(\mu) + \int_\Omega \varphi_\lambda \, d\mu}{\log \alpha} \quad . \tag{35}$$

Proof. Let $\Gamma_m = \Gamma_m(\mu)$ be as in the statement of Lemma 3. In order that a reduced semi-infinite word $\omega = x_1 x_2 \cdots$ be an element of $\Lambda \cap \Omega_\mu$ it is necessary that $x_1 x_2 \cdots x_m \in \mathscr{H}_m \cap \Gamma_m$ for infinitely many integers m, where as before $\mathscr{H}_m$ is the set of all vertices at distance m from the root 1 that are ever visited by a particle of the branching random walk. Define

$$\Lambda_m(\mu) = \{\omega = x_1 x_2 \cdots \in \Lambda \cap \Omega_\mu \mid x_1 x_2 \cdots x_m \in \mathscr{H}_m \cap \Gamma_m\};$$

then for each $m \geq 1$

$$\Lambda \cap \Omega_\mu \subset \cup_{n \geq m} \Lambda_n(\mu) \quad . \tag{36}$$

Thus, $\cup_{n \geq m} \Lambda_n(\mu)$ is a covering of $\Lambda_n(\mu)$ by sets of diameter no larger than α^m. Consequently, to bound the Hausdorff dimension of $\Lambda_n(\mu)$ it is enough to bound the cardinality of $\mathscr{H}_m \cap \Gamma_m$. By the same argument as in the proof of Lemma 2,

$$E|\mathscr{H}_m \cap \Gamma_m| \leq \sum_{i_1 i_2 \cdots i_m \in \Gamma_m} \prod_{j=1}^m F_{i_j}(\lambda) = \sum_{x \in \Gamma_m} \exp\{\sum_{j=1}^m \varphi_\lambda(\sigma^j x)\} \quad .$$

By inequality (32), $|\Gamma_m| \leq \exp\{m(h(\mu) + \varepsilon)\}$. Moreover, by (33), for all words $x \in \Gamma_m$, $\sum_{j=1}^m \varphi_\lambda(\sigma^j x) \leq m(\int \varphi_\lambda \, d\mu + \varepsilon)$. Hence, for all sufficiently large m, the expected cardinality of $\mathscr{H}_m \cap \Gamma_m$ is no greater than $\exp\{m(h(\mu) + \int \varphi_\lambda \, d\mu + 2\varepsilon)\}$. The Borel-Cantelli lemma therefore implies that with probability 1, eventually

$$|\mathscr{H}_m \cap \Gamma_m| \leq \exp\{m(h(\mu) + \int \varphi_\lambda \, d\mu + 3\varepsilon)\} \quad . \tag{37}$$

It follows that if $h(\mu) + \int \varphi_\lambda \, d\mu + 3\varepsilon < 0$ then eventually $\mathscr{H}_m \cap \Gamma_m$ is empty. This in turn implies that $\Lambda_m(\mu)$ is empty, and so by (36), $\Lambda \cap \Omega_\mu = \emptyset$. If $h(\mu) + \int \varphi_\lambda \, d\mu \geq 0$ then inequality (37) implies that for every $n \geq 1$ and all m sufficiently large, the set $\Lambda_n(\mu)$ is covered by $\exp\{m(h(\mu) + \int \varphi_\lambda \, d\mu + 3\varepsilon)\}$ sets of diameter α^m. Thus, since $\varepsilon > 0$ can be chosen arbitrarily small,

$$\delta_H(\Lambda_m(\mu)) \leq -(h(\mu) + \int \varphi_\lambda \, d\mu)/\log \alpha \quad .$$

Since $\Lambda \cap \Omega_\mu \subset \cup_m \Lambda_m(\mu)$, inequality (35) follows. □

6. Lower bounds

To complete the proofs of the formulas (9) and (14) for the Hausdorff dimensions of the random sets Λ and $\Lambda \cap \Omega_\mu$, we must establish the reverse inequalities to those of Corollary 3 and Proposition 5. Our strategy will be to exhibit subsets of Λ and $\Lambda \cap \Omega_\mu$ whose Hausdorff dimensions approach the desired bounds. These subsets will be the limit sets of labelled Galton–Watson trees embedded in the branching random walk; the calculation of their Hausdorff dimensions will be accomplished by appealing to a theorem of HAWKES [5], [15] and an extension due to Lalley and Sellke [11].

6.1. Hawkes' theorem

To any Galton–Watson chain there is associated a genealogical tree τ, which we shall call a *Galton–Watson tree*. For a detailed description, see [5]. In a nutshell: Vertices of τ are arranged in levels $V_0, V_1, V_2, \ldots$; the vertices of level V_n represent the individuals of the nth generation of the corresponding Galton–Watson chain. Edges of the tree connect vertices corresponding to parent-child pairs; thus, there are edges only between vertices of successive levels. The limit set Λ_{GW} of the tree is the set of *ends*, where an end is defined to be an *infinite* path that starts at the root (the unique vertex in V_0) and visits every level V_n exactly once. (The limit set is empty on the event that the Galton–Watson chain dies out.) For each $\alpha \in (0, 1)$, the d_α-distance between two ends γ, γ' is defined to be α^n, where $n = n(\gamma, \gamma')$ is the index of the last level V_n where the paths γ and γ' touch. The metric d_α makes Λ_{GW} a compact metric space.

Hawkes' theorem. *If the offspring distribution has mean $\mu > 1$ and finite second moment then almost surely on the event of nonextinction, the limit set Λ_{GW} of the Galton–Watson tree τ has Hausdorff dimension* (*in metric d_α*)

$$\delta_H(\Lambda_{GW}) = -\frac{\log \mu}{\log \alpha} \ .$$

Hawkes discusses only the case $\alpha = 1/2$, but the result and its proof are valid for all $\alpha \in (0, 1)$. See [15] for another proof, and the appendix of [10] for a sharper result.

6.2. Extension of Hawkes' theorem

Let $\mathscr{B}$ be an arbitrary finite set. A *labelled Galton–Watson process* with label space $\mathscr{B}$ is determined by a probability distribution Q on the set $2^{\mathscr{B}}$ of subsets of $\mathscr{B}$. Each individual ζ, regardless of its type, produces a random set $\mathscr{O}_\zeta$ of offspring, with distribution Q, and the offspring sets of different individuals are conditionally independent, as in an unlabelled Galton–Watson process. Observe that a labelled Galton–Watson process is a multi-type Galton–Watson process in which (i) all types have the same offspring distribution, and (ii) the offspring distribution is constrained to allow at most one individual of each label. Because of (i), if Z_n is the cardinality of the nth generation, then the sequence $\{Z_n\}_{n\geq 0}$ constitutes an ordinary Galton–Watson process; and if the expected cardinality of a random set chosen according to the distribution Q is greater than 1, then $\{Z_n\}_{n\geq 0}$ is supercritical.

A *labelled Galton–Watson tree* τ is the genealogical tree associated with a labelled Galton–Watson process, with vertices assigned the labels of the corresponding individuals. If the underlying Galton–Watson process is supercritical, then with positive probability the tree τ is infinite. On this event, each end of τ will be naturally identified with a unique sequence $\omega = x_1 x_2 \cdots$, where x_n is the label of the nth level vertex through which the end passes. (NOTE: The root vertex need not be labelled.) Thus, the set $\partial\tau$ of ends of τ is naturally embedded in the sequence space $\mathscr{B}^{\mathbb{N}}$.

For each label $i \in \mathscr{B}$, let $q_i = \sum_{F\subset\mathscr{B}:i\in F} Q(F)$ be the probability that label i is included in a random set with distribution Q. Define a function $\psi : \mathscr{B}^{\mathbb{N}} \to \mathbb{R}$ by

$$\psi(x_1 x_2 \cdots) = \log q_{x_1} \quad .$$

For any ergodic, shift-invariant probability measure μ on the sequence space $\mathscr{B}^{\mathbb{N}}$, define $\mathscr{B}^{\mathbb{N}}_{\mu}$ to be the set of μ*-generic* sequences, i.e., the set of all sequences ω such that for every continuous function $f : \mathscr{B}^{\mathbb{N}} \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(\sigma^i \omega) = \int f \, d\mu \quad . \tag{38}$$

Extended Hawkes' theorem . [11] *Let τ be the labelled Galton–Watson tree attached to a supercritical labelled Galton–Watson process with label set $\mathscr{B}$ and offspring distribution Q, and let μ be any ergodic, σ-invariant probability measure on $\mathscr{B}^{\mathbb{N}}$. If $h(\mu) + \int \psi \, d\mu < 0$ then with probability 1,*

$$\partial\tau \cap \mathscr{B}^{\mathbb{N}}_{\mu} = \emptyset \quad . \tag{39}$$

If $h(\mu) + \int \psi \, d\mu \geq 0$ then, almost surely on the event of nonextinction, the Hausdorff dimension of $\partial\tau \cap \mathscr{B}^{\mathbb{N}}_{\mu}$ in the metric d_α is

$$\delta_H(\partial\tau \cap \mathscr{B}^{\mathbb{N}}_{\mu}) = -\frac{h(\mu) + \int \psi \, d\mu}{\log \alpha} \quad . \tag{40}$$

6.3. *Embedded Galton–Watson trees τ_r*

The construction of embedded Galton–Watson trees τ_r in the branching random walk is similar to the analogous construction for isotropic contact processes [10]. The *offspring* of a vertex x in τ_r will be vertices y at distance r from x such that a particle of the branching random walk located at x gives rise to a descendant particle whose trail begins with a *downward path* (defined below) from x to y. Thus, all vertices of τ_r will be vertices of the tree $\mathscr{T}$ that are visited by particles of the branching random walk. Consequently, the limit set Λ_{τ_r} of the Galton–Watson tree τ_r will be a subset of the limit set Λ of the branching random walk.

Fix a generator $a \in \mathscr{A}$, and consider the subtree $\mathscr{T}^* = \mathscr{T} - \mathscr{T}(a^{-1})$ of $\mathscr{T}$ each of whose vertices except the root vertex 1 is represented by a word $x \in \mathscr{G}$ beginning with a letter $x_1 \neq a^{-1}$. The tree $\mathscr{T}^*$ may be arranged in *levels* $\mathscr{L}_0, \mathscr{L}_1, \mathscr{L}_2, \ldots$, with $x \in \mathscr{L}_n$ if and only if $|x| = n$; thus, there are $(2d-1)^n$ vertices in level $\mathscr{L}_n$. For $n \geq 1$, define $\mathscr{L}_n^*$ to be the subset of $\mathscr{L}_n$ containing those vertices whose word representation $x = x_1 x_2 \cdots x_n$ terminates in the letter $x_n = a$. Note that for any vertex $x \in \mathscr{L}_n^*$, the set of nearest neighbors of x in $\mathscr{L}_{n+1}$ is $\{xy : y \in \mathscr{L}_1\}$; thus, for $x \in \mathscr{L}_n^*$, the tree $\mathscr{T}(x)$ is the left translate of the tree $\mathscr{T}^*$ by the group element x.

Let x be a vertex in level $\mathscr{L}_n$ and let y be a vertex contained in $\mathscr{T}(x)$. Since the word x is a prefix of the word y, the vertex y must lie in a level $\mathscr{L}_{n+m}$ at greater depth than vertex x. A *downward path* from x to y is a finite path in $\mathscr{T}(x)$ that begins at x and first enters level $\mathscr{L}_{n+m}$ at y, where it terminates.

Definition of $\boldsymbol{\tau_r}$: Fix an integer $r \geq 1$. Define generations $V_n \subset \mathscr{L}_{nr}^*$ and distinguished particles ζ_x associated with the vertices $x \in V_n$ inductively as follows:

(a) $V_0 = \{1\}$ and ζ_1 is the initial particle of the branching random walk.
(b) For each $x \in V_n$, the *offspring* of x are those vertices $y \in \mathscr{L}_{nr+r}^*$ such that at least one particle of the branching random walk initiated by particle ζ_x has a trail that begins with a downward path from x to y. For each offspring vertex y of x, the *parent* of y is x.
(c) V_{n+1} is the set of all offspring of vertices in V_n.
(d) For each vertex $y \in V_{n+1}$ with parent $x \in V_n$, the distinguished particle ζ_y is the first particle of the branching random walk initiated by ζ_x to arrive at y via a downward path from x to y.

For each vertex $y \neq 1$ of the tree τ_r define the *label* of y to be the word $x^{-1}y$, where x is the parent of y. Observe that each label is a reduced word of length r that ends in the letter a, so the set of labels is finite.

Proposition 6. *τ_r is a labelled Galton–Watson tree.*

Proof. The particles ζ_x associated with different vertices $x \in V_n$ are distinct, so by the Superposition Property the branching random walks initiated by the particles $\zeta_x, x \in V_n$ are mutually independent, and independent of the pre-$\mathscr{L}_{nr}$ history of the branching random walk initiated at ζ_1. Since all vertices of V_n have final letter a, the branching random walks initiated by the particles $\zeta_x, x \in V_n$ are all "oriented" (relative to the level structure of the tree $\mathscr{T}^*$) the same way. Consequently, the offspring distributions for vertices $x \in V_n$ (viewed as probability distributions on the set of labels) are all the same as that of the initial vertex $1 \in V_0$. Thus, τ_r is a labelled Galton–Watson tree. □

Note that for such an embedded Galton–Watson tree τ_r, the relation between the metrics d_α^τ for the tree τ_r and d_α for the full tree $\mathscr{T}$ is:

$$d_\alpha^\tau(x, y) = d_{\alpha^r}(x, y) \quad . \tag{41}$$

6.4. Properties of the embedded G.W. Trees

Our strategy now will be to show that the limit sets Λ_{τ_r} of the embedded Galton–Watson trees τ_r, which are contained in the limit set Λ of the branching random walk, have Hausdorff dimensions converging to $\delta(\lambda)$, and that their intersections with the sets Ω_μ have Hausdorff dimensions converging to $\delta(\lambda;\mu)$. The argument rests on the fact that particles of the branching random walk that reach a vertex x at a large distance from the root 1 tend to do so by moving essentially along the geodesic segment from 1 to x.

Recall that $Z_+^x(x)$ is the number of particles ultimately frozen at x in the modification of the branching random walk with freezing at vertex x. For each integer $r \geq 0$, define $Z_+^{x,r}(x)$ to be the number of particles counted in $Z_+^x(x)$ whose trails remain within distance r of the geodesic segment from 1 to x.

Lemma 4.

$$\lim_{r\to\infty} \inf_{x\in\mathscr{G}-\{1\}} \left(\frac{EZ_+^{x,r}(x)}{EZ_+^x(x)}\right)^{1/|x|} = 1$$

Proof. By the monotone convergence theorem, $EZ_+^{x,r}(x)/EZ_+^x(x) \to 1$ as $r \to \infty$ for each *fixed* vertex $x \neq 1$, and, in particular, uniformly for $x \in \mathscr{A}$. Fix $\varepsilon > 0$, and choose $r = r_\varepsilon$ so large that $EZ_+^{x,r}(x) \geq (1-\varepsilon)EZ_+^x(x)$ for each $x \in \mathscr{A} - \{1\}$.

Now consider an arbitrary vertex $x = i_1 i_2 \cdots i_m$. To reach x, a particle must follow a trail γ that passes through each of the vertices $x_l = i_1 i_2 \cdots i_l$ on the geodesic segment from 1 to x. Let γ_l be the segment of the trail γ between the times of first passage to x_l and x_{l+1}, respectively. If for each l the trail segment γ_l stays within distance r of the edge $[x_l, x_{l+1}]$, then the concatenation γ remains within distance r of the geodesic segment from 1 to x. Thus, by the Superposition Property,

$$EZ_+^{x,r}(x) = \prod_{j=1}^{m} EZ_+^{i_j,r}(i_j) \ .$$

But by Lemma 1 and equation (20),

$$EZ_+^{x}(x) = \prod_{j=1}^{m} EZ_+^{i_j}(i_j) \ .$$

Thus, for every vertex $x \neq 1$, $EZ_+^{x,r}(x)/EZ_+^x(x) \geq (1-\varepsilon)^{|x|}$. □

Lemma 5. *For each vertex $x \neq 1$ and every integer $r \geq 1$, the random variable $Z_+^{x,r}(x)$ has finite variance.*

Proof. Consider yet another modification of the branching random walk in which particles are *killed* upon reaching the set of vertices at distances greater than r from the geodesic segment from 1 to x. The total number N of particles ever born in this modification stochastically dominates $Z_+^{x,r}(x)$. Thus, to prove the lemma, it suffics to prove that N has finite variance.

In the modified branching random walk, particles can exist only at vertices in a finite subset F, and so the branching random walk may be viewed as a multi-type Galton–Watson process Y_n with type space F. Since the *unmodified* branching random walk survives only weakly, the multi-type Galton–Watson process Y_n is subcritical. (To see that it cannot be critical, note that increasing r by one would strictly increase the Malthusian parameter, so the corresponding multi-type Galton–Watson process with r replaced by $r+1$ would be supercritical. But if this were the case then clearly the unmodified branching random walk would survive *strongly* with positive probability, contradicting the standing hypothesis (8) that $\lambda \leq R$.) By Proposition 11 of the Appendix, the total number of particles ever born in a subcritical multi-type Galton–Watson process has finite second moment, provided that the offspring distributions have finite second moments. This is certainly the case here, because the offspring distribution in the unmodified branching random walk has a geometric distribution. □

For $x \in \mathscr{L}_n$, define $J(x)$ to be the event that among the particles counted in $Z_+^x(x)$ there is at least one whose trail is a downward path from the root 1 to x. Note that $Z_+^x(x) \geq 1$ on the event $J(x)$, so $P(J(x)) \leq EZ_+^x(x)$.

Proposition 7. *As $m \to \infty$,*

$$\left(\min_{x \in \mathscr{L}_m} \frac{P(J(x))}{EZ_+^x(x)} \right)^{1/m} \longrightarrow 1 \ . \tag{42}$$

Proof. Fix $\varepsilon > 0$ small, and choose r so large that for all vertices $x \neq 1$,

$$EZ_+^{x,r}(x) \geq (1-\varepsilon)^{|x|} EZ_+^x(x) \ . \tag{43}$$

By Lemma 4, such an integer r exists. Now consider a vertex $x \in \mathscr{L}_m$, where $m = (n+2)r$. Let $x_1, x_2, \dots, x_{n+2}$ be the vertices in levels $\mathscr{L}_r, \mathscr{L}_{2r}, \dots, \mathscr{L}_{nr+2r}$ on the geodesic segment from 1 to x, and for each $i \leq n+1$, set $y_i = x_i^{-1} x_{i+1}$. The event $J(x)$ occurs if there is a particle whose trail (i) proceeds from 1 to x_1 without first exiting $\cup_{j=0}^r \mathscr{L}_j$; then (ii) travels from each x_i to the subsequent x_{i+1}, $1 \leq i \leq n+1$, remaining within distance r of the geodesic segment from x_i to x_{i+1}; and finally (iii) proceeds directly along the geodesic segment from x_{n+1} to x_{n+2}.

Define random variables $Y_0, Y_1, Y_2, \dots, Y_n$ inductively as follows. If there is a particle of the branching random walk whose trail proceeds from 1 to x_1 without first exiting $\cup_{j=0}^r \mathscr{L}_j$, then $Y_0 = 1$; otherwise $Y_0 = 0$. If $Y_0 = 0$ then $Y_j = 0$ for all $j \geq 0$. If $Y_0 = 1$ then there is at least one particle whose trail reaches x_1 without first exiting $\cup_{j=0}^r \mathscr{L}_j$; upon reaching x_1, the first such particle initiates a branching random walk starting at x_1. Suppose that particles of this offspring branching random walk are frozen upon reaching x_2; define Y_1 to be the number of particles ultimately frozen at x_2 whose trails remain within distance r of the geodesic segment from x_1 to x_2. Those particles counted in Y_1 initiate branching random walks starting at x_2. Suppose that particles of these offspring branching random walks are frozen upon reaching x_3; define Y_2 to be the number of particles ultimately frozen at x_3 whose trails remain within distance r of the geodesic segment from x_2 to x_3. Define $Y_3, Y_4, \dots, Y_n$ similarly, by induction.

By the Superposition Property, the random variables $Y_0, Y_1, Y_2, \dots, Y_{n-1}$ are the first n terms of a *Galton–Watson process in a varying environment* (see the Appendix below), i.e.,

$$E(s^{Y_{k+1}} \mid Y_0, Y_1, \dots, Y_k) = \varphi_{y_k}(s)^{Y_k} \quad ,$$

where $\varphi_y(s)$ is the probability generating function of the random variable $Z_+^{y,r}(y)$. There are only finitely many words y of length r, so there are only finitely many possible offspring distributions. For each possible word y, the mean offspring number $\varphi_y'(1)$ is strictly less than 1, because $Z_+^{y,r}(y)$ is stochastically dominated by $Z_+^y(y)$, by construction, and $EZ_+^y(y) < 1$ by Lemma 1 and Corollary 1. Moreover, the random variable $Z_+^{y,r}(y)$ has finite variance, by Lemma 5, and $P\{Z_+^{y,r}(y) = 1\} > 0$ (because it is certainly possible, with positive probability, for exactly one particle in a branching random walk to follow a trail along the geodesic segment from 1 to y and for all other particle trails to avoid the subtree $\mathscr{T}(a^{-1})$ forever). Thus, by Proposition 13 of the Appendix,

$$\lim_{n\to\infty} \frac{P_{y_1 y_2 \dots}\{Y_n = 1\}}{\prod_{j=1}^n EZ_+^{y_j,r}(y_j)} = \alpha(y_1 y_2 \cdots) \tag{44}$$

uniformly for all infinite sequences $y_1 y_2 \cdots$ of words of length r, and the limit $\alpha(y_1 y_2 \cdots)$ is uniformly bounded away from 0.

The event $J(x)$ contains the event $\{Y_n \geq 1\} \cap L(x_{n+1}, x_{n+2})$, where $L(x_{n+1}, x_{n+2})$ is the event that the first particle to reach x_{n+1} among the particles counted in Y_n has a descendant particle whose trail follows the geodesic segment from x_{n+1} to x_{n+2}. Since $L(x_{n+1}, x_{n+2})$ has positive (conditional) probability, the result (42) follows from the choice of r (in particular, inequality (43)) and the uniformity in (44). □

Corollary 4. *As $m \to \infty$,*

$$\left(\sum_{x \in \mathscr{L}_m} P(J(x)) \right)^{1/m} \longrightarrow \theta(\lambda) \quad . \tag{45}$$

Proof. By Proposition 7, it suffices to show that $(\sum_{x\in\mathscr{L}_m} EZ_+^x(x))^{1/m} \to \theta(\lambda)$. But by Lemma 1 and equation (20), $\sum_{x\in\mathscr{L}_m} EZ_+^x(x) = \mathbf{u}_a^t M_1(\lambda)^m \mathbf{1}$, where $\mathbf{1}$ is the vector all of whose entries are 1 and $\mathbf{u}_a$ is the vector with entry 0 in the a^{-1} slot and all other entries 1. Since $M_1(\lambda)$ is a Perron-Frobenius matrix whose lead eigenvalue is $\theta(1; \lambda) = \theta(\lambda)$, the result follows. □

Corollary 5. *The mean offspring numbers μ_k for the Galton–Watson trees τ_k satisfy*

$$\lim_{k\to\infty} \mu_k^{1/k} = \theta(\lambda) \ . \tag{46}$$

Proof. The mean offspring number μ_k for the Galton–Watson tree τ_k is, by construction,

$$\mu_k = \sum_{x\in\mathscr{L}_k^*} P(J(x)) \ .$$

This sum differs from the sum in relation (45) above only in that the index set $\mathscr{L}_k$ is replaced by the smaller index set $\mathscr{L}_k^*$. (Recall that $\mathscr{L}_k^*$ consists of all those vertices of $\mathscr{L}_k$ whose word representations end in the letter a.) Consider a vertex $x \in \mathscr{L}_k$; on the event $J(x)$, the vertex x is visited by a particle ζ of the branching random walk whose trail is a downward path from the root 1 to x. The particle ζ engenders its own offspring branching random walk, starting from vertex x. Clearly, there is positive probability ρ, independent of x, that this offspring branching random walk, if allowed to run for two time units, would produce an offspring particle at a vertex $y \in \mathscr{L}_{k+2}^* \cap \mathscr{T}(x)$. Consequently,

$$\sum_{x\in\mathscr{L}_{k+2}^*} P(J(x)) \geq \rho \sum_{x\in\mathscr{L}_k} P(J(x)) \ .$$

The result therefore follows from Corollary 4. □

Corollary 6. *Let μ be an ergodic, shift-invariant probability measure on the space Ω of infinite reduced words $x = x_1x_2\cdots$ from the alphabet $\mathscr{A}$. Then*

$$\lim_{k\to\infty} \frac{1}{k}\int_\Omega \log P(J(x_1x_2\cdots x_k))\,d\mu(x) = \int_\Omega \varphi_\lambda(x)\,d\mu(x) \ . \tag{47}$$

Proof. Recall that $P(J(y)) \leq EZ_+^y(y)$ for every finite reduced word y. Consequently, by Proposition 7, Lemma 1, and equation (20),

$$\lim_{k\to\infty}\left(\frac{P(J(x_1x_2\cdots x_k))}{\prod_{j=1}^k F_{x_j}(\lambda)}\right)^{1/k} = 1$$

uniformly for $x \in \Omega$, and so

$$\lim_{k\to\infty}\left(\frac{1}{k}\int_\Omega \log P(J(x_1x_2\cdots x_k))\,d\mu(x) - \frac{1}{k}\int_\Omega \log\prod_{j=1}^k F_{x_j}(\lambda)\,d\mu(x)\right) = 0 \ .$$

The result now follows by the shift invariance of μ and the definition of φ_λ. □

6.5. Lower bounds for $\delta(\lambda)$ and $\delta(\lambda;\mu)$

The following corollary, along with Corollary 3, completes the proof of equation (9).

Corollary 7. *With probability one,*

$$\delta(\lambda) \geq -\log\theta(\lambda)/\log\alpha \quad .$$

Proof. Since the Hausdorff dimension $\delta_H(\Lambda)$ is almost surely constant (see Remark (C) at the end of section 1), it suffices to show that for any real number $\delta^* \leq -\log\theta(\lambda)/\log\alpha$, there is positive probability that $\delta_H(\Lambda) \geq \delta^*$. Consider the embedded Galton–Watson trees τ_k. By Hawkes' Theorem, on the set of nonextinction the limit set Λ_k of τ_k has Hausdorff dimension, in the metric d_α, equal to

$$\delta_k = -\frac{\log\mu_k^{1/k}}{\log\alpha} \quad .$$

Since the probability of nonextinction is positive, it follows that with positive probability the set Λ has a subset Λ_k of dimension δ_k. Finally, by relation (46), $\delta_k \geq \delta^*$ for all sufficiently large k, so with positive probability the Hausdorff dimension of Λ is at least δ^*. □

Together with Proposition 5, the following statement completes the proof of equation (14).

Corollary 8. *For any ergodic, shift-invariant probability measure μ on Ω,*

$$\delta(\lambda;\mu) \geq -(h(\mu) + \int \varphi_\lambda\, d\mu)/\log\alpha \quad . \tag{48}$$

Proof. If the right-hand side is negative, then the inequality is trivially true. Consider the limit set Λ_k of the Galton–Watson tree τ_k: by the Extended Hawkes Theorem, almost surely on the event $\Lambda_k \neq \emptyset$, the intersection of Λ_k with Ω_μ has Hausdorff dimension (in the metric d_α)

$$-\frac{h(\mu) + k^{-1}\int_\Omega \log P(J(x_1x_2\cdots x_k))\, d\mu(x)}{\log\alpha}$$

Since Λ_k is a subset of Λ it follows that this is a lower bound for the Hausdorff dimension of $\Lambda \cap \Omega_\mu$. The result now follows from Corollary 6. □

7. Backscattering inequalities

7.1. The inequality $\delta(\lambda) \leq \frac{1}{2}\delta_H(\Omega)$

We refer to this inequality as a "backscattering inequality" because it was suggested, and may be proved, by the "backscattering argument" explained (in different contexts) in [9] and [10]. In fact, it was this argument that first led the authors to conjecture Theorem 2. The following argument, which does not use the backscattering heuristic, is both short and relatively elementary. Another argument, based on the Gibbs Variational Principle, will be given below.

Proposition 8. *For all $\lambda \leq R$,*

$$\theta(\lambda)^2 \leq \theta(2;\lambda)(2d-1) \leq 2d-1 \ , \tag{49}$$

with strict inequality $\theta(\lambda)^2 < 2d-1$ holding except possibly at $\lambda = R$.

Proof. Recall that $\theta(\lambda) = \theta(1;\lambda)$ and $\theta(2;\lambda)$ are the lead (Perron-Frobenius) eigenvalues of the matrices $M_1(\lambda)$ and $M_2(\lambda)$, respectively. Consequently, if $\mathbf{1}$ is the $2d$-vector with all entries 1 then

$$\theta(\rho;\lambda) = \lim_{m\to\infty} (\mathbf{1}^t M_\rho(\lambda)^m \mathbf{1})^{1/m} \ .$$

By the Cauchy-Schwartz inequality,

$$\begin{aligned}
\left(\mathbf{1}^t M_1(\lambda)^m \mathbf{1}\right)^{1/m} &= \left(\sum_{i_1 i_2 \cdots i_m} \prod_{j=1}^m F_{i_j}(\lambda)\right)^{1/m} \\
&\leq \left(\sum_{i_1 i_2 \cdots i_m} \prod_{j=1}^m F_{i_j}(\lambda)^2\right)^{1/2m} \left(\sum_{i_1 i_2 \cdots i_m} 1\right)^{1/2m} \\
&\longrightarrow \sqrt{\theta(2;\lambda)(2d-1)} \ ,
\end{aligned}$$

as $m \to \infty$, since the number of reduced words $i_1 i_2 \cdots i_m$ of length m is $2d(2d-1)^{m-1}$. Letting $m \to \infty$ shows that $\theta(1;\lambda)^2 \leq \theta(2;\lambda)(2d-1)$. That $\theta(2;\lambda) \leq 1$, with strict inequality except when $\lambda = R$, follows from Proposition 4 and Corollary 2. □

Corollary 9. *For all $\lambda \leq R$, with probability 1,*

$$\delta(\lambda) \leq \frac{1}{2}\delta_H(\Omega) \ , \tag{50}$$

with strict inequality except possibly at $\lambda = R$.

Proof. This is an immediate consequence of the preceding proposition and relation (9), as the Hausdorff dimension of Ω in the metric d_α is $-\log(2d-1)/\log\alpha$. □

7.2. The Gibbs variational principle

The *Gibbs Variational Principle* states that for any ergodic, σ-invariant probability measure μ on Ω and any Hölder continuous function $\varphi : \Omega \to \mathbb{R}$, if $h(\mu)$ is the entropy of the measure-preserving system (Ω, σ, μ) and Pressure(φ) is the thermodynamic "pressure" of the function φ then

$$h(\mu) + \int_{\Omega} \varphi \, d\mu \leq \text{Pressure}(\varphi) \ , \tag{51}$$

with strict inequality holding unless $\mu = \mu_\varphi$ is the *Gibbs state* with potential function φ. See [3], chapters 1, 2, for the relevant definitions and complete proofs. In the special cases $\varphi = \varphi_\lambda$ and $\varphi = 2\varphi_\lambda$, where $\varphi_\lambda : \Omega \to \mathbb{R}$ is defined by $\varphi_\lambda(x_1 x_2 \cdots) = \log F_{x_1}(\lambda)$, the relevant values of the pressure functional are given by

$$\text{Pressure}(\rho\varphi_\lambda) = \lim_{m\to\infty} \frac{1}{m} \log \left(\sum_{i_1 i_2 \cdots i_m} \prod_{j=1}^{m} F_{i_j}(\lambda)^\rho \right) = \log \theta(\rho; \lambda) \ . \tag{52}$$

Since the potential functions depend on only the first coordinate, the relevant Gibbs states $\mu_{\rho\varphi_\lambda}$ are the distributions of the stationary Markov chains with transition probabilities

$$(\mathbb{P}(\rho; \lambda))_{ij} = \frac{F_j(\lambda)^\rho}{\theta(\rho; \lambda)} \frac{v_j}{v_i} (1 - \delta_i(j^{-1})) \ ,$$

where v is the lead right eigenvalue of the Perron-Frobenius matrix $M(\rho; \lambda)$ and δ is the Kronecker delta function. Recall that when $\rho = 2$ and $\lambda = R$, this Gibbs state coincides with the measure μ_* of Theorem 2.

The following statement completes the proof of Theorem 2.

Proposition 9. *For every ergodic, σ-invariant probability measure μ on Ω,*

$$\delta(\lambda; \mu) \leq \frac{1}{2} \delta_H(\Omega_\mu) \tag{53}$$

almost surely, with strict inequality unless $\lambda = R$ and $\mu = \mu_*$.

Proof. Fix λ, and let $\varphi = \varphi_\lambda$. By the Gibbs Variational Principle, (52) and the formula (14) for the Hausdorff dimension $\delta(\lambda; \mu)$ of $\Lambda \cap \Omega_\mu$,

$$\begin{aligned} h(\mu) + 2 \int \varphi \, d\mu &\leq \log \theta(2; \lambda) \implies \\ h(\mu) + \int \varphi \, d\mu &\leq \frac{1}{2}(h(\mu) + \log \theta(2; \lambda)) \implies \\ \delta(\lambda; \mu) &\leq -\frac{1}{2}(h(\mu) + \log \theta(2; \lambda)) / \log \alpha; \end{aligned}$$

moreover, strict inequality holds unless μ coincides with the Gibbs state $\mu_{2\varphi}$ with potential function 2φ. Recall that $\theta(2;\lambda)$ is strictly increasing in λ, since the generating functions $F_i(\lambda)$ are strictly increasing, and, by Corollary 2, $\theta(2;R)=1$. Hence,

$$\delta(\lambda;\mu) \leq -\frac{1}{2}h(\mu)/\log\alpha$$

with strict inequality unless $\lambda = R$ and $\mu = \mu_*$. Since the Hausdorff dimension of Ω_μ in the metric d_α is $-h(\mu)/\log\alpha$, by a well known theorem of Billingsley, the proposition follows. $\square$

Together with the result of Corollary 9, the following proposition completes the proof of Theorem 1.

Proposition 10. *Strict inequality holds in* (50) *at* $\lambda = R$ *except when the branching random walk is isotropic, i.e., when the step distribution of the marginal random walk satisfies* $p_i = p_j$ *for all* $i, j \in \mathscr{A}$.

Proof. It suffices to prove that

$$\theta(R)^2 < \theta(2;R)(2d-1) \tag{54}$$

except when the branching random walk is isotropic. By formula (52) and the Gibbs Variational Principle (51), if μ_ρ is the Gibbs state with potential function $\rho\varphi_R$ then

$$2\log\theta(1;R) = 2h(\mu_1) + 2\int \varphi_R\, d\mu_1 \text{ and} \tag{55}$$

$$\log\theta(2;R) \geq h(\mu_1) + 2\int \varphi_R\, d\mu_1 \;, \tag{56}$$

with strict inequality in (56) except when the Gibbs states μ_1 and μ_2 coincide. By Theorem 1.28 of [3], this occurs when the difference of the potential functions φ_R and $2\varphi_R$ is (co-)homologous to a constant function, equivalently, if and only if there exist constants $C > 0$ and $\varepsilon > 0$ such that

$$\varepsilon < \frac{1}{C^m}\frac{\prod_{j=1}^m F_{i_j}(R)^2}{\prod_{j=1}^m F_{i_j}(R)} < \frac{1}{\varepsilon}$$

for every $m \geq 1$ and all finite reduced words $i_1 i_2 \cdots i_m$ of length m from the alphabet $\mathscr{A}$. Clearly, this occurs if and only if all of the values F_i, $i \in \mathscr{A}$, are the same. But by Proposition 2, this is the case if and only if the branching random walk is isotropic. $\square$

8. Appendix: subcritical Galton–Watson processes

Proposition 11. *Let $Z_n = (Z_n^{(1)}, Z_n^{(2)}, \dots, Z_n^{(k)})$ be a subcritical multi-type Galton–Watson process with type set $\mathscr{K} = \{1, 2, \dots, k\}$ and offspring distributions $(F_{ij})_{i,j\in\mathscr{K}}$. Let P^i, E^i denote the probability and expectation operators under which $Z_0^{(j)} = \delta_{ij}$ a.s., where δ is Kronecker's delta. If each offspring distribution F_{ij} has finite variance, then for each $i \in \mathscr{K}$,*

$$E^i \left(\sum_{n=0}^{\infty} \sum_{i\in\mathscr{K}} Z_n^{(i)} \right)^2 < \infty \ . \tag{57}$$

Proof. First we shall give the proof for Galton–Watson processes with only one particle type, after which we shall indicate how the result for multi-type processes may be deduced from the corresponding result for Galton–Watson processes with only one particle type.

(1) Let Z_n be a subcritical Galton–Watson process for which the offspring distribution has mean $\mu < 1$ and variance $\sigma^2 < \infty$. Then

$$\begin{aligned}
E(\sum_{n=0}^{\infty} Z_n)^2 &\le 2E \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} Z_n Z_{m+n} \\
&= 2E \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} Z_n^2 \mu^m \\
&= 2(1-\mu)^{-1} \sum_{n=0}^{\infty} E Z_n^2 \\
&= 2(1-\mu)^{-1} \sum_{n=0}^{\infty} (\mu^{2n} + \frac{\sigma^2 \mu^{n-1}(1-\mu^n)}{1-\mu}) \\
&< \infty \ .
\end{aligned}$$

(2) To prove the result for an arbitrary subcritical multi-type Galton–Watson process $(Z_n)_{n\ge 0}$ it suffices to prove it for the multi-type Galton–Watson process $(Z_{nm})_{n\ge 0}$, where $m \ge 1$ is a fixed but arbitrary positive integer, by an elementary argument using the triangle inequality. Since the original process $(Z_n)_{n\ge 0}$ is subcritical, there exists a constant $\rho < 1$ such that for all sufficiently large $m \ge 1$,

$$\sum_{i\in\mathscr{K}} \sum_{j\in\mathscr{K}} E^i Z_m^{(j)} \le \rho^m \tag{58}$$

(see, e.g., [1], section V.2). Let ξ have the same distribution as the convolution of the distributions of $Z_m^{(j)}$ under the measures $P^{(i)}$, $i \in \mathscr{K}$. Then by (58), $E\xi < 1$; and since by hypothesis all of the distributions F_{ij} have finite variances, so does ξ. Clearly ξ stochastically dominates the total number of particles in the mth generation under any of the probability measures $P^{(i)}$. It follows that if Y_n is the Galton–Watson process with offspring distribution equal to the law of ξ then

$$E^{(i)}(\sum_{n=0}^{\infty}\sum_{i\in\mathscr{K}} Z_{nm}^{(i)})^2 \leq E(\sum_{n=0}^{\infty} Y_n)^2 < \infty \quad . \qquad \square$$

Proposition 12. *Let* $\{Z_n\}_{n\geq 0}$ *be a subcritical Galton–Watson process whose offspring distribution has finite second moment, expectation* $0 < \mu < 1$, *and support containing* $\{1\}$. *Then*

$$\lim_{n\to\infty} \frac{P\{Z_n = 1\}}{\mu^n} > 0 \tag{59}$$

Note. A well known theorem of YAGLOM (see [1], section 1.8) states that for any subcritical Galton–Watson process Z_n, the conditional distribution of Z_n given that $Z_n \geq 1$ converges weakly to a limit as $n \to \infty$.

Proof of Proposition 12 . Let $\varphi(s)$ be the moment generating function of the offspring distribution, and let $\varphi_n(s)$ be its nth iterate. Then $EZ_n = \mu^n = \varphi_n'(1)$ and

$$P\{Z_n = 1\} = \varphi_n'(0) = \prod_{k=0}^{n-1} \varphi'(\varphi_k(0)) \quad .$$

Since the offspring distribution attaches positive probability to 1, the derivative $\varphi'(s) \geq \varphi'(0) > 0$ for all $s \in [0,1]$. Since the offspring distribution has a finite second moment, $\varphi''(s) \leq \varphi''(1) < \infty$ for all $s \in [0,1]$. Since $\mu = \varphi'(1) < 1$, the iterates $\varphi_k(0)$ converge to 1 exponentially fast, and so $\sum_{k=0}^{\infty}(1-\varphi_k(0)) < \infty$. Consequently,

$$\sum_{k=1}^{\infty}(\varphi'(1) - \varphi'(\varphi_k(0))) \leq \varphi''(1)\sum_{k=1}^{\infty}(1-\varphi_k(0)) < \infty \quad ,$$

and so Weierstrass' theorem on infinite products implies that

$$\lim_{n\to\infty} \frac{P\{Z_n = 1\}}{\mu^n} = \lim_{n\to\infty} \prod_{k=0}^{n-1} \frac{\varphi'(\varphi_k(0))}{\varphi'(1)}$$

exists and is positive. $\square$

The next result is a generalization of Proposition 12 to Galton–Watson processes in varying environments. Let $\mathscr{G} = \{G_1, G_2, \dots, G_k\}$ be a finite set of offspring distributions with moment generating functions $\varphi_j(s)$. An *environment* is a semi-infinite sequence $\xi = i_1 i_2 \cdots$ valued in the set of indices $[k] = \{1, 2, \dots, k\}$. A *Galton–Watson process with environment* $\xi = i_1 i_2 \cdots$ is a sequence of nonnegative integer-valued random variables Z_n such that (i) $Z_0 = 1$, and (ii) for each $s \in [0,1]$,

$$E_\xi(s^{Z_{n+1}} \mid Z_0, Z_1, \dots, Z_n) = \varphi_{i_{n+1}}(s)^{Z_n} \quad . \tag{60}$$

Proposition 13. *Assume that each of the offspring distributions G_j satisfies the hypotheses of Proposition 12: in particular, its mean satisfies $\mu_j < 1$; its variance is finite; and its support contains* 1*. Assume that under P_ξ the sequence Z_n is a Galton–Watson process with environment $\xi = i_1 i_2 \cdots$. Then*

$$\lim_{n\to\infty} \frac{P_\xi\{Z_n = 1\}}{\prod_{j=1}^{n} \mu_{i_j}} \stackrel{\Delta}{=} \alpha(\xi) \tag{61}$$

uniformly for $\xi \in [k]^{\mathbb{N}}$, and

$$\inf_{\xi \in [k]^{\mathbb{N}}} \alpha(\xi) > 0 \quad . \tag{62}$$

Proof. For each environmental sequence $\xi = i_1 i_2 \cdots$, define a nondecreasing sequence of real numbers $s_n = s_n^\xi \in [0, 1]$ inductively by $s_0 = 0$ and $s_n = \varphi_{i_n}(s_{n-1})$. As in the proof of Proposition 12,

$$\frac{P_\xi\{Z_n = 1\}}{\prod_{j=1}^{n} \mu_{i_j}} = \prod_{j=1}^{n} \frac{\varphi'_{i_j}(s_j^\xi)}{\varphi'_{i_j}(1)} \tag{63}$$

Let $\psi(s) = as + 1 - a$ be the linear function with slope $a = \max_j \mu_j < 1$ that satisfies $\psi(1) = 1$. Then each of the generating functions $\varphi_j(s)$ satisfies $\varphi_j(s) \geq \psi(s)$ for all $s \in [0, 1]$, because each φ_j is convex and satisfies $\varphi_j(1) = 1$ and $\varphi'_j(1) = \mu_j \leq a = \psi'(1)$. Consequently, if s_n^* is the nondecreasing sequence defined by $s_0^* = 0$ and $s_n^* = \psi(s_{n-1}^*)$, then $s_n^\xi \geq s_n^*$ for every environmental sequence ξ and every $n \geq 1$. It is routine to verify that $s_n^* = 1 - a^n$, so the sequence s_n^* converges to 1 exponentially fast. It follows that s_n^ξ converges to 1 exponentially fast *uniformly in* ξ; in particular, for each $m \geq 1$ and each ξ,

$$\sum_{n=m}^{\infty} (1 - s_n^\xi) \leq \sum_{n=m}^{\infty} (1 - s_n^*) = \sum_{n=m}^{\infty} a^n \quad . \tag{64}$$

By hypothesis, each of the distributions G_j has finite second moment, so the second derivative $\varphi''_j(1)$ is finite. Moreover, each of the distributions G_j attaches positive mass to 1, so $\varphi'_j(0) > 0$; since each φ_j is convex, it follows that the functions $\varphi'_j(s)$ are uniformly bounded away from 0. Denote by $\varepsilon > 0$ the infimum. Then for each $s \in [0, 1]$ and each $j \in [k]$,

$$0 \leq 1 - \frac{\varphi'_j(s)}{\varphi'_j(1)} \leq \frac{1}{\varepsilon} \max_{i \in [k]} \varphi''_i(1)(1 - s) \quad . \tag{65}$$

The convergence in (61) follows from (63)–(65). Uniformity in ξ follows from (64), as does the assertion (62). ◻

References

1. Athreya, K., Ney, P.: Branching Processes. Springer-Verlag (1972)
2. Benjamini, I., Peres, Y.: Markov chains indexed by trees. Annals of Probability **22**, 219–243 (1994)
3. Bowen, R.: Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Springer Lecture Notes in Math. **470**, (1975)
4. Gerl, P., Woess, W.: Local limits and harmonic functions for nonisotropic random walks on free groups. Probability Theory and Related Fields **71**, 341–355 (1986)
5. Hawkes, J.: Trees generated by a simple branching process. J. London Math. Soc.(2) **24**, 373–384 (1981)
6. Kendig, K.: Elementary Algebraic Geometry. Springer-Verlag (1977)
7. Lalley, S.: Finite range random walk on a homogeneous tree. Annals of Probability **21**, 2087–2130 (1993)
8. Lalley, S.: Percolation on Fuchsian Groups. Annales Inst. H. Poincaré Probab. Statist **34**, 151–177 (1996)
9. Lalley, S., Sellke, T.: Hyperbolic branching Brownian motion. Probability Theory and Related Fields **108**, 171–192 (1997)
10. Lalley, S., Sellke, T.: Limit set of a weakly supercritical contact process on a homogeneous tree. Annals of Probability **6**, 644–657 (1998)
11. Lalley, S., Sellke, T.: An extension of Hawkes' theorem on the Hausdorff dimension of a Galton–Watson tree. To appear, Probability Theory and Related Fields (MS1246) (1998)
12. Lang, S.: Algebra, 2nd ed. Addison-Wesley, Reading MA (1984)
13. Liggett, T.: Multiple transition points for the contact process on the binary tree. Annals of Probability **24**, 1675–1710 (1996)
14. Liggett, T.: Branching random walks and contact processes on homogeneous trees. Probability Theory and Related Fields **106**, 495–519 (1996)
15. Lyons, R.: Random walks and percolation on trees. Annals of Probability **18**, 931–958 (1990)
16. Pemantle, R.: The contact process on trees. Annals of Probability **20**, 2089–2116 (1992)
17. Stacey, A.: The existence of an intermediate phase for the contact process on trees. Annals of Probability **24**, 1711–1726 (1996)