A Central Limit Theorem for biased random walks on Galton-Watson trees

Yuval Peres* Ofer Zeitouni†

June 22, 2006

Abstract

Let $\mathcal{T}$ be a rooted Galton-Watson tree with offspring distribution $\{p_k\}$ that has $p_0 = 0$, mean $m = \sum kp_k > 1$ and exponential tails. Consider the λ-biased random walk $\{X_n\}_{n\geq 0}$ on $\mathcal{T}$; this is the nearest neighbor random walk which, when at a vertex v with d_v offspring, moves closer to the root with probability $\lambda/(\lambda + d_v)$, and moves to each of the offspring with probability $1/(\lambda+d_v)$. It is known that this walk has an a.s. constant speed $\mathtt{v} = \lim_n |X_n|/n$ (where $|X_n|$ is the distance of X_n from the root), with $\mathtt{v} > 0$ for $0 < \lambda < m$ and $\mathtt{v} = 0$ for $\lambda \geq m$. For all $\lambda \leq m$, we prove a quenched CLT for $|X_n| - n\mathtt{v}$. (For $\lambda > m$ the walk is positive recurrent, and there is no CLT.) The most interesting case by far is $\lambda = m$, where the CLT has the following form: for almost every $\mathcal{T}$, the ratio $|X_{[nt]}|/\sqrt{n}$ converges in law as $n \to \infty$ to a deterministic multiple of the absolute value of a Brownian motion. Our approach to this case is based on an explicit description of an invariant measure for the walk from the point of view of the particle (previously, such a measure was explicitly known only for $\lambda = 1$) and the construction of appropriate harmonic coordinates.

AMS Subject classification: primary 60K37, 60F05. Secondary 60J80, 82C41.

1 Introduction and statement of results

Let $\mathcal{T}$ be a rooted Galton-Watson tree with offspring distribution $\{p_k\}$. That is, the numbers of offspring d_v of vertices $v \in \mathcal{T}$ are i.i.d. random variables, with $P(d_v = k) = p_k$. Throughout this paper, we assume that $p_0 = 0$, and that $m := \sum kp_k > 1$. In particular, $\mathcal{T}$ is almost surely an infinite tree. For technical reasons, we also assume the existence of exponential moments, that is the existence of some $\beta > 1$ such that $\sum \beta^k p_k < \infty$. We let $|v|$ stand for the distance of a vertex v from the root of $\mathcal{T}$, and let o denote the root of $\mathcal{T}$.

*Dept. of Mathematics and Dept. of Statistics, University of California, Berkeley. Research partially supported by MSRI and by NSF grants #DMS-0104073 and #DMS-0244479

†Department of Mathematics, University of Minnesota, and Depts. of Mathematics and of Electrical Engineering, Technion. Research partially supported by MSRI and by NSF grants #DMS-0302230 and DMS-0503775.

We are interested in λ-biased random walks on the tree $\mathcal{T}$. These are Markov chains $\{X_n\}_{n\geq 0}$ with $X_0 = o$ and transition probabilities

$$P_{\mathcal{T}}(X_{n+1} = w | X_n = v) = \begin{cases} \lambda/(\lambda + d_v)\,, & \text{if } v \text{ is an offspring of } w\,, \\ 1/(\lambda + d_v)\,, & \text{if } w \text{ is an offspring of } v\,. \end{cases}$$

Let GW denote the law of Galton-Watson trees. Lyons [13] showed that

- If $\lambda > m$, then for GW-almost every $\mathcal{T}$, the random walk $\{X_n\}$ is positive recurrent.
- if $\lambda = m$, then for GW-almost every $\mathcal{T}$, the random walk $\{X_n\}$ is null recurrent.
- if $\lambda < m$, then for GW-almost every $\mathcal{T}$, the random walk $\{X_n\}$ is transient.

In the latter case, $\lambda < m$, it was later shown in [16] and [17] that $|X_n|/n \to \mathtt{v} > 0$ almost surely, with a deterministic $\mathtt{v} = \mathtt{v}(\lambda)$ (an explicit expression for $\mathtt{v}$ is known only for $\lambda = 1$).

Our interest in this paper is mainly in the critical case $\lambda = m$. Then, $|X_n|/n$ converges to 0 almost surely. Our main result is the following.

Theorem 1 *Assume $\lambda = m$. Then, there exists a deterministic constant $\sigma^2 > 0$ such that for* GW*-almost every $\mathcal{T}$, the processes $\{|X_{\lfloor nt \rfloor}|/\sqrt{\sigma^2 n}\}_{t\geq 0}$ converges in law to the absolute value of a standard Brownian motion.*

Theorem 1 is proved in Section 6 by coupling λ-biased walks on GW trees to λ-biased walks on auxiliary trees, which have a marked ray emanating from the root. The ergodic theory of walks on such trees turns out (in the special case of $\lambda = m$) to be particularly nice. We develop this model and state the Central Limit Theorem (CLT) for it, Theorem 2, in Section 2. The proof of Theorem 2, which is based on constructing appropriate martingales and controlling the associated corrector, is developed in Sections 3, 4 and 5.

We conclude by noting that when $\lambda > m$, the biased random walk is positive recurrent, and no CLT limit is possible. On the other hand, [17] proved that when $\lambda < m$ and the walk is transient, there exists a sequence of stationary regeneration times. Analyzing these regeneration times, one deduces a quenched invariance principle with a proper deterministic centering, see Theorem 3 in Section 7 for the statement. We note in passing that this improves the annealed invariance principle derived in [20] for $\lambda = 1$.

2 A CLT for trees with a marked ray

We consider infinite trees $\mathcal{T}$ with one (semi)-infinite directed path, denoted Ray, starting from a distinguished vertex, called the **root** and denoted o. For vertices $v, w \in \mathcal{T}$, we let $d(v, w)$ denote the length of the (unique) geodesic connecting v and w (we consider the geodesic as containing both v and w, and its length as the number of vertices in it minus one). A vertex w is an offspring of a vertex v if

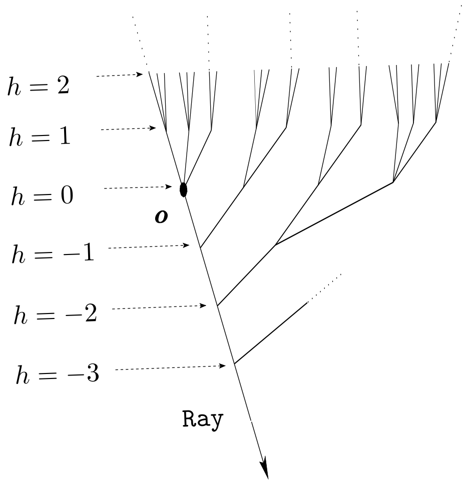


Figure 1: Tree, Ray and horocycle distance

$d(v,w)=1$ and either $d(w,\texttt{Ray})>d(v,\texttt{Ray})$ or $v,w\in\texttt{Ray}$ and $d(v,o)>d(w,o)$. In particular, the root is an offspring of its unique neighbor on Ray. For any vertex $v\in\mathcal{T}$, we let d_v denote the number of offspring of v.

For v a vertex in $\mathcal{T}$, let $R_v\in\texttt{Ray}$ denote the intersection of the geodesic connecting v to Ray with Ray, that is $d(v,R_v)=d(v,\texttt{Ray})$. For $v_1,v_2\in\mathcal{T}$, let $h(v_1,v_2)$ denote the horocycle distance between v_1 and v_2 (possibly negative), which is defined as the unique function $h(v_1,v_2)$ which equals to $d(x,v_2)-d(x,v_1)$ for all vertices x such that both v_1 and v_2 are descendants of x. (A vertex $w\in\mathcal{T}$ is a descendant of v if the geodesic connecting w to v contains an offspring of v.) We also write $h(v)=h(o,v)$; The quantity $h(v)$, which may be either positive or negative, is the *level* to which v belongs, see Figure 1. Let $D_n(v)$ denote the descendants of v in $\mathcal{T}$ at distance n from v. Explicitly,

$$D_n(v)=\{w\in\mathcal{T}:d(w,v)=h(w)-h(v)=n\}\,. \tag{1}$$

We let $Z_n(v)=|D_n(v)|$ be the number of descendants of v at level $h(v)+n$. Then $\{Z_n(v)/m^n\}_{n\geq 1}$ forms a martingale and converges a.s., as $n\to\infty$, to a random variable denoted W_v. Moreover, W_v has exponential tails, and there are good bounds on the rate of convergence, see [1].

Motivated by [15], we next describe a measure on the collection of trees with marked rays, which we denote by IGW. Fix a vertex o (the root) and a semi-infinite ray, denoted Ray, emanating from it. Each vertex $v\in\texttt{Ray}$ with $v\neq o$ is assigned independently a size-biased number of offspring, that is $P_{\texttt{IGW}}(d_v=k)=kp_k/m$, one of which is identified with the descendant of v on Ray. To each offspring of $v\neq o$ not on Ray, and to o, one attaches an independent Galton-Watson tree of offspring distribution $\{p_k\}_{k\geq 1}$. The resulting random tree $\mathcal{T}$ is distributed according to IGW. An alternative characterization of IGW is obtained as follows, see [15] for a similar construction.

Lemma 1 *Consider the measure Q_n on rooted trees with root r, obtained from* $\mathtt{GW}$ *by size-biasing with respect to $|D_n(r)|$ (that is, $dQ_n/d\mathtt{GW} = |D_n(r)|/m^n$). Choose a vertex $o \in D_n(r)$ uniformly, creating a (finite) ray from o to the root of the original tree, and extend the ray from r to obtain an infinite ray, creating thus a random rooted tree with marked ray emanating from the new root o. Call* $\mathtt{IGW}_n$ *the distribution thus obtained. Then,* $\mathtt{IGW}$ *is the weak limit of* $\mathtt{IGW}_n$.

Sometimes, we also need to consider trees where the root has no ancestors. Often, these will be distributed according to the Galton-Watson measure $\mathtt{GW}$. There is however another important measure that we will use, described in [15], namely the **size-biased** measure $\widehat{\mathtt{GW}}$ corresponding to $\mathtt{GW}$. It is defined formally by $d\widehat{\mathtt{GW}}/d\mathtt{GW} = W_o$. An alternative construction of $\widehat{\mathtt{GW}}$ is by sampling, size-biased, a particular trunk.

We let $\{X_n\}$ denote the λ-biased random walk on the tree $\mathcal{T}$, where $\lambda = m$. Explicitly, given a tree $\mathcal{T}$, X_n is a Markov process with $X_0 = o$ and transition probabilities

$$P_{\mathcal{T}}(X_{n+1} = u | X_n = v) = \begin{cases} \lambda/(\lambda + d_v)\,, & \text{if } \ 1 = d(u,v) = h(u,v) \\ 1/(\lambda + d_v)\,, & \text{if } 1 = d(u,v) = h(v,u) \\ 0\,, & \text{else}\,. \end{cases}$$

That is, the walker moves with probability $\lambda/(\lambda + d_v)$ toward the ancestor of v and with probability $1/(\lambda + d_v)$ toward any of the offspring of v. We recall that the model of λ-biased random walk on a rooted tree is reversible, and possesses an electric network interpretation, where the conductance between $v \in D_n(o)$ and an offspring $w \in D_{n+1}(o)$ of v is λ^{-n} (see e.g. [14] for this representation, and [9] for general background on reversible random walks interpreted in electric networks terms). With a slight abuse of notation, we let $P^v_{\mathcal{T}}$ denote the law, conditional on the given tree $\mathcal{T}$ and $X_0 = v$, on the path $\{X_n\}$. We refer to this law as the *quenched* law. Our main result for the $\mathtt{IGW}$ trees is the following.

Theorem 2 *Under* $\mathtt{IGW}$*, the horocycle distance satisfies a quenched invariance principle. That is, for some deterministic $\sigma^2 > 0$ (see (10) below for the value of σ), for* $\mathtt{IGW}$*-a.e. $\mathcal{T}$, the processes $\{h(X_{\lfloor nt \rfloor})/\sqrt{\sigma^2 n}\}_{t \geq 0}$ converge in distribution to a standard Brownian motion.*

3 Martingales, stationary measures, and proof of Theorem 2

The proof of Theorem 2 takes the bulk of this paper. We describe here the main steps.

- In a first step, we construct in this section a martingale M_t, whose increments consist of the normalized population size $W_{X_{t+1}}$ when $h(X_{t+1}) - h(X_t) = 1$ and $-W_{X_t}$ otherwise. (Thus, the increments of the martingale depend on the "environment as seen from the particle"). This martingale provides "harmonic coordinates" for the random walk, in the spirit of [12] and, more recently, [21] and [3].

- In the next step, we prove an invariance principle for the martingale M_t. This involves proving a law of large numbers for the associated quadratic variation. It is at this step that it turns out that `IGW` is not so convenient to work with, since the environment viewed from the point of view of the particle is not stationary under `IGW`. We thus construct a small modification of `IGW`, called `IGWR`, which is a reversing measure for the environment viewed from the point of view of the particle, and is absolutely continuous with respect to `IGW` (see Lemma 2). This step uses crucially that $\lambda = m$. Equipped with the measure `IGWR`, it is then easy to prove an invariance principle for M_t, see Corollary 1.

- In the final step, we introduce the corrector Z_t, which is the difference between a constant multiple $1/\eta$ of the harmonic coordinates M_t and the position of the random walk, X_t. As in [3], we seek to show that the corrector is small, see Proposition 1. The proof of Proposition 1 is postponed to Section 4, and is based on estimating the time spent by the random walk at any given level.

In the sequel (except in Section 6), we often use the letters s, t to denote time, reserving the letter n to denote distances on the tree $\mathcal{T}$. Set $M_0 = 0$ and, if $X_t = v$ for a vertex v with parent u and offspring $Y_1, \ldots, Y_{d_v}$, set

$$M_{t+1} - M_t = \begin{cases} -W_v, & X_{t+1} = u \\ W_{Y_j}, & X_{t+1} = Y_j \,. \end{cases}$$

Quenched (i.e., given the realization of the tree), M_t is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$, as can be seen by using the relation $W_v = \sum_{j=1}^{d_v} W_{Y_j}/m$. Also, for $v \in \mathcal{T}$, let g_v denote the geodesic connecting v with `Ray` (which by definition contains both v and R_v), and set

$$S_v = \begin{cases} \sum_{u \in g_v, u \neq o} W_u, & \text{if}\, R_v = o\,, \\ \sum_{u \in g_v, u \neq R_v} W_u - \sum_{u \in \mathtt{Ray}, 0 \geq h(u) > h(R_v)} W_u, & \text{if}\, R_v \neq o\,. \end{cases}$$

Then, $M_t = S_{X_t}$.

Set $\eta = E_{GW} W_o^2 (= E_{\widehat{GW}} W_o)$ and $Z_t = M_t/\eta - h(X_t)$. Fix

$$\alpha = 1/3, \epsilon_0 < 1/100\,, \delta \in (1/2 + \alpha + 4\epsilon_0, 1 - 4\epsilon_0). \tag{2}$$

(The reason for the particular choice of constants here will become clearer in the course of the proof.) For any integer t, let τ_t denote an integer valued random variable, independent of $\mathcal{T}$ and $\{X_s\}_{s \geq 0}$, uniformly chosen in $[t, t + \lfloor t^\delta \rfloor]$. We prove in Section 4 the following estimate, which shows that M_t/η is close to $h(X_t)$. The variable τ_t is introduced here for technical reasons as a smoothing device, that allows us to consider occupation measures instead of pointwise in time estimates on probabilities.

Proposition 1 *With the above notation, for any $\epsilon < \epsilon_0$,*

$$\lim_{t\to\infty} P^o_{\mathcal{T}}(|Z_{\tau_t}| \geq \epsilon\sqrt{t}) = 0\,, \quad \texttt{IGW} - a.s. \tag{3}$$

Further,

$$\lim_{t\to\infty} P^o_{\mathcal{T}}(\sup_{r,s\leq t, |r-s|<t^\delta} |h(X_r) - h(X_s)| > t^{1/2-\epsilon}) = 0\,,\ \texttt{IGW} - a.s. \tag{4}$$

The interest in the martingale M_t is that we can prove for it a full invariance principle. Toward this end, one needs to verify that the normalized quadratic variation process

$$V_t = \frac{1}{t}\sum_{i=1}^{t} E^o_{\mathcal{T}}\left((M_{i+1} - M_i)^2 | \mathcal{F}_i\right) \tag{5}$$

converges `IGW`-a.s. Note that if $X_i = v$ with offspring $Y_1, \ldots, Y_{d_v}$ then

$$\begin{aligned} E^o_{\mathcal{T}}\left[(M_{i+1} - M_i)^2 | \mathcal{F}_i\right] &= \frac{m}{m+d_v} W_v^2 + \frac{1}{m+d_v}\sum_{j=1}^{d_v} W_{Y_j}^2 \qquad (6) \\ &= \frac{1}{m+d_v}\sum_{j=1}^{d_v} W_{Y_j}^2 + \frac{1}{m(m+d_v)}\left(\sum_{j=1}^{d_v} W_{Y_j}\right)^2 =: \mu_v^2\,. \end{aligned}$$

It turns out that to ensure the convergence of V_t, it is useful to introduce a new measure on trees, denoted `IGWR`, which is absolutely continuous with respect to the measure `IGW`, and such that the "environment viewed from the point of view of the particle" becomes stationary under that measure, see Lemma 2 below. The measure `IGWR` is similar to `IGW`, except at the root. The root o has an infinite path v_j of ancestors, which all possess an independent number of offspring which is size-biased, that is

$$P(d_{v_j} = k) = kp_k/m\,, \quad \text{for all}\, j, k > 0.$$

The number of offspring at the root itself is independent of the variables just mentioned, and possesses a distribution which is the average of the original and the size biased laws, that is:

$$P(d_o = k) = (m+k)p_k/(2m)\,, \quad \text{for all}\, k > 0.$$

All other vertices have the original offspring law. All these offspring variables are independent. In other words, $d\texttt{IGWR}/d\texttt{IGW} = (m+d_o)/2d_o$. Consequently, we can use the statements "`IGW`-a.s." and "`IGWR`-a.s." interchangeably.

For v a neighbor of o, let $\theta^v\mathcal{T}$ denote the tree which is obtained by shifting the location of the root to v and adding or erasing one edge from `Ray` in the only way that leaves an infinite ray emanating from the new root. We also write, for an arbitrary vertex $w \in \mathcal{T}$ with geodesic $g_w = (v_1, v_2, \ldots, v_{|w|-1}, w)$ connecting o to w, the shift $\theta^w\mathcal{T} = \theta^w \circ \theta^{v_{|w|-1}} \circ \ldots \theta^{v_1}\mathcal{T}$. Finally, we set $\mathcal{T}_t = \theta^{X_t}\mathcal{T}$. It

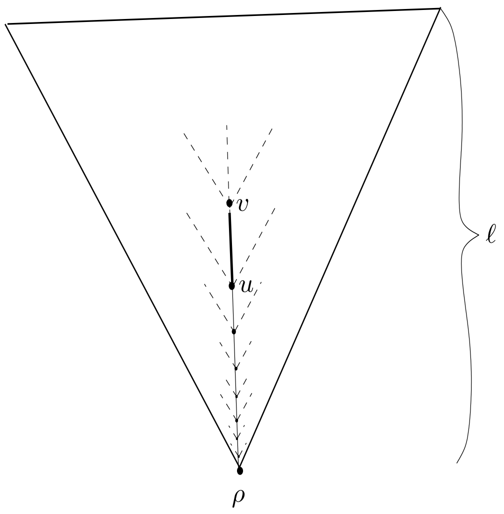


Figure 2: The finite tree $\mathcal{T}_F$

is evident that $\mathcal{T}_t$ is a Markov process, with the location of the random walk being frozen at the root, and we write $P_{\mathcal{T}}(\cdot)$ for its transition density, that is $P_{\mathcal{T}}(A) = P_{\mathcal{T}}(\mathcal{T}_1 \in A)$. What is maybe surprising at first is that `IGWR` is reversing for this Markov process. That is, we have.

Lemma 2 *The Markov process $\mathcal{T}_t$ with initial measure* `IGWR` *is stationary and reversible.*

Proof of Lemma 2 Suppose that $\mathcal{T}_0$ is picked from IGWR, and $\mathcal{T}_1$ is obtained from it by doing one step (starting with $X_0 = o$) of the critically biased walk on $\mathcal{T}_0$, then moving the root to X_1 and adjusting Ray accordingly. We must show that the ordered pair $(\mathcal{T}_0, \mathcal{T}_1)$ has the same law as $(\mathcal{T}_1, \mathcal{T}_0)$.

Let $\mathcal{T}_F$ be finite tree of depth ℓ rooted at ρ, and let u, v be adjacent internal nodes of $\mathcal{T}_F$, at distance k and $k+1$, respectively, from ρ (see figure 2).

Let $A(\mathcal{T}_F, u)$ be the cylinder set of infinite labeled rooted trees $\mathcal{T}$ in the support of `IGWR` which locally truncate to $\mathcal{T}_F$ rooted at u, that is, the connected component of the root of $\mathcal{T}$ among levels between $-k$ and $\ell - k$ in $\mathcal{T}$ is identical to $\mathcal{T}_F$ once the root of $\mathcal{T}$ is identified with u, and `Ray` in $\mathcal{T}$ goes through the vertex identified with ρ in $\mathcal{T}$. Let $\{w : \rho \le w < u\}$ denote the set of vertices on the path from ρ (inclusive) to u (exclusive) in $\mathcal{T}_F$. Then

$$P_{\mathtt{IGWR}}[A(\mathcal{T}_F, u)] = P_{\mathtt{GW}}(\mathcal{T}_F) \prod_{\{w:\rho \le w < u\}} \left[\frac{d_w}{m} \cdot \frac{1}{d_w}\right] \frac{m + d_u}{2m}\,, \tag{7}$$

where the factors d_w/m and $(m+d_u)/(2m)$ come from the density of the $IGWR$ offspring distributions with respect to the GW offspring distribution, and the factors $1/d_w$ comes from the uniformity in the choice of `Ray`. Thus

$$P_{\mathtt{IGWR}}[A(\mathcal{T}_F, u)] = P_{\mathtt{GW}}(\mathcal{T}_F) m^{-k-1}(m + d_u)/2\,, \tag{8}$$

and similarly

$$P_{\mathtt{IGWR}}[A(\mathcal{T}_F, v)] = P_{\mathtt{GW}}(\mathcal{T}_F) m^{-k-2}(m + d_v)/2\,. \tag{9}$$

Since the transition probabilities for the critically biased random walk are $p(u, v) = 1/(m + d_u)$ and $p(v, u) = m/(m + d_v)$, we infer from (8) and (9) that

$$P_{\mathtt{IGWR}}[A(\mathcal{T}_F, u)]p(u, v) = P_{\mathtt{IGWR}}[A(\mathcal{T}_F, v)]p(v, u)$$

as required.

□

With V_t as in (5), the following corollary is of crucial importance.

Corollary 1

$$V_t \to E_{\mathtt{IGWR}}\mu_0^2 =: \sigma^2\eta^2\,,\ \mathtt{IGWR} - a.s. \tag{10}$$

Proof of Corollary 1 That IGWR is absolutely continuous with respect to IGW is obvious from the construction. By Lemma 2, IGWR is invariant and reversible under the Markov dynamics induced by the process $\mathcal{T}_t$. Thus, (10) holds as soon as one checks that $\mu_0 \in L^2(\mathtt{IGWR})$, which is equivalent to checking that with v_i denoting the offspring of o, it holds that $(\sum_{i=1}^{d_o} W_{v_i})^2 \in L^1(\mathtt{IGWR})$. This in turn is implied by $E_{\mathtt{GW}}(W_o^2) < \infty$, which holds due to [1]. □

Proof of Theorem 2 In what follows, we consider a fixed $\mathcal{T}$, with the understanding that the statements hold true for IGW almost every such tree. Due to (10) and the invariance principle for the Martingale M_t, see [4, Theorem 14.1], it holds that for IGWR almost every $\mathcal{T}$, $\{M_{\lfloor nt\rfloor}/\sqrt{\eta^2\sigma^2 n}\}_{t\geq 0}$ converges in distribution, as $n \to \infty$, to a standard Brownian motion. Further, by [4, Theorem 14.4], so does $\{M_{\tau_{nt}}/\sqrt{\eta^2\sigma^2 n}\}_{t\geq 0}$. By (3), it then follows that the finite dimensional distributions of the process $\{Y_t^n\}_{t\geq 0} = \{h(X_{\tau_{nt}})/\sqrt{\sigma^2 n}\}_{t\geq 0}$ converge, as $n \to \infty$, to those of a standard Brownian motion. On the other hand, due to (4), the sequence of processes $\{Y_t^n\}_{t\geq 0}$ is tight, and hence converges in distribution to standard Brownian motion. Applying again [4, Theorem 14.4], we conclude that the sequence of processes $\{h(X_{\lfloor nt\rfloor})/\sqrt{\sigma^2 n}\}_{t\geq 0}$ converges in distribution to a standard Brownian motion, as claimed. □

4 Proof of Proposition 1

Proof of Proposition 1 For any tree with root o, we write D_n for $D_n(o)$, c.f. (1). Recall that $E_{\widehat{\mathtt{GW}}} W_o = \eta$. For $\epsilon > 0$, let $A_n^\epsilon = A_n^\epsilon(\mathcal{T}) = \{v \in D_n : |n^{-1}S_v - \eta| > \epsilon\}$, noting that for GW or $\widehat{\mathtt{GW}}$ trees, $S_v = \sum_{u\in g_o, u\neq o} W_u$. We postpone for a moment the proof of the following.

Lemma 3 *For any $\epsilon > 0$ there exists a deterministic $\nu = \nu(\epsilon) > 0$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log P_{\widehat{\mathtt{GW}}}\left(\frac{1}{n}\log\frac{|A_n^\epsilon|}{|D_n|} > -\nu\right) \leq -\nu/2\,, \tag{11}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \log P_{\mathtt{GW}} \left(\frac{1}{n} \log \frac{|A_n^\epsilon|}{|D_n|} > -\nu \right) \le -\nu/2\,. \qquad (12)$$

Turning our attention to trees governed by the measure $\mathtt{IGW}$, for any vertex $w \in \mathcal{T}$ we set

$$S_w^{\mathtt{Ray}} = \sum_{v\in\mathcal{T}\setminus\mathtt{Ray}: v \text{ is on the geodesic connecting } w \text{ and } \mathtt{Ray}} W_v\,.$$

Let $B_n^\epsilon(\mathcal{T}) = \{w \in \mathcal{T} : d(w, \mathtt{Ray}) = n, |n^{-1}S_w^{\mathtt{Ray}} - \eta| > \epsilon\}$, and set

$$Q_t(\mathcal{T}) = \{w \in \mathcal{T} : d(w, \mathtt{Ray}) \le t^\alpha\}. \qquad (13)$$

The following proposition will be proved in Section 5.

Proposition 2

$$\limsup_{t\to\infty} P_{\mathcal{T}}^o(X_{\tau_t} \in Q_t(\mathcal{T})) = 0\,,\ \mathtt{IGW} - a.s.\,. \qquad (14)$$

We can now prove the following.

Lemma 4 *With the preceding notation, it holds that for any $\epsilon > 0$,*

$$\lim_{t\to\infty} P_{\mathcal{T}}^o(X_{\tau_t} \in \cup_m B_m^\epsilon(\mathcal{T})) = 0\,, \quad \mathtt{IGW} - a.s.$$

Proof of Lemma 4 By (14),

$$a_t := P_{\mathcal{T}}^o(X_{\tau_t} \in Q_t(\mathcal{T})) \to_{t\to\infty} 0\,,\ \mathtt{IGW} - a.s. \qquad (15)$$

Letting $\gamma_m^\epsilon = \min(t : X_t \in B_m^\epsilon(\mathcal{T}))$, we have (using $t + \lceil t^\delta \rceil \le 2t$),

$$P_{\mathcal{T}}^o(X_{\tau_t} \in \cup_m B_m^\epsilon(\mathcal{T})) \le a_t + \sum_{\ell=t^\alpha}^{2t} P_{\mathcal{T}}^o(\gamma_\ell^\epsilon \le 2t)\,. \qquad (16)$$

Consider the excursions of $\{X_i\}$ down the $\mathtt{GW}$ trees whose starting points are offspring of a vertex in $\mathtt{Ray}$, where an excursion is counted between visits to such a starting point. The event $\{\gamma_\ell^\epsilon \le 2t\}$ implies that of the first $2t$ such excursions, there is at least one excursion that reaches level $\ell - 1$ below the corresponding starting point, at a vertex v with $|\ell^{-1}S_v - \eta| > \epsilon$. Therefore, with $\tau_o = \min\{t > 0 : X_t = o\}$, for ℓ large so that $\{x > 0 : |\ell^{-1}x - \eta| > \epsilon\} \subset \{x > 0 : |(\ell-1)^{-1}x - \eta| > \epsilon/2\}$,

$$P_{\mathtt{IGW}}^o(\gamma_\ell^\epsilon \le 2t) \le 2t P_{\mathtt{GW}}^o(\bar\gamma_{\ell-1}^{\epsilon/2} \le 2t \wedge \tau_o)\,, \qquad (17)$$

where we set for a $\mathtt{GW}$ rooted tree, $\bar\gamma_\ell^{\epsilon/2} = \min\{i > 0 : X_i \in A_\ell^{\epsilon/2})\}$. But, for a $\mathtt{GW}$ rooted tree, the conductance $\mathcal{C}(o \leftrightarrow A_\ell^{\epsilon/2})$ from the root to the vertices in $A_\ell^{\epsilon/2}$

is at most $\lambda^{-\ell}|A_\ell^{\epsilon/2}|$. Note that with $Z_n := |D_n|m^{-n}$ it holds that $E_{\mathtt{GW}}(Z_n) = 1$ and

$$E_{\mathtt{GW}}(Z_{n+1}^2) = E_{\mathtt{GW}}(Z_n^2) + \frac{E_{\mathtt{GW}}(d_o^2 - d_o)}{\lambda^2}(E_{\mathtt{GW}}(Z_n))^2$$

and hence $E_{\mathtt{GW}}(Z_\ell^2) \le c\ell$ for some deterministic constant c. Therefore,

$$P^o_{\mathtt{GW}}(\bar\gamma^{\epsilon/2}_{\ell-1} \le \tau_o) \le E_{\mathtt{GW}}(\mathcal{C}(o \leftrightarrow A^{\epsilon/2}_{\ell-1})) \le E_{\mathtt{GW}}(\lambda^{-\ell+1}|A^{\epsilon/2}_{\ell-1}|) = E_{\mathtt{GW}}(Z_{\ell-1}\frac{|A^{\epsilon/2}_{\ell-1}|}{|D_{\ell-1}|})$$

$$\le [E_{\mathtt{GW}}(Z^2_{\ell-1})]^{1/2}[E_{\mathtt{GW}}((\frac{|A^{\epsilon/2}_{\ell-1}|}{|D_{\ell-1}|})^2))]^{1/2} \le e^{-\nu(\epsilon/2)\ell/4}.$$

for ℓ large, where Lemma 3 was used in the last inequality. Combined with (17), we conclude that

$$\sum_{\ell=t^\alpha}^{2t} P^o_{\mathtt{IGW}}(\gamma^\epsilon_\ell \le 2t) \le e^{-\nu(\epsilon/2)t^\alpha/8}.$$

By Markov's inequality and the Borel-Cantelli lemma, this implies that

$$\limsup_{t\to\infty} e^{\nu(\epsilon/2)t^\alpha/16} \sum_{\ell=t^\alpha}^{2t} P^o_{\mathcal{T}}(\gamma^\epsilon_\ell \le 2t) = 0\,, \quad \mathtt{IGW} - a.s.$$

Substituting in (16) and using (15), one concludes the proof of Lemma 4. □

Proof of Lemma 3 Recall the construction of the measures $\widehat{\mathtt{GW}}$ and $\widehat{\mathtt{GW}}_*$, see [15, Pg 1128]. Note that $\widehat{\mathtt{GW}}_*$ is a measure on rooted trees with a marked ray emanating from the root. We let v_n^* denote the marked vertex at distance n from the root.

By [15, (2.1),(2.2)], and denoting by $\mathcal{T}_n$ the first n generations of the tree $\mathcal{T}$, it holds that

$$\widehat{\mathtt{GW}}_*(v_n^* \in A_n^\epsilon) = E_{\widehat{\mathtt{GW}}}\left(\frac{1}{|D_n|}\sum_{v\in D_n} P_{\widehat{\mathtt{GW}}}(v \in A_n^\epsilon|\mathcal{T}_n)\right).$$

We show below that there exists $\delta_1 = \delta_1(\epsilon) > 0$ such that

$$\widehat{\mathtt{GW}}_*(v_n^* \in A_n^\epsilon) \le e^{-2\delta_1 n}. \tag{18}$$

We assume that (18) has been proved, and complete the proof of the lemma. By Markov's inequality, (18) implies that

$$P_{\widehat{\mathtt{GW}}}\left(E_{\widehat{\mathtt{GW}}}\left(\frac{|A_n^\epsilon|}{|D_n|} \mid \mathcal{T}_n\right) \ge e^{-\delta_1 n}\right) \tag{19}$$

$$= \quad P_{\widehat{\mathtt{GW}}}\left(\frac{1}{|D_n|}\sum_{v\in D_n} P_{\widehat{\mathtt{GW}}}(v \in A_n^\epsilon \mid \mathcal{T}_n) \ge e^{-\delta_1 n}\right) \le \frac{e^{-2\delta_1 n}}{e^{-\delta_1 n}} = e^{-\delta_1 n}.$$

We thus get

$$\begin{aligned}
&P_{\widehat{\mathtt{GW}}}\left(\frac{|A_n^\epsilon|}{|D_n|} > e^{-\delta_1 n/2}\right) = E_{\widehat{\mathtt{GW}}}\left(P_{\widehat{\mathtt{GW}}}\left(\frac{|A_n^\epsilon|}{|D_n|} > e^{-\delta_1 n/2} \mid \mathcal{T}_n\right)\right)\\
\leq\ & E_{\widehat{\mathtt{GW}}}\left(E_{\widehat{\mathtt{GW}}}\left(\frac{|A_n^\epsilon|}{|D_n|} \mid \mathcal{T}_n\right) e^{\delta_1 n/2}\right)\\
\leq\ & e^{-\delta_1 n/2} + e^{\delta_1 n/2} P_{\widehat{\mathtt{GW}}}\left(E_{\widehat{\mathtt{GW}}}\left(\frac{|A_n^\epsilon|}{|D_n|} \mid \mathcal{T}_n\right) \geq e^{-\delta_1 n}\right) \leq 2e^{-\delta_1 n/2},
\end{aligned}$$

where Markov's inequality was used in the first inequality and (19) in the last. This proves (11). While (12) could be proved directly, one notes that, with $r > 1$ such that $p_1 m^{r-1} < 1$,

$$\begin{aligned}
&P_{\mathtt{GW}}\left(\frac{1}{n}\log\frac{|A_n^\epsilon|}{|D_n|} > -\nu\right) = E_{\widehat{\mathtt{GW}}}\left(W_o^{-1}\mathbf{1}_{\frac{1}{n}\log\frac{|A_n^\epsilon|}{|D_n|} > -\nu}\right)\\
\leq\ & (E_{\widehat{\mathtt{GW}}} W_o^{-r})^{1/r}\left(P_{\widehat{\mathtt{GW}}}\left(\frac{1}{n}\log\frac{|A_n^\epsilon|}{|D_n|} > -\nu\right)\right)^{1-1/r},
\end{aligned}$$

where Hölder's inequality with exponent $r > 1$ was used. Since $E_{\widehat{\mathtt{GW}}} W_o^{-r} = E_{\mathtt{GW}}(W_o^{-(r-1)}) < \infty$ by [18, Theorem 1], (12) follows from (11).

It remains to prove (18). We use the following: Since

$$(E_{\widehat{\mathtt{GW}}} e^{\xi W_o})^2 \leq E_{\mathtt{GW}}(W_o^2) E_{\mathtt{GW}} e^{2\xi W_o} < \infty$$

for some $\xi > 0$, where the last inequality is due to [1], it follows that there exists a $\xi > 0$ such that

$$E_{\widehat{\mathtt{GW}}_*} e^{\xi W_o} = E_{\widehat{\mathtt{GW}}} e^{\xi W_o} < \infty. \tag{20}$$

For a marked vertex v_k^*, we let $\tilde{Z}_n^{v_k^*}$ denote the size of the subset of vertices in $D_n(v_k^*)$ whose ancestral line does not contain v_{k+1}^*, and we define $\tilde{W}_k$ as the a.s. limit (as $n \to \infty$) of $\tilde{Z}_n^{v_k^*}/m^n$, which exists by the standard martingale argument. Note that by construction, for $k < n$, with $W_k = W_{v_k^*}$,

$$W_k = \tilde{W}_k + \frac{\tilde{W}_{k+1}}{m} + \ldots + \frac{\tilde{W}_{n-1}}{m^{n-k-1}} + \frac{W_n}{m^{n-k}}. \tag{21}$$

Therefore,

$$S_{v_n^*} = \sum_{k=0}^{n-1} \tilde{W}_k C_k + W_n C_n,$$

where $C_k = 1 + 1/m + (1/m)^2 + \ldots + (1/m)^k$. Due to (20), we have the existence of a $\delta_2 > 0$ such that

$$P_{\widehat{\mathtt{GW}}_*}(|W_n C_n| > \epsilon n/4) \leq P_{\widehat{\mathtt{GW}}}(|W_o| > (1 - 1/m)\epsilon n/4) \leq e^{-\delta_2 n}. \tag{22}$$

Also,

$$P_{\widehat{\mathtt{GW}}_*}(\sum_{k=0}^{n-1}\tilde{W}_k[C_\infty - C_k] > \epsilon n/4) = P_{\widehat{\mathtt{GW}}_*}(\sum_{k=0}^{n-1}\frac{\tilde{W}_k}{m^{k+1}(1-1/m)} > \epsilon n/4)$$
$$\leq nP_{\widehat{\mathtt{GW}}}(\tilde{W}_o > c_{\epsilon,m}n) \leq e^{-\delta_2 n}\,, \qquad (23)$$

for some constant $c_{\epsilon,m}$, where (20) was used in the second inequality. On the other hand,

$$\eta = E_{\widehat{\mathtt{GW}}_*}W_k = E_{\widehat{\mathtt{GW}}}[C_\infty \tilde{W}_0]\,,$$

where the first equality follows from the construction of $\widehat{\mathtt{GW}}$ and the definition of η, and the second from (21). The random variables $\tilde{W}_k$ are i.i.d. by construction under $\widehat{\mathtt{GW}}_*$. Therefore, using (22) and (23),

$$P_{\widehat{\mathtt{GW}}_*}\left(\left|\frac{S_{v_n^*}}{n} - \eta\right| > \epsilon\right) \leq 2e^{-\delta_2 n} + P_{\widehat{\mathtt{GW}}_*}\left(\frac{1}{n}\sum_{k=0}^{n-1}[C_\infty \tilde{W}_k - \eta] > \frac{\epsilon}{2}\right).$$

Standard large deviations (applied to the sum of i.i.d. random variables $\tilde{W}_k$ that possess exponential moments) together with (20) now yield (18) and complete the proof of Lemma 3. □

Continuing with the proof of Proposition 1, let v_n denote the vertex on $\mathtt{Ray}$ with $h(v_n) = -n$. By the same construction as in the course of the proof of Lemma 3, it holds that

$$S_{v_n}/n \to_{n\to\infty} -\eta\,, \quad \mathtt{IGW}-a.s.. \qquad (24)$$

Let $R_t = R_{X_{\tau_t}}$. Note that $S_{X_{\tau_t}} = -S_{R_t} + S^{\mathtt{Ray}}_{X_{\tau_t}}$. Thus,

$$|Z_{\tau_t}| \leq |S_{R_t}/\eta + |h(R_t)|| + |S^{\mathtt{Ray}}_{X_{\tau_t}}/\eta - h(R_t, X_{\tau_t})|\,.$$

Note that since the random walk restricted to $\mathtt{Ray}$ is transient, $h(R_t) \to_{t\to\infty} -\infty$, and hence by (24), $S_{R_t}/\eta|h(R_t)| \to -1$. Therefore, for any positive ϵ_1, for all large t, using that $\tau_t \leq 2t$, it follows that $|S_{R_t}/\eta + |h(R_t)|| \leq \epsilon_1 \sup_{s\leq 2t}|M_s|$. Similarly, for any $\epsilon_1 < \epsilon$, on the event $X_{\tau_t} \notin \cup_m B_m^{\epsilon_1}(\mathcal{T})$, it holds that for large t, $|S^{\mathtt{Ray}}_{X_{\tau_t}}/\eta - h(R_t, X_{\tau_t})| \leq \sup_{s\leq 2t}|M_s|$ for all t large. Thus, for such ϵ_1, $|Z_{\tau_t}| \leq 2\epsilon_1 \sup_{s\leq 2t}|M_s|$ for all t large. From Lemma 4,

$$\limsup_{t\to\infty} P^o_{\mathcal{T}}(X_{\tau_t} \in \cup_m B_m^{\epsilon_1}(\mathcal{T}) = 0\,. \qquad (25)$$

But, since the normalized increasing process V_t is $\mathtt{IGWR}$-a.s. bounded, standard Martingale inequalities imply that

$$\lim_{\epsilon_1\to 0}\limsup_{t\to\infty} P^o_{\mathcal{T}}(\sup_{s\leq t}|M_t| > \epsilon\sqrt{t}/2\epsilon_1) = 0\,.$$

It follows that

$$\lim_{t\to\infty} P^o_{\mathcal{T}}(|Z_{\tau_t}| \ge \epsilon\sqrt{t}) = 0\,,$$

as claimed.

The proof of (4) is provided in Section 5, see (35). This completes the proof of Proposition 1. □

5 Auxiliary computations and proof of (4)

We begin by an a-priori annealed estimate on the displacement of the random walk in a GW tree.

Lemma 5 *For any $u, t \ge 1$, it holds that*

$$P^o_{\mathtt{GW}}(|X_i| \ge u \text{ for some } i \le t) \le 4te^{-u^2/2t}\,. \tag{26}$$

Proof of Lemma 5 Throughout, we write $|v| = d(v, o)$. Let $\mathcal{T}_u$ denote the truncation of the tree $\mathcal{T}$ at level u, and let $\mathcal{T}^*$ denote the graph obtained from $\mathcal{T}_u$ by adding an extra vertex (denoted o^*) and connecting it to all vertices in D_u. Let X^*_s denote the random walk on $\mathcal{T}^*$, with

$$P_{\mathcal{T}}(X^*_{i+1} = w | X^*_i = v) = \begin{cases} P_{\mathcal{T}}(X_{i+1} = w | X_i = v), & \text{if } v \notin D_u, \\ 1/2, & \text{if } v \in D_u \text{ and } d(v,w) = 1, \\ 1/|D_u|, & \text{if } v = o^* \text{ and } d(v,w) = 1\,. \end{cases}$$

Then,

$$\begin{aligned} & P^o_{\mathtt{GW}}(|X_i| \ge u \text{ for some } i \le t) = P^o_{\mathtt{GW}}(|X^*_i| = u \text{ for some } i \le t) \\ \le\ & \sum_{i=1}^{t} P^o_{\mathtt{GW}}(|X^*_i| = u) \le 2\sum_{i=1}^{t+1} P^o_{\mathtt{GW}}(|X^*_i| = o^*)\,. \end{aligned} \tag{27}$$

By the Carne-Varopoulos bound, see [7, 23], [14, Theorem 12.1],

$$P^o_{\mathcal{T}}(|X^*_i| = o^*) \le 2\sqrt{\lambda^{-u}|D_u|/d_o}\, e^{-u^2/2i}\,.$$

Hence, since $E_{GW}|D_u| = \lambda^u$,

$$2\sum_{i=1}^{t+1} P^o_{\mathtt{GW}}(|X^*_i| = o^*) \le 4te^{-u^2/2t}\,.$$

Combining the last estimate with (27), we get (26). □

We get the following.

Corollary 2 *It holds that*

$$P^o_{\mathtt{IGWR}}(|h(X_i)| \ge u \text{ for some } i \le t) \le 8t^3 e^{-(u-1)^2/2t}\,. \tag{28}$$

and

$$P^o_{\mathtt{IGW}}(|h(X_i)| \ge u \text{ for some } i \le t) \le 16t^3 e^{-(u-1)^2/2t}\,. \tag{29}$$

Proof of Corollary 2 We begin by estimating $P^o_{\mathtt{IGWR}}(h(X_i) \geq u)$. Note that, decomposing according to the last visit to the level 0,

$$\begin{aligned} & P^o_{\mathtt{IGWR}}(h(X_i) \geq u) \\ \leq\ & P^o_{\mathtt{IGWR}}(\exists j < i:\ h(X_i) - h(X_j) \geq u\,, h(X_t) - h(X_j) > 0\,\forall t \in \{j+1, \ldots, i\}) \\ \leq\ & \sum_{j=0}^{i-1} P^o_{\mathtt{IGWR}}(h(X_i) - h(X_j) \geq u\,, h(X_t) - h(X_j) > 0\,\forall t \in \{j+1, \ldots, i\})\,. \end{aligned}$$

Using the stationarity of IGWR, we thus get

$$\begin{aligned} & P^o_{\mathtt{IGWR}}(h(X_i) \geq u) \\ \leq\ & \sum_{j=0}^{i-1} P^o_{\mathtt{IGWR}}(h(X_{i-j}) \geq u, h(X_s) > 0\ \forall s \in \{1, \ldots, i-j\})\,, \\ \leq\ & i \max_{r \leq i} P^o_{\mathtt{IGWR}}(h(X_r) \geq u, h(X_s) > 0\ \forall s \in \{1, \ldots, r\}). \end{aligned} \tag{30}$$

On the other hand, for $r, u > 1$,

$$P^o_{\mathtt{IGWR}}(h(X_r) \geq u, h(X_s) > 0\ \forall s \in \{1, \ldots, r\}) \leq P^o_{\mathtt{GW}}(h(X_r) \geq u-1)\,, \tag{31}$$

because reaching level u before time r and before returning to the root or visiting Ray requires reaching level u from one of the offspring of the root before returning to the root. Substituting in (30) we get

$$P^o_{\mathtt{IGWR}}(h(X_i) \geq u) \leq i \max_{r \leq i} P^o_{\mathtt{GW}}(h(X_r) \geq u-1) \leq 4i^2 e^{-(u-1)^2/2i}\,, \tag{32}$$

where (26) was used in the last inequality. It follows from the above that

$$P^o_{\mathtt{IGWR}}(h(X_i) \geq u \text{ for some } i \leq t) \leq 4t^3 e^{-(u-1)^2/2t}\,. \tag{33}$$

Recall the process $\mathcal{T}_s = \theta^{X_s}\mathcal{T}$, which is reversible under $P_{\mathtt{IGWR}}$, and note that $h(X_i) - h(X_0)$ is a measurable function, say H, of $\{\mathcal{T}_j\}_{0 \leq j \leq i}$ (we use here that for IGWR-almost every $\mathcal{T}$, and vertices $v, w \in \mathcal{T}$, one has $\theta^v \mathcal{T} \neq \theta^w \mathcal{T}$. Further, with $\widehat{\mathcal{T}}_j := \mathcal{T}_{i-j}$, it holds that $H(\{\widehat{\mathcal{T}}_j\}_{0 \leq j \leq i}) = -H(\{\mathcal{T}_j\}_{0 \leq j \leq i})$. Therefore,

$$P^o_{\mathtt{IGWR}}(h(X_i) \leq -u) = P^o_{\mathtt{IGWR}}(h(X_i) \geq u)\,.$$

Applying (32), one concludes that

$$P^o_{\mathtt{IGWR}}(h(X_i) \leq -u \text{ for some } i \leq t) \leq 4t^3 e^{-(u-1)^2/2t}\,. \tag{34}$$

Together with (33), the proof of (28) is complete. To see (29), note that IGW is absolutely continuous with respect to IGWR, with Radon-Nikodym derivative uniformly bounded by 2. □

We can now give the

Proof of (4) The increments $h(X_{i+1}) - h(X_i)$ are stationary under $P^o_{\mathtt{IGWR}}$. Therefore, by (28), for any ϵ and $r, s \leq t$ with $|r-s| \leq t^\delta$,

$$P^o_{\mathtt{IGWR}}(|h(X_r) - h(X_s)| > t^{1/2-\epsilon}) = P^o_{\mathtt{IGWR}}(|h(X_{r-s})| > t^{1/2-\epsilon}) \leq 8t^3 e^{-t^{1-\delta-2\epsilon}}\,.$$

Therefore, by Markov's inequality, for all t large,

$$P_{\mathtt{IGWR}}\left(P_{\mathcal{T}}^{o}\left(|h(X_{r-s})| > t^{1/2-\epsilon}\right) \geq t^{-2}e^{-t^{1-\delta-\epsilon}}\right) \leq e^{-t^{1-\delta-\epsilon}}.$$

Consequently,

$$P_{\mathtt{IGWR}}\left(P_{\mathcal{T}}^{o}\left(\sup_{r,s\leq t,|r-s|<t^{\delta}} |h(X_r) - h(X_s)| > t^{1/2-\epsilon}\right) \geq e^{-t^{1-\delta-\epsilon}}\right) \leq e^{-t^{1-\delta-\epsilon}}.$$

It follows that

$$\limsup_{t\to\infty} \frac{P_{\mathcal{T}}^{o}\left(\sup_{r,s\leq t,|r-s|<t^{\delta}} |h(X_r) - h(X_s)| > t^{1/2-\epsilon}\right)}{e^{-t^{1-\delta-\epsilon}}} \leq 1\,, \quad \mathtt{IGWR} - a.s., \tag{35}$$

completing the proof of (4) since the measures IGWR and IGW are mutually absolutely continuous. □

We next control the expected number of visits to D_n during one excursion from the root of a GW tree. We recall that $T_o = \min\{n \geq 1 : X_n = o\}$.

Lemma 6 *Let $\mathcal{N}_o(n) = \sum_{i=1}^{T_o} \mathbf{1}_{X_i \in D_n}$. There exists a constant C independent of n such that*

$$E_{\mathtt{GW}}^{o}(\mathcal{N}_o(n)|d_o) \leq C d_o \quad \text{and} \quad E_{\widehat{\mathtt{GW}}}^{o}(\mathcal{N}_o(n)|d_o) \leq C d_o\,. \tag{36}$$

Further,

$$\limsup_{n\to\infty} E_{\mathcal{T}}^{o}(\mathcal{N}_o(n)) < \infty\,, \quad \mathtt{GW} - a.s. \tag{37}$$

Proof of Lemma 6 We begin by conditioning on the tree $\mathcal{T}$, and fix a vertex $v \in D_n$. Let Γ_v denote the number of visits to v before T_o. Then,

$$E_{\mathcal{T}}^{o}(\Gamma_v) = P_{\mathcal{T}}^{o}(T_v < T_o) E_{\mathcal{T}}^{v}(\Gamma_v)\,.$$

Note that the walker performs, on the ray connecting o and v, a biased random walk with holding times. Therefore, by standard computations,

$$P_{\mathcal{T}}^{o}(T_v < T_o) = \frac{1}{d_o[1 + \lambda + \lambda^2 + \ldots + \lambda^{n-1}]}\,,$$

and, when starting at v, Γ_v is a Geometric random variable with parameter $\lambda^n/[(\lambda+d_v)(1+\lambda+\lambda^2+\ldots+\lambda^{n-1})]$. Therefore, for some deterministic constant C,

$$E_{\mathcal{T}}^{o}(\Gamma_v) \leq C\lambda^{-n} d_v\,.$$

Thus,

$$E_{\mathcal{T}}^{o}(\mathcal{N}_o(n)) \leq C \sum_{v\in D_n} \lambda^{-n} d_v\,. \tag{38}$$

Since the random variables d_v are i.i.d., independent of D_n, and possess exponential moments, and since $|D_n|\lambda^{-n} \to_{n\to\infty} W_o < \infty$, it holds that

$$\limsup_{n\to\infty} \sum_{v\in D_n} \lambda^{-n} d_v < \infty \,.$$

Together with (38), this proves (37). Further, it follows from (38) that

$$E^o_{\mathtt{GW}}\left(\mathcal{N}_o(n)\,\middle|\, d_o\right) \le C\lambda^{-n} E_{\mathtt{GW}}\left(|D_n|\,\middle|\, d_o\right) = C d_o \,.$$

The proof for $\widehat{\mathtt{GW}}$ is similar. □

We return to IGW trees. Recall that $Q_t(\mathcal{T}) = \{w \in \mathcal{T} : d(w, \mathtt{Ray}) \le t^\alpha\}$, and set $N_t(\alpha) = \sum_{i=1}^t \mathbf{1}_{X_i \in Q_t(\mathcal{T})}$.

Lemma 7 *For each $\epsilon > 0$ it holds that for all t large enough,*

$$E^o_{\mathtt{IGW}}(N_t(\alpha)) \le t^{1/2+\alpha+\epsilon} \,. \tag{39}$$

Proof of Lemma 7 Let $U_t = \min\{h(X_i) : i \le t\}$ and $t_\epsilon = \lceil t^{1/2+\epsilon/4}\rceil$. By (29), for t large,

$$P^o_{\mathtt{IGW}}(U_t \le -t_\epsilon) \le 16t^3 e^{-t^{\epsilon/2}/3} \,. \tag{40}$$

Let $\xi_i = \min\{s : h(X_s) = -i\}$. It follows from (40) that for all t large,

$$\begin{aligned} E^o_{\mathtt{IGW}}(N_t(\alpha)) &\le 1 + E^o_{\mathtt{IGW}}(N_t(\alpha); U_t > -t_\epsilon) \\ &\le 1 + E^o_{\mathtt{IGW}}(N_t(\alpha); \xi_{t_\epsilon} \ge t) \,. \end{aligned} \tag{41}$$

For all $k \ge 0$, let v_k be the unique vertex on Ray satisfying $h(v_k) = -k$, and set $d_k = d_{v_k}$. We next claim that there exists a constant $C_1 = C_1(\epsilon)$ independent of t such that, with

$$\Upsilon_{t,\epsilon} := \{\max_{k\in[0,t_\epsilon]} d_k \le C_1(\log t_\epsilon)\} \,,$$

it holds that

$$P_{\mathtt{IGW}}(\Upsilon^c_{t,\epsilon}) \le \frac{1}{t} \,. \tag{42}$$

Indeed, with $\beta' = 1 + (\beta - 1)/2 > 1$,

$$\begin{aligned} P_{\mathtt{IGW}}(\Upsilon^c_{t,\epsilon}) &\le t_\epsilon P_{\mathtt{IGW}}(d_0 > C_1 \log t_\epsilon) \\ &\le \frac{t_\epsilon}{m} \sum_{j=C_1\log t_\epsilon}^\infty j p_j \le \frac{t_\epsilon (\beta')^{-C_1 \log t_\epsilon}}{m} \sum_{j=1}^\infty j p_j (\beta')^j \,, \end{aligned} \tag{43}$$

from which (42) follows if C_1 is large enough since $\sum \beta^j p_j < \infty$ by assumption. Combined with the fact that $N_t(\alpha) \le t$ and (41), we conclude that for such C_1,

$$E^o_{\mathtt{IGW}}(N_t(\alpha)) \le 2 + E^o_{\mathtt{IGW}}(N_t(\alpha); \xi_{t_\epsilon} \ge t; \Upsilon_{t,\epsilon}) \,. \tag{44}$$

For the next step, let $\theta_0 = 0$ and, for $\ell \geq 1$, let θ_ℓ denote the ℓ-th visit to $\mathtt{Ray}$, that is $\theta_\ell = \min\{t > \theta_{\ell-1} : X_t \in \mathtt{Ray}\}$. Let $H_\ell = X_{\theta_\ell}$ denote the skeleton of X_i on $\mathtt{Ray}$. Note that $h_\ell = h(H_\ell)$ is a (biased) random walk in random environment with holding times; that is,

$$P(h_{\ell+1} = j | h_\ell = k) = \begin{cases} \lambda/(\lambda + d_k)\,, & j = k-1, \\ 1/(\lambda + d_k)\,, & j = k+1, \\ (d_k - 1)/(\lambda + d_k)\,, & j = k\,. \end{cases} \tag{45}$$

Let h^*_ℓ denote the homogeneous Markov chain on $\mathbb{Z}$ with $h^*_0 = 0$ and transitions as in (45) corresponding to a homogeneous environment with $d_k = C_1 \log t_\epsilon$, and set $\eta_i = \min\{\ell : h_\ell = -i\}$ and $\eta^*_i = \min\{\ell : h^*_\ell = -i\}$. The chain h^*_ℓ possesses the same drift as the chain h_ℓ, and on the event $\Upsilon_{t,\epsilon}$, its holding times dominate those of the latter chain. Therefore,

$$\mathbf{1}_{\Upsilon_{t,\epsilon}} P^o_{\mathcal{T}}(\eta_{t_\epsilon} > m) \leq P(\eta^*_{t_\epsilon} > m)\,.$$

Further, setting $\bar\theta_0 = 0$ and, for $j \geq 1$, using $\bar\theta_j = \min\{i > \bar\theta_{j-1} : h^*_i \neq h^*_{\bar\theta_{j-1}}\}$ to denote the successive jump time of the walk h^*_i, one can write

$$\eta^*_i = \sum_{j : \bar\theta_j < \eta^*_i} G_j$$

where the G_j are independent geometric random variables with parameter $(\lambda + 1)/(\lambda + C_1 \log t_\epsilon)$ that represent the holding times. Therefore, for any constants C_2, C_3 independent of ϵ and t,

$$P(\eta^*_{t_\epsilon} > C_2 t_\epsilon (\log t_\epsilon)^2) \leq P(\bar\theta_{C_3 t_\epsilon} < \eta^*_{t_\epsilon}) + P(\sum_{j=1}^{C_3 t_\epsilon} G_j > C_2 t_\epsilon (\log t_\epsilon)^2)\,.$$

The event $\{\bar\theta_{C_3 t_\epsilon} < \eta^*_{t_\epsilon}\}$ has the same probability as the event that a biased nearest neighbor random walk on $\mathbb{Z}$ started at 0, with probability $\lambda/(\lambda+1)$ to increase at each step, does not hit t_ϵ by time $C_3 t_\epsilon$. Because $\lambda > 1$, choosing $C_3 = C_3(\epsilon)$ large, this probability can be made exponentially small in t_ϵ, and in particular bounded above by $1/t$ for t large. Fix such a C_3. Now,

$$P(\sum_{j=1}^{C_3 t_\epsilon} G_j > C_2 t_\epsilon (\log t_\epsilon)^2) \leq C_3 t_\epsilon P(G_1 > C_2 (\log t_\epsilon)^2 / C_3)\,.$$

By choosing $C_2 = C_2(\epsilon)$ large, one can make this last term smaller than $1/t$. Therefore, with such a choice of C_2 and C_3, and writing $\widehat\Upsilon_{t,\epsilon} = \Upsilon_{t,\epsilon} \cap \{\eta_{t_\epsilon} < C_2 t_\epsilon (\log t_\epsilon)^2\}$, we obtain from (44) that for all t large,

$$E^o_{\mathtt{IGW}}(N_t(\alpha)) \leq 4 + E^o_{\mathtt{IGW}}(N_t(\alpha); \xi_{t_\epsilon} \geq t; \widehat\Upsilon_{t,\epsilon})\,. \tag{46}$$

On the event $\Upsilon_{t,\epsilon}$, all excursions $\{X_\ell,\ \ell = \eta_{i-1}, \ldots, \eta_i - 1\}$ away from $\mathtt{Ray}$ that start at $v \in \mathtt{Ray}$ with $h(v) > -t_\epsilon$ are excursions into GW-trees where the degree of the root is bounded by $C_1(\log t_\epsilon) - 1$. Therefore,

$$E^o_{\mathtt{IGW}}\left(\sum_{\ell=\eta_{i-1}}^{\eta_i} \mathbf{1}_{X_\ell \in Q_t(\mathcal{T})}; \Upsilon_{t,\epsilon}, h(X_{\eta_{i-1}}) > -t_\epsilon\right) \tag{47}$$
$$\leq \max_{d \leq C_1(\log t_\epsilon)-1} E^o_{\widehat{\mathtt{GW}}}\left(\sum_{\ell=0}^{T_o} \mathbf{1}_{h(X_\ell)\leq t^\alpha} | d_o = d\right)$$
$$= \max_{d \leq C_1(\log t_\epsilon)-1} \left(\sum_{j=0}^{t^\alpha} E^o_{\widehat{\mathtt{GW}}}(\mathcal{N}_o(j) | d_o = d)\right).$$

Therefore, for all t large,

$$\begin{aligned}
& E^o_{\mathtt{IGW}}(N_t(\alpha); \xi_{t_\epsilon} \geq t; \widehat{\Upsilon}_{t,\epsilon})) \\
\leq\ & E^o_{\mathtt{IGW}}\left(\sum_{i=1}^{C_2 t_\epsilon (\log t_\epsilon)^2} \mathbf{1}_{\{h(X_{\eta_{i-1}})>-t_\epsilon\}} \sum_{\ell=\eta_{i-1}}^{\eta_i} \mathbf{1}_{X_\ell \in Q_t(\mathcal{T})}; \Upsilon_{t,\epsilon}\right) \\
\leq\ & C_2 t_\epsilon (\log t_\epsilon)^2 \max_{d \leq C_1(\log t_\epsilon)-1} \left(\sum_{j=0}^{t^\alpha} E^o_{\widehat{\mathtt{GW}}}(\mathcal{N}_o(j) | d_o = d)\right) \\
\leq\ & t^{1/2+\alpha+\epsilon/2},
\end{aligned} \tag{48}$$

where the second inequality uses (47), and (36) was used in the last inequality. Combined with (46), this completes the proof of Lemma 7. □

Corollary 3 *For each $\epsilon > 0$ there exists a $t_1 = t_1(\mathcal{T}, \epsilon) < \infty$ such that for all $t \geq t_1$,*

$$E^o_{\mathcal{T}} N_t(\alpha) \leq t^{1/2+\alpha+2\epsilon}, \ \mathtt{IGW} - a.s.. \tag{49}$$

Proof of Corollary 3 From Lemma 7 and Markov's inequality we have

$$P_{\mathtt{IGW}}(E^o_{\mathcal{T}} N_t(\alpha) > c_\epsilon t^{1/2+\alpha+3\epsilon/2}) \leq t^{-\epsilon/2}.$$

Therefore, with $t_k = 2^k$, it follows from Borel-Cantelli that there exists an $k_1 = k_1(\mathcal{T}, \epsilon)$ such that for $k > k_1$,

$$E^o_{\mathcal{T}} N_{t_k}(\alpha) \leq c_\epsilon t_k^{1/2+\alpha+3\epsilon/2}, \ \mathtt{IGW} - a.s..$$

But for $t_k < t < t_{k+1}$ one has that $N_t(\alpha) \leq N_{t_{k+1}}(\alpha)$. The claim follows. □

Proof of Proposition 2 Note that the number of visits of X_i to $Q_t(\mathcal{T})$ between time $i = t$ and $i = t + \lceil t^\delta \rceil$ is bounded by $N_{t+\lceil t^\delta \rceil}(\alpha)$. Therefore,

$$P^o_{\mathcal{T}}(X_{\tau_t} \in Q_t(\mathcal{T})) = \frac{1}{t^\delta} \sum_{i=t}^{t+\lceil t^\delta \rceil} P^o_{\mathcal{T}}(X_i \in Q_t(\mathcal{T})) \leq \frac{1}{t^\delta} E^o_{\mathcal{T}}(N_{t+\lceil t^\delta \rceil}(\alpha)).$$

Applying Corollary 3 with our choice of ϵ_0, see (2), it follows that for all $t > t_1(\mathcal{T}, \epsilon_0)$, for IGW-almost every $\mathcal{T}$,

$$P^o_{\mathcal{T}}(X_{\tau_t} \in Q_t(\mathcal{T})) \le \frac{(t + \lceil t^\delta \rceil)^{1/2+\alpha+3\epsilon_0}}{t^\delta} \le \frac{1}{t^{\epsilon_0}} .$$

□

6 From IGW to GW: Proof of Theorem 1

Our proof of Theorem 1 is based on constructing a shifted coupling between the random walk $\{X_n\}$ on a GW tree and a random walk $\{Y_n\}$ on an IGW tree. We begin by introducing notation. For a tree (finite or infinite, rooted or not) $\mathcal{T}$, we let $\mathcal{L}\mathcal{T}$ denote the collection of leaves of $\mathcal{T}$, that is of vertices of degree 1 in $\mathcal{T}$ other than the root. We set $\mathcal{T}^o = \mathcal{T} \setminus \mathcal{L}\mathcal{T}$. For two trees $\mathcal{T}_1, \mathcal{T}_2$ with roots (finite or infinite) and a vertex $v \in \mathcal{L}\mathcal{T}_1$, we let $\mathcal{T}_1 \circ^v \mathcal{T}_2$ denote the tree obtained by gluing the root of $\mathcal{T}_2$ at the vertex v. Note that if $\mathcal{T}_1$ has an infinite ray emanating from the root, and $\mathcal{T}_2$ is a finite rooted tree, then $\mathcal{T}_1 \circ^v \mathcal{T}_2$ is a rooted tree with a marked infinite ray emanating from the root.

Given a GW tree $\mathcal{T}$ and a path $\{X_n\}$ on the tree, we construct a family of finite trees $\mathcal{T}_i$ and of finite paths $\{u^i_n\}$ on $\mathcal{T}_i$ as follows. Set $\tau_0 = 0, \eta_0 = 0$, and let $\mathcal{U}_0$ denote the rooted tree consisting of the root o and its offspring. For $i \ge 1$, let

$$\begin{aligned}
\tau_i &= \min\{n > \eta_{i-1} : X_n \in \mathcal{L}\mathcal{U}_{i-1}\}, \quad \text{(Excursion start)} \\
\eta_i &= \min\{n > \tau_i : X_n \in \mathcal{U}^o_{i-1}\}, \quad \text{(Excursion end)} \\
v_i &= X_{\tau_i}, \quad \text{(Excursion start location)} . \qquad (50)
\end{aligned}$$

We then set

$$\mathcal{V}_i = \{v \in \mathcal{T} : X_n = v \, \text{for some } n \in [\tau_i, \eta_i)\} ,$$

define $\overline{\mathcal{V}}_i = \mathcal{V}_i \cup \{v \in \mathcal{T} : v \text{ is an offspring of some } w \in \mathcal{V}_i\}$ and let $\mathcal{T}_i$ denote the rooted subtree of $\mathcal{T}$ with vertices in $\overline{\mathcal{V}}_i$ and root v_i. We also define the path $\{u^i_n\}^{\eta_i - \tau_i - 1}_{n=0}$ by $u^i_n = X_{n+\tau_i}$, noting that u^i_n is a path in $\mathcal{T}_i$. Finally, we set

$$\mathcal{U}_i = \mathcal{U}_{i-1} \circ^{v_i} \mathcal{T}_i . \qquad (51)$$

Note that $\mathcal{U}_i$ is a tree rooted at o since $v_i \in \mathcal{L}\mathcal{U}_{i-1}$. Further, by the GW-almost sure recurrence of the biased random walk on $\mathcal{T}$, it holds that $\mathcal{T} = \lim_i \mathcal{U}_i$.

Next, we construct an IGW tree $\widehat{\mathcal{T}}$ with root o and an infinite ray, denoted Ray, emanating from the root, and a (λ-biased) random walk $\{Y_n\}$ on $\widehat{\mathcal{T}}$, as follows. First, we choose a vertex denoted o and a semi-infinite directed path Ray emanating from it. Next, we let each vertex $v \in$ Ray have d_v offspring, where $P(d_v = k) = kp_k/m$, and the $\{d_v\}_{v \in \text{Ray}}$ are independent. For each vertex $v \in$ Ray, $v \ne o$, we identify one of its offspring with the vertex $w \in$ Ray that satisfies $d(w, o) = d(v, o) - 1$, and write $\widehat{\mathcal{U}}_0$ for the resulting tree with root o and marked ray Ray.

Set next $\hat{\tau}_0 = \hat{\eta}_0 = 0$. We start a λ-biased random walk Y_n on $\widehat{\mathcal{U}}_0$ with $Y_0 = o$, and define

$$\hat{\tau}_1 = \min\{n > 0 : Y_n \in \mathcal{L}\widehat{\mathcal{U}}_0\}\,.$$

Let $\hat{v}_1 = Y_{\hat{\tau}_1}$. We now set $\widehat{\mathcal{U}}_1 = \widehat{\mathcal{U}}_0 \circ^{\hat{v}_1} \mathcal{T}_1$ and $\hat{\eta}_1 = \hat{\tau}_1 + \eta_1 - \tau_1$, and for $\hat{\tau}_1 \le n \le \hat{\eta}_1 - 1$, set $Y_n = u^i_{n-\hat{\tau}_1}$. Finally, with $\hat{w}_1$ the ancestor of $\hat{v}_1$, we set $Y_{\hat{\eta}_1} = \hat{w}_1$.

The rest of the construction proceeds similarly. For $i > 1$, start a λ-biased random walk $\{Y_n\}_{n \ge \hat{\eta}_{i-1}}$ on $\widehat{\mathcal{U}}_{i-1}$ with $Y_{\hat{\eta}_{i-1}} = \hat{w}_{i-1}$ and define

$$\begin{aligned}
\hat{\tau}_i &= \min\{n > \hat{\eta}_{i-1} : Y_n \in \mathcal{L}\widehat{\mathcal{U}}_{i-1}\}, \quad \text{(Excursion start)}\,,\\
\hat{v}_i &= Y_{\hat{\tau}_i}, \quad \text{(Excursion start location)}\,,\\
\hat{\eta}_i &= \hat{\tau}_i + \eta_i - \tau_i\,, \quad \text{(Excursion end)}\,,\\
\widehat{\mathcal{U}}_i &= \widehat{\mathcal{U}}_{i-1} \circ^{\hat{v}_i} \mathcal{T}_i\,, \quad \text{(Extended tree)},\\
Y_n &= X_{n-\hat{\tau}_i}, n \in [\hat{\tau}_i, \hat{\eta}_i) \quad \text{(Random walk path during excursion)}\,,\\
Y_{\hat{\eta}_i} &= \hat{w}_i = \text{ancestor of } \hat{v}_i\,.
\end{aligned} \tag{52}$$

Finally, with $\widehat{\mathcal{U}} = \lim_i \widehat{\mathcal{U}}_i$, define the tree $\widehat{\mathcal{T}}$ by attaching to each vertex of $\mathcal{L}\widehat{\mathcal{U}}$ an independent Galton-Watson tree, thus obtaining an infinite tree with root o and infinite ray emanating from it. The construction leads immediately to the following.

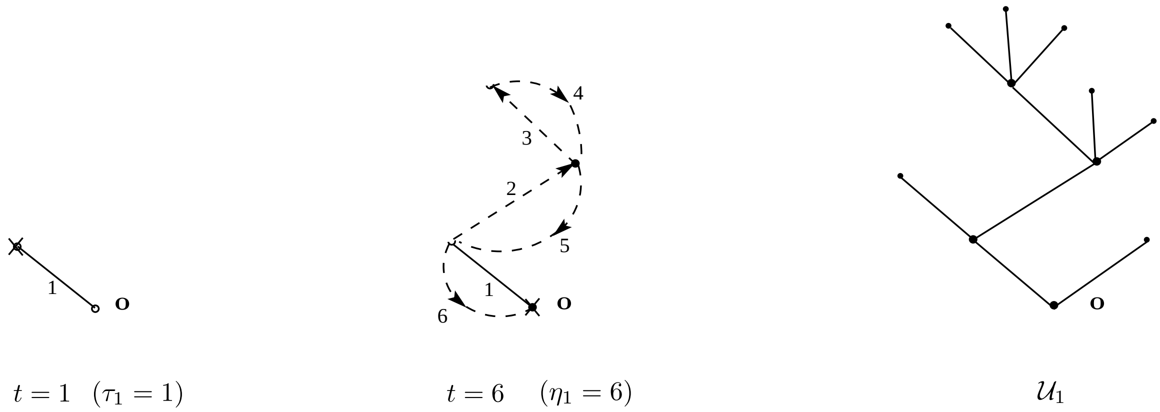
Lemma 8 *a) The tree $\widehat{\mathcal{T}}$ with root o and marked ray* `Ray` *is distributed according to* `IGW`*.*
b) Conditioned on $\widehat{\mathcal{T}}$, the law of $\{Y_n\}$ is the law of a λ-biased random walk on $\widehat{\mathcal{T}}$.

Let $\mathcal{R}_n = h(Y_n) - \min_{i=1}^n h(Y_i) \ge 0$. Due to Theorem 2, for `IGW`-almost all $\widehat{\mathcal{T}}$, the process $\mathcal{R}_{\lfloor nt \rfloor}/\sqrt{n}$ converges to a Brownian motion reflected at its running minimum, which possesses the same law as the absolute value of a Brownian motion, see e.g. [11, Theorem 6.17]. Our efforts are therefore directed toward estimating the relation between the processes $\{X_n\}$ and $\{\mathcal{R}_n\}$. Toward this end, let $I_n = \max\{i : \tau_i \le n\}$ and $\widehat{I}_n = \max\{i : \hat{\tau}_i \le n\}$ measure the number of excursions started by the walks $\{X_n\}$ and $\{Y_n\}$ before time n, and set $\Delta_n = \sum_{i=1}^{I_n} (\tau_i - \eta_{i-1})$, and $\widehat{\Delta}_n = \sum_{i=1}^{\hat{I}_n} (\hat{\tau}_i - \hat{\eta}_{i-1})$. Set also $B_n = \max_{s<t\le n: Y_s \in \mathtt{Ray}, Y_t \in \mathtt{Ray}} (h(Y_t) - h(Y_s))$ (B_n measures the maximal amount the random walk $\{Y_n\}$ backtracks, that is moves against the drift, along `Ray` before time n). Next set, recalling (13),

$$\begin{aligned}
\Delta_n^\alpha &= \sum_{i=1}^{I_n} \sum_{t \in [\eta_{i-1}, \tau_i)} \mathbf{1}_{|X_t| \le n^\alpha}\,,\\
\widehat{\Delta}_n^\alpha &= \sum_{i=1}^{\widehat{I}_n} \sum_{t \in [\hat{\eta}_{i-1}, \hat{\tau}_i)} \mathbf{1}_{Y_t \in Q_{n^\alpha}(\widehat{\mathcal{T}})}\,.
\end{aligned} \tag{53}$$

Clearly, $\Delta_n^\alpha \le \Delta_n$ and $\widehat{\Delta}_n^\alpha \le \widehat{\Delta}_n$. We however can say more.

(a) GW side.



(b) IGW side.

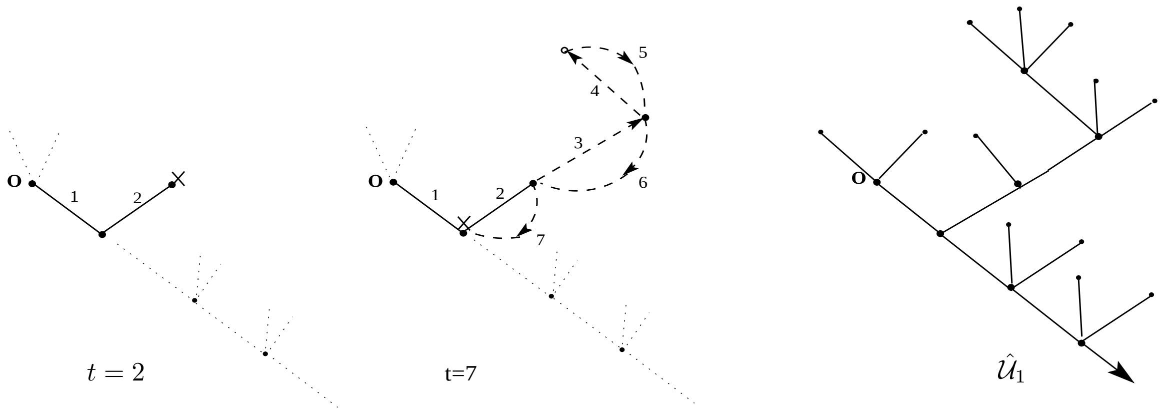


Figure 3: The coupling between the `GW` and `IGW` walks. X marks the location of the walker.

Lemma 9 *Let* $A_n = \{\Delta_n^\alpha = \Delta_n\}$ *and* $\widehat{A}_n = \{\widehat{\Delta}_n^\alpha = \widehat{\Delta}_n\}$. *Then,*

$$\lim_{n\to\infty} P_{\mathcal{T}}^o(A_n^c) = 0\,, \quad \texttt{GW}-a.s., \tag{54}$$

and

$$\lim_{n\to\infty} P_{\mathcal{T}}^o(\widehat{A_n}^c) = 0\,, \quad \texttt{IGW}-a.s.. \tag{55}$$

Further,

$$\limsup \frac{|\Delta_n|}{n} = 0\,, \quad \texttt{GW}-a.s.\,, \tag{56}$$

and

$$\limsup \frac{|\widehat{\Delta}_n|}{n} = 0\,, \quad \texttt{IGW}-a.s.\,, \tag{57}$$

Finally,

$$\limsup \frac{B_n}{\sqrt{n}} = 0\,, \quad \texttt{IGW}-a.s.\,, \tag{58}$$

We postpone for the moment the proof of Lemma 9. Note that on the event $A_n \cap \widehat{A}_n$, one has

$$\min_{s:|s-n|\le \Delta_n+\widehat{\Delta}_n} |\,|X_n| - \mathcal{R}_s| \le 2n^\alpha + B_n\,. \tag{59}$$

(To see that, note that the position $|X_n|$ consists of sums of excursions $\{u^i_\cdot\}$, up to an error coming from the parts of the path not contained in these excursions, all contained in a distance at most n^α from the root. Similarly, for some s with $|s-n| \le \Delta_n + \widetilde{\Delta}_n$, $\mathcal{R}_s$ consists of the sum of the same excursions, up to an error coming from the parts of the path not contained in these excursions, which sum up to a total distance of at most n^α from Ray in addition to the amount B_n of backtracking along Ray.)

In view of Lemma 9, the convergence in distribution (for IGW-almost every $\widehat{\mathcal{T}}$) of $\mathcal{R}_{\lfloor nt\rfloor}/\sqrt{n}$ to reflected Brownian motion, together with (59), complete the proof of Theorem 1. $\square$

Proof of Lemma 9 Consider a rooted tree $\mathcal{T}$ distributed according to GW, and a random walk path $\{X_t\}_{t\ge 0}$ with $X_0 = o$ on it. We introduce some notation. For $k \ge 1$, let $a_k = \sum_{j=1}^k \tau_j$, $b_k = \sum_{j=1}^{k-1} \eta_j$, and $J_k = [a_k - b_k + k, a_{k+1} - b_{k+1} + k]$ (the length of J_k is the time spent by the walk between the k-th and the $k+1$-th excursions). For $s \in J_k$, we define $t(s) = \eta_k + s - (a_k - b_k + k)$. Finally, we set $\widetilde{X}_0 = 0, \widetilde{X}_1 = X_{\tau_1} = X_1$, and $\widetilde{X}_s = X_{t(s)}$ (note that the process $\widetilde{X}_s$ travels on vertices "off the coupled excursions"). Note that even conditioned on $\mathcal{T}$, the nearest neighbor process $\{\widetilde{X}_s\}_{s\ge 0}$ on $\mathcal{T}$ is neither Markovian nor progressively measurable with respect to its natural filtration. To somewhat address this issue, we define the filtration $\mathcal{G}_s = \sigma(X_i, i \le t(s))$, and note that conditioned on $\mathcal{T}$, $\{\widetilde{X}_s\}_{s\ge 0}$ is progressively measurable with respect to the filtration $\mathcal{G}_s$.

The statement (54) will follow as soon as we prove the statement

$$\lim_{n\to\infty} P_{\mathcal{T}}^o\Big(\max_{s\in\cup_{k=1}^{I_n} J_k} |\widetilde{X}_s| \ge n^\alpha\Big) = 0\,, \quad \texttt{GW}-a.s., \tag{60}$$

The proof of (60) will be carried out in several steps. The first step allows us to control the event that the time spent by the process X_t inside excursions is short. The proof is routine and postponed.

Lemma 10 *For all $\epsilon > 0$,*

$$\lim_{n\to\infty} P_{\mathcal{T}}^o\Big(\sum_{i=1}^{n^{1/2+\epsilon}} (\eta_i - \tau_i) < n\Big) = 0\,, \quad \mathtt{GW} - a.s.. \tag{61}$$

Further, with

$$\widetilde{T}_n = \min\{t : W_{X_t} > (\log n)^2\}\,,$$

it holds that

$$\lim_{n\to\infty} n P_{\mathcal{T}}^o(\widetilde{T}_n \le n) = 0\,, \quad \mathtt{GW} - a.s. \tag{62}$$

Our next step involves "coarsening" the process $\{\widetilde{X}_s\}$ by stopping it at random times $\{\Theta_i\}$ in such a way that if the stopped process has increased its distance from the root between two consecutive stopping times, with high probability one of the intervals J_k has been covered. More precisely, define $\Theta_0 = 0$, and for $i \ge 1$,

$$\Theta_i = \min\{s > \Theta_{i-1} : \Big||\widetilde{X}_s| - |\widetilde{X}_{\Theta_{i-1}}|\Big| = \lfloor (\log n)^{3/2} \rfloor\}\,.$$

We emphasize that the Θ_i depend on n, although this dependence is suppressed in the notation. The following lemma, whose proof is again routine and postponed, explains why this coarsening is useful.

Lemma 11 *With the notation above,*

$$\lim_{n\to\infty} P_{\mathcal{T}}^o(\text{for some } k \le I_n,\, \Theta_{i-1}, \Theta_i \in J_k,\, |\widetilde{X}_{\Theta_i}| > |\widetilde{X}_{\Theta_{i-1}}|) = 0\,, \quad \mathtt{GW} - a.s. \tag{63}$$

We have now prepared all needed preliminary steps. Fix $\epsilon > 0$. Note first that due to (11) and the Borel-Cantelli lemma, for all n large, $|A_{n^\alpha}^\epsilon| \le |D_{n^\alpha}| e^{-\nu(\epsilon) n^\alpha}$, $\mathtt{GW}$-a.s. On the other hand, since $E_{\mathtt{GW}}|D_{n^\alpha}| = m^{n^\alpha}$, Markov's inequality and the Borel-Cantelli lemma imply that for all n large, $|D_{n^\alpha}| \le m^{n^\alpha} e^{\nu(\epsilon) n^\alpha/2}$, $\mathtt{GW}$-a.s. Combining these facts, it holds that for all n large,

$$|A_{n^\alpha}^\epsilon| \le m^{n^\alpha} e^{-\nu(\epsilon) n^\alpha/2}, \quad \mathtt{GW} - a.s.\,. \tag{64}$$

For any vertex $v \in D_{n^\alpha}$, by considering the trace of the random walk on the path connecting o and v it follows that

$$P_{\mathcal{T}}^o(X_t = v \text{ for some } t \le n) \le 1 - (1 - \lambda^{-n^\alpha})^n \le n\lambda^{-n^\alpha}\,, \mathtt{GW} - a.s.$$

Using this and (64) in the first inequality, and (62) in the second, we get

$$\begin{aligned} &\limsup_{n\to\infty} P_{\mathcal{T}}^o\big(\max_{s\in\cup_{k=1}^{I_n} J_k} |\widetilde{X}_s| \ge n^\alpha\big) && (65)\\ \le\ &\limsup_{n\to\infty} P_{\mathcal{T}}^o(\exists s \in \cup_{k=1}^{I_n} J_k : |\widetilde{X}_s| = n^\alpha, S_{\widetilde{X}_s} \ge \eta n^\alpha/2) \\ \le\ &\limsup_{n\to\infty} P_{\mathcal{T}}^o(\exists s \in \cup_{k=1}^{I_n} J_k : |\widetilde{X}_s| = n^\alpha, S_{\widetilde{X}_s} \ge \eta n^\alpha/2, t(s) \le \widetilde{T}_n)\,. \end{aligned}$$

We next note that by construction,

$$|\{i \in \{1,\ldots,\ell\} : |\widetilde{X}_{\Theta_i}| > |\widetilde{X}_{\Theta_{i-1}}|\}| \geq \ell/2\,.$$

Hence, with $P_{\mathcal{T}}$ probability approaching 1 as n goes to infinity, $t(\Theta_{2n^{1/2+\epsilon}}) > n$ because of (61) and Lemma 11. From this and (65), we conclude that

$$\begin{aligned}
&\limsup_{n\to\infty} P^o_{\mathcal{T}}(\max_{s\in\cup_{k=1}^{I_n} J_k} |\widetilde{X}_s| \geq n^\alpha) \\
\leq\ & \limsup_{n\to\infty} \sum_{i=1}^{2n^{1/2+\epsilon}} P^o_{\mathcal{T}}\Big(|\widetilde{X}_{\Theta_i}| \geq n^\alpha - (\log n)^2, S_{\widetilde{X}_{\Theta_i}} \geq \eta n^\alpha/2 - (\log n)^4, \\
&\qquad \widetilde{T}_n > t(\Theta_i)\Big)\,.
\end{aligned}$$

On the event $\widetilde{T}_n > t(\Theta_i)$ it holds that $|S_{\widetilde{X}_{\Theta_i}} - S_{\widetilde{X}_{\Theta_{i-1}}}| \leq (\log n)^4$. Therefore, decomposing according to return times of $\widetilde{X}_{\Theta_i}$ to the root,

$$\begin{aligned}
&\limsup_{n\to\infty} P^o_{\mathcal{T}}(\max_{s\in\cup_{k=1}^{I_n} J_k} |\widetilde{X}_s| \geq n^\alpha) \qquad (66)\\
\leq\ & \limsup_{n\to\infty} \sum_{i=0}^{2n^{1/2+\epsilon}} \sum_{j=i+1}^{2n^{1/2+\epsilon}} P^o_{\mathcal{T}}\Big(|\widetilde{X}_{\Theta_j}| \geq n^\alpha - (\log n)^2, \widetilde{X}_{\Theta_i} = o, \\
&\qquad S_{\widetilde{X}_{\Theta_j}} \geq \eta n^\alpha/2 - (\log n)^4, \\
&\qquad |\widetilde{X}_{\Theta_k}| > 0 \text{ and } |S_{\widetilde{X}_{\Theta_k}} - S_{\widetilde{X}_{\Theta_{k-1}}}| \leq (\log n)^4 \text{ for } i < k \leq j\Big) \\
=:\ & \limsup_{n\to\infty} \sum_{i=0}^{2n^{1/2+\epsilon}} \sum_{j=i+1}^{2n^{1/2+\epsilon}} P_{i,j,n}\,.
\end{aligned}$$

Fixing i, set for $t \geq 1$, $\widetilde{M}_t = S_{\widetilde{X}_{\Theta_{i+t}}}$. Introduce the random time

$$\begin{aligned}
K_n = \min\{t > 1 : X_s = o \text{ for some } s \in [t(\Theta_{i+1}), t(\Theta_{i+t})] \\
\text{or } |\widetilde{M}_t - \widetilde{M}_{t-1}| \geq (\log n)^4\}\,,
\end{aligned}$$

and the filtration $\widetilde{\mathcal{G}}_t = \mathcal{G}_{\Theta_{i+t}}$. The crucial observation is that $\{\widetilde{M}_{t\wedge K_n} - \widetilde{M}_1\}$ is a supermartingale for the filtration $\widetilde{\mathcal{G}}_t$, with increments bounded in absolute value by $(\log n)^4$ for all $t < K_n$, and bounded below by $-(\log n)^4$ even for $t = K_n$ (it fails to be a martingale due to the "defects" at the boundary of each of the intervals J_k, at which times r the conditional expectation of the increment $S_{\widetilde{X}_{r+1}} - S_{\widetilde{X}_r}$ is negative). Let $\widetilde{M}'_t = \widetilde{M}_t$ if $t < K_n$ or $t = K_n$ but $\widetilde{M}_t < \widetilde{M}_{t-1} + (\log n)^4$, and $\widetilde{M}'_t = \widetilde{M}_{K_n-1}$ otherwise. That is, $\widetilde{M}'_t - \widetilde{M}_1$ is a truncated version of the supermartingale $\widetilde{M}_{t\wedge K_n} - \widetilde{M}_1$. It follows that for some non-negative process a_t, $\{\widetilde{M}'_t - \widetilde{M}_1 + a_t\}$ is a martingale with increments bounded for all $t \leq K_n$ by $2(\log n)^4$. Therefore, by Azuma's inequality [2], for $j \leq n^{1/2+\epsilon}$, and all n large,

$$P_{i,j,n} \leq P^o_{\mathcal{T}}\Big(\max_{1 \leq k \leq 2n^{1/2+\epsilon}} [\widetilde{M}'_k - \widetilde{M}_1] \geq \eta n^\alpha/3 \Big) \leq e^{-n^{2\alpha}/n^{1+3\epsilon}} .$$

Since this estimate did not depend on i or j, together with (66), this completes the proof of (60), and hence of (54). The proofs of (55) and (58) are similar and omitted.

We next turn to the proof of (57). Recall that from Lemma 7, for any $\epsilon > 0$, and all $n > n_0(\epsilon)$,

$$P^o_{\mathtt{IGW}}(N_n(\alpha) \geq n^{1/2+\alpha+2\epsilon}) \leq n^{-\epsilon} .$$

Therefore, noting the monotonicity of $N_n(\alpha)$ in n, an application of the Borel-Cantelli lemma (to the sequence $n_k = 2^k$) shows that

$$\frac{N_n(\alpha)}{n^{1/2+\alpha+3\epsilon}} \to_{n\to\infty} 0 , \quad \mathtt{IGW} - a.s.$$

Since ϵ can be chosen such that $1/2 + \alpha + 3\epsilon < 1$, c.f. (2), and $\widehat{\Delta}^\alpha_n \leq N_n(\alpha)$, (57) follows.

We finally turn to the proof of (56). In what follows, we let $C_i = C_i(\mathcal{T})$ denote constants that may depend on $\mathcal{T}$ (but not on n). Let $T_\epsilon(n) = \min\{t : |X_t| = n^{1/2+\epsilon}\}$. By Lemma 5,

$$P^o_{\mathtt{GW}}(T_\epsilon(n) \leq n) \leq 4ne^{-n^{2\epsilon}/2} .$$

In particular, by the Borel-Cantelli lemma, for $\mathtt{GW}$-almost every $\mathcal{T}$,

$$P^o_{\mathcal{T}}(T_\epsilon(n) \leq n) \leq C_4(\mathcal{T})e^{-n^\epsilon} . \tag{67}$$

Let $\mathcal{C}_{o,\ell}$ denote the conductance between the root and D_ℓ. That is, define a unit flow f on $\mathcal{T}$ as a collection of non-negative numbers $f_{v,w}$, with $v \in \mathcal{T}$ and $w \in \mathcal{T}$ an offspring of v, such that Kirchoff's current law hold: $1 = \sum_{w\in D_1} f_{o,w}$ and $f_{v,w} = \sum_{w':w' \text{ is an offspring of } w} f_{w,w'}$. Then,

$$\mathcal{C}^{-1}_{o,\ell} = \inf_{f: f \text{ is a unit flow}} \sum_{i=0}^{\ell-1} \sum_{v\in D_i} \sum_{w:w \text{ is an offspring of } v} f^2_{v,w}\lambda^i .$$

By [19, Theorem 2.2], for $\mathtt{GW}$-almost every $\mathcal{T}$ there exists a constant $C_5(\mathcal{T})$ and a unit flow f such that

$$\sum_{v\in D_i} \sum_{w:w \text{ is an offspring of } v} f^2_{v,w} \leq C_5(\mathcal{T})\lambda^{-i} .$$

It follows that

$$\mathcal{C}^{-1}_{o,\ell} \leq C_5(\mathcal{T})\ell. \tag{68}$$

On the other hand, by standard theory, see [14, Exercise 2.47], for a given tree $\mathcal{T}$, with $L_o(j)$ denoting the number of visits to the root before time j,

$$E^o_{\mathcal{T}} L_o(T_\epsilon(n)) = d_o \mathcal{C}^{-1}_{o,n^{1/2+\epsilon}}\,.$$

Hence, $E^o_{\mathcal{T}} L_o(T_\epsilon(n)) \le d_o C_5(\mathcal{T}) n^{1/2+\epsilon}$. By Lemma 6, we also have that $E^o_{\mathcal{T}}(\mathcal{N}_o(\ell)) \le C_6(\mathcal{T})$, for any ℓ. Thus, using $\bar{N}_n(\alpha) = \sum_{t=1}^n \mathbf{1}_{\{|X_t|\le n^\alpha\}}$,

$$E^o_{\mathcal{T}}(\bar{N}_n(\alpha); T_\epsilon(n) \ge n) \le E^o_{\mathcal{T}} L_o(T_\epsilon(n)) E^o_{\mathcal{T}}\left(\sum_{\ell=0}^{n^\alpha} \mathcal{N}_o(\ell)\right) \le d_o C_5(\mathcal{T}) C_6(\mathcal{T}) n^{1/2+\epsilon+\alpha}\,.$$

It follows from this that

$$E^o_{\mathcal{T}}(\bar{N}_n(\alpha)) \le n P^o_{\mathcal{T}}(T_\epsilon(n) \le n) + d_o C_5(\mathcal{T}) C_6(\mathcal{T}) n^{1/2+\epsilon+\alpha}\,.$$

Using (67) and the fact that $\bar{N}_n(\alpha) \ge \Delta^\alpha_n$, together with (54), completes the proof of (56), and hence of Lemma 9. $\square$

Proof of Lemma 10: We note first that under the annealed measure $\mathtt{GW}$, the random times $(\eta_i - \tau_i)$, which denote the length of the excursions, are i.i.d., and for all x,

$$P^o_{\mathtt{GW}}(\eta_i - \tau_i \ge x) \ge \frac{1}{\lambda+1} P^o_{\mathtt{GW}}(T_o \ge x),$$

where $T_o = \min\{t \ge 1 : X_t = o\}$ denotes the first return time of X_t to o.

Throughout, the constants $C_i(\mathcal{T})$, that depend only on the tree $\mathcal{T}$, are as in the proof above. Let $x_t = t^{1/2+\epsilon/2}$ and set $T_z = \min\{t : |X_t| = z\}$. Then,

$$P^o_{\mathcal{T}}(T_o \ge t) \ge P^o_{\mathcal{T}}(T_{x_t} < T_o) P^o_{\mathcal{T}}(T_{x_t} \ge t | T_{x_t} < T_o)\,. \tag{69}$$

Note however that $P^o_{\mathcal{T}}(T_{x_t} < T_o)$ is bounded by the effective conductance between the root and D_{x_t}, which by (68) is bounded below by $C_5(\mathcal{T}) x_t^{-1}$. In particular,

$$P^o_{\mathcal{T}}(T_{x_t} < T_o) \ge \frac{C_5(\mathcal{T})}{x_t} \tag{70}$$

On the other hand, using (70) and the Carne-Varopoulos bound (see [14, Theorem 12.1], [7, 23]) in the second inequality,

$$P^o_{\mathcal{T}}(T_{x_t} < t | T_{x_t} < T_o) \le \frac{P^o_{\mathcal{T}}(T_{x_t} < t)}{P^o_{\mathcal{T}}(T_{x_t} < T_o)} \le C_7(\mathcal{T}) x_t e^{-t^{2\epsilon}} \tag{71}$$

It follows that for all t large,

$$P^o_{\mathcal{T}}(T_{x_t} \ge t | T_{x_t} < T_o) > 1/2\,,$$

implying with (69) and (70) that for all t large,

$$P^o_{\mathcal{T}}(T_o \ge t) \ge \frac{C_5(\mathcal{T})}{2t^{1/2+\epsilon/2}}\,. \tag{72}$$

It follows that for some deterministic constant C and all t large,

$$P^o_{\mathtt{GW}}(T_o \geq t) \geq \frac{C}{t^{1/2+\epsilon/2}}\,. \tag{73}$$

Hence,

$$\begin{aligned} P^o_{\mathtt{GW}}(\sum_{i=1}^{n^{1/2+\epsilon}} (\eta_i - \tau_i) < n) &\leq \left(1 - \frac{P^o_{\mathtt{GW}}(T_o \geq n)}{\lambda+1}\right)^{n^{1/2+\epsilon}} \\ &\leq \left(1 - \frac{C}{n^{1/2+\epsilon/2}}\right)^{n^{1/2+\epsilon}} \leq e^{-Cn^{\epsilon/2}}\,. \end{aligned}$$

An application of the Borel-Cantelli lemma yields (61).

To see (62), note that by time n the walker explored at most n distinct sites. We say that t is a fresh time if $X_s \neq X_t$ for all $s < t$. Then,

$$P^o_{\mathtt{GW}}(W_{X_t} \geq (\log n)^2,\, t \quad \text{is a fresh time}) \leq P^o_{\mathtt{GW}}(W_o \geq (\log n)^2) \leq e^{-c(\log n)^2}\,,$$

by the tail estimates on W_o, see [1]. Therefore,

$$\begin{aligned} &P^o_{\mathtt{GW}}(W_{X_t} \geq (\log n)^2,\, \text{for some } t \leq n) \\ &\leq \sum_{t=0}^{n} P^o_{\mathtt{GW}}(W_{X_t} \geq (\log n)^2,\, t \quad \text{is a fresh time}) \leq (n+1)e^{-c(\log n)^2}\,, \end{aligned}$$

from which (62) follows by an application of the Borel-Cantelli lemma. □

Proof of Lemma 11: Let G_n denote the event inside the probability in the left hand side of (63). The event G_n implies the existence of times $t_0 < t_1 < t_2 \leq n$ and vertices u, v such that $X_{t_0} = u = X_{t_2}$, $X_{t_1} = v$, and $|v| = |u| - \lfloor (\log n)^{3/2} \rfloor$. Thus, using the Markov property,

$$P^o_{\mathcal{T}}(G_n) \leq |\{(t_0, t_1) : t_0 < t_1 \leq n\}| \max_{\substack{u,v \in \mathcal{T}: \\ |v| = |u| - \lfloor (\log n)^{3/2} \rfloor}} P^v(X_t = u \,\text{for some } t \leq n)\,.$$

Noting that for each fixed u, v as above, the last probability is dominated by the probability of a λ-biased (toward 0) random walk on $\mathbb{Z}_+$ reflected at 0 to hit location $\lfloor (\log n)^{3/2} \rfloor$ before time n, we get

$$P^o_{\mathcal{T}}(G_n) \leq n^2 e^{-c(\log n)^{3/2}}\,,$$

for some $c > 0$, which implies (63). □

7 The transient case

Recall that when $\lambda < m$, it holds that $|X_n|/n \to_{n\to\infty} \mathtt{v} > 0$, $\mathtt{GW}$-a.s., for some non-random $\mathtt{v} = \mathtt{v}(\lambda)$ (see [17]). Our goal in this section is to prove the following:

Theorem 3 *Assume $\lambda < m$ and $p_0 = 0, \sum_k \beta^k p_k < \infty$ for some $\beta > 1$. Then, there exists a deterministic constant $\sigma^2 > 0$ such that for* GW*-almost every $\mathcal{T}$, the processes $\{(|X_{\lfloor nt \rfloor}| - nt\mathtt{v})/\sqrt{\sigma^2 n}\}_{t \geq 0}$ converges in law to standard Brownian motion.*

Before bringing the proof of Theorem 3, we need to derived an *annealed* invariance principle, see Corollary 4 below. The proof of the latter proceeds via the study of regeneration times, which are defined as follows: we set

$$\tau_1 := \inf\{t : |X_t| > |X_s| \text{ for all } s < t, \text{ and } |X_u| \geq |X_t| \text{ for all } u \geq t\},$$

and, for $i \geq 1$,

$$\tau_{i+1} := \inf\{t > \tau_i : |X_t| > |X_s| \text{ for all } s < t, \text{ and } |X_u| \geq |X_t| \text{ for all } u \geq t\}.$$

We recall (see [17]) that under the assumptions of the theorem, there exists $\mathtt{GW}$-a.s. an infinite sequence of regeneration times $\{\tau_i\}_{i\geq 1}$, and the sequence $\{(|X_{\tau_{i+1}}| - |X_{\tau_i}|), (\tau_{i+1} - \tau_i)\}_{i \geq 1}$ is i.i.d. under the $\mathtt{GW}$ measure, and the variables $|X_{\tau_2}| - |X_{\tau_1}|$ and $|X_{\tau_1}|$ possess exponential moments (see [8, Lemma 4.2] for the last fact). A key to the proof of an annealed invariance principle is the following

Proposition 3 *When $\lambda < m$, it holds that $E_{\mathtt{GW}}((\tau_2 - \tau_1)^k) < \infty$ for all integer k.*

Proof of Proposition 3: By coupling with a biased (away from 0) simple random walk on $\mathbb{Z}_+$, the claim is trivial if $\lambda < 1$. The case $\lambda = 1$ is covered in [20, Theorem 2]. We thus consider in the sequel only $\lambda \in (1, m)$. Let $T_o = \inf\{t > 0 : X_t = o\}$ denote the first return time to the root and $T_n = \min\{t > 0 : |X_t| = n\}$ denote the hitting time of level n. Let $o' \in D_1$ be an arbitrary offspring of the root. By [8, (4.25)], the law of $\tau_2 - \tau_1$ under $\mathtt{GW}$ is identical to the law of τ_1 for the walk started at v, under the measure $\mathtt{GW}^v(\cdot | T_o = \infty)$. Therefore,

$$E^o_{\mathtt{GW}}((\tau_2 - \tau_1)^k) = E^{o'}_{\mathtt{GW}}(\tau_1^k \,| T_o = \infty) = \frac{E^{o'}_{\mathtt{GW}}(\tau_1^k \,; T_o = \infty)}{P^o_{\mathtt{GW}}(T_o = \infty)}$$

where in the last equality we used that $P^o_{\mathtt{GW}}(T_o = \infty) = P^{o'}_{\mathtt{GW}}(T_o = \infty)$. Thus, with c denoting a deterministic constant whose value may change from line to line,

$$\begin{aligned}
E^o_{\mathtt{GW}}((\tau_2 - \tau_1)^k) &\leq c \sum_{n=1}^{\infty} E^{o'}_{\mathtt{GW}}(\tau_1^k \,; |X_{\tau_1}| = n, T_o = \infty) \\
&= c \sum_{n=1}^{\infty} E^{o'}_{\mathtt{GW}}(T_n^k \,; |X_{\tau_1}| = n, T_o = \infty) \\
&\leq c \sum_{n=1}^{\infty} E^{o'}_{\mathtt{GW}}(T_n^{2k} \,; T_o = \infty)^{1/2} P^{o'}_{\mathtt{GW}}(|X_{\tau_1}| = n)^{1/2} \\
&\leq c \sum_{n=1}^{\infty} e^{-n/c} E^o_{\mathtt{GW}}(T_n^{2k} \,; T_o = \infty)^{1/2},
\end{aligned}$$

where the last inequality is due to the above mentioned exponential moments on $|X_{\tau_1}|$. Therefore,

$$E^o_{\mathtt{GW}}((\tau_2-\tau_1)^k) \le c\sum_{n=1}^{\infty} e^{-n/c} n^{10k} \left(\sum_{j=0}^{\infty}(j+1)^{2k} P^o_{\mathtt{GW}}(T_n > jn^{10}\,;\, T_o=\infty)\right)^{1/2}. \tag{74}$$

We proceed by estimating the latter probability. For $j\ge 1$, let

$$\mathcal{A}_{1,j,n} = \{\text{there exists a } t\le jn^{10} \text{ such that}\, d_{X_t} \ge (\log jn^{10})^2\}\,.$$

Note that by the assumption $\sum \beta^k p_k < \infty$ for some $\beta>1$, there exists a constant c such that for all j and all n large,

$$P^o_{\mathtt{GW}}(\mathcal{A}_{1,j,n}) \le e^{-c(\log(jn^{10}))^2} \le e^{-c(\log n^{10})^2 - c(\log j)^2}\,, \tag{75}$$

We next recall that t is a fresh time for the random walk if $X_s\neq X_t$ for all $s<t$. Let $N_{j,n} := |\{t\le jn^{10} : t \text{ is a fresh time}\}|$ (i.e., $N_{j,n}$ is the number of distinct vertices visited by the walk up to time jn^{10}). Set

$$\mathcal{A}_{2,j,n} = \{N_{j,n} < \sqrt{jn^{10}}\}\cap\{T_o=\infty\}\,.$$

Note that on the event $\mathcal{A}_{2,j,n}\cap\mathcal{A}^c_{1,j,n}$ there is a time $t\le jn^{10}$ and a vertex v with $d_v\le(\log(jn^{10}))^2$ such that $X_t=v$ and v is subsequently visited $\sqrt{jn^{10}}$ times with no visit at the root. Considering the trace of the walk on the ray connecting v and o, and conditioning on $X_t=v$, the last event has a probability bounded uniformly (in t,v) by $(1-c/(\log(jn^{10}))^2)^{\sqrt{jn^{10}}}$, since $\lambda>1$. Hence, for all n large, using (75),

$$\begin{aligned} P^o_{\mathtt{GW}}(\mathcal{A}_{2,j,n}) &\le e^{-c(\log(jn^{10}))^2} + jn^{10}\left(1-\frac{c}{(\log(jn^{10}))^2}\right)^{\sqrt{jn^{10}}} \\ &\le e^{-c(\log n^{10})^2 - c(\log j)^2} + jn^{10} e^{-(jn^{10})^{1/4}}\,. \end{aligned} \tag{76}$$

The event $\mathcal{A}^c_{2,j,n}\cap\{T_o=\infty\}$ entails the existence of at least $j^{1/2}n^3$ fresh times which are at distance at least n^2 from each other. Letting $t_1=\min\{t>0 : t \,\text{is a fresh time}\}$ and

$$t_i = \min\{t > t_{i-1}+n^2 : t\,\text{is a fresh time}\}\,,$$

we observe that if $|X_{t_i}|<n$ then $P^{X_{t_i}}_{\mathtt{GW}}(T_n<n^2|\mathcal{F}_{t_i})>c>0$ (since from each fresh time, the walk has under the GW measure a strictly positive probability to escape with positive speed without backtracking to the fresh point). Thus,

$$P^o_{\mathtt{GW}}(T_n > jn^{10}\,,\, T_o=\infty, \mathcal{A}^c_{2,j,n}) \le (1-c)^{j^{1/2}n^3}\,. \tag{77}$$

Combining (76) and (77), we conclude that

$$\sum_{j=0}^{\infty}(j+1)^{2k} P^o_{\mathtt{GW}}(T_n > jn^{10}\,,\, T_o=\infty) \le c\,.$$

Substituting in (74), the lemma follows. $\square$

A standard consequence of Proposition 3 and the regeneration structure (see e.g. [22, Theorem 4.1],[24, Theorem 3.5.24]) is the following:

Corollary 4 *There exists a constant σ^2 such that, under the annealed measure $\mathtt{GW}$, the process $\{(|X_{\lfloor nt\rfloor}| - n\mathrm{v}t)/\sqrt{\sigma^2 n}\}_{t\geq 0}$ converges in distribution to a Brownian motion.*

Proof of Theorem 3: Our argument is based on the technique introduced by Bolthausen and Sznitman in [5], as developed in [6]. Let $B^n_t = B^n_t(|X|.) = (|X_{\lfloor nt\rfloor}| - nt\mathrm{v})/\sqrt{n}$, and let $\mathbb{B}^n_t(|X|.)$ denote the polygonal interpolation of $(k/n) \to B^n_{k/n}$. Consider the space $\mathcal{C}_T$ of continuous functions on $[0,T]$, endowed with the distance $d_T(u,u') = \sup_{t\leq T}|u(t)-u'(t)|\wedge 1$. By [5, Lemma 4.1], Theorem 3 will follow from Corollary 4 once we show that for all bounded by 1 Lipschitz function F on $\mathcal{C}_T$ with Lipschitz constant 1, and $b\in(1,2]$,

$$\sum_k \mathrm{var}_{\mathtt{GW}}\left(E^o_{\mathcal{T}}[F(\mathbb{B}^{\lfloor b^k\rfloor})]\right) < \infty\,. \tag{78}$$

In the sequel, fix b and F as above. For the same tree $\mathcal{T}$, let $X^1_\cdot$ and $X^2_\cdot$ be independent λ-biased random walks on $\mathcal{T}$, and set $\mathbb{B}[i,k]_t = \mathbb{B}^{\lfloor b^k\rfloor}_t(|X^i|.)$ and $\mathbb{B}[i,k,s]_t = \mathbb{B}^{\lfloor b^k\rfloor}_t(|X^i|_{\cdot+s} - |X^i|_s)$, $i=1,2$. Set

$$\tau^{i,k} = \min\{t > \lfloor b^{k/4}\rfloor : t \text{ is a regeneration time for } X^i\}$$

$$\mathcal{A}^1_k := \{\{X^1_s, s\leq \tau^{1,k}\}\cap X^2_{\tau^{2,k}} = \emptyset\},\quad \mathcal{A}^2_k := \{\{X^2_s, s\leq \tau^{2,k}\}\cap X^1_{\tau^{1,k}} = \emptyset\},$$
$$\mathcal{A}_k = \mathcal{A}^1_k\cap\mathcal{A}^2_k\,,$$
$$\mathcal{B}^i_k := \{\tau^{i,k}\leq b^{k/3}\}.$$

Note that on the event $\mathcal{A}^1_k$, the paths $\{X^1_s, s\geq\tau^{1,k}\}$ and $\{X^2_s, s\geq \tau^{2,k}\}$ can intersect only if $X^2_{\tau^{2,k}}$ is a descendant of $X^1_{\tau^{1,k}}$. Applying the same reasoning for the symmetric event $\mathcal{A}^2_k$, we conclude that on the event $\mathcal{A}_k$, these two paths do not intersect.

By construction, for any path $X.$ on $\mathcal{T}$, the path $\mathbb{B}^{\lfloor b^k\rfloor}(|X|.)$ is Lipschitz with Lipschitz constant bounded by $b^{k/2}$. Hence, since

$$\max_t |\mathbb{B}[i,k]_t - \mathbb{B}[i,k,\tau^{i,k}]_t| \leq \frac{\tau^{i,k}}{b^{k/2}}$$

and using the fact that F is a Lipschitz function with Lipschitz constant 1, we have that on the event $\mathcal{B}^i_k$, $|F(\mathbb{B}[i,k]) - F(\mathbb{B}[i,k,\tau^{i,k}])| \leq b^{k/3}/b^{k/2}$, and thus, since $|F|\leq 1$,

$$\begin{aligned}
&\mathrm{var}_{\mathtt{GW}}\left(E^o_{\mathcal{T}}[F(\mathbb{B}^{\lfloor b^k\rfloor})]\right)\\
=\;& E_{\mathtt{GW}}[F(\mathbb{B}[1,k])F(\mathbb{B}[2,k])] - E_{\mathtt{GW}}[F(\mathbb{B}[1,k])]E_{\mathtt{GW}}[F(\mathbb{B}[2,k])]\\
\leq\;& 4P_{\mathtt{GW}}((\mathcal{B}^1_k)^c) + 4b^{k/3-k/2} + E_{\mathtt{GW}}[F(\mathbb{B}[1,k,\tau^{1,k}])F(\mathbb{B}[2,k,\tau^{2,k}])]\\
&\quad - E_{\mathtt{GW}}[F(\mathbb{B}[1,k,\tau^{1,k}])]E_{\mathtt{GW}}[F(\mathbb{B}[2,k,\tau^{2,k}])]\,.
\end{aligned}$$

Conditioning on the event $\mathcal{A}_k$ and using again that $|F| \le 1$, we get

$$\begin{aligned}\mathrm{var}_{\mathtt{GW}}\left(E^o_{\mathcal{T}}[F(\mathbb{B}^{\lfloor b^k \rfloor})]\right) &\le 4P_{\mathtt{GW}}((\mathcal{B}^1_k)^c) + 4P_{\mathtt{GW}}(\mathcal{A}^c_k) + 4b^{k/3-k/2} \\ &\quad + E_{\mathtt{GW}}[F(\mathbb{B}[1,k,\tau^{1,k}])F(\mathbb{B}[2,k,\tau^{2,k}])|\mathcal{A}_k] \\ &\quad - E_{\mathtt{GW}}[F(\mathbb{B}[1,k,\tau^{1,k}])|\mathcal{A}_k]E_{\mathtt{GW}}[F(\mathbb{B}[2,k,\tau^{2,k}])|\mathcal{A}_k]\,.\end{aligned}$$

Conditioned on the event $\mathcal{A}_k$, the paths $\mathbb{B}[1,k,\tau^{1,k}]$ and $\mathbb{B}[2,k,\tau^{2,k}]$ are independent under the GW measure. Therefore, we conclude that

$$\mathrm{var}_{\mathtt{GW}}\left(E^o_{\mathcal{T}}[F(\mathbb{B}^{\lfloor b^k \rfloor})]\right) \le 4(P_{\mathtt{GW}}(\mathcal{A}^c_k) + P_{\mathtt{GW}}((\mathcal{B}^1_k)^c) + b^{k/3-k/2})\,. \qquad (79)$$

Let τ^j_i denote the successive regeneration times for $X^j_\cdot$, $j = 1,2$. The event $\{(\mathcal{B}^1_k)^c\} \cap \{\tau^1_1 \le b^{k/4}\}$ implies that at least one of the first $b^{k/4}$ inter-regeneration times $\tau^1_{i+1} - \tau^1_i$ is larger than $b^{k/3}$. Therefore,

$$\begin{aligned}P_{\mathtt{GW}}((\mathcal{B}^1_k)^c) &\le P_{\mathtt{GW}}(\tau^1_1 > b^{k/4}) + P_{\mathtt{GW}}(\text{one of } (\tau^1_{i+1} - \tau^1_i)_{i=1}^{b^{k/4}} \text{ is larger than } b^{k/3}) \\ &\le P_{\mathtt{GW}}(\tau^1_1 > b^{k/4}) + b^{k/4}P_{\mathtt{GW}}(\tau^1_2 - \tau^1_1 > b^{k/3}) \\ &\le P_{\mathtt{GW}}(\tau^1_1 > b^{k/4}) + b^{k/4-k/3}E_{\mathtt{GW}}(\tau^1_2 - \tau^1_1) \\ &\le P_{\mathtt{GW}}(\tau^1_1 > b^{k/4}) + cb^{k/4-k/3}\,.\end{aligned}$$

where Markov's inequality was used in the third step. Let $T_\ell = \min\{t > 0 : |X_t| = \ell\}$. Let Y_t be a nearest neighbor random walk on $\mathbb{Z}_+$ with $P(Y_{t+1} = Y_t - 1|Y_t) = \lambda/(\lambda+1)$ whenever $Y_t \ne 0$. $Y_\cdot$ and $X_\cdot$ can be constructed on the same probability space, such that $T_\ell \le \min\{t > 0 : Y_t = \ell\} =: T^Y_\ell$ for all ℓ. On the other hand, using the Markov property, for any constant c and all ℓ large,

$$P(T^Y_\ell > e^{c\ell}) \le \left(1 - \left(\frac{1}{1+\lambda}\right)^\ell\right)^{e^{c\ell}/\ell}$$

In particular, there exists a $c_1 = c_1(\lambda) > 0$ such that $P_{\mathtt{GW}}(T_\ell > e^{c_1\ell}) \le e^{-\ell/c_1}$ (better bounds are available but not needed). Thus, for some deterministic constants $c_i = c_i(\lambda, b) > 0$, $i \ge 2$, and all k large,

$$\begin{aligned}P_{\mathtt{GW}}(\tau_1 > b^{k/4}/2) &\le P_{\mathtt{GW}}(|X_{\tau_1}| > c_2k) + P_{\mathtt{GW}}(\tau_1 > b^{k/4}/2, |X_{\tau_1}| \le c_2k) \\ &\le P_{\mathtt{GW}}(|X_{\tau_1}| > c_2k) + P_{\mathtt{GW}}(T_{c_2k} > b^{k/4}/2) \le e^{-c_3k}, \quad (80)\end{aligned}$$

where we have used the above mentioned fact that $|X_{\tau_1}|$ possesses exponential moments. We conclude that with $c_4 \le c_3$,

$$P_{\mathtt{GW}}((\mathcal{B}^1_k)^c) \le b^{-c_4k}\,.$$

It remains to estimate $P_{\mathtt{GW}}(\mathcal{A}^c_k) \le 2P_{\mathtt{GW}}((\mathcal{A}^1_k)^c)$. Let

$$\mathcal{C}'_{k,i} := \{\tau^i_1 < b^{k/4}/2\},\ \mathcal{C}''_{k,i} := \{\tau^i_{\lfloor b^{k/8} \rfloor} < b^{k/4}\},\ \mathcal{C}_k := \mathcal{C}'_{k,1} \cap \mathcal{C}'_{k,2} \cap \mathcal{C}''_{k,1} \cap \mathcal{C}''_{k,2}\,.$$

Using (80), it follows that $P_{\mathtt{GW}}(\mathcal{C}'^{c}_{k,1}) \le b^{-c_3 k}$. On the other hand, the event $\mathcal{C}'_{k,1} \cap (\mathcal{C}''_{k,1})^c$ implies that the sum of the difference $\tau^1_{i+1} - \tau^1_i$, $i = 1, \ldots, \lfloor b^{k/8} \rfloor$, is larger than $b^{k/4}/2$, and hence, by Markov's inequality,

$$P_{\mathtt{GW}}(\mathcal{C}'_{k,1} \cap (\mathcal{C}''_{k,1})^c) \le 2b^{k/8} \frac{E_{\mathtt{GW}}(\tau^1_2 - \tau^1_1)}{b^{k/4}} \le b^{-c_5 k},$$

for some deterministic constant $c_5 < c_4$. Since the same estimates are valid also for $C'_{k,2}$ and $C''_{k,2}$ replacing $C'_{k,1}$ and $C''_{k,1}$, it follows that

$$P_{\mathtt{GW}}(\mathcal{C}^c_k) \le 4b^{-c_5 k}. \tag{81}$$

On the other hand, let $\mathcal{Z}^i$ denote the collection of vertices in $D_{\lfloor b^{k/8} \rfloor}$ hit by $X^i_\cdot$. On $\mathcal{C}_k$ there are at most $b^{k/4}$ vertices in $\mathcal{Z}^1$. The event $(\mathcal{A}^1_k)^c \cap \mathcal{C}_k$ implies that the path X^2 intersected the path X^1 at a distance at least $\lfloor b^{k/8} \rfloor$ from the root, and this has to happen before time $\tau^2_{\lfloor b^{k/8} \rfloor}$, i.e. before time $b^{k/4}$, for otherwise $\mathcal{Z}^1 \cap \mathcal{Z}^2 = \emptyset$. Therefore,

$$\begin{aligned} P_{\mathtt{GW}}((\mathcal{A}^1_k)^c \cap \mathcal{C}_k) &\le E_{\mathtt{GW}} P^o_{\mathcal{T}}(X^2_\cdot \text{ visits } \mathcal{Z}^1 \text{ before time } b^{k/4}) \qquad (82)\\ &\le b^{k/4} E_{\mathtt{GW}} \max_{v \in D_{\lfloor b^{k/8} \rfloor}} P^o_{\mathcal{T}}(X^2_\cdot \text{ visits } v \text{ before time } b^{k/4}). \end{aligned}$$

When $\lambda > 1$, there exists a constant $c_6 < c_5$ such that uniformly in $v \in D_{\lfloor b^{k/8} \rfloor}$,

$$P^o_{\mathcal{T}}(X^2_\cdot \text{ visits } v \text{ before time } b^{k/4}) \le b^{k/4} e^{-c_6 b^{k/8}}.$$

On the other hand, even when $p_1 > 0$, Lemma 2.2 of [8] shows that there exists a $\beta > 0$ such that with $M_v = |\{w \text{ is an ancestor of } v : d_w \ge 2\}|$, it holds that

$$\limsup_{\ell \to \infty} P_{\mathtt{GW}}(\min_{v \in D_\ell} M_v/\ell < \beta) < 0.$$

It immediately follows, reducing c_6 if necessary, that when $\lambda \le 1$, for all k large,

$$E_{\mathtt{GW}} \max_{v \in D_{\lfloor b^{k/8} \rfloor}} P^o_{\mathcal{T}}(X^2_\cdot \text{ visits } v \text{ ever}) \le e^{-c_6 b^{k/8}}.$$

Substituting in (82), we conclude that whenever $\lambda < m$,

$$P_{\mathtt{GW}}(\mathcal{A}^c_k \cap \mathcal{C}_k) \le 2P_{\mathtt{GW}}((\mathcal{A}^1_k)^c \cap \mathcal{C}_k) \le e^{-c_7 b^{k/8}}.$$

Together with (81), (80), and (79), we conclude that (78) holds and thus conclude the proof of Theorem 3. $\square$

Acknowledgment We thank Nina Gantert for asking the question that led to this work, and for many useful discussions. We thank Nathan Levy for a careful reading of several earlier versions of this paper.

References

[1] K. B. Athreya, Large deviation rates for branching processes. I. Single type case, *Ann. Appl. Probab.* **4** (1994), pp. 779–790.

[2] K. Azuma, Weighted sums of certain dependent random variables, *Tôhoku Math. J.* **19** (1967), pp. 357–367.

[3] N. Berger and M. Biskup, *Quenched invariance principle for simple random walk on percolation clusters*, http://front.math.ucdavis.edu/math.PR/0503576.

[4] P. Billingsley, *Convergence of Probability Measures*, second edition, Wiley (1999).

[5] E. Bolthausen and A.-S. Sznitman, On the static and dynamic points of view for certain random walks in random environment, *Methods Appl. Anal.* **9** (2002), pp. 345–375.

[6] E. Bolthausen, A.-S. Sznitman, and O. Zeitouni, Cut points and diffusive random walks in random environments, *Ann. Inst. H. Poincare* **39** (2003), pp. 527–555.

[7] T. K Carne, A transmutation formula for Markov chains, *Bull Sci. Math.* **109** (1985), pp. 399–405.

[8] A. Dembo, N. Gantert, Y. Peres and O. Zeitouni, Large deviations for random walks on Galton-Watson trees: averaging and uncertainty, *Prob. Th. Rel. Fields* **122** (2001), pp. 241–288.

[9] P. G. Doyle and J. L. Snell, *Random walks and electric networks*, Carus Mathematical Monographs, 22, Mathematical Association of America, Washington, DC, (1984).

[10] T. E. Harris, Branching processes, *Ann. Math. Statist.* **41** (1948), pp. 474–494.

[11] I. Karatzas and S. Shreve, *Brownian motion and stochastic calculus*, second edition, Springer (1988).

[12] S. M. Kozlov, The method of averaging and walks in inhomogeneous environments, *Russian Math. Surveys* **40** (1985) pp. 73–145.

[13] R. Lyons, Random walks and percolation on trees, *Ann. Probab.* **18** (1990), 931–958.

[14] R. Lyons with Y. Peres, *Probability on trees and networks.* Available at http://mypage.iu.edu/˜rdlyons/prbtree/prbtree.html

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

[16] R. Lyons, R. Pemantle and Y. Peres, Ergodic theory on Galton-Watson trees: speed of random walk and dimension of harmonic measure, *Ergodic Theory Dynam. Systems* **15** (1995), pp. 593–619.

[17] R. Lyons, R. Pemantle and Y. Peres, Biased random walks on Galton-Watson trees, *Probab. Theory Related Fields* **106** (1996), pp. 249–264.

[18] P. E. Ney and A. N. Vidyashankar, Harmonic moments and large deviation rates for supercritical branching processes, *Annals Appl. Probab.* **13** (2003), pp. 475–489.

[19] R. Pemantle and Y. Peres, Galton-Watson trees with the same mean have the same polar sets, *Annals Probab.* **23** (1995), pp. 1102–1124.

[20] D. Piau, Théorème central limite fonctionnel pour une marche au hasard en environment aléatoire, *Annals Probab.* **26** (1998), pp. 1016–1040.

[21] V. Sidoravicius and A.-S. Sznitman, Quenched invariance principles for walks on clusters of percolation or among random conductances, *Probab. Theory Related Fields* **129** (2004), pp. 219–244.

[22] A.-S. Sznitman, Slowdown estimates and central limit theorem for random walks in random environment, *J. Eur. Math Soc.* **2** (2000), pp. 93–143.

[23] N. Th. Varopoulos, Long range estimates for Markov chains, *Bull. Sci. Math.* **109** (1985), pp. 225–252.

[24] O. Zeitouni, Random walks in random environment, XXXI Summer school in probability, St Flour (2001). *Lecture notes in Math.* **1837** (Springer) (2004), pp. 193–312.