On a random walk that grows its own tree

Daniel Figueiredo[*]   Giulio Iacobelli[†]   Roberto Oliveira[‡]
Bruce Reed[§]   Rodrigo Ribeiro[¶]

Abstract

Random walks on dynamic graphs have received increasingly more attention from different academic communities over the last decade. Despite the relatively large literature, little is known about random walks that construct the graph where they walk while moving around. In this paper we study one of the simplest conceivable discrete time models of this kind, which works as follows: before every walker step, with probability p a new leaf is added to the vertex currently occupied by the walker. The model grows trees and we call it the Bernoulli Growth Random Walk (BGRW). We show that the BGRW walker is transient and has a well-defined linear speed $c(p) > 0$ for any $0 < p \le 1$. Moreover, we show that the tree as seen by the walker converges (in a suitable sense) to a random tree that is one-ended. Some natural open problems about this tree and variants of our model are collected at the end of the paper.

1 Introduction

Random walks on graphs that change over time have received much attention over the past decades. Within this context, a large body of work assumes only edge (or node) weights change over time while the graph structure (i.e., edge set) remains constant. Examples include reinforced random walks and random walks in random environments [1, 8, 9, 10, 15, 11, 6, 7]. A much smaller line of work assumes that the graph structure (edges and nodes) changes over time. However, such works generally assume graph dynamics to be independent of the walker [3, 12, 17] (an exception is [14]).

This work explores a novel model where the random walk constructs its own graph, mutually coupling the walker and graph dynamics. This model is defined as follows:

---
[*]UFRJ. E-mail: daniel@land.ufrj.br
[†]UFRJ. E-mail: giulio@im.ufrj.br
[‡]IMPA. E-mail: rimfo@impa.br
[§]McGill University. E-mail: breed@cs.mcgill.ca
[¶]Puc-Chile. E-mail: rribeiro@impa.br

(0) start with a finite tree with the walker sitting on one of the vertices;

(1) with probability p, add and connect a new leaf to the current location of the walker;

(2) let the walker take one step on the current graph;

(3) go to step 1.

We refer to this model as BGRW (Bernoulli Growth Random Walk). Note that our model may be seen as a sequence of pairs $(T_n, X_n)_{n \in \mathbb{N}}$, where T_n is a tree and X_n is the walker's position on the tree T_n. A more formal definition is given in Section 2. This model is a variant of the Non-Restart Random Walk (NRRW) proposed and studied in [2, 13]. There, the initial graph is a vertex with a loop edge, and new leafs are created every s steps for a fixed $s > 0$. Prior to [2], other models of walks creating graphs had been proposed, but they all had periodic *restart* step where the random walker would jump to a uniform vertex in the tree [5, 19].

A fundamental question on dynamic graphs is the recurrent or transient nature of the walker. Figure 1 shows simulated sample paths from the BGRW model for different values of p. The depicted trees have exactly the same number of nodes (i.e., regardless the value of p the simulations stopped when the number of nodes reaches $n = 5000$). For larger values of p the generated trees are very slim and long, as p decreases the trees become fatter and shorter. Intuitively, with large p the random walk can escape more easily, while for small p the random walk wanders more. Our findings show that for any fixed p, the random walk escapes with a positive speed $c(p) > 0$.

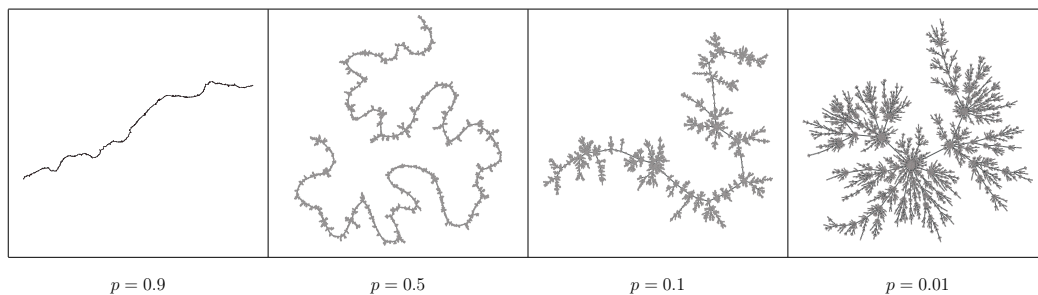| $p = 0.9$ | $p = 0.5$ | $p = 0.1$ | $p = 0.01$ |

Figure 1: Trees with $n = 5000$ nodes generated by simulating BGRW using different values for p.

Figure 1 naturally raises a question regarding the tree from the perspective of the walker as it walks around. This "point of view of the particle" has proven to be powerful in the investigation of many examples of random processes on random environments, see [18] and references therein. In our case, this approach gives fine information about the speed of the walk, as well as about the topology of the environment the walker constructs around itself. More specifically, we show that the tree around the walker converges (in a suitable sense) to an infinite one-ended random tree whose law does not depend on the initial states, i.e., the initial graph is in some sense "forgotten". Remarkably, we also show that the frequency with which the walker sees any given tree of finite height around itself is strictly positive for any value of $p \in (0, 1)$.

1.1 Main results

Our first main result, proved in Section 5, shows that our process has a well-defined, linear[1] asymptotic speed.

---

[1] We also use the term "positive speed" to mean linear speed.

Theorem 1.1 (Positive speed). *For each $0 < p \leq 1$, there exists a well-defined speed $0 < c(p) \leq p$ for the BGRW process with parameter p. That is, for any initial condition (T_0, x_0) of the process, the distance between X_n and x_0 satisfies:*

$$\lim_{n \to \infty} \frac{\operatorname{dist}_{T_n}(X_n, x_0)}{n} = c(p), \ \mathbb{P}_{T_0, x_0, p}\text{-almost surely.}$$

Our second main result is to understand the environment from the perspective of the random walk. To make this precise, we consider pairs (T_t, X_t) as random elements of the space $\mathcal{T}_*$ of *rooted locally finite trees* considered up to isomorphisms, with the *local topology*. The definition of $\mathcal{T}_*$ and its local metric (which makes it a Polish space) are recalled in Section 6.

For each $t \in \mathbb{N}$, one may define an element $[T_t, X_t] \in \mathcal{T}_*$ as the equivalence class of the tree T_t rooted at X_t. In what follows, $[T_t, X_t]_h$ is the subtree of height h consisting of X_t and all nodes of T_t within distance at most h from X_t. The next theorem counts how many times the tree of radius h around X_t takes a certain shape.

Theorem 1.2 (Convergence of the tree as seen by the walker). *For each $0 < p \leq 1$, there exists a probability distribution P_p over rooted trees such that, for any initial condition (T_0, x_0) and any finite rooted tree $[S, x] \in \mathcal{T}_*$ with height $h > 0$,*

$$\frac{1}{n+1} \sum_{t=0}^{n} \mathbb{1}_{\{[T_t, X_t]_h = [S, x]\}} \to P_p(\{[T, o] \in \mathcal{T}_* \ : \ [T, o]_h = [S, x]\}), \quad \mathbb{P}_{T_0, x_0, p}\text{-almost surely.}$$

This is essentially a mixing property of the BGRW when viewed as a Markov chain over $\mathcal{T}_*$. Our proof reveals that P_p is the unique stationary measure of this chain with the following property: if the process is started from P_p, then almost surely, for all $\ell \geq 1$, the walker will eventually be at the tip of a path of length ℓ. Theorem 1.2 follows from a stronger statement, Theorem 7.3 in Section 7.

The next theorem gives us some information on the probability measure P_p. Recall that an infinite rooted tree is one-ended if it contains a single infinite path starting from the root.

Theorem 1.3 (Support of P_p). *Given $0 < p \leq 1$, let P_p be the limit measure in Theorem 1.2. Then P_p is supported on one-ended infinite trees. If $0 < p < 1$, any rooted tree $[S, x]$ (with a certain height h) satisfies*

$$P_p(\{[T, o] \in \mathcal{T}_* \ : \ [T, o]_h = [S, x]\}) > 0.$$

Moreover, the degree of the root has exponential tails under P_p: that is, there exist $c, C > 0$ such that for all $k \geq 1$:

$$P_p(\{[T, o] \in \mathcal{T}_* \ : \ d_T(o) \geq k\}) \leq C\, e^{-ck},$$

where $d_T(o)$ denotes the degree (number of neighbors) of the root in T.

Theorem 1.3 is in fact a corollary of Theorem 7.4, which we present in Section 7.

1.2 Intuition and some difficulties

Behind our theorems, there is a probabilistic mechanism that explains our claims.

1. Given a constant ℓ, the BGRW will often create new induced paths of length ℓ. That is, there will be frequent times when the walker is at the "tip" of a path of length ℓ.

2. On the other hand, the probability of backtracking on a path of length ℓ by more than $\ell/2$ steps goes to zero very fast with ℓ.

The moments when BGRW creates a long path and does not backtrack by more than half its length are "local regeneration times" in the following precise sense: the sequence of neighborhoods of radius $\ell/2$ seen by the walker from that point on are independent of the past.

As it turns out, there are difficulties in establishing the intuitive picture drawn by items 1 and 2 above. Creating paths looks easy enough: all one needs is to create a new leaf and jump to it ℓ consecutive times. However, in order for this to happen often, the walker needs to come across lots of nodes of low degree. For this to take place, it is necessary that nodes are usually not visited too many times. To prove that returns to a vertex are unlikely to be numerous, one needs to show that the walk moves away fast enough from any vertex. The upshot is that a weaker form of *positive speed* is needed before we can justify the above items.

Proving that $\operatorname{dist}_{T_n}(X_n, X_0)$ grows *at least linearly with time* for any positive p will thus be the first goal in our proof. In essence the proof consists in establishing a coupling between the random walk in BGRW and a biased random walk on $\mathbb{Z}$. While this can be readily achieved for p sufficiently large (e.g., for $p = 1$ see [13], but also for $p > 2/3$, more generally), establishing such a coupling for all positive p turns out to be surprisingly laborious. We do that by resorting to a grass-roots argument which we believe closely mimic the actual behaviour of the walker in BGRW. Given the preliminary result on the growth of $\operatorname{dist}_{T_n}(X_n, X_0)$, we can move on to establishing the convergence of the tree as seen by the walker. The latter result, in particular, implies the existence of a well-defined positive speed. It will be crucial for us that this tree – or rather, the corresponding empirical measure – converges almost surely to an invariant measure P_p for the dynamics on rooted trees, for all "nice" initial conditions. There is a high-level similarity to the work of Lyons, Pemantle and Peres [16] on random walk on Galton-Watson trees, where the existence of a speed relies on ergodic-theoretic arguments in the space $\mathcal{T}_*$. However, in that paper the stationary measure can be described explicitly (leading to an explicit formula for the speed), which is not the case in our setting. Our analysis will also require more quantitative estimates on the convergence.

1.3 Organization and main proof steps

The remainder of the paper is organized as follows. In Section 2 we define our process formally. One important conceptual point will be to define it on locally finite trees from the start.

The proof of linear growth of $\operatorname{dist}_{T_n}(X_n, X_0)$ begins in Section 3, where we show that, for any positive M, it is very likely that $\operatorname{dist}_{T_m}(X_m, X_0) \geq \log^M n$ for some $m \leq n$. This requires comparisons with simple one dimensional random walk. The argument continues in Section 4, where we prove that backtracking on a long path, if it happens at all, typically takes a really long time. For this we introduce a simplified "loop process" that only keeps track of the path itself along the way. Combining polylog distances and long times to backtrack, we prove in Section 5 that our process has positive drift away from X_0, and derive some consequences of this fact for the degrees in the tree.

We switch gears for the remainder of the paper, as our arguments become more abstract. In Section 6 we extend the definition of our process to the space of rooted locally finite trees. The statement in Theorem 1.2, on the convergence of the tree as seen by the walker, can then be stated as a weak convergence result for the empirical measure of this process. Tightness and weak convergence criteria are discussed in this section, and tightness is proven right away.

Section 5 begins to study the loss-of-memory mechanism described above, whereby long paths are created and never backtracked on. Since we deal with infinite trees, we need some condition on the initial distribution to ensure that this takes place. We use

this in Section 6 to prove that the averages of "local functions" along the trajectories of the process always converge almost surely. This is basically the last ingredient we need to prove stronger versions of Theorems 1.2 and 1.3 in Section 7. One key point is that the limiting measure P_p is a stationary distribution for our process.

The paper wraps up with Section 8, with some final comments, and an Appendix containing technical results.

2 Definition of the model

2.1 Preliminaries

All trees in this paper are locally finite (all vertices have finite degree). Given a tree T, we let $V(T)$ and $E(T)$ denote its vertex and edge sets. For $x, y \in V(T)$, $d_T(x)$ denotes the degree (number of neighbors) of x in T and $\mathrm{dist}_T(x, y)$ is the shortest-path distance between x and y.

We let Ω denote the set of all pairs (T, x), where T is locally finite tree with $V(T) \subset \mathbb{N}$ and $\mathbb{N} \backslash V(T)$ infinite, and $x \in V(T)$. This set Ω can be described as a subset of a product space $\{0, 1\}^{\mathbb{N}} \times \{0, 1\}^{\binom{\mathbb{N}}{2}} \times \mathbb{N}$ with a natural σ-field; we omit the details.

Given a probability measure μ over some space, we use the symbol $U \sim \mu$ to mean that U is a random element with law μ.

2.2 Definition of the process

Let $p \in (0, 1]$. The BGRW process with parameter p is a Markov chain with transition kernel K_p. To define this kernel, given (T, x), we sample $(T', x') \sim K_p((T, x), \cdot)$ as follows.

1. Let $m = \min \mathbb{N} \backslash V(T)$. With probability p, set $V(T') = V(T) \cup \{m\}$ and $E(T') = E(T) \cup \{\{m, x\}\}$. Otherwise, set $T' = T$.
2. Conditionally on the above, let x' be a uniformly chosen neighbor of x in T'.

It is easy to see that this does define a valid Markov transition kernel on Ω. We use $(T_t, X_t)_{t \geq 0}$ to denote a trajectory of K_p and $\mathbb{P}_{\mu,p}$ and $\mathbb{E}_{\mu,p}$ to denote probabilities and expectations when $(T_0, X_0) \sim \mu$. If μ is a point mass on (S_0, x_0), we replace μ by S_0, x_0 in the subscripts.

Remark 2.1. Most of the time we will ignore this formal definition of the process and stick with the informal version presented in the introduction. We will later need to define this process on the set of rooted trees up to isomorphisms. See Section 6 for details.

3 Nontrivial distance from the root

In this section we show the following property of our process: for any $y \in T_0$, and any constant $M > 0$, the random walker will most likely reach distance $\geq \log^M n$ from y in at most n steps.

Lemma 3.1. *For any $M > 0$ and $0 < p \leq 1$ there exists $n_0 = n_0(p, M) \in \mathbb{N}$, depending only on p and M such that, for all $n \geq n_0$, all finite trees T_0 and all $x_0, y \in T_0$,*

$$\mathbb{P}_{T_0, x_0, p}\left(\exists m \leq n, \mathrm{dist}_{T_m}(X_m, y) \geq \log^M n\right) \geq 1 - e^{-n^{1/4}}.$$

For the proof of this lemma, we need the following definition.

Definition 3.2. *Say that $M > 0$ is admissible if the conclusion of Lemma 3.1 holds for this specific value of M. That is, M is admissible if, for any $0 < p \leq 1$, there exists $n_0 = n_0(p, M) \in \mathbb{N}$, depending only on p and M such that, for all $n \geq n_0$, all finite trees T_0 and all $x_0, y \in T_0$,*

$$\mathbb{P}_{T_0, x_0, p}\left(\exists m \leq n, \mathrm{dist}_{T_m}(X_m, y) \geq \log^M n\right) \geq 1 - e^{-n^{1/4}}.$$

Clearly, if M is admissible, so are all $0 < M' < M$. Lemma 3.1 follows from the next two propositions.

Proposition 3.3 ("Easy"; proof in subsection 3.1). *$M = \frac{1}{2}$ is admissible.*

Proposition 3.4 ("Harder"; proof in subsection 3.3). *If $M \geq \frac{1}{2}$ is admissible, so is $M + 1/2$.*

The proof of Proposition 3.4 relies on an intermediate result. It basically says that, when M is admissible and x_0 is "far" from y, then it is likely that the distance from the walker to y will increase by at least one unit by time n. This probability is large enough that we are likely to see many such increases in a small time window.

Claim 3.5 (Small growth in distance; proof in subsection 3.2). *Assume $M \geq \frac{1}{2}$ is admissible (cf. Definition 3.2). Then there exists $n_1(p, M) \in \mathbb{N}$ such that, for $n \geq n_1(p, M)$, the following property holds. Take a finite tree T and $x, y \in T$ with $\mathrm{dist}_T(x, y) \geq \log^M n$. Then*

$$\mathbb{P}_{T,x,p} \left(\mathrm{dist}_{T_t}(X_t, y) = \mathrm{dist}_T(x, y) + 1 \text{ for some } t \leq n \right) \geq 1 - 2 \frac{(\log \log n)^2}{\log^M n}.$$

As we will see, the "harder" Proposition 3.4 follows from this claim and the assumption that M is admissible applied to time $\approx \sqrt{n}$ instead of n.

Throughout this section we will use the following simple and standard lemma, which we prove in the Appendix for completeness.

Lemma 3.6 (Proof in Appendix A). *Suppose $(I_j)_{j \in \mathbb{N} \setminus \{0\}}$ are indicator random variables. Assume μ is such that $\mathbb{P}(I_1 = 1) \geq \mu$ and*

$$\forall j > 1 \,:\, \mathbb{P}(I_j = 1 \mid I_1, \ldots, I_{j-1}) \geq \mu.$$

Then for any $k, m \in \mathbb{N} \setminus \{0\}$

$$\mathbb{P}\left(\text{at least } k \text{ consecutive } 1\text{'s in the sequence } (I_j)_{j=1}^m \right) \geq 1 - (1 - \mu^k)^{\lfloor m/k \rfloor}.$$

3.1 The "easy" proposition

Proof of Proposition 3.3. It suffices to show that there exists $n_0 = n_0(p) \in \mathbb{N}$ such that, for all $n \geq n_0$, for any initial tree T_0 and any vertices $x_0, y \in T_0$,

$$\mathbb{P}_{T_0,x_0,p} \left(\exists t \leq n \,:\, \mathrm{dist}_{T_t}(X_t, y) \geq \log^{\frac{1}{2}} n \right) \geq 1 - e^{-n^{1/4}}.$$

To do this, we define, for each time $1 \leq t \leq n$,

$$I_t := \mathbb{1}\{\mathrm{dist}_{T_t}(X_t, y) = \mathrm{dist}_{T_{t-1}}(X_{t-1}, y) + 1\}.$$

Notice that we have the following inclusion of events

$$\left\{ \exists t \leq n \,:\, \mathrm{dist}_{T_t}(X_t, y) \geq \log^{\frac{1}{2}} n \right\} \supset \left\{ \text{at least } \lceil \log^{\frac{1}{2}} n \rceil \text{ consecutive } 1\text{'s in } (I_t)_{t=1}^n \right\}.$$

So we will apply Lemma 3.6. The point is that, by the Markov property:

$$\mathbb{P}_{T_0,x_0,p}(I_t = 1 \mid I_1, \ldots, I_{t-1}) \geq \inf_{T,x} \mathbb{P}_{T,x,p}(I_1 = 1) \geq \frac{p}{2}. \tag{3.1}$$

Indeed, the value $p/2$ is achieved when x is a leaf of T. In that case $I_1 = 1$ only occurs when a new neighbor is created for x (with probability p) and then the walker jumps to that neighbor (with probability $1/2$). If x is not a leaf, then the probability of the latter event is strictly larger than $1/2$.

Combining (3.1) with Lemma 3.6, we obtain:

$$\mathbb{P}_{T_0,x_0,p}\left(\exists t \leq n \,:\, \mathrm{dist}_{T_t}(X_t,y) \geq \log^{\frac{1}{2}} n\right) \geq 1 - \left(1 - \left(\frac{p}{2}\right)^{\lceil \log^{\frac{1}{2}} n \rceil}\right)^{\left\lceil \frac{n}{\lceil \log^{\frac{1}{2}} n \rceil} \right\rceil}. \qquad (3.2)$$

Now, observe that

$$\left(1 - \left(\frac{p}{2}\right)^{\lceil \log^{\frac{1}{2}} n \rceil}\right)^{\left\lceil \frac{n}{\lceil \log^{\frac{1}{2}} n \rceil} \right\rceil}$$

$$\leq \exp\left\{-\exp\left\{\log n \left(1 - \frac{\log\left(\frac{2}{p}\right)\left(\sqrt{\log n}+1\right)}{\log n} - \frac{\log\log n}{\log n}\right)\right\}\right\}.$$

Choosing n_0 sufficiently large such that $\frac{\log\left(\frac{2}{p}\right)\left(\sqrt{\log n}+1\right)}{\log n} + \frac{\log\log n}{\log n} < \frac{3}{4}$, the lower bound (3.2) is larger than $1 - e^{-n^{1/4}}$ for all $n \geq n_0$. For later purposes we point out that, $n_0(p)$ is non-increasing as a function of p. $\qquad\square$

3.2 A key claim

We now come to the proof of Claim 3.5, which connects the "easy" and "hard" proposition in the proof.

Proof of Claim 3.5. Consider the vertex y_* on the unique path from x to y with $\mathrm{dist}_T(x,y_*) = \lceil \log^M n \rceil - 1$. We recall that the admissibility of M implies, for all $0 < p \leq 1$, the existence of n_0 depending only on M and p such that:

$$\forall n \geq n_0 \,:\, \mathbb{P}_{T,x,p}(\exists t \leq n \,:\, \mathrm{dist}_{T_t}(X_t,y_*) \geq \log^M n) \geq 1 - e^{-n^{1/4}}. \qquad (3.3)$$

Now let F_n (for failure) denote the event that $\mathrm{dist}_{T_t}(X_t,y) \leq \mathrm{dist}_T(x,y)$ for all $t \leq n$. Let τ_{y_*} the hitting time of y_*

$$\tau_{y_*} := \inf\{t \in \mathbb{N} \,:\, X_t = y_*\} \in \mathbb{N} \cup \{+\infty\}.$$

Define τ_x^{+k} the k-th return time to x. That is, we set $\tau_x^{+0} := 0$, and then (recursively for $k \in \mathbb{N}\backslash\{0\}$):

$$\tau_x^{+k} := \begin{cases} +\infty, & \text{if } \tau_x^{+(k-1)} = +\infty; \\ \inf\{t > \tau_x^{+(k-1)} \,:\, X_t = x\} \in \mathbb{N} \cup \{+\infty\}, & \text{otherwise.} \end{cases}$$

Note that, for any k, we may upper bound the probability of F_n by three terms, which we will bound separately.

$$\mathbb{P}_{T,x,p}(F_n) \leq \mathbb{P}_{T,x,p}(F_n \cap \{\tau_{y_*} > n\}) + \mathbb{P}_{T,x,p}(F_n \cap \{\tau_x^{+k} < \tau_{y_*} \leq n\}) + \mathbb{P}_{T,x,p}(\tau_x^{+k} > \tau_{y_*}). \qquad (3.4)$$

We start with the first term in the RHS. Note that if $\tau_{y_*} > n$, then the unique path from X_t to y passes through y_* at all times $t \leq n$. Therefore,

$$\forall t \leq n \,:\, \mathrm{dist}_{T_t}(X_t,y) - \mathrm{dist}_T(x,y) = \mathrm{dist}_{T_t}(X_t,y_*) - \mathrm{dist}_T(x,y_*).$$

In particular, when $F_n \cap \{\tau_{y_*} > n\}$ holds, we must have

$$\forall t \leq n \,:\, \mathrm{dist}_{T_t}(X_t,y_*) \leq \mathrm{dist}_T(x,y_*) < \log^M n, \text{ i.e. the event in (3.3) does not hold.}$$

We conclude:

$$\mathbb{P}_{T,x,p}(F_n \cap \{\tau_{y_*} > n\}) \le e^{-n^{1/4}}. \tag{3.5}$$

We now consider the second term in the RHS of (3.4). In order for $F_n \cap \{\tau_x^{+k} < \tau_{y_*} \le n\}$ to take place, it must be that X_t returns at least k times to x before visiting y_* *but* never gets to jump to a neighbor x' of x with $\text{dist}(x', y) = \text{dist}(x, y) + 1$. Now, at each return to x, the probability that X_t jumps to such a neighbor *conditionally on the process up to that point* is at least $p/2$: the probability of creating a leaf and then not jumping in the direction of y (similarly to what we did in Equation (3.1)). We deduce:

$$\mathbb{P}_{T,x,p}(F_n \cap \{\tau_x^{+k} < \tau_{y_*} \le n\}) \le \left(1 - \frac{p}{2}\right)^k. \tag{3.6}$$

Finally, we come to the third term in the RHS of (3.4). Consider the walk X_t at the sequence of time steps that it spends on the path from x to y, up to the time τ_{y_*}. The resulting process is a simple random walk on the path with potential delays and reflecting barriers: indeed, since our graph is a tree at all times, whenever the random walk leaves the path, it must return to it (if it does return) at the same point that was last visited. Now, in order that $\tau_x^{+k} > \tau_{y_*}$, it must be that this simple random walk on the path hits y_* before returning k times to x. Since the path has length $\text{dist}_T(x, y) = \lceil \log^M n \rceil - 1$, the probability of this happening is $\le k/(\lceil \log^M n \rceil - 1)$, and we obtain:

$$\mathbb{P}_{T,x,p}(\tau_x^{+k} > \tau_{y_*}) \le \frac{k}{\lceil \log^M n \rceil - 1}. \tag{3.7}$$

Combining the terms in (3.5), (3.6) and (3.7), we obtain a bound in (3.4):

$$\begin{aligned}
\mathbb{P}_{T,x,p}(F_n) &\le e^{-n^{1/4}} + \left(1 - \frac{p}{2}\right)^k + \frac{k}{\lceil \log^M n \rceil - 1} \\
&\le \frac{\lceil (\log\log n)^2 \rceil}{\log^M n} \left[e^{-n^{1/4}} \frac{\log^M n}{\lceil (\log\log n)^2 \rceil} + e^{-\frac{p}{2}(\log\log n)^2} \frac{\log^M n}{\lceil (\log\log n)^2 \rceil} + \frac{\log^M n}{\log^M n - 1} \right],
\end{aligned} \tag{3.8}$$

where the last inequality follows by choosing $k := \lceil (\log\log n)^2 \rceil$. Choosing $n_1 = n_1(p, M)$ sufficiently large in (3.8) such that $e^{-n^{1/4}} \frac{\log^M n}{\lceil (\log\log n)^2 \rceil} + e^{-\frac{p}{2}(\log\log n)^2} \frac{\log^M n}{\lceil (\log\log n)^2 \rceil} + \frac{\log^M n}{\log^M n - 1} < 2$ we obtain that, for all $n \ge n_1$,

$$\mathbb{P}_{T,x,p}(F_n) \le 2\frac{(\log\log n)^2}{\log^M n}. \tag{3.9}$$

This proves the claim. Note that $n_1(p, M)$ is non-increasing in p. $\qquad\square$

3.3 The "harder" proposition

We now prove Proposition 3.4, thus finishing the proof of Lemma 3.1.

Proof of Proposition 3.4. Fix an admissible $M \ge \frac{1}{2}$. Define (with hindsight) two time lengths:

$$t_* := \lceil \log^{M+1/2} n \rceil, \ \ell_* := \lfloor \sqrt{n} \rfloor. \tag{3.10}$$

Both of these time lengths depend on n, but we omit this dependency from the notation to avoid clutter. The definitions of these quantities will probably seem misterious, but we comment on them in due time; see Remarks 3.7, 3.8 and 3.9 below.

The fact that M is admissible and $\ell_* \to +\infty$ when $n \to +\infty$ implies that, for any $0 < p \le 1$, there exists $n_2(p, M)$ such that if $n \ge n_2(p, M)$, we have the following properties for any finite tree T_0 and any $x_0, y \in T_0$.

1. By the definition of admissibility (Definition 3.2) applied with ℓ_* replacing n:

$$\mathbb{P}_{T_0,x_0,p}\left(\exists m \leq \ell_*, \mathrm{dist}_{T_m}(X_m,y) \geq 2^{-M}\log^M n\right) \geq 1 - e^{-\ell_*^{1/4}} \geq 1 - e^{-\frac{1}{2}n^{1/8}}. \quad (3.11)$$

2. By Claim 3.5 applied with ℓ_* replacing n, if T is finite and $x, y \in T$ with $\mathrm{dist}_T(x,y) \geq 2^{-M}\log^M n$, then:

$$\mathbb{P}_{T,x,p}\left(\mathrm{dist}_{T_t}(X_t,y) = \mathrm{dist}_T(x,y) + 1 \text{ for some } t \leq \ell_*\right) \geq 1 - 2\frac{(\log\log n)^2}{2^{-M}\log^M n}. \quad (3.12)$$

Remark 3.7 (ℓ_* is large enough)**.** In the above we used the fact that $\log \ell_* = \log\lfloor\sqrt{n}\rfloor$ to guarantee that the two events under consideration have high probability. We will later need that $n/\ell_* t_*$ is large; see Remark 3.8 below.

We now define a sequence of stopping times σ_j and indicators I_j. Intuitively, we will want that $\mathrm{dist}_{T_{\sigma_j}}(X_{\sigma_j},y) > \mathrm{dist}_{T_{\sigma_{j-1}}}(X_{\sigma_{j-1}},y)$, and we will signal such a success by setting $I_j = 1$. More formally, for $j = 0$, $\sigma_j = 0$ and $I_j = 0$. We define σ_j and I_j for $j > 0$, with the following two choices.

(a) *When $X_{\sigma_{j-1}}$ is too close to y:* if $\mathrm{dist}_{T_{\sigma_{j-1}}}(X_{\sigma_{j-1}},y) < 2^{-M}\log^M n$, we let:

$$
\begin{aligned}
\sigma_j &:= \inf\{t \geq \sigma_{j-1} : \mathrm{dist}_{T_t}(X_t,y) \geq 2^{-M}\log^M n \text{ or } t - \sigma_{j-1} = \ell_*\}, \\
I_j &= \mathbb{1}\{\mathrm{dist}_{T_{\sigma_j}}(X_{\sigma_j},y) \geq 2^{-M}\log^M n\}.
\end{aligned}
$$

(b) *When $X_{\sigma_{j-1}}$ is not too close to y:* if $\mathrm{dist}_{T_{\sigma_{j-1}}}(X_{\sigma_{j-1}},y) \geq 2^{-M}\log^M n$, we set

$$
\begin{aligned}
\sigma_j &:= \inf\{t \geq \sigma_{j-1} : \mathrm{dist}_{T_t}(X_t,y) = \mathrm{dist}_{T_{\sigma_{j-1}}}(X_{\sigma_{j-1}},y) + 1 \text{ or } t - \sigma_{j-1} = \ell_*\}, \\
I_j &= \mathbb{1}\{\mathrm{dist}_{T_{\sigma_j}}(X_{\sigma_j},y) = \mathrm{dist}_{T_{\sigma_{j-1}}}(X_{\sigma_{j-1}},y) + 1\}.
\end{aligned}
$$

We note some basic properties of these new random variables. First, we have the deterministic bound $\sigma_j - \sigma_{j-1} \leq \ell_*$. In particular, $\sigma_j \leq \ell_* j$ for all j. Second, we observe that, if we have $m + 1$ consecutive successes, i.e. if $I_{j_0} = I_{j_0+1} = \cdots = I_{j_0+m} = 1$ for some $j_0, m \in \mathbb{N}\setminus\{0\}$, then:

$$\mathrm{dist}_{T_{\sigma_{j_0}}}(X_{\sigma_{j_0}},y) \geq 2^{-M}\log^M n,$$

and

$$\mathrm{dist}_{T_{\sigma_{j_0+m}}}(X_{\sigma_{j_0+m}},y) = \mathrm{dist}_{T_{\sigma_{j_0+m-1}}}(X_{\sigma_{j_0+m-1}},y) + 1 = \cdots = \mathrm{dist}_{T_{\sigma_{j_0}}}(X_{\sigma_{j_0}},y) + m.$$

We thus arrive at the following crucial observation (recall the time lengths t_*, ℓ_* in (3.10)).

Observation 3.1. Assume that there are $t_* + 1$ consecutive ones in the sequence $(I_j)_{j=1}^{\lfloor n/\ell_*\rfloor}$. Then, there exists $t \leq n$ such that

$$\mathrm{dist}_{T_t}(X_t,y) \geq t_* \geq \log^{M+1/2} n.$$

Indeed, if we have $t_* + 1$ consecutive ones in this sequence there exists j_0 with $(j_0 + t_*)\ell_* \leq n$ and $I_{j_0} = I_{j_0+1} = \cdots = I_{j_0+t_*} = 1$. What we are after is to show that the probability of $t_* + 1$ consecutive ones is very likely. The upshot is that we may apply Lemma 3.6 above to control the probability that the random walk reaches distance $\log^{M+1/2} n$ from y in n time steps, i.e. that $M + 1/2$ is admissible.

Remark 3.8. Note that for the lemma to be effective we need $n/\ell_* \gg 1$ so that there are "many indicators" to consider, and also that $n/\ell_* t_* \gg 1$. It will later become clear that (for t_* polylogarithmic) it suffices that $n/\ell_* \geq n^{1/4+c+o(1)}$ for some $c > 0$.

In order to apply Lemma 3.6, we must lower bound $\mathbb{P}_{T_0,x_0,p}(I_1 = 1)$ as well as $\mathbb{P}_{T_0,x_0,p}(I_j = 1 \mid I_1, \ldots, I_{j-1})$. By the strong Markov property, these quantities are lower bounded by

$$\inf_{T,x} \mathbb{P}_{T,x,p}(I_1 = 1).$$

Crucially, conditions **(a)** and **(b)** in the definition of σ_1 and I_1 correspond precisely to the situations in the two bounds (3.11) and (3.12) (respectively). It follows that, for any $0 < p \leq 1$, there exists a $n_3 = n_3(p, M)$ such that, for all $n \geq n_3$,

$$\inf_{T,x} \mathbb{P}_{T,x,p}(I_1 = 1) \geq \mu_n := 1 - 2 \frac{(\log \log n)^2}{2^{-M} \log^M n}.$$

Lemma 3.6 implies that:

$$\mathbb{P}_{T_0,x_0,p}\left(\text{at least } t_* + 1 \text{ consecutive ones in } (I_j)_{j=1}^{\lfloor n/\ell_* \rfloor}\right) \geq 1 - (1 - \mu_n^{t_*+1})^{\left\lfloor \frac{\lfloor n/\ell_* \rfloor}{t_*+1} \right\rfloor}.$$

What is left to show is that, for any $0 < p \leq 1$, there exists $n_4 = n_4(p, M)$ such that

$$\forall n \geq n_4 : (1 - \mu_n^{t_*+1})^{\left\lfloor \frac{\lfloor n/\ell_* \rfloor}{t_*+1} \right\rfloor} \leq \exp(-n^{1/4}), \tag{3.13}$$

which would imply

$$\forall n \geq n_4 : \mathbb{P}_{T_0,x_0,p}\left(\text{at least } t_* \text{ consecutive ones in } (I_j)_{j=1}^{\lfloor n/\ell_* \rfloor}\right) \geq 1 - \exp(-n^{1/4}).$$

The latter bound, which is uniform in T_0, x_0, together with Observation 3.1, will prove that $M + 1/2$ is admissible.

Remark 3.9. The important thing here is that, because $t_* = \log^{M+1/2} n$, the probability of $t_* + 1$ consecutive ones, i.e. $\mu_n^{t_*+1}$, goes to 0 slowly with n. Remark 3.8 shows we are considering "polynomially many indicators", so such not-too-small probabilities make a long run of ones very likely.

Let us show that for our choices of t_* and ℓ_* we can find an n_4 sufficiently large such that Equation (3.13) holds. We point out that we will implicitly assume $n_4 > n_3$ so to guarantee that $\inf_{T,x} \mathbb{P}_{T,x,p}(I_1 = 1) \geq \mu_n$. It is for this reason that n_4 will depend on p (in particular, n_4 will be non-increasing in p).

$$(1 - \mu_n^{t_*+1})^{\left\lfloor \frac{\lfloor n/\ell_* \rfloor}{t_*+1} \right\rfloor} \leq \exp\left(-\mu_n^{t_*+1} \frac{n^{1/2}}{2(\log^{M+1/2} n + 1)}\right).$$

Since $t_* = \lceil \log^{M+1/2} n \rceil$, for sufficiently large n we have

$$\mu_n^{t_*+1} \geq \exp\left(-\left(2(\log \log n)^2 2^{M+1} + o((\log \log n)^2)\right) \sqrt{\log n}\right) \mu_n = n^{-o(1)} \mu_n,$$

where we are using that $\left(1 - \frac{b_n}{a_n}\right)^{a_n} \approx e^{-b_n - o(b_n)}$ for sufficiently large n whenever $a_n, b_n \to +\infty$ and $b_n = o(a_n)$. Thus,

$$(1 - \mu_n^{t_*+1})^{\left\lfloor \frac{\lfloor n/\ell_* \rfloor}{t_*+1} \right\rfloor} \leq \exp\left(-n^{1/4} \frac{n^{1/4-o(1)} \mu_n}{2(\log^{M+1/2} n + 1)}\right) \leq \exp\left(-n^{1/4}\right),$$

for large enough n. $\qquad\qquad\square$

4 The loop process, or why it's hard to go back

Now we stop the discussion about the BGRW process, to introduce a simpler process which will help us to understand how long the walker X stays on specific subgraphs of the random trees $\{T_n\}_{n \in \mathbb{N}}$. Roughly speaking, a loop process on an initial graph G is a random walker such that at each step adds a loop to its position according to a coin and then chooses uniformly one edge of its current position to walk on. In other words, the process is quite similar to the BGRW but here the walker adds loops instead of leaves, which makes it possible for it to stand still.

We will be particularly interested in the loop process over specific graphs which we define before the definition of the process itself. We call a finite graph $\mathcal{B}$ a *backbone* of length ℓ if $\mathcal{B}$ is a path of length ℓ having a loop attached to its $\ell + 1$-th vertex and possibly to its other vertices, see Figure 2.
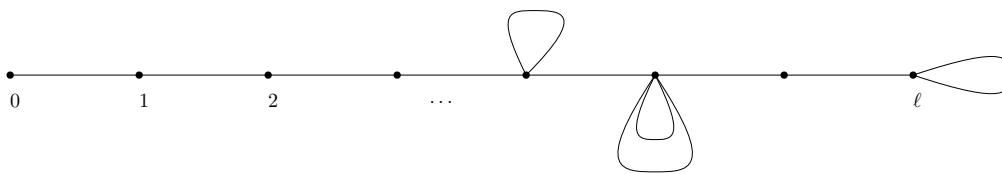

Figure 2: A *backbone* of length ℓ.

In this section, we will abuse the graph terminology saying *degree* of a vertex even though we do not count loops twice. We then reserve the special notation $\deg_t(i)$ to denote the number of edges attached to vertex i at time t.

The model has one parameter $p \in [0, 1]$ which is the parameter of a sequence $\{Z_t\}_{t \in \mathbb{N}}$ of *i.i.d* random variables with law $\mathrm{Ber}(p)$. The loop process on a backbone of length ℓ is a Markov chain $\{(\mathcal{B}_t, X_t^{\mathrm{loop}})\}_{t \in \mathbb{N}}$ where $\mathcal{B}_t$ denotes the resultant backbone at time t and X_t^{loop} is one of its $\ell + 1$ vertices. The loop process is also defined inductively according to the update rules below

1. Obtain $\mathcal{B}_{t+1}$ from $\mathcal{B}_t$ by adding a new loop to X_t^{loop} if $Z_{t+1} = 1$;

2. Choose uniformly one edge attached to X_t^{loop} in $\mathcal{B}_{t+1}$. If the chosen edge is a loop set X_{t+1}^{loop} as X_t^{loop}. Otherwise, X_{t+1}^{loop} becomes the neighbor of X_t^{loop} corresponding to the chosen edge.

We stress out to the index $t + 1$ of $\mathcal{B}$ in (2). It means that we may add a new loop at (1) and then choose it at (2).

For a fixed backbone $\mathcal{B}$ of length ℓ and $i \in \{0, 1, \cdots, \ell\}$ we let $\mathbb{P}_i(\cdot)$ denote the law of $\{(\mathcal{B}_t, X_t^{\mathrm{loop}})\}_{t \in \mathbb{N}}$ when $(\mathcal{B}_0, X_0^{\mathrm{loop}}) \equiv (\mathcal{B}, i)$.

Once we have defined the loop process, we are interested on the time it takes to go from one end of the backbone to another. More precisely, we would like to obtain bounds on the stopping time below

$$\eta_0^{\mathrm{loop}} := \inf \left\{ t \geq 1 \,\middle|\, X_t^{\mathrm{loop}} = 0 \right\},$$

when the process starts from ℓ. The next lemma gives us some estimates.

Lemma 4.1. *There exist positive constants c_1 and c_2 depending on p only, such that, for all integer K*

$$\mathbb{P}_\ell \left(\eta_0^{\mathrm{loop}} \leq e^K \right) \leq 1 - \left(1 - \frac{1}{\ell} \right)^{c_2 K} + e^{-c_1 K}.$$

Proof of Lemma 4.1. We need some notation and definitions. Consider the following sequence of stopping times

$$\tau_0 \equiv 0,$$
$$\tau_k := \inf\left\{t > \tau_{k-1} | X_t^{\mathrm{loop}} \neq X_{\tau_{k-1}}^{\mathrm{loop}}\right\}.$$

Observe that the probability of X_t^{loop} leaving its current position is at least $1/(t+C)$ (where C is a constant depending on $\mathcal{B}_0$ only) since the degree of a vertex at time t is at most t plus the number of edges in $\mathcal{B}_0$. This implies that τ_k is finite a.s. for all k. This allow us to define the process

$$Y_k := X_{\tau_k}^{\mathrm{loop}}.$$

Note that by the strong Markov Property, $\{Y_k\}_k$ is a simple random walk on $\{0, 1, \cdots, \ell\}$ with reflecting barriers. Regarding the process $\{Y_k\}_k$ we let σ be the following stopping time

$$\sigma := \inf\{k > 0 | Y_k = 0\}.$$

Observe that $\eta_0^{\mathrm{loop}} = \tau_\sigma$.

We prove the lemma by showing that X^{loop} spends at least $\exp\{K\}$ steps on the vertex ℓ. More precisely, we prove that the degree of ℓ at time τ_σ is at least $\exp\{K\}$, *w.h.p.* To do this, first observe that the degree of a vertex may be written in terms of Y_k and $\Delta\tau_k := \tau_{k+1} - \tau_k$ as follows

$$\deg_{\tau_n}(\ell) = 2 + \sum_{k=0}^{n-1} \mathbb{1}\{Y_k = \ell\} \left(\sum_{m=\tau_k}^{\tau_{k+1}-1} Z_m\right) = 2 + \sum_{k=0}^{n-1} \mathbb{1}\{Y_k = \ell\} \mathrm{Bin}\left(\Delta\tau_k, p\right).$$

Also notice that if $Y_k = \ell$, then the number of steps X^{loop} spends on ℓ is exactly $\Delta\tau_k$, which satisfies the following stochastic ordering

$$\Delta\tau_k \geq \mathrm{Geo}\left(1/\deg_{\tau_k}(\ell)\right).$$

To see the bound above, consider the random variable which counts the number of steps X^{loop} spends on ℓ by choosing only the loops which were already attached on ℓ when X^{loop} arrived at ℓ. This random variable is clearly (stochastically) smaller than $\Delta\tau_k$ and follows a geometric distribution of parameter $1/\deg_{\tau_k}(\ell)$. Thus, we have

$$\deg_{\tau_n}(\ell) \geq \sum_{k=0}^{n-1} \mathbb{1}\{Y_k = \ell\} \mathrm{Bin}\left(\mathrm{Geo}\left(1/\deg_{\tau_k}(\ell)\right), p\right).$$

Regarding the random variables $\mathrm{Bin}\left(\mathrm{Geo}\left(1/\deg_{\tau_k}(\ell)\right), p\right)$, we claim that

Claim 4.2. *For all $\varepsilon \in (0,1)$, there exists a positive constant $q = q(\varepsilon, p)$ such that*

$$\mathbb{P}_\ell\left(Y_k = \ell, \mathrm{Bin}\left(\mathrm{Geo}\left(1/\deg_{\tau_k}(\ell)\right), p\right) \geq \varepsilon p \deg_{\tau_k}(\ell) \big| \mathcal{F}_{\tau_k}\right) \geq q\mathbb{1}\{Y_k = \ell\}, \quad \mathbb{P}_\ell\text{-a.s.}$$

Proof of the claim: To simplify our writing, write

$$G_k := \mathrm{Geo}\left(1/\deg_{\tau_k}(\ell)\right); \; d_k := \deg_{\tau_k}(\ell),$$

and let $\tilde{\mathcal{F}}_{\tau_k}$ be the σ-algebra generated by G_k and $\mathcal{F}_{\tau_k}$. Now, by Chernoff bounds, we have that

$$\mathbb{P}_\ell\left(Y_k = \ell, \mathrm{Bin}\left(G_k, p\right) \geq \varepsilon p d_k, G_k \geq d_k \big| \tilde{\mathcal{F}}_{\tau_k}\right) \geq \left(1 - e^{-\frac{(1-\varepsilon)^2}{2} p G_k}\right) \mathbb{1}\{Y_k = \ell, G_k \geq d_k\}$$

$$\geq \left(1 - e^{-\frac{(1-\varepsilon)^2}{2} p d_k}\right) \mathbb{1}\{Y_k = \ell, G_k \geq d_k\}.$$

Recall that d_k is greater than or equal to 2 for all k. So, taking the conditional expectation *wrt* $\mathcal{F}_{\tau_k}$ on the above inequality yields

$$\mathbb{P}_\ell \left(Y_k = \ell, \mathrm{Bin}\left(G_k, p\right) \geq \varepsilon p d_k, G_k \geq d_k | \mathcal{F}_{\tau_k} \right) \geq \left(1 - e^{-(1-\varepsilon)^2 p} \right) \mathbb{P}_\ell \left(Y_k = \ell, G_k \geq d_k | \mathcal{F}_{\tau_k} \right)$$
$$\geq \left(1 - e^{-(1-\varepsilon)^2 p} \right) e^{-3/2} \mathbb{1}\{Y_k = \ell\},$$

which proves the claim. ∎

The above claim tells us that conditionally on the past, every time $\{Y_k\}_k$ visits ℓ it has probability at least q of increasing the degree of ℓ by a factor of at least $1 + \varepsilon p$. This points out that the degree of ℓ must be at least exponential of the number of visits ℓ receives from $\{Y_k\}_k$. So, let $N_\sigma(\ell)$ the number of visits made by Y to ℓ before it reaches vertex 0. Since Y is a simple random walk on $\{0, 1, \cdots, \ell\}$, $N_\sigma(\ell)$ follows a geometric distribution of parameter $1/\ell$. Moreover, the random variable W that counts how many times we have successfully multiplied the degree of ℓ by $1 + \varepsilon p$ may be written as follows

$$W := \sum_{k=0}^{\sigma} \mathbb{1}\{Y_k = \ell\} \mathbb{1}\{\mathrm{Bin}\left(\mathrm{Geo}\left(1/\deg_{\tau_k}(\ell)\right), p\right) \geq \varepsilon p \deg_{\tau_k}(\ell)\},$$

and dominates a random variable distributed as $\mathrm{Bin}(N_\sigma(l), q)$. Consequently, by Chernoff bounds

$$\mathbb{P}_\ell \left(W \leq \frac{K}{\log(1 + \varepsilon p)} \right) \leq \mathbb{P}_\ell \left(\mathrm{Bin}(N_\sigma(\ell), q) \leq \frac{K}{\log(1 + \varepsilon p)} \right)$$
$$\leq \mathbb{P}_\ell \left(\mathrm{Bin}(N_\sigma(\ell), q) \leq \frac{K}{\log(1 + \varepsilon p)} \bigg| N_\sigma(\ell) \geq \frac{2K}{q \log(1 + \varepsilon p)} \right)$$
$$+ \mathbb{P}_\ell \left(N_\sigma(\ell) < 2q^{-1} K / \log(1 + \varepsilon p) \right)$$
$$\leq \exp\{-c_1 K\} + 1 - \left(1 - \frac{1}{\ell} \right)^{c_2 K},$$

where $c_1 = (4 \log(1 + \varepsilon p))^{-1}$, $c_2 = 8c_1/q$ and $q = \left(1 - e^{-(1-\varepsilon)^2 p} \right) e^{-3/2}$. Finally, observe that if $W \geq K / \log(1 + \varepsilon p)$, then $\deg_{\tau_\sigma}(\ell) \geq 2e^K$ which implies that τ_σ is at least this amount, finishing the proof. ∎

The following special case of Lemma 4.1 will be particularly useful to our proposes.

Corollary 4.3. *There exists a positive constant C depending on p only, such that*

$$\mathbb{P}_\ell \left(\eta_0^{loop} \leq e^{\sqrt{\ell}} \right) \leq \frac{C}{\sqrt{\ell}}.$$

Proof of Corollary 4.3. Letting $K = \sqrt{\ell}$ in Lemma 4.1 and using that $\left(1 - \frac{1}{\ell} \right)^{c_2 \sqrt{\ell}} \approx e^{-c_3/\sqrt{\ell}}$ and $1 - e^{-x} \leq x$, choosing properly the constants the corollary follows. ∎

4.1 Coupling the BGRW and the loop process

In this subsection we construct a coupling of the BGRW and loop processes in such way that the loop process is always closer to the root than the walker X. For this, let T be a rooted locally finite tree of height at least $2(\ell + 1)$, x a vertex such that $\mathrm{dist}_T(x, \mathrm{root}) \geq \ell$ and $d_T(x) \geq 2$ and y its ancestor at distance ℓ. Since T is a tree, there exists only one path $\mathcal{P}$ connecting x to y. With this in mind, we define a graph operation $\mathcal{B}$ which associates to each pair (T, x) and ancestor y satisfying the aforementioned conditions a backbone $\mathcal{B}(T, y, x)$ of length ℓ as follows:

1. delete all vertices of T whose distance from $\mathcal{P}$ is at least 2;
2. replace each edge $xy \in E(T)$ with $x \in \mathcal{P}$ and $y \notin \mathcal{P}$ with a loop edge at x (so each edge stemming out of the path becomes a loop).
3. label the vertices on $\mathcal{P}$ by their distance from y (so y gets label 0, its neighbor on $\mathcal{P}$ gets label 1 and so on).

Figure 3 below gives a concrete example of $\mathcal{B}$ in action when y is taken as the root:
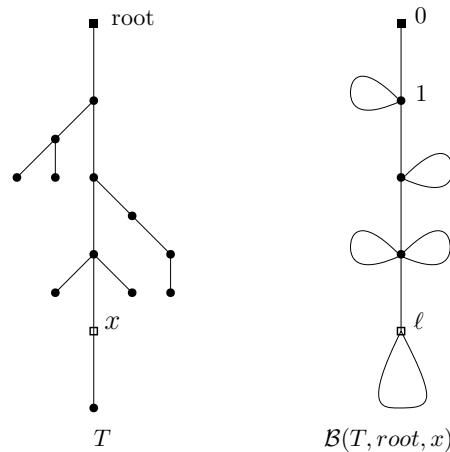

Figure 3: Example of a pair (T, x) transformed into a backbone.

So far, we have shown that the BGRW is capable of reaching long distances – powers of $\log n$ – away from the root. Now, we would like to argue that once it has gone so far, it takes too long to return. More specifically, if the initial condition is T_0, x_0 and y is one ancestor of x_0, we would like to obtain lower bounds on the following stopping time

$$\eta_y := \inf\{n \geq 1 | X_n = y\}.$$

The way we bound η_y from below is comparing it with η_0^{loop}, which we recall its definition

$$\eta_0^{\mathrm{loop}} := \inf\left\{t \geq 1 \Big| X_t^{\mathrm{loop}} = 0\right\}.$$

The next proposition tells us that we may couple the BGRW and the loop process in such way that η_y is greater than η_0^{loop} almost surely.

Proposition 4.4 (Coupling). *Let T_0 be a rooted locally finite tree, x_0 one of its vertices different from the root and y an ancestor of x_0. Then, there exists a coupling P of $\{(T_n, X_n)\}_{n \in \mathbb{N}}$ starting from (T_0, x_0) and a loop process $\{(\mathcal{B}_n, X_n^{loop})\}_{n \in \mathbb{N}}$ starting from $(\mathcal{B}(T_0, y, x_0), x_0)$ such that*

$$\mathrm{P}\left(\eta_y \geq \eta_0^{loop}\right) = 1.$$

Proof of Proposition 4.4. Let $\mathcal{P}$ denote the path connecting x_0 to its ancestor y on T_0 and ℓ its length. Also consider the following sequence of stopping times,

$$\zeta_0 \equiv 0; \quad \zeta_k := \inf\{m > \zeta_{k-1} | X_m \in \mathcal{P}\},$$

and let $\{W_k\}_{k \in \mathbb{N}}$ be a sequence of *i.i.d.* r.v's independent of the process BGRW, such that $W_k \sim \mathrm{Ber}(p)$. We couple a loop process $\{(\mathcal{B}_k, X_k^{\mathrm{loop}})\}_{k \in \mathbb{N}}$ to the BGRW inductively as follows. Start by defining

$$(\mathcal{B}_0, X_0^{\mathrm{loop}}) := (\mathcal{B}(T_0, y, x_0), x_0),$$

and assume we have defined $\{(\mathcal{B}_i, X_i^{\mathrm{loop}})\}_{i=0}^{k-1}$ in such way that this random vector is distributed as $k-1$ steps of a loop process starting from $(\mathcal{B}_0, \ell)$. Now, we define $(\mathcal{B}_k, X_k^{\mathrm{loop}})$

- If $\zeta_{k-1} < \infty$, we set $\mathcal{B}_k = \mathcal{B}(T_{\zeta_{k-1}+1}, y, x_0)$;
- If $\zeta_{k-1} = \infty$, we make use of our independent sequence $\{W_k\}_{k \in \mathbb{N}}$. We modify $\mathcal{B}_{k-1}$ adding a loop to X_{k-1}^{loop} whether $W_k = 1$;

Observe that if ζ_{k-1} is finite, then $X_{k-1}^{\mathrm{loop}} = X_{\zeta_{k-1}}$. In fact, when $X_{k-1}^{\mathrm{loop}} = X_{\zeta_{k-2}}$ we know that $X_{\zeta_{k-2}}$ has jumped outside $\mathcal{P}$. Thus, the only way of X coming back to $\mathcal{P}$ is through $X_{\zeta_{k-2}}$, which implies $X_{\zeta_{k-1}} = X_{\zeta_{k-2}}$. On the other hand, when $X_{k-1}^{\mathrm{loop}} = X_{\zeta_{k-2}+1}$ we know that $X_{\zeta_{k-2}}$ has moved on $\mathcal{P}$ implying that $\zeta_{k-1} = \zeta_{k-2} + 1$. Consequently, regardless the finiteness of ζ_{k-1} – since W_k is independent of the BGRW, $\mathcal{B}_k$ is obtained by adding a loop on X_{k-1}^{loop} independently of the whole past and with probability p.

To define X_k^{loop} we proceed in the following way

- If $\zeta_{k-1} < \infty$,

    - we set $X_k^{\mathrm{loop}} = X_{\zeta_{k-1}+1}$, if $X_{\zeta_{k-1}}$ moves on $\mathcal{P}$;
    - otherwise, when $X_{\zeta_{k-1}}$ jumps outside $\mathcal{P}$, we let X_k^{loop} be $X_{\zeta_{k-1}}$.

- In case $\zeta_{k-1} = \infty$, we select uniformly an edge of X_{k-1}^{loop} on $\mathcal{B}_k$ to walk X_{k-1}^{loop} through.

By definition of the above coupling, we obtain that $\eta_y \geq \eta_0^{\mathrm{loop}}$. The best scenario would be that in which the BGRW only walks over $\mathcal{P}$, in this case the stopping times are equals. $\qquad\square$

As a straightforward consequence of the above coupling, we restate Corollary 4.3 in terms of the hitting time η_y.

Corollary 4.5. *Let T_0 be a rooted locally finite tree, x_0 one of its vertices different from the root and y an ancestor of x_0 at distance ℓ. Then there exists a constant C depending on p only, such that*

$$\mathbb{P}_{T_0, x_0, p}\left(\eta_y \leq e^{\sqrt{\ell}}\right) \leq \frac{C}{\sqrt{\ell}}.$$

5 Positive drift and its consequences

In this section we show that the BGRW has a positive drift away from the root. In particular, we show that

$$\liminf_{n \to \infty} \frac{\mathrm{dist}_{T_n}(X_n, \mathrm{root})}{n} > 0,$$

almost surely, which implies the *transience* of the walker X. We do that by tracking the distance of X from the *root* at random times and comparing it to a right biased simple random walk on $\mathbb{Z}$. Furthermore, this comparison with the right biased simple random walk allows us to improve the results given in Proposition 3.4 and Corollary 4.5. Specifically, we can now prove

- the walker achieves distances of order n in n steps *w.h.p*;
- the probability of the walker going back long distances is exponentially rare.

Intuitively, the main message of this section is that if we look at the distance of X from the root properly normalized and at some "random times" we see a random walk on the line that dominates a right biased simple random walk. Let us begin by formally define what we mean by "random times". For a fixed positive integer r, we define three stopping conditions from a vertex x_0.

1. $\mathrm{dist}_{T_n}(X_n, \mathrm{root}) - \mathrm{dist}_{T_0}(x_0, \mathrm{root}) = r$;

2. $\mathrm{dist}_{T_n}(X_n, \mathrm{root}) - \mathrm{dist}_{T_0}(x_0, \mathrm{root}) = -r$;

3. X walks $\exp\{\sqrt{r}\}$ steps and none of the previous conditions occurs.

We say $(1) - (3)$ occurred from X_m whenever one of the three stopping conditions occurred when we put x_0 as X_m. We must point out that if $\mathrm{dist}_{T_m}(X_m, \mathrm{root}) < r$, then stopping condition (2) cannot be attained.

We define our sequence of stopping times as follows:

$$\sigma_0 \equiv 0;$$
$$\sigma_k := \inf\left\{ m > \sigma_{k-1}^{(r)} \mid (1) - (3) \text{ occurs from } X_{\sigma_{k-1}} \right\}.$$

From the definition of $(1) - (3)$ follows that, for all k, $\sigma_k^{(r)}$ is bounded from above by $\exp\{\sqrt{r}\}k$. To avoid clutter, when r is fixed we suppress the upper script from the nation of $\sigma_k^{(r)}$.

Lemma 5.1 (Coupling to the biased random walk). *Let T_0 be a rooted locally finite tree and x_0 one of its vertices. For any $q \in (1/2, 1)$ there exists $r = r(p, q)$ depending on p and q only, such that the process $\{\mathrm{dist}_{T_{\sigma_k}}(X_{\sigma_k}, \mathrm{root})/r\}_{k \geq 0}$ can be coupled with a simple random walk $\{S_k\}_{k \geq 0}$ on $\mathbb{Z}$, whose probability of jumping to the right is equal to q, in such way that*

$$\mathbb{P}\left(\mathrm{dist}_{T_{\sigma_k}}(X_{\sigma_k}, \textit{root}) \geq rS_k, \forall k \right) = 1,$$

when the process $\{S_k\}_{k \geq 0}$ starts from $\lfloor \mathrm{dist}_{T_0}(x_0, \textit{root})/r \rfloor$.

Proof of Lemma 5.1. We begin by a few notations. To simplify our writing, put $d_k := \mathrm{dist}_{T_{\sigma_k}}(X_{\sigma_k}, \mathrm{root})$. Note that the process $\{d_k\}_{k \geq 0}$ is a non-markovian random walk on the half line whose increments belong to the interval $[-r, r]$. Let $\{U_k\}_{k \geq 0}$ be a sequence of *i.i.d* random variables independent of the BGRW and following uniform distribution on $[0, 1]$. Also let $\tilde{\mathcal{F}}_k$ be the σ-algebra generated by the BGRW process up to time σ_k and let $\mathcal{H}_k$ be the σ-algebra encoding all information of $\tilde{\mathcal{F}}_k$ and all the uniform random variables up to time k.

Regarding the increments of $\{d_k\}_{k \geq 0}$, we claim

Claim 5.2. *There exist positive constants r and C depending on p only, such that for all k*

$$\inf_{T_0, x_0} \mathbb{P}_{T_0, x_0, p}\left(\Delta d_{k+1} = r \middle| \tilde{\mathcal{F}}_k \right) > 1 - \frac{C}{\sqrt{r}}.$$

Proof of the claim: First of all, observe that $\Delta d_{k+1} = r$ if, and only if, σ_{k+1} stops because of condition (1). Also, by the strong Markov Property, it is enough to prove the claim for σ_1. Said that, let us first assume we start from an initial condition (T_0, x_0) with $\mathrm{dist}_{T_0}(x_0, \mathrm{root}) \geq r$. We derive the desired lower bound by proving upper bounds for the probability of the events $\{\Delta d_1 = -r\}$ and $\{|\Delta d_1| < r\}$. The former occurs if, and only if, σ_1 stops because condition (2) and the latter because of condition (3).

Observe that if σ_1 stops because of (2) (which can happen given that $\mathrm{dist}_{T_0}(x_0, \mathrm{root}) \geq r$) then the walker X visited the ancestor y of x_0 at distance r, spending at most $\exp\{\sqrt{r}\}$ steps. So, using Corollary 4.5 we have

$$\mathbb{P}_{T_0, x_0, p}\left(\sigma_1 \text{ stops because of } (2) \right) \leq \mathbb{P}_{T_0, x_0, p}\left(\eta_y \leq e^{\sqrt{r}} \right) \leq \frac{C'}{\sqrt{r}}.$$

Note that the above upper bound holds for all possible pairs of a rooted locally finite tree T_0 and one of its vertex x_0 at distance greater than r from the root.

Finally, if X stops because of (3) then X has walked for $e^{\sqrt{r}}$ steps and has neither visited the ancestor y of x_0 at distance r, nor has increased its distance from x_0 by r. This is the same of observing a process $\tilde{X}$ on the subtree T_y that in $e^{\sqrt{r}}$ steps does not get at distance $2r$ from the root y. Applying Proposition 3.4 (page 6) with $n = e^{\sqrt{r}}$ and $M = 4$ we obtain that there exists $r_0 = r(p)$ such that for all $r \geq r_0$ we have

$$\inf_{T_0, x_0} \mathbb{P}_{T_0, x_0, p} \left(\exists m \leq e^{\sqrt{r}}, \operatorname{dist}_{T_m}(\tilde{X}_m, \tilde{root}) > r^2 \right) \geq 1 - \exp\{-e^{\sqrt{r}/4}\}.$$

Choosing r large enough so that $r \geq r_0$, $\exp\{-e^{\sqrt{r}/4}\} \leq C'/\sqrt{r}$ and $r^2 > 2r$ we conclude that

$$\sup_{T_0, x_0} \mathbb{P}_{T_0, x_0, p} \left(\sigma_1 \text{ stops because of } (3) \right) \leq \exp\{-e^{\sqrt{r}/4}\}.$$

To drop our assumption that we start at distance greater than r from the root, just recall that when this is not the case the condition (2) has probability zero and the above upper bound for condition (3) still holds. This implies that

$$\inf_{T_0, x_0} \mathbb{P}_{T_0, x_0, p} \left(\sigma_1 \text{ stops because of } (1) \right) \geq 1 - \frac{C}{\sqrt{r}},$$

which combined with the strong Markov Property proves the claim.　∎

Since we may increase r, we choose it large enough so

$$1 - \frac{C}{\sqrt{r}} \geq q.$$

Now we couple the process $\{d_k/r\}_{k \geq 1}$ and $\{S_k\}_{k \geq 1}$ in the following way. Set

$$S_0 := \left\lfloor \frac{\operatorname{dist}_{T_0}(x_0, \text{root})}{r} \right\rfloor,$$

and assume we have defined $\{S_j\}_{j=0}^{k-1}$ in such way that it has the distribution of $k-1$ steps of a q-right biased random walk. Let Q_{k-1} denote

$$Q_{k-1} := \mathbb{P}_{T_0, x_0, p}(\Delta d_k = r | \tilde{\mathcal{F}}_{k-1}),$$

and recall from Claim 5.2 that $Q_k \geq q$ for all k. We then define S_k in the following way

$$S_k = \begin{cases} S_{k-1} + 1 & \text{if } \Delta d_k = r \text{ and } U_k \leq q/Q_{k-1}, \\ S_{k-1} - 1 & \text{otherwise.} \end{cases}$$

In words, if at time σ_k the walker X increased its distance from the root by r, then S_k jumps to the right with probability q/Q_{k-1}. In this way, whenever the process $\{d_k/r\}_{k \geq 0}$ jumps back (at most one unit), the SRW also jumps back one unit.

Now, we show that the process $\{S_k\}_{k \geq 0}$ does have the distribution of a q-right biased random walk. We start by checking that the increments are 1 with probability q or -1 with probability $1 - q$. By the definition of S we have

$$\begin{aligned} \mathbb{P}_{T_0, x_0, p} \left(\Delta S_k = 1 | \mathcal{H}_{k-1} \right) &= \mathbb{P}_{T_0, x_0, p} \left(\Delta d_{k+1} = r, U_k \leq q/Q_{k-1} | \mathcal{H}_{k-1} \right) \\ &= \mathbb{P}_{T_0, x_0, p}(\Delta d_k = r | \mathcal{H}_{k-1}) \frac{q}{Q_{k-1}} \\ &= q, \end{aligned} \tag{5.1}$$

since U_k is independent of $\mathcal{H}_{k-1}$ and Q_{k-1} is measurable with respect to $\mathcal{H}_{k-1}$. Moreover, the equality below holds

$$Q_{k-1} = \mathbb{P}_{T_0, x_0, p}(\Delta d_k = r | \tilde{\mathcal{F}}_{k-1}) = \mathbb{P}_{T_0, x_0, p}(\Delta d_k = r | \mathcal{H}_{k-1}),$$

since $\mathcal{H}_{k-1}$ is $\tilde{\mathcal{F}}_{k-1}$ added of information independent of the whole process $\{(T_k, X_k)\}_{k \geq 1}$. Equation (5.1) allows us to derive the independence of all the increments of our right biased random walk $\{S_k\}_{k \geq 1}$. For any fixed set of indexes $k_1 < k_2 < \cdots < k_j$ and any vector $(a_1, \cdots, a_j) \in \{-1, 1\}^j$ we have

$$\mathbb{P}\left(\Delta S_{k_1} = a_1, \cdots, \Delta S_{k_j} = a_j\right) = \mathbb{E}\left[\mathbb{1}_{\{\Delta S_{k_1} = a_1\}} \cdots \mathbb{1}_{\{\Delta S_{k_{j-1}} = a_{j-1}\}} \mathbb{E}\left[\mathbb{1}_{\{\Delta S_{k_j} = a_j\}} \Big| \mathcal{H}_{k_j - 1}\right]\right]$$
$$= \mathbb{P}\left(\Delta S_{k_1} = a_1, \cdots, \Delta S_{k_{j-1}} = a_{j-1}\right) \mathbb{P}\left(\Delta S_{k_j} = a_j\right),$$

which implies the independence of the increments ΔS_k. So, $\{S_k\}_{k \geq 0}$ is distributed as simple random walk on $\mathbb{Z}$ with probability $q > 1/2$ of jumping to the right starting at the left of d_0/r. And, by construction, we have that $d_k/r \geq S_k$ for all k. $\qquad\square$

As a consequence of the coupling above, we prove the ballisticity of the walker.

Proposition 5.3 (Ballisticity of the walker). *There exists a constant $c > 0$ depending on p only, such that*

$$\liminf_{n \to \infty} \frac{\text{dist}_{T_n}(X_n, \text{root})}{n} \geq c, \ \mathbb{P}_{T_0, x_0, p}\text{-almost surely,}$$

for all initial conditions (T_0, x_0).

Proof of Proposition 5.3. Given a $q > 1/2$ and $\{S_k\}_{k \geq 0}$ a q-right biased simple random walk on $\mathbb{Z}$, Lemma 5.1 guarantees the existence of a positive constant r, depending on p and q, so that, for any initial condition on the BGRW,

$$\text{dist}_{T_{\sigma_k}}(X_{\sigma_k}, \text{root}) \geq r S_k, \ \text{a.s.}$$

By the Strong Law of Large Numbers we also have that

$$\lim_{k \to \infty} \frac{S_k}{k} = 2q - 1 = \mu > 0, \ \ \text{a.s.}$$

This implies that

$$\liminf_{k \to \infty} \frac{\text{dist}_{T_{\sigma_k}}(X_{\sigma_k}, \text{root})}{k} \geq r\mu, \ \ \text{a.s.} \tag{5.2}$$

On the other hand, by the definition of the stopping times σ_k we have that, for all k, the following inequality holds

$$|\sigma_{k+1} - \sigma_k| \leq \exp\{\sqrt{r}\},$$

almost surely, as well as

$$\left|\text{dist}_{T_{\sigma_k}}(X_{\sigma_k}, \text{root}) - \text{dist}_{T_{\sigma_{k-1}}}(X_{\sigma_{k-1}}, \text{root})\right| \leq r,$$

almost surely. Thus, if n is an integer such that $n \in [\sigma_k, \sigma_{k+1}]$ we have

$$\frac{\text{dist}_{T_n}(X_n, \text{root})}{n} \geq \frac{\text{dist}_{T_{\sigma_{k+1}}}(X_{\sigma_{k+1}}, \text{root}) - r}{e^{\sqrt{r}}(k+1)},$$

which combined with (5.2) proves that

$$\liminf_{n \to \infty} \frac{\text{dist}_{T_n}(X_n, \text{root})}{n} \geq c, \ \ \text{a.s,}$$

for some positive constant c depending on p only. $\qquad\square$

Remark 5.4 (A lower bound for the speed)**.** From the proof of Proposition 5.3 one can derive a lower bound for the speed $c(p)$ of the form $\exp\{-C'p^{-C'}\}$, for some universal constant C'. In our argument, given $q > 1/2$ we choose r satisfying

$$1 - \frac{C}{\sqrt{r}} \geq q, \tag{5.3}$$

and obtain that

$$\liminf_{n\to\infty} \frac{\mathrm{dist}_{T_n}(X_n, \mathrm{root})}{n} \geq \frac{r(2q-1)}{e^{\sqrt{r}}}. \tag{5.4}$$

Since q can be chosen arbitrarily close to $1/2$ and independent of p, to obtain a lower bound for the velocity as a function of p, one must know how C depends on p. By inspection of the proof of Lemma 5.1, the constant C is coming from Corollary 4.3 and it is obtained by setting $K = \sqrt{\ell}$ in the expression $1 - \left(1 - \frac{1}{\ell}\right)^{c_2 K} + \exp\{-c_1 K\}$, where $c_1 = (4\log(1 + \varepsilon p))^{-1}$, $c_2 = 8c_1/q$, $q = \left(1 - e^{-(1-\varepsilon)^2 p}\right)e^{-3/2}$ and ε is an auxiliary parameter in $(0, 1)$. Choosing $q = 2/3$, $\varepsilon = 1/2$ and using that $(1 - 1/x)^x \geq e^{-3/2}$ for all $x \geq 2$, one obtains that $C = 3c_2$, and Equation (5.3) is satisfied with $\sqrt{r} = 9c_2$. Finally, replacing this expression for r in Equation (5.4) one obtains a lower bound of the form $\exp\{-C'p^{-C'}\}$.

5.1 Controlling returns and degrees

The coupling gives us a picture of the dynamics of the walker: it is moving away from the root at linear speed. Before moving on, we use the coupling to show that vertices are not visited many times and (as a result) degrees in our tree tend to be small.

Lemma 5.5. *Let (T_0, x_0) be the initial state of the dynamics and $y \in T_0$ be given. Then for any $0 < p \leq 1$, the number $v(y)$ of visits to y,*

$$v(y) := |\{0 \leq t < +\infty \,:\, X_t = y\}|,$$

satisfies

$$\forall k \geq 0 \,:\, \mathbb{P}_{T_0,x_0,p}(v(y) \geq k) \leq C\, e^{-\beta\, k},$$

and

$$\mathbb{P}_{T_0,x_0,p}(v(y) > 0) \leq C\, e^{-\beta\, \mathrm{dist}_{T_0}(x_0, y)},$$

for some $\beta, C > 0$ depend only on p. Moreover, if $p_0 > 0$ and $p_0 \leq p \leq 1$, C and α can be chosen uniformly in p_0.

Proof of Lemma 5.5. Recall from the coupling we just constructed that we may choose r large enough so that the process $\{\mathrm{dist}_{T_{\sigma_k}}(X_{\sigma_k}, y)/r\}_k$ dominates a right biased random walk $\{S_k\}_k$ on $\mathbb{Z}$ starting on $s_0 := \lfloor \frac{\mathrm{dist}_{T_0}(x_0,y)}{r} \rfloor$. We count the visits to y per time interval $(\sigma_{k-1}, \sigma_k]$ (with a possible additional visit at time 0).

$$v(y) := \mathbb{1}_{\{X_0 = y\}} + \sum_{k=1}^{\infty} \sum_{t=\sigma_{k-1}+1}^{\sigma_k} \mathbb{1}_{\{X_t = y\}}.$$

Recalling that $\sigma_k - \sigma_{k-1} \leq e^{\sqrt{r}}$, we may rewrite the expression as:

$$v(y) \leq \mathbb{1}_{X_0 = y} + e^{\sqrt{r}} \sum_{k=1}^{\infty} \mathbb{1}_{\{X_t = y \text{ for some } \sigma_{k-1} < t \leq \sigma_k\}}.$$

Now, by the definition of the stopping time sequence $\{\sigma_{k+1}\}_k$, for $t \in [\sigma_k, \sigma_{k+1}]$ then

$$|\mathrm{dist}_{T_{\sigma_k}}(X_{\sigma_k}, y) - \mathrm{dist}_{T_t}(X_t, y)| \leq r.$$

As a result, if $X_t = y$ for some $\sigma_{k-1} \leq t \leq \sigma_k$, then $S_k \leq \mathrm{dist}_{T_{\sigma_k}}(X_{\sigma_k}, y)/r \leq 1$. In particular,

$$v(y) \leq \mathbb{1}_{\{X_0 = y\}} + e^{\sqrt{r}} \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k \leq 1\}}.$$

The RHS is (up to a constant) the number of visits of a right-biased random walk started from $S_0 \geq 0$ to the interval $(-\infty, 1]$. This number has an exponential tail that only depends on the bias. The first inequality follows because we can choose $r = r(p_0)$ to guarantee a bias of $1/3$ (say) for all $p_0 \leq p \leq 1$. The second inequality also follows (perhaps with changes to C, β) once we realize that $S_0 \geq \mathrm{dist}_{T_0}(x_0, y)/r$. $\qquad\square$

Lemma 5.6. *Given $p_0 > 0$, there exist constants $C, \alpha > 0$ depending only on p_0 such that, for all $p_0 \leq p \leq 1$, all finite (T_0, x_0), and all $n, k \geq 1$,*

$$\sum_{t=0}^{n} \mathbb{P}_{T_0, x_0, p}(d_{T_t}(X_t) \geq k) \leq C\left(n + |V(T_0)|\right) e^{\alpha\,(\Delta(T_0) - k)},$$

where $d_{T_t}(X_t)$ denotes the degree of vertex X_t in T_t.

Proof of Lemma 5.6. Assume without loss of generality that $V(T) = \{1, \ldots, \ell\}$. Also let $\ell + 1, \ell + 2, \ldots$ be the vertices that the BGRW process started from T, x creates. Finally, we let $v_t(i)$ denote the number of visits to i up to time t, so that $v_t(i) = 0$ if $i > t + \ell$. Note that $d_t(v_i) \geq k$ implies that $v_n(i) \geq k - \Delta(T)$, as the degree only grows at times when i is visited. Therefore, for all $0 \leq t \leq n$, $\alpha > 0$ and $k \geq \ell + 1$:

$$\mathbb{P}_{T, o, p}(d_{T_t}(X_t) \geq k) = \sum_{i=1}^{\ell + n} \mathbb{P}_{T, o, p}(X_t = i, d_{T_t}(i) \geq k)$$

$$\leq e^{\alpha\,(\Delta(T) - k)} \sum_{i=1}^{\ell + n} \mathbb{E}_{T, o, p}[\mathbb{1}_{\{X_t = i\}} \exp(\alpha\, v_n(i))].$$

As a consequence:

$$\sum_{t=0}^{n} \mathbb{P}_{T, o, p}(d_{T_t}(X_t) \geq k) \;\leq\; e^{\alpha\,(\Delta(T) - k)} \sum_{t=0}^{n} \sum_{i=1}^{\ell + n} \mathbb{E}_{T, o, p}[\mathbb{1}_{\{X_t = i\}} \exp(\alpha\, v_n(i))]$$

$$(\text{use } \textstyle\sum_{t \leq n} \mathbb{1}_{\{X_t = i\}} = V_n(i)) \;\leq\; e^{\alpha\,(\Delta(T) - k)} \sum_{i=1}^{\ell + n} \mathbb{E}_{T, o, p}[v_n(i)\, \exp(\alpha\, v_n(i))].$$

By Lemma 5.5, $v_n(i) \leq v(i)$ has an exponential tail uniformly in i, n, T_0, x_0 and $p \in [p_0, 1]$. Thus we may take $\alpha = \beta/2$ (with β from Lemma 5.5) and adjust C to obtain our result. $\quad\square$

6 The local point of view on infinite trees

In this section we start our study of the environment as seen by the walker. From now on, we consider the pair (T_t, X_t) as random elements of the space of *rooted locally finite trees*.

6.1 Rooted graphs and trees

We recall the definition of rooted graphs, rooted trees and the local topology on these objects. All notions we need are defined and studied in Bordenave's lecture notes [4].

A *rooted graph* is a pair (G, o) where G is a connected, locally finite graph and o is vertex of G. Note that any rooted graph must have a countable vertex set which

we may assume to be a subset of $\mathbb{N}$ or $\mathbb{Z}$. Two rooted graphs (G, o) and (G', o') are rooted-isomorphic – denoted by $(G, o) \simeq (G, o')$ – if there is a bijection of the vertex sets of G and G' that maps o to o' and preserves edges. We let $[G, o]$ denote the *equivalence class* of (G, o) under rooted isomorphisms, and let $\mathcal{G}_*$ denote the set of all such $[G, o]$.

Given $h \in \mathbb{N}$ and a rooted graph (G, o), $(G, o)_h$ denotes the graph obtained by retaining only the vertices of G within graph-theoretic distance h from o and the edges between those vertices. Clearly, $(G, o) \simeq (G', o')$ implies $(G, o)_h \simeq (G'o')_h$ for all $h > 0$, and we may define $[G, o]_h$ as the equivalence class of $(G, o)_h$. The *distance* between $[G, o], [G', o'] \in \mathcal{G}_*$ is defined by:

$$\rho([G, o], [G', o']) := \frac{1}{1 + \sup\{h \in \mathbb{N} \ : \ [G, o]_h = [G', o']_h\}}.$$

One can show that $(\mathcal{G}_*, \rho)$ is a Polish metric space.

The set $\mathcal{T}_* \subset \mathcal{G}_*$ of rooted trees is the set of equivalence classes $[T, o]$ where T is a locally finite tree. This is a closed subset of $\mathcal{G}_*$ and is therefore a Polish space with the metric ρ.

Our BGRW dynamics (see, Section 2.2) may be naturally extended to the set $\mathcal{T}_*$ of random rooted trees. The idea is that the state $[T_t, X_t]$ describes the tree "rooted at the position of the walker". With some abuse of notation, we use K_p to denote the Markov transition kernel of our process over this space as well.

6.2 Empirical measures and local functions

Much of the remainder of the paper will be spent dealing with the tree as viewed by the walker. More precisely, we will study the *empirical measure of the tree around the walker*.

Definition 6.1. *Given a realization $[T_t, X_t]$ of the process K_p, we let $\widehat{P}_n$ denote the empirical measure, that is, the random probability measure over $\mathcal{T}_*$ given by:*

$$\widehat{P}_n = \frac{1}{n+1} \sum_{t=0}^{n} \delta_{[T_t, X_t]}.$$

Thus for a given element $[T, v] \in \mathcal{T}_*$ and $h \in \mathbb{N}$,

$$(n + 1)\widehat{P}_n([G, o] \in \mathcal{T}_* \ : \ [G, o]_h = [T, v]_h),$$

counts the number of times $0 \le t \le n$ at which the ball of radius h around X_t in T_t is isomorphic to $[T, v]_h$.

In this section we show the convergence of empirical averages of *bounded local functions* on the space $\mathcal{T}_*$ under suitable assumptions on $[T_0, x_0]$. A specific instance of this general result easily implies that the walker has a well-defined speed. In Section 7 we show that $\widehat{P}_n \Rightarrow P_p$ almost surely, where P_p is an invariant measure for the process K_p. In the proof of such results, *local functions* will play an important role.

Definition 6.2 (Local function). *A function $\psi : \mathcal{T}_* \to \mathbb{R}$ is said to be h-local if $\psi([T, o]) = \psi([T, o]_h)$ for all $[T, o] \in \mathcal{T}_*$. A function ψ is local if it is h-local for some $h \in \mathbb{N}$.*

To avoid clutter, with a slight abuse of notation, we will write $\psi(G, o)$ instead of $\psi([G, o])$.

In order to prove the convergence of empirical averages we restrict ourselves on measures on $\mathcal{T}_*$ which have some nice properties. To define this class of measures we first define certain stopping times. Given $\ell \in \mathbb{N}\backslash\{0\}$, let Q_ℓ be the path of length ℓ, i.e. the graph with vertex set $\{0, 1, \ldots, \ell\}$ and edges $\{(i - 1), i\}$ ($1 \le i \le \ell$). We define τ_ℓ as the first time when the graph around X_t in T_t is Q_ℓ.

$$\tau_\ell := \inf\{t \ge 0 \ : \ [T_t, X_t]_\ell = [Q_\ell, \ell]\}.$$

Definition 6.3. *A measure μ over pairs (T, x) is called p-escapable if $\tau_\ell < +\infty$ $\mathbb{P}_{\mu,p}$-almost surely for every $\ell \geq 1$.*

We can now state the main theorem of this section.

Theorem 6.4. *For each parameter $0 < p \leq 1$ and each integer $h \geq 1$ and all h-local function $\psi : \mathcal{T}_* \to \mathbb{R}$, there exists a constant $\mathcal{M}_p \psi$ such that*

$$\lim_{n \to +\infty} \widehat{P}_n \psi = \mathcal{M}_p \psi \quad \mathbb{P}_{\mu,p}\text{-almost surely},$$

for all p-escapable distribution μ on $\mathcal{T}_*$.

As we will see in Section 7, this result implies that "the tree seen by the walker" converges to an infinite random tree when $n \to +\infty$.

The proof of Theorem 6.4 relies on the following results. The first lemma shows that the empirical measures of processes started from two p-escapable measures are asymptotically the same, i.e. the initial distribution is "forgotten". The second lemma shows that the empirical measures of local functions when our process starts from specific p-escapable measures, corresponding to $\delta_{[T_0, x_0]}$ with T_0 finite, converges to a constant.

Lemma 6.5 (Forgetfulness of empirical measures; proof in §6.5)**.** *Let μ, ν be two p-escapable measures. Let h be a positive integer and $\psi : \mathcal{T}_* \to \mathbb{R}$ be a bounded h-local function. If there exists a constant $\mathcal{M}_p \psi$ such that $\widehat{P}_n \psi \to \mathcal{M}_p \psi$ $\mathbb{P}_{\nu,p}$-almost surely, then we also have $\widehat{P}_n \psi \to \mathcal{M}_p \psi$ $\mathbb{P}_{\mu,p}$-almost surely.*

The above statement allows us to restrict ourselves to subclasses of initial distributions and then extend the results to the whole class of p-escapable distributions. This procedure greatly reduces the amount of work in our proofs. The next lemma is a finite version of Theorem 6.4.

Lemma 6.6 (proof in §6.6)**.** *Consider an initial condition (T_0, x_0) with T_0 finite. Fix $0 < p \leq 1$ and a bounded h-local function $\psi : \mathcal{T}_* \to \mathbb{R}$ (with h a positive integer). Then there exists a constant $\mathcal{M}_p \psi$ such that $\widehat{P}_n \psi \to \mathcal{M}_p \psi$, $\mathbb{P}_{T_0, x_0, p}$-almost surely.*

Combining the two lemmas gives us a straightforward proof of Theorem 6.4.

Proof of Theorem 6.4. Let $[T_0, x_0]$ be a finite rooted tree and consider $\nu = \delta_{[T_0, x_0]}$. By Lemma 6.6, we have that $\widehat{P}_n \psi \to \mathcal{M}_p \psi$, $\mathbb{P}_{\nu,p}$-almost surely. But, by Lemma 5.6, ν is p-escapable. Therefore, by Lemma 6.5, $\widehat{P}_n \psi \to \mathcal{M}_p \psi$, $\mathbb{P}_{\mu,p}$-almost surely. $\qquad\square$

Now we know how Theorem 6.4 follows from our lemmas we organize the remainder of this section as follows. In the next subsection we prove Theorem 1.1 as an application of Theorem 6.4. Then, in Subsection 6.4, we discuss the concept of p-escapable trees and give quantitative criteria for escapabality. Finally, in Subsections 6.5 and 6.6, we prove Lemma 6.5 and Lemma 6.6, respectively.

6.3 Existence of the speed

In this section we prove a stronger version of Theorem 1.1 as an application of Theorem 6.4. We prove that the random walker in the BGRW model has a well-defined positive speed for any p-escapable initial distribution on $\mathcal{T}_*$. This result illustrates how the our results regarding the "point of view of the particle" can give finer information about the walk.

Theorem 6.7 (Linear speed of the walker)**.** *For each $0 < p \leq 1$, there exists a constant $0 < c(p) \leq p$ such that, for any p-escapable distribution μ on $\mathcal{T}_*$,*

$$\lim_{n \to \infty} \frac{\mathrm{dist}_{T_n}(X_n, \mathrm{root})}{n} = c(p) \ \mathbb{P}_{\mu,p}\text{-almost surely.}$$

Proof of Theorem 6.7. Our strategy is to prove that the distance from the root at time n may be written as a sum of a bounded term, a martingale, and a sum of a 1-local function computed along the trajectory. The result will then follow from Theorem 6.4.

Fix a p-escapable measure μ and define the following bounded 1-local function:

$$\psi(T, x) := \frac{p(d_T(x) - 1)}{d_T(x) + 1} + \frac{(1 - p)(d_T(x) - 2)}{d_T(x)}.$$

Note that:

$$\mathbb{E}_{\mu,p}\left[\Delta\mathrm{dist}_{T_n}(X_n, \mathrm{root}) \mid \mathcal{F}_n\right] = \psi(T_n, X_n) + (1 - \psi(T_n, X_n))\mathbb{1}_{\{X_n=\mathrm{root}\}}.$$

Thus

$$\mathrm{dist}_{T_{n+1}}(X_{n+1}, \mathrm{root}) = M_{n+1} + \sum_{j=0}^{n} \psi(T_j, X_j) + C_{n+1},$$

where $\{M_j\}_j$ is a martingale with bounded increments and

$$C_{n+1} := \sum_{j=0}^{n} (1 - \psi(T_n, X_n))\mathbb{1}_{\{X_n=\mathrm{root}\}},$$

is bounded by the total number of visits to the root, which is a.s. finite because our process is transient.

Since the martingale has bounded increments, Azuma's inequality implies that $M_n/n \to 0$ almost surely. Since C_n is a.s. bounded, $C_n/n \to 0$ almost surely as well. Theorem 6.4 gives us that

$$\lim_{n\to+\infty} \frac{\mathrm{dist}_{T_{n+1}}(X_{n+1}, \mathrm{root})}{n + 1} = \lim_{n\to\infty} \frac{1}{n + 1} \sum_{j=0}^{n} \psi(T_j, X_j) = c(p) := \mathcal{M}_p(\psi),\ \mathbb{P}_{\mu,p}\text{-}a.s,$$

with $\mathcal{M}_p(\psi)$ is a constant depending on p and ψ only (but not on μ). Proposition 5.3 implies $c(p) > 0$. $\qquad\square$

6.4 Escapable trees and mixing

The concept of escapable trees is closely related to the forgetfulness of the walker. A tree T is p-escapable if the measure $\delta_{(T,x)}$ is p-escapable for every $x \in V(T)$. Note that, by the Markov property it is enough to show that $\delta_{(T,x)}$ is p-escapable for a particular $x \in V(T)$. In this section we prove a quantitative criteria for checking "escapability". Let us notice that not all trees are p-escapable. The next example gives a useful example to keep in mind.

Example 6.8 (A hard tree). Consider an infinite rooted tree T_{hard}, with a root node x_{hard}, such that each node at distance h from the root has $g(h) > 0$ children, with $1/g(h)$ summable. Consider the BGRW from $(T_{\mathrm{hard}}, x_{\mathrm{hard}})$. One can show via the Borel-Cantelli Lemma that there is a positive probability that $\mathrm{dist}_{T_t}(X_t, x_{\mathrm{hard}}) = t$ for all $t \in \mathbb{N}$ (i.e. the walker simply "walks down the tree" for eternity). For the same reason, for any $\ell \in \mathbb{N}\setminus\{0\}$ there is a positive probability that $\tau_\ell = +\infty$.

The next lemma explains one reason why p-escapability is important to us. The lemma gives a quantitative criterion for "escapability". In a way, it says that what makes the tree in Example 6.8 non-escapable is that X_t sees very large degrees along the way.

Lemma 6.9 (Quantitative escapability). *Assume μ is a starting measure for which there exist $C, D, \alpha > 0$ such that, for all $n, k \geq 1$,*

$$\sum_{t=0}^{n} \mathbb{P}_{\mu,p}(d_{T_t}(X_t) \geq k) \leq C (n + D) e^{-\alpha k}. \tag{6.1}$$

Then μ is p-escapable and

$$\mathbb{E}_{\mu,p}[\tau_\ell \wedge n] \leq \frac{\ell 2^{\ell-1}}{p^\ell} \left\lceil \frac{2\log(n+D) + \log C}{\alpha} \right\rceil + 1. \qquad (6.2)$$

Moreover, the following event holds with probability $\geq 1 - 1/n^{16}$: for all times $1 \leq t \leq n$,

$$\mathbb{E}_{T_t,X_t,p}[\tau_\ell \wedge n] \leq \frac{\ell 2^{\ell-1}}{p^\ell} \left\lceil \frac{\log(2n+D) + 17\log n + \log C}{\alpha} \right\rceil + 1. \qquad (6.3)$$

In particular, all measures $\mu = \delta_{(T,x)}$, with T a finite tree, satisfy the above inequalities, with $C \leq e^{\alpha \Delta(T)}$ by virtue of Lemma 5.6.

So Lemma 6.9 also guarantees that not only μ is p-escapable, but also that, with high probability, each intermediate $[T_t, X_t]$ satisfies a quantitative escapability property. One important point is that the assumptions of Lemma 6.9 always apply to initial trees that are finite, by virtue of Lemma 5.6.

Proof of Lemma 6.9. We first prove inequality (6.2) for $\mathbb{E}_{\mu,p}[\tau_\ell \wedge n]$, noting that it implies p-escapability because, assuming it, we can show:

$$\mathbb{P}_{\mu,p}(\tau_\ell = +\infty) \leq \frac{\mathbb{E}_{\mu,p}[\tau_\ell \wedge n]}{n} \leq \frac{1}{n}\left(\frac{\ell 2^{\ell-1}}{p^\ell} \left\lceil \frac{2\log(n+D) + \log C}{\alpha} \right\rceil + 1 \right) \overset{n \to +\infty}{\to} 0.$$

Fix some $\Delta > 0$. Let η_Δ be the first time the walker "sees" a node with degree $\geq \Delta$,

$$\eta_\Delta := \inf\left\{ m \geq 0 \mid d_{T_m}(X_m) \geq \Delta \right\},$$

and consider the event A_i in which at time $i\ell$ the walker creates a new leaf, jumps to it and grows a path of length $\ell - 1$ from it in such way that at time $i\ell + \ell - 1$ the walker finds itself at the tip of a path of length ℓ. Now we define another stopping time

$$\tilde{\tau}_\ell := \inf\left\{ i \geq 0 \mid A_i \text{ occurs } \right\},$$

and note that $\tau_\ell \leq \ell \tilde{\tau}_\ell$.

By the definition of the event A_i, we have that

$$\mathbb{P}_{\mu,p}(A_i | \mathcal{F}_{i\ell}) = \frac{p}{d_{T_{i\ell}}(X_{i\ell})} \left(\frac{p}{2} \right)^{\ell-1}.$$

Thus, by the definition of the stopping time η_Δ, we also have

$$\mathbb{1}\{\eta_\Delta > i\ell\}\mathbb{P}_{\mu,p}(A_i | \mathcal{F}_{i\ell}) \geq \frac{p}{\Delta}\left(\frac{p}{2} \right)^{\ell-1} \mathbb{1}\{\eta_\Delta > i\ell\}.$$

Using the above inequality and noticing that $\mathbb{1}\{\eta_\Delta > k\ell\} \leq \mathbb{1}\{\eta_\Delta > (k-1)\ell\}$, we obtain

$$\mathbb{P}_{\mu,p}\left(\tilde{\tau}_\ell \geq k, \eta_\Delta > k\ell\right) = \mathbb{E}_{\mu,p}\left[\mathbb{E}_{\mu,p}\left[\mathbb{1}_{A_k^c} | \mathcal{F}_{k\ell} \right] \mathbb{1}\{\eta_\Delta > k\ell\} \prod_{i=0}^{k-1} \mathbb{1}_{A_i^c} \right]$$

$$\leq \left(1 - \frac{p}{\Delta}\left(\frac{p}{2} \right)^{\ell-1} \right) \mathbb{P}_{\mu,p}\left(\tilde{\tau}_\ell \geq k-1, \eta_\Delta > (k-1)\ell \right).$$

Then, proceeding by induction, we obtain the following upper bound

$$\mathbb{P}_{\mu,p}\left(\tilde{\tau}_\ell \geq k, \eta_\Delta > k\ell\right) \leq \left(1 - \frac{p}{\Delta}\left(\frac{p}{2} \right)^{\ell-1} \right)^k.$$

Since

$$\mathbb{P}_{\mu,p}\left(\tau_\ell \wedge n \geq k\ell\right) \leq \begin{cases} \mathbb{P}_{\mu,p}\left(\tilde{\tau}_\ell \geq k\right), & k\ell \leq n \\ 0, & \text{otherwise;} \end{cases}$$

we have:

$$\mathbb{P}_{\mu,p}\left(\tau_\ell \wedge n \geq k\ell\right) \leq \mathbb{P}_{\mu,p}\left(\tilde{\tau}_\ell \geq k, \eta_\Delta > k\ell\right) + \mathbb{P}_{\mu,p}\left(\eta_\Delta \leq k\ell\right),$$

and consequently,

$$\mathbb{E}_{\mu,p}\left[\tau_\ell \wedge n\right] \leq \sum_{j=0}^{\lfloor \frac{n}{\ell} \rfloor} \ell \mathbb{P}_{\mu,p}\left(\tau_\ell \wedge n \geq j\,\ell\right) \leq \frac{\ell\,2^{\ell-1}}{p^\ell}\Delta + n\,\mathbb{P}_{\mu,p}\left(\eta_\Delta \leq n\right). \tag{6.4}$$

But, by our condition (7.1),

$$\mathbb{P}_{\mu,p}\left(\eta_\Delta \leq n\right) \leq \sum_{t=0}^{n} \mathbb{P}_{\mu,p}(d_{T_t}(X_t) \geq \Delta) \leq C\,(n+1)\,e^{-\alpha\,\Delta}, \tag{6.5}$$

so

$$\mathbb{E}_{\mu,p}\left[\tau_\ell \wedge n\right] \leq \frac{\ell\,2^{\ell-1}}{p^\ell}\Delta + C\,(n+1)^2 e^{-\alpha\,\Delta}.$$

We may choose

$$\Delta = \left\lceil \frac{2\log(n+1) + \log C}{\alpha} \right\rceil,$$

to finish the proof of (6.2).

To prove the last statement in the lemma, we go back to (6.5) and note that was the only step in that proof where we used (6.1). In particular, we can go back to (6.4) and obtain:

$$\mathbb{E}_{T_t,X_t,p}\left[\tau_\ell \wedge n\right] \leq \frac{\ell\,2^{\ell-1}}{p^\ell}\Delta + n\,\mathbb{P}_{T_t,X_t,p}(\eta_\Delta \leq n).$$

In particular, if Δ is given and we consider the event

$$\mathcal{G}(\Delta) := \bigcap_{t=0}^{n} \{\mathbb{P}_{T_t,X_t,p}(\eta_\Delta \leq n) \leq 1/n)\},$$

we have that, when $\mathcal{G}(\Delta)$ holds, then:

$$\mathbb{E}_{T_t,X_t,p}[\tau_\ell \wedge n] \leq \frac{\ell\,2^{\ell-1}}{p^\ell}\Delta + 1, \tag{6.6}$$

for each $0 \leq t \leq n$. To finish, we show that $\mathbb{P}_{\mu,p}(\mathcal{G}(\Delta)) \geq 1 - n^{-16}$ for an appropriate choice of Δ. Indeed, by union bound and Markov's inequality:

$$1 - \mathbb{P}_{\mu,p}(\mathcal{G}(\Delta)) \leq n \sum_{t=0}^{n} \mathbb{E}_{\mu,p}[\mathbb{P}_{T_t,X_t,p}(\eta_\Delta \leq n)].$$

Now by the Simple Markov property, for $0 \leq t \leq n$

$$\mathbb{E}_{\mu,p}[\mathbb{P}_{T_t,X_t,p}(\eta_\Delta \leq n)] = \mathbb{P}_{\mu,p}(\exists\, t \leq s \leq t+n \,:\, d_{T_s}(X_s) \geq \Delta) \leq \mathbb{P}_{\mu,p}(\eta_\Delta \leq 2n).$$

And again, by condition (6.1),

$$1 - \mathbb{P}_{\mu,p}(\mathcal{G}) \leq n\,\mathbb{P}_{\mu,p}(\eta_\Delta \leq 2n) \leq n \sum_{t=0}^{2n} \mathbb{P}_{\mu,p}(d_t(X_t) \geq \Delta) \leq C\,n\,(2n+D)\,e^{-\alpha\,\Delta}.$$

Taking

$$\Delta := \left\lceil \frac{\log(2n+D) + 17\log n + \log C}{\alpha} \right\rceil,$$

guarantees $\mathbb{P}_{\mu,p}(\mathcal{G}(\Delta)) \geq 1 - n^{-16}$. Plugging this choice of Δ into (6.6) gives (6.3) inside the event $\mathcal{G}(\Delta)$. $\qquad\square$

6.5 Forgetfulness of empirical measures: proof of Lemma 6.5

The proof of the lemma will follow as a consequence of another result that highlight the relation between "escapabability" and forgetfulness. As we will see, it implies an approximated renewal property of the BGRW process in the sense that after a random time it may be coupled to a BGRW process starting from a semi-infinite path. Then, both walkers see around them the same tree structure as long as this coupling works.

For this purpose, we will need additional notation. Let Q^* be a semi-infinite path, i.e.,

$$Q^* = \{v_0, v_1, v_2, \dots\}, \tag{6.7}$$

rooted at v_0 and consider a trajectory $(T_t^*, X_t^*)_{t \geq 0}$ of BGRW started from (Q^*, v_0) with empirical measures $\widehat{P}_n^*$. Also, for each pair of integers h and n, let $\pi(h, n)$ be

$$\pi(h, n) := \mathbb{P}_{Q_0^*, v_0, p}\left(\exists m \leq n, \ X_m^* = v_h\right). \tag{6.8}$$

Lemma 6.10. *Let $\psi : \mathcal{T}_* \to \mathbb{R}$ be a h-local and bounded and μ be a p-escapable distribution. Consider a trajectory $(T_t, X_t)_{t \geq 0}$ of BGRW starting from μ, with empirical measures $\widehat{P}_n$. Then,*

(a) for every $\varepsilon \in (0, 1)$, there exists a coupling of the processes $(T_t, X_t)_{t \geq 0}$ and $(T_t^*, X_t^*)_{t \geq 0}$ such that:

$$\mathbb{P}\left(\lim_{n \to +\infty} |(\widehat{P}_n - \widehat{P}_n^*)\psi| = 0\right) \geq 1 - \varepsilon.$$

(b)

$$\left|\mathbb{E}_{\mu, p}\left[\widehat{P}_n \psi\right] - \mathbb{E}_{Q^*, v_0, p}\left[\widehat{P}_n^* \psi\right]\right| \leq 2\|\psi\|_\infty \pi(h, n) + \frac{2\|\psi\|_\infty \mathbb{E}_{\mu, p}[\tau_{2h} \wedge n]}{n + 1}.$$

Proof of Lemma 6.10. Part (a). Take an $\ell \in \mathbb{N}$ bigger than $2h$. Consider a chain started from measure μ. Since μ is p-escapable, the stopping time τ_ℓ is finite $\mathbb{P}_{\mu, p}$-almost-surely.

Let $y_0 = X_{\tau_\ell}$, y_1, ..., y_ℓ be the unique vertices at distances $0, 1, 2, \dots, \ell$ at time τ_ℓ. Define a random time $\tau'_{\ell, h}$ as follows:

$$\tau'_{\ell, h} := \inf\{t \geq 0 \ : \ X_{\tau_\ell + t} = y_{\ell - h}\}.$$

Notice that this time may well be infinite. In fact, since $\text{dist}(X_{\tau_\ell}, y_{\ell - h}) = \ell - h$, the second statement in Lemma 5.5, combined with the strong Markov property and the p-escapability of μ, implies that

$$\mathbb{P}_{\mu, p}\left(\tau'_{\ell, h} < +\infty\right) \leq C\, e^{-\beta\,(\ell - h)}.$$

Note that τ_ℓ is the time X_t is at the tip of a path of length ℓ and $\tau_\ell + \tau'_{\ell, h}$ is the first time at which X_t comes within distance h of the other end of that path.

Now consider the process (T_t^*, X_t^*) started on vertex v_0 of the infinite path $v_0, v_1, \dots$. Let

$$\tau^*_{\ell, h} := \inf\{t \geq 0 \ : \ X_t^* = v_{\ell - h}\}.$$

This time may also be infinite. It is easy to see that one can couple (T_t^*, X_t^*) to $(T_{\tau_\ell + t}, X_{\tau_\ell + t})$ so that $\tau_{\ell, h} = \tau'_{\ell, h}$ and $[T_t^*, X_t^*]_h = [T_{\tau_\ell + t}, X_{\tau_\ell + t}]_h$ for all $0 \leq t \leq \tau'_{\ell, h}$. Under this coupling, if $\tau'_{\ell, h} > n$ and $\tau_\ell \leq n$, then

$$\widehat{P}_n \psi \quad = \quad \frac{1}{n + 1} \sum_{i=0}^{n} \psi(T_i, X_i)$$

$$\text{(coupling and } h\text{-locality)} = \frac{1}{n + 1} \sum_{i=0}^{\tau_\ell \wedge n - 1} \psi(T_i, X_i) + \frac{1}{n + 1} \sum_{t=0}^{n - \tau_\ell \wedge n} \psi(T_t^*, X_t^*)$$

$$= \quad \widehat{P}_n^* \psi + \frac{1}{n + 1} \sum_{i=0}^{\tau_\ell \wedge n - 1} \left(\psi(T_i, X_i) - \psi(T_{n - \tau_\ell \wedge n + i + 1}, X_{n - \tau_\ell \wedge n + i + 1})\right). \tag{6.9}$$

Since $\tau_\ell < +\infty$ is bounded and $\widehat{P}_n\psi \leq \|\psi\|_\infty$, this implies that, under the coupling:

$$\tau'_{\ell,\rho} > n \Rightarrow |\widehat{P}_n\psi - \widehat{P}^*_n\psi| \leq \frac{2\|\psi\|_\infty\,(\tau_\ell \wedge n)}{n+1} \to 0.$$

So, for any fixed ℓ, we have

$$\mathbb{P}\left(\lim_{n\to+\infty} |(\widehat{P}_n - \widehat{P}^*_n)\psi| = 0\right) \geq \mathbb{P}(\tau'_{\ell,h} = +\infty) \geq 1 - C\,e^{-\beta\,(\ell-h)}.$$

In particular, this can be made greater than $1 - \varepsilon$ with an appropriate choice of ℓ. This proves part (a).

Part (b). Put $\ell = 2h$. From Equation (6.9) and the fact that under the coupling $\tau'_{\ell,h} = \tau^*_{\ell,h}$, we may obtain that

$$\left|\widehat{P}_n\psi - \widehat{P}^*_n\psi\right| = \left|\widehat{P}_n\psi - \widehat{P}^*_n\psi\right| \mathbb{1}\{\tau'_{\ell,h} \leq n\} + \left|\widehat{P}_n\psi - \widehat{P}^*_n\psi\right| \mathbb{1}\{\tau'_{\ell,h} > n\}$$

$$\leq 2\|\psi\|_\infty \mathbb{1}\{\tau^*_{\ell,h} \leq n\} + \frac{2\|\psi\|_\infty\,(\tau_\ell \wedge n)}{n+1},$$

taking the expected value both sides it is enough to proves part (b). $\qquad\square$

Now we see how Lemma 6.5 follows immediately from the above result.

Proof of Lemma 6.5. From part (a) of Lemma 6.10, one sees that, given a deterministic $c \in \mathbb{R}$,

$$\mathbb{P}_{\mu,p}(\widehat{P}_n\psi \to c) = 1 \Leftrightarrow \mathbb{P}_{Q^*,v_0,p}(\widehat{P}^*_n\psi \to c) = 1.$$

This holds for *any* p-escapable measure μ. In particular, if ν is *also* p-escapable the same must hold with ν replacing μ. $\qquad\square$

6.6 Convergence of average of local functions: proof of Lemma 6.6

The proof of Lemma 6.6 will follow essentially from the lemma below that states that, when started from a finite initial condition, the average of local functions is close to its expected value but started from the semi-infinite path Q^* introduced in (6.7).

Lemma 6.11. *Given $p_0 > 0$, there exist constants $C, \alpha > 0$ depending only on p_0 such that, for all $p_0 \leq p \leq 1$, all finite (T_0, x_0), all $-1 \leq \psi \leq 1$ h-local, all $\varepsilon > 0$, and all $1 \leq q < n$, $\lfloor n/q \rfloor > 0$,*

$$\mathbb{P}_{T_0,x_0,p}\left(\left|\widehat{P}_n\psi - \mathbb{E}_{Q^*,v_0,p}\left[\widehat{P}_q\psi\right]\right| \geq \varepsilon + \frac{2q}{n} + err(q,h)\right) \leq 2\exp\{-\varepsilon^2\lfloor n/q \rfloor/32\} + \frac{1}{q^{16}},$$

where

$$err(q,h) := 4\pi(h,q) + \frac{h2^{2h}}{q\,p^{2h}}\left\lceil \frac{\log(2q + |V(T_0)|) + 17\log q + \alpha\Delta(T_0)\log C}{\alpha}\right\rceil + \frac{1}{q}.$$

Proof of Lemma 6.11. Let $n = kq + s$, with $k \geq 1$. Obseve that, for every $k, q \geq 1$, we can write

$$\widehat{P}_{kq-1}\psi = \frac{1}{kq}\sum_{i=0}^{kq-1}\psi(T_i, X_i) = \frac{1}{k}\left(\frac{1}{q}\sum_{i=0}^{q-1}\psi(T_i, X_i) + \cdots + \frac{1}{q}\sum_{i=kq-q}^{kq-1}\psi(T_i, X_i)\right)$$

$$= \frac{1}{k}\sum_{j=1}^{k}S_j,$$

where, $S_j := \dfrac{1}{q} \sum\limits_{i=(j-1)q}^{jq-1} \psi(T_i, X_i)$, for $j \in \{1, \cdots, k\}$. Let us denote by $\{\tilde{\mathcal{F}}_j\}_{j \geq 0}$ the following filtration:

$$\tilde{\mathcal{F}}_0 = \mathcal{F}_0; \quad \tilde{\mathcal{F}}_j = \mathcal{F}_{jq-1}.$$

By the definition of S_j, the process $\{S_j\}_{j \geq 1}$ is adapted to the filtration $\{\tilde{\mathcal{F}}_j\}_{j \geq 0}$. Moreover, the process $\{M_k\}_{k \geq 1}$ defined as

$$M_k := \sum_{j=1}^{k} S_j - \mathbb{E}_{T_0, x_0, p}\left[S_j \Big| \tilde{\mathcal{F}}_{j-1}\right] = M_{k-1} + S_k - \mathbb{E}_{T_0, x_0, p}\left[S_k \Big| \tilde{\mathcal{F}}_{k-1}\right],$$

is a mean zero martingale with respect to $\{\tilde{\mathcal{F}}_j\}_{j \geq 0}$, whose increments are bounded by 2, since S_j is an average of random variables bounded by 1.

With all these definitions, by triangle inequality, we have that

$$\left|\widehat{P}_n \psi - \mathbb{E}_{Q^*, v_0, p}\left[\widehat{P}_q \psi\right]\right| \leq \left|\widehat{P}_n \psi - \widehat{P}_{kq-1} \psi\right| \quad (a)$$

$$+ \left|\widehat{P}_{kq-1}(\psi) - \frac{1}{k} \sum_{j=1}^{k} \mathbb{E}_{T_0, x_0, p}\left[S_j \Big| \tilde{\mathcal{F}}_{j-1}\right]\right| \quad (b)$$

$$+ \left|\frac{1}{k}\left(\sum_{j=1}^{k} \mathbb{E}_{T_0, x_0, p}\left[S_j \Big| \tilde{\mathcal{F}}_{j-1}\right] - \mathbb{E}_{Q^*, v_0, p}\left[\widehat{P}_q \psi\right]\right)\right| \quad (c).$$

As far as (a) is concerned, we observe that the following deterministic bound holds:

$$\left|\widehat{P}_n \psi - \widehat{P}_{kq-1} \psi\right| = \left|\left(\frac{kq}{n+1} - 1\right)\frac{1}{kq}\sum_{i=0}^{kq-1} \psi(T_i, X_i) + \frac{1}{n+1}\sum_{i=kq}^{kq+s} \psi(T_i, X_i)\right|$$

$$\leq \frac{2\|\psi\|_\infty (s+1)}{n+1} < \frac{2q}{n},$$

since, by hypothesis, $\|\psi\|_\infty \leq 1$. Thus, to prove the claim, it suffices to show that

$$\mathbb{P}_{T_0, x_0, p}\Big((b) + (c) \geq \varepsilon + err(q, h)\Big) \leq 2\exp\{-\varepsilon^2 k / 32\} + \frac{1}{q^{16}}. \qquad (6.10)$$

Using the simple observation that

$$\mathbb{P}_{T_0, x_0, p}\Big((b) + (c) \geq \varepsilon + err(q, h)\Big)$$

$$\leq \mathbb{P}_{T_0, x_0, p}\left((b) \geq \frac{\varepsilon + err(q, h)}{2}\right) + \mathbb{P}_{T_0, x_0, p}\left((c) \geq \frac{\varepsilon + err(q, h)}{2}\right),$$

we will prove Equation (6.10) by showing

$$(1): \qquad \mathbb{P}_{T_0, x_0, p}\left((b) \geq \frac{\varepsilon}{2}\right) \leq 2\exp\{-\varepsilon^2 k / 32\},$$

$$(2): \qquad \mathbb{P}_{T_0, x_0, p}\left((c) > \frac{err(q, h)}{2}\right) \leq \frac{1}{q^{16}}.$$

To show (1), we observe that the martingale M_k can be written as the following difference

$$M_k = k\widehat{P}_{kq-1}(\psi) - \sum_{j=1}^{k} \mathbb{E}_{T_0, x_0, p}\left[S_j \Big| \tilde{\mathcal{F}}_{j-1}\right].$$

Thus, (b) is equal to $|M_k/k|$ and we apply Azuma's inequality which gives us:

$$\mathbb{P}_{T_0,x_0,p}\left(|M_k| \geq \frac{\varepsilon k}{2}\right) \leq 2\exp\{-\varepsilon^2 k/32\}.$$

To show item (2), we first observe that, by the Simple Markov Property, for all $j \in \{2, \cdots, k\}$ the following identity holds

$$\mathbb{E}_{T_0,x_0,p}\left[S_j \Big| \tilde{\mathcal{F}}_{j-1}\right] = \frac{1}{q}\sum_{i=1}^{q} \mathbb{E}_{T_{(j-1)q-1}, X_{(j-1)q-1}, p}[\psi(T_i, X_i)] = \mathbb{E}_{T_{(j-1)q-1}, X_{(j-1)q-1}, p}\left[\widehat{P}_q(\psi)\right].$$

Then, by Part (b) of Lemma 6.10 we have

$$\left|\mathbb{E}_{T_0,x_0,p}\left[S_j \Big| \tilde{\mathcal{F}}_{j-1}\right] - \mathbb{E}_{Q^*,v_0,p}\left[\widehat{P}_q(\psi)\right]\right| \leq 2\pi(h,q) + \frac{2\mathbb{E}_{T_{(j-1)q-1}, X_{(j-1)q-1}, p}\left[\tau_{2h} \wedge q\right]}{q},$$

which implies

$$(c) \leq 2\pi(h,q) + \frac{1}{k}\sum_{j=1}^{k} \frac{2\mathbb{E}_{T_{(j-1)q-1}, X_{(j-1)q-1}, p}\left[\tau_{2h} \wedge q\right]}{q}.$$

Therefore, it is enough to show that the following event

$$\underbrace{\left\{\sum_{j=0}^{k-1} \frac{2\mathbb{E}_{T_{jq-1}, X_{jq-1}, p}\left[\tau_{2h} \wedge q\right]}{k\,q} > \frac{h2^{2h}}{q\,p^{2h}}\left\lceil\frac{\log(2q + |V(T_0)|) + 17\log q + \alpha\Delta(T_0)\log C}{\alpha}\right\rceil + \frac{1}{q}\right\}}_{=:B},$$

happens with probability at most q^{-16}.

By Lemma 6.9 (using the constants coming from Lemma 5.6) we know that, if we define

$$A_t := \left\{\mathbb{E}_{T_t, X_t, p}[\tau_{2h} \wedge q] \leq \frac{h2^{2h}}{p^{2h}}\left\lceil\frac{\log(2q + |V(T_0)|) + 17\log q}{\alpha} + \Delta(T_0)\log C\right\rceil + 1\right\},$$

it holds that

$$\mathbb{P}_{T_0,x_0,p}\left(\bigcap_{t=0}^{q} A_t\right) \geq 1 - \frac{1}{q^{16}},$$

which implies $\mathbb{P}_{T_0,x_0,p}(B) \leq \frac{1}{q^{16}}$, and concludes the proof. $\qquad\square$

The next step towards the proof of Lemma 6.6 is to derive from the above lemma three small results which combined will give us the proof. The first corollary is a more convenient version of Lemma 6.11.

Corollary 6.12. *Given $p_0 > 0$ and (T_0, x_0) finite, there exists n_0 and a constant c depending only on p_0, T_0 such that, for all $n \geq n_0$, for all $p \in [p_0, 1]$ we can find values for h and q such that, for all $-1 \leq \psi \leq 1$ h-local*

$$\mathbb{P}_{T_0,x_0,p}\left(\left|\widehat{P}_n\psi - \mathbb{E}_{Q^*,v_0,p}\left[\widehat{P}_q(\psi)\right]\right| > \frac{c}{2\log n}\right) \leq \frac{2}{n^8}.$$

Proof of Corollary 6.12. We use Lemma 6.11, suitably choosing the values of q, h and ε. We begin observing that by Lemma 5.5 we have that

$$\pi(h,q) \leq \frac{C_1}{\log n}.$$

Choosing $q = n^{1/2}$ and $h = \left\lfloor \frac{\log n}{8 \log(2/p)} \right\rfloor$, we obtain that

$$err(q, h) = 4\pi(h, q) + \frac{4h2^{2h}}{q\,p^{2h}} \left\lceil \frac{\log(2q + |V(T_0)|) + 17\log q + \alpha\Delta(T_0)\log C}{\alpha} \right\rceil + \frac{1}{q}$$

$$\leq \frac{4C_1}{\log n} + \frac{\log n}{2n^{1/4}\log(2/p)} \left(\frac{\log 2n^{1/2} + |V(T_0)| + 9\log n}{\alpha} + \Delta(T_0)\log C \right) + \frac{1}{n^{1/2}}$$

$$\leq \frac{4C_1}{\log n} + \frac{\log^2 2n}{\alpha\,n^{1/4}} + \frac{|V(T_0)|\log n}{\alpha n^{1/4}} + \frac{9\log^2 n}{\alpha n^{1/4}} + \frac{\log n\,\Delta(T_0)\log C}{n^{1/4}} + \frac{1}{n^{1/2}}.$$

Given that the constants C, C_1, α only depend on p_0, and T_0 is finite, there exists a constant c and a sufficiently large n_0, both depending on p_0 and T_0 only, such that

$$err(q, h) \leq \frac{c}{6\log n}.$$

Setting $\varepsilon = c/6\log^{-1} n$ and since $2q/n = 2n^{-1/2} \leq c/6\log^{-1} n$ for sufficiently large n, by Lemma 6.11 we have that

$$\mathbb{P}_{T_0,x_0,p} \left(\left| \widehat{P}_n\psi - \mathbb{E}_{Q^*,v_0,p}\left[\widehat{P}_q\psi \right] \right| > \frac{c}{2\log n} \right) \leq 2e^{-\varepsilon^2/32\sqrt{n}} + \frac{1}{q^{16}} \leq \frac{2}{n^8},$$

for sufficiently large n (how large depends on ε which only depends on p_0, T_0). □

Corollary 6.13. *Given $p_0 > 0$ and (T_0, x_0) finite, there exists n_0 and a constant c depending only on p_0, T_0 such that, for all $n \geq n_0$, for all $p \in [p_0, 1]$ and for all $-1 \leq \psi \leq 1$ h-local:*

$$\mathbb{P}_{T_0,x_0,p} \left(\left| \widehat{P}_m\psi - \widehat{P}_{m'}\psi \right| > \frac{c}{\log n}, \text{ for some } m, m' \in \{n, n+1, \cdots, n^2\} \right) \leq \frac{4}{n^4}.$$

Proof of Corollary 6.13. By union bound, it suffices to show that there exists n_0 such that for all $n \geq n_0$ and for all $m, m' \in \{n, n+1, \cdots, n^2\}$,

$$\mathbb{P}_{T_0,x_0,p} \left(\left| \widehat{P}_m\psi - \widehat{P}_{m'}\psi \right| > \frac{c}{\log n} \right) \leq \frac{4}{n^8}.$$

From Corollary 6.12 we obtain that for every $m \in \{n, n+1, \cdots, n^2\}$

$$\mathbb{P}_{T_0,x_0,p} \left(\left| \widehat{P}_m\psi - \mathbb{E}_{Q^*,v_0,p}\left[\widehat{P}_q(\psi) \right] \right| > \frac{c}{2\log n} \right) \leq \frac{2}{n^8},$$

and the claim follows by triangle inquality. □

Corollary 6.14. *Given $p_0 > 0$, for all $p \in [p_0, 1]$, for all finite (T_0, x_0) and for all h-local function ψ such that $-1 \leq \psi \leq 1$, $\{P_n\psi\}_{n\in\mathbb{N}}$ is a Cauchy sequence $\mathbb{P}_{T_0,x_0}$-almost surely.*

Proof of Corollary 6.14. Using Corollary 6.13, together with a Borel-Cantelli argument implies that $\mathbb{P}_{T_0,x_0,p}$ almost surely there exists n_0 such that, for all $n \geq n_0$ and for all $m, m' \in \{n, n+1, \cdots, n^2\}$ we have

$$\left| \widehat{P}_m\psi - \widehat{P}_{m'}\psi \right| \leq \frac{c}{\log n}.$$

For any $\delta > 0$, we can choose n_1 large enough so that $\frac{c}{\log n_1} < \delta/2$ and $n_1 \geq n_0$. Now, for $m, m' \geq n_1$, assuming $m \leq m'$, there exists k_0 such that $m' \in \{m^{2^{k_0}}, \cdots, m^{2^{k_0+1}}\}$. Thus,

$$\left| \widehat{P}_m\psi - \widehat{P}_{m'}\psi \right| \leq \sum_{j=0}^{k_0-1} \left| \widehat{P}_{m^{2^j}}\psi - \widehat{P}_{m^{2^{j+1}}}\psi \right| + \left| \widehat{P}_{m^{2^{k_0}}}\psi - \widehat{P}_{m'}\psi \right|$$

$$\leq \sum_{j=0}^{k_0} \frac{c}{(2^j\log m)} \leq \frac{2c}{\log m} \leq \frac{2c}{\log n_1} < \delta. \qquad \square$$

Now we are able to prove Lemma 6.6.

Proof of Lemma 6.6. From Corollary 6.14 we have that $\{\widehat{P}_n\psi\}_{n\in\mathbb{N}}$ converges $\mathbb{P}_{T_0,x_0}$-almost surely. Thus, in order to prove the result, we prove first that the sequence $\{\widehat{P}_n\psi\}_{n\in\mathbb{N}}$ started from any finite initial condition always converges to the same limit. Finally, we prove that this limit must be a constant.

Consider two independent BGRW processes $(T_t, X_t)_{t\in\mathbb{N}}$ and $(T'_t, X'_t)_{t\in\mathbb{N}}$ starting from two finite initial conditions (T_0, x_0) and (T'_0, x'_0) respectively. Let $\{\widehat{P}_n\psi\}_{n\in\mathbb{N}}$ and $\{\widehat{P}'_n\psi\}_{n\in\mathbb{N}}$ be the averages associated to each process. The proof follows from (6.6) and triangle inequality, which combined give us that

$$\mathbb{P}\left(\left|\widehat{P}_n\psi - \widehat{P}'_n\psi\right| \geq \frac{c}{\log n}\right) \leq \frac{4}{n^8},$$

for some positive constant c depending on p, T_0 and T'_0 only. Then, an application of Borel-Cantelli Lemma allows us to conclude that both processes converge to the same limit $\mathcal{M}_p\psi$. Once we have that, we finally observe that since the BGRW processes are independent, the events $\{\mathcal{M}_p\psi \in A\}$ and $\{\mathcal{M}_p\psi \in B\}$ are independents for all Borel sets on the real line. This implies that $\mathcal{M}_p\psi$ is independent of itself, thus, constant almost surely. □

7 Weak convergence of empirical measures

In this section we prove Theorem 1.2 and Theorem 1.3 in the Introduction. In fact, we prove slightly stronger statements that only require the initial measure to be p-escapable, which in particular applies to all initial measures supported on finite trees (cf. Lemma 6.9). Specifically, we will show that, for any p-escapable initial measure μ, there exits a measure P_p on the space $\mathcal{T}_*$ such that, when $n \to +\infty$,

$$\widehat{P}_n := \frac{1}{n+1}\sum_{t=0}^n \delta_{[T_t, X_t]} \Rightarrow P_p \qquad \mathbb{P}_{\mu,p}\text{-almost surely},$$

$$P_n := \frac{1}{n+1}\sum_{t=0}^n \mu K_p^t \Rightarrow P_p,$$

where K_p denotes the Markov transition kernel of the BGRW process over $\mathcal{T}_*$.

In our proof we use the next proposition, which gives a convenient criterion for checking weak convergence of distributions over $\mathcal{T}_*$. Before stating the proposition we need to define a specific kind of local function. Given $h \in \mathbb{N}$ and $[T, o] \in \mathcal{T}_*$, let $D_h([T, o])$ denote the largest degree of a vertex in T within distance h from o:

$$D_h(T, o) := \max\{d_T(x) \,:\, x \in V(T), \, \mathrm{dist}_T(x, o) \leq h\}.$$

One can check that $D_h : \mathcal{T}_* \to \mathbb{R}$ is $(h+1)$-local.

Proposition 7.1 (proof in Appendix B.2). *Suppose $\{Q_n\}_{n\in\mathbb{N}}$ is a sequence of probability measures over $\mathcal{T}_*$. Then:*

1. Tightness: *If $\lim_{k\to+\infty} \limsup_{n\in\mathbb{N}} Q_n(\{[T, o] \in \mathcal{T}_* \,:\, D_h(T, o) \geq k\}) = 0$ for any $h \geq 0$, then $\{Q_n\}_{n\in\mathbb{N}}$ is tight.*

2. Weak convergence: *Assume $\{Q_n\}_{n\in\mathbb{N}}$ satisfies item (1). Then there exists a countable family $\mathcal{S}$ of bounded local functions such that, if $\lim_n Q_n f$ exists for all $f \in \mathcal{S}$, then $Q_n \Rightarrow Q$ for some Q that is uniquely defined by:*

$$Qf = \lim_n Q_n f, \, f \in \mathcal{S}.$$

We begin with addressing tightness for $\{P_n\}_{n\in\mathbb{N}}$. The next lemma shows Cèsaro-mean-type sequences of measures generated by our Markov chain are tight when the initial measure satisfy the quantitative criterion of p-escapability defined in Lemma 6.9.

Lemma 7.2. *Assume μ is a measure over $\mathcal{T}_*$ with the following property: there exist $C, \alpha, D > 0$ such that for all $n, k \geq 1$:*

$$\sum_{t=0}^{n} \mathbb{P}_{\mu,p}(d_{T_t}(X_t) \geq k) \leq C\,(n+D)\,e^{-\alpha\,k}. \tag{7.1}$$

Then the sequence

$$P_n := \frac{1}{n+1}\sum_{t=0}^{n}\mu K_p^t,$$

satisfies the tightness criterion in Proposition 7.1. In fact, the following quantitative estimate is satisfied: for all $h, k \geq 1, n \in \mathbb{N}$:

$$P_n(\{[G,o]\in\mathcal{T}_* \,:\, D_h(G,o)\geq k\}) \leq C\,\frac{(n+D+h)}{n+1}\left(\frac{(k+1)^{h+1}-1}{k}\right)e^{-\alpha\,k}.$$

Proof of Lemma 7.2. Note that:

$$(n+1)P_n(\{[G,o]\in\mathcal{T}_* \,:\, D_h(G,o)\geq k\}) = \sum_{t=0}^{n}\mathbb{P}_{\mu,p}(D_h([T_t,X_t])\geq k).$$

So it suffices to prove the following quantitative estimate: for all $h, n, k \geq 1$,

$$\textbf{Goal: } \sum_{t=0}^{n}\mathbb{P}_{\mu,p}(D_h(T_t,X_t)\geq k) \leq C\,(n+D+h)\left(\frac{(k+1)^{h+1}-1}{k}\right)e^{-\alpha\,k},$$

which is true for $h = 0$ by our assumption (7.1).

Consider $h > 0$ and assume that we have proven that, for each $j \leq h-1$, there exists $C_j > 0$ depending only on j and p, and an exponent α as in the base case, such that

$$\textbf{(induction hyp.) } \forall m \in \mathbb{N} \,:\, \sum_{t=0}^{m}\mathbb{P}_{\mu,p}(D_j(T_t,X_t)\geq k) \leq C_j\,(m+D+j)\,e^{-\alpha\,k},$$

for all $k \geq 1$ (again, this is true for $j = 0$ with $C_0 = C$). We *claim* that a similar statement holds for $j = h$ with $C_h = C_0 + kC_{h-1}$. A simple induction would then imply our goal:

$$C_h = C_0\sum_{i=0}^{h}(k+1)^i = \frac{(k+1)^{h+1}-1}{k}\,C.$$

To obtain our bound, note that

$$\sum_{t=0}^{n}\mathbb{P}_{\mu,p}(D_h(T_t,X_t)\geq k) \;\;\leq\;\; \sum_{t=0}^{n}\mathbb{P}_{\mu,p}(d_{T_t}(X_t)\geq k)$$
$$+\sum_{t=0}^{n}\mathbb{P}_{\mu,p}(D_h(T_t,X_t)\geq k, d_{T_t}(X_t)\leq k).$$

The first term in the RHS is $\leq C_0\,(n+D)\,e^{-\alpha\,k}$ by the base case. For the second term, we observe that $D_h(T_t,X_t)\geq k$ – i.e. some vertex at distance h from X_t has degree $\geq k$ – means $D_{h-1}(T_t,v)\geq k$ for one of the neighbors of X_t in T_t. Now, if $d_{T_t}(X_t)\leq k$, there is a chance of at least $1/(k+1)$ that $X_{t+1} = v$ and thus $D_{h-1}(T_{t+1},X_{t+1})\geq k$. We conclude:

$$\frac{\mathbb{P}_{\mu,p}(D_h(T_t,X_t)\geq k, d_{T_t}(X_t)\leq k)}{k+1} \leq \mathbb{P}_{\mu,p}(D_{h-1}(T_{t+1},X_{t+1})\geq k).$$

So:

$$
\begin{aligned}
\sum_{t=0}^{n} \mathbb{P}_{\mu,p}(D_h(T_t, X_t) \geq k) \quad &\leq \quad C_0\,(n+D)\,e^{-\alpha\,k} \\
&\quad +(k+1)\sum_{t=0}^{n}\mathbb{P}_{\mu,p}(D_{h-1}(T_{t+1}, X_{t+1}) \geq k) \\
\text{(induction hyp. for } j = r-1, m = n+1) \quad &\leq \quad C_0\,(n+D)\,e^{-\alpha\,k} \\
&\quad +(k+1)C_{h-1}\,(n+h+D)\,e^{-\alpha k} \\
&\leq \quad C_h\,(n+h+D)\,e^{-\alpha\,k},
\end{aligned}
$$

where $C_h := C + (k+1)\,C_{h-1}$ as desired. $\qquad\square$

In the next result – a strengthening of Theorem 1.2 –, we prove convergence of empirical measures $\widehat{P}_n$ and Cesàro-style convergence of the chain K_p from any initial measure that is p-escapable. We also characterize the limiting probability measure of the chain as a stationary distribution for K_p.

Theorem 7.3 (Convergence of the empirical measure; proof in §7.1)**.** *Given $p \in (0, 1]$, there exists a probability measure P_p on the space $\mathcal{T}_*$ of rooted trees such that, for any p-escapable initial measure μ,*

$$
\widehat{P}_n \Rightarrow P_p \;\; \mathbb{P}_{\mu,p}\text{-almost surely} \quad \text{and} \quad \frac{1}{n+1}\sum_{t=0}^{n} \mu K_p^t \Rightarrow P_p \;\; \text{when } n \to +\infty.
$$

For $0 < p_0 \leq p \leq 1$, the measure P_p satisfies:

$$
\forall k, h \geq 1 : P_p(\{[T, o] \in \mathcal{T} \,:\, D_h([T, o]) \geq k\}) \leq C\,\frac{(k+1)^{h+1} - 1}{k}\,e^{\alpha\,(2-k)}, \qquad (7.2)
$$

where $C, \alpha > 0$ only depend on p_0. Moreover, P_p is an invariant measure for K_p.

Note that the measure P_p itself is p-escapable, by Lemma 6.9 combined with (7.2). That is, P_p is the unique p-escapable invariant measure for K_p. On the other hand, the hard tree in Example 6.8 shows that one does not have convergence to P_p from all initial measures.

Our next result extends Theorem 1.3 from the Introduction.

Theorem 7.4 (Support of P_p; proof in §7.2)**.** *Given $0 < p \leq 1$, let P_p be the limit measure in Theorem 7.3. Then P_p is supported on one-ended infinite trees. Moreover, if $0 < p < 1$, given a finite rooted tree (S_0, x_0) of diameter $\leq h$,*

$$
P_p(\{[T, o] \in \mathcal{T}_* \,:\, [T, o]_h = [S_0, x_0]\}) > 0.
$$

In particular, this shows that, if μ is such that $\widehat{P}_n \Rightarrow P_p \;\mathbb{P}_{\mu,p}$-a.s., then $\mathbb{P}_{\mu,p}(\tau_\ell < +\infty) = 1$ i.e. μ is p-escapable. In this sense, p-escapability is a necessary condition in Theorem 7.3.

7.1 Weak convergence of empirical measures: proof of Theorem 7.3

We present here the proof of convergence to the stationary measure.

Proof of Theorem 7.3. Fix μ and p. Theorem 6.4 guarantees that

$$
\mathbb{P}_{\mu,p}(\widehat{P}_n\psi \to \mathcal{M}_p\psi) = 1,
$$

for all bounded local functions, where $\mathcal{M}_p\psi$ does not depend on μ. If we write:

$$P_n := \frac{1}{n+1} \sum_{t=0}^{n} \mu K_p^t,$$

we also have

$$P_n\psi = \mathbb{E}_{\mu,p}[\widehat{P}_n\psi] \to \mathcal{M}_p\psi, \tag{7.3}$$

by the Bounded Convergence Theorem.

We claim that the sequences $\{P_n\}_n$ and $\{\widehat{P}_n\}_n$ are *tight* (almost surely, in the second case). More specifically, we will apply the criterion in Proposition 7.1, part (1). In fact, fixing $k, h \in \mathbb{N}$, we may take

$$\phi_{k,h}(T,x) := \mathbb{1}_{\{D_h(T,x) \geq k\}} \ ([T,x] \in \mathcal{T}_*),$$

which is a $(h+1)$-local function, so that

$$P_n\phi_{k,h} \to \mathcal{M}_p\phi_{k,h}, \ \text{ and } \ \widehat{P}_n\phi_{k,h} \to \mathcal{M}_p\phi_{k,h} \quad \mathbb{P}_{\mu,p}\text{-a.s.}$$

Now, to evaluate the limit in the RHS, we replace the initial measure μ with a measure μ_0 supported on a finite tree with two vertices. Applying Lemma 7.2 in conjunction with Lemma 5.6, and letting $n \to +\infty$, we obtain:

$$\mathcal{M}_p\phi_{k,h} \leq C \left(\frac{(k+1)^{h+1} - 1}{k} \right) e^{\alpha(2-k)}.$$

Since the RHS of this inequality goes to 0 as $k \to +\infty$, we obtain the required tightness.

We now prove that the weak convergence criterion of Proposition 7.1 (part (2)) is also satisfied, which asks for the convergence of expectations of all bounded local functions. For P_n this is immediate from (7.3). For $\widehat{P}_n$ there is the slight issue that we have:

$$\forall \text{ bounded local } \psi \ : \ \mathbb{P}_{\mu,p}(\widehat{P}_n\psi \to \mathcal{M}_p(\psi)) = 1,$$

and we now need to "move the quantifier $\forall$ inside the probability". However, Proposition 7.1 shows that we only need to worry about a countable family of ψ, and there is no problem in moving the quantifier inside for a countable family.

The upshot is that, assuming this claim, we have that Proposition 7.1 assures that P_n and $\widehat{P}_n$ converge weakly (almost surely, in the second case) to the *same* probability measure P_p, which is uniquely characterized by the fact that $P_p\psi = \mathcal{M}_p\psi$. In particular, this measure satisfies the property that:

$$\forall h, k \geq 1 \ : \ P_p(\{[T,o] \ : \ D_h(T,o) \geq k) \leq C \left(\frac{(k+1)^{h+1} - 1}{k} \right) e^{\alpha(2-k)}.$$

Finally, we show that P_p is an invariant measure for K_p. To this aim, we need the following fact assuring that if ψ is bounded Lipschitz, so is $K_p(\psi)$.

Fact 7.5. *Given a bounded measurable function $\psi : \mathcal{T}_* \to \mathbb{R}$, let*

$$K_p(\psi)([T,o]) := \int_{\mathcal{T}_*} \psi([T',o']) \, K_p([T,o], d[T',o']).$$

If ψ is bounded and Lipschitz, so is $K_p(\psi)$.

Proof sketch. This follows from the fact that, if $[T,o]$ and $[S,v]$ satisfy $\rho([T,o],[S,v]) \leq 1/(1+r)$ for some $r > 0$, one can couple $[T',o'] \sim K([T,o],\cdot)$ and $[S',v'] \sim K([S,v],\cdot)$ so that $\rho([T',o'],[S',v']) \leq 1/r$. $\qquad\square$

So we may apply P_p to the function $K_p(\psi)$ and obtain:

$$P_p K_p(\psi) = \lim_{n \to +\infty} P_n K_p(\psi) = \lim_{n \to +\infty} \frac{1}{n+1} \sum_{t=0}^{n} \mu K_p^{t+1}(\psi)$$

$$= \lim_{n \to +\infty} \frac{1}{n+1} \sum_{t=1}^{n+1} \mu K_p^{t}(\psi) = P_p(\psi).$$

Since ψ is an arbitrary bounded Lipschitz function, this implies $P_p K_p = P_p$. So P_p is a stationary measure for the chain K_p. $\qquad\square$

7.2 On the support of the limiting distribution: proof of Theorem 7.4

Proof of Theorem 7.4. That P_p is supported on infinite trees is immediate. We prove next the final statement in the theorem, namely:

$$P_p(\{[T,o] \in \mathcal{T}_* \ : \ [T,o]_h = [S,x_0])\}) > 0.$$

To see this, recall that (S, x_0) has height h. Let x_1 be a leaf node of S at distance h from x_0. One may traverse the tree S by first walking h steps from x_1 to x_0 and then doing a depth-first search on S from x_0, which will end back at x_0.

This traversal gives us a recipe for "creating" S inside T_{t_0}, for a certain t_0. Namely, all we need is that:

1. The walk creates leaves and jumps to them for h consecutive steps (this corresponds to moving from x_1 to x_0;
2. The walk then follows the sequence of steps given by the DFS in S, adding leaves as needed.

The above sequence of steps has positive probability and, if followed, ensures $[T_{t_0}, X_{t_0}]_h = [S, x_0]$. By stationarity of the process

$$P_p(\{[T,o] \in \mathcal{T}_* \ : \ [T,o]_h = [S,x_0])\}) = \mathbb{P}_{P_p,p}([T_{t_0}, X_{t_0}]_h = [S, x_0]) > 0.$$

We now prove one-endedness. Denote by $O(s)$ the set of all rooted trees $[T, o]$ that satisfy the following two properties:

1. T is infinite (but locally finite, as all other trees in $\mathcal{T}_*$);
2. there exists a $h \geq s$ such that all paths of length $\geq h$ starting from o pass through a single vertex at distance s from o.

One may check that the set of one-ended trees is precisely $\cap_{s \in \mathbb{N}} O(s)$. Therefore, it is enough to show that $P_p(O(s)) = 1$ for all $s \in \mathbb{N}$.

So fix $s \in \mathbb{N}$ and take a trajectory $[T_t, X_t]$ of BGRW started from P_p. Since P_p is stationary,

$$\forall n \in \mathbb{N} \ : \ \mathbb{P}_{P_p,p}([T_n, X_n] \in O(s)) = P_p(O(s)).$$

We will show that, for all $\varepsilon > 0$, one can choose $n \in \mathbb{N}$ so that the LHS above is $\geq 1 - \varepsilon$, which suffices to finish the proof.

Recall that P_p is p-escapable, as we commented after the statement of Theorem 7.3. This means that $\tau_\ell < +\infty$ almost surely for all $\ell \in \mathbb{N}$. Take $\ell \geq s$. We will follow the reasoning in the proof Lemma 6.5 – more specifically the Claim at the end of the proof – and argue as follows. Let $X_{\tau_\ell} = v_0, v_1, \ldots, v_\ell$ be the (induced) path of length ℓ around X_{τ_ℓ} at time τ_ℓ. Define

$$\tau'_{\ell,s} := \inf\{t \geq 0 \ : \ X_{\tau_\ell + t} = v_{\ell-s}\}.$$

Notice that if $\tau_\ell \leq n < \tau_\ell + \tau'_{\ell,s}$, then $[T_n, X_n] \in O(s)$ almost surely. Indeed, $\tau_\ell \leq n < \tau_\ell + \tau'_{\ell,s}$ has the following consequences:

1. X_n and v_ℓ lie on "opposite sides" of $v_{\ell-s}$, so $\mathrm{dist}_{T_n}(X_n, v_\ell) \geq \mathrm{dist}_{T_n}(v_{\ell-s}, v_\ell) = s$.
2. The subtree consisting of $v_{\ell-s}$ and all nodes on the same side as X_n is finite, as it consists of v_i for $0 \leq i \leq \ell - s$ plus the new nodes added between times τ_ℓ and n.
3. The subtree consisting of $v_{\ell-s}$ and all nodes on the same side as v_ℓ is a.s. infinite, as it contains the tree T_0 at time 0.

Combining these properties, we see that all long enough paths in T_n that start from X_n must pass through v_ℓ, which lies at distance greater or equal to s from X_n. This implies $[T_n, X_n] \in O(s)$, since these same paths must pass through the unique vertex at distance s from X_n that lies on the path from X_n to v_ℓ. We have thus shown that if $\tau_\ell \leq n < \tau_\ell + \tau'_{\ell,s}$, then $[T_n, X_n] \in O(s)$, as claimed.

We deduce:
$$\mathbb{P}_{P_p,p}([T_n, X_n] \in O(s)) \geq \mathbb{P}_{P_p,p}(\tau_\ell \leq n < \tau_\ell + \tau'_{\ell,s}).$$

Since
$$\mathbb{P}_{P_p,p}(\tau_\ell \leq n < \tau_\ell + \tau'_{\ell,s}) \geq 1 - \mathbb{P}_{P_p,p}(\tau_\ell > n) - \mathbb{P}_{P_p,p}(\tau'_{\ell,s} < +\infty),$$

we may observe as in the proof of Lemma 6.5 that
$$\mathbb{P}_{P_p,p}(\tau'_{\ell,s} < +\infty) \leq C e^{-\beta(\ell-s)} \leq \frac{\varepsilon}{2},$$

if ℓ is large enough. Since P_p is p-escapable, we may ensure that $\mathbb{P}_{P_p,p}(\tau_\ell > n) \leq \frac{\epsilon}{2}$ by taking n large enough (in terms of ℓ). So for some choice of ℓ, n we obtain the desired inequality:
$$\mathbb{P}_{P_p,p}(\tau_\ell \leq n < \tau_\ell + \tau'_{\ell,s}) \geq 1 - \varepsilon. \qquad \square$$

8 Final comments

Our analysis leaves open many problems about the BGRW process and its variants. We briefly discuss some of these.

Problem 8.1. Is the speed $c(p)$ given by Theorem 1.1 a monotone function of p?

It is possible to show using our techniques that $p \mapsto c(p)$ is continuous. Monotonicity is much less clear, though our (limited) simulations suggest it should hold.

Problem 8.2. Consider a model where $p = p_n$ decreases with n, and find a threshold function for transience.

That is, how fast can p_n decay while maintaining transience? Clearly, our arguments break down in this regime.

Problem 8.3. Consider models with cycles.

A "cheap" way to create cycles would be to connect the current vertex to some uniformly random vertex at distance K, with K a random variable. If K has light tails, it should still be possible to use the local topology to study the structure of our model.

A A simple lemma on dependent indicators

We prove here Lemma 3.6.

Proof of Lemma 3.6. For $r = 1, \ldots, \lfloor m/k \rfloor$ define blocks events
$$F_r := \{\forall j = (r-1)k+1, (r-1)k+2 \ldots, rk : I_j = 1\}.$$

Note that k consecutive indices j appear in each F_r, so:

$$\mathbb{P}\left(\text{ at least } k \text{ consecutive 1's in the sequence } (I_j)_{j=1}^m\right) \geq \mathbb{P}\left(\bigcup_{r=1}^{\lfloor m/k \rfloor} F_r\right).$$

Moreover, the values of indices j involved in events $F_r, F_{r'}$ are disjoint for $r \neq r'$. With this in mind one may easily show (via our assumptions) that $\mathbb{P}(F_1) \geq \mu^k$ and $\mathbb{P}(F_r \mid F_1^c \cap \ldots F_{r-1}^c) \geq \mu^k$.

$$1 - \mathbb{P}\left(\bigcup_{r=1}^{\lfloor m/k \rfloor} F_r\right) = \mathbb{P}\left(\bigcap_{r=1}^{\lfloor m/k \rfloor} F_r^c\right) \leq (1 - \mu^k)^{\lfloor m/k \rfloor}. \qquad \square$$

B Some facts on weak convergence and the local topology

B.1 General facts

In this section, (M, d) is a Polish metric space and $\mathrm{Prob}(M, d)$ is the set of all probability measures over the Borel σ-field over M. We need criteria to ascertain the weak convergence of a sequence $\{P_n\}_{n \in \mathbb{N}} \subset \mathrm{Prob}(M, d)$.

B.1.1 Criterion for tightness

Proposition B.1. *Suppose we are given a doubly-indexed family*

$$F_k^{(r)} \ : \ r \in \mathbb{N}, \ k \in \mathbb{N},$$

of closed subsets of M with the following two properties.

1. For each $r \in \mathbb{N}$, the sets $F_k^{(r)}$ are increasing and satisfy

$$F_k^{(r)} \nearrow M \text{ as } k \to +\infty.$$

2. For all choices of $k_1, k_2, k_3, \cdots \in \mathbb{N}$, the set

$$K(\{k_r\}_{r=1}^{+\infty}) := \bigcap_{r=1}^{+\infty} F_{k_r}^{(r)},$$

is compact.

Assume $\{P_n\}_{n \in \mathbb{N}} \subset \mathrm{Prob}(M, d)$ is such that

$$\forall r \in \mathbb{N} \ : \ \lim_{k \to +\infty} \liminf_{n \in \mathbb{N}} P_n(F_k^{(r)}) = 1.$$

Then $\{P_n\}_{n \in \mathbb{N}}$ is tight.

Proof of Proposition B.1. Given $\epsilon > 0$, we will argue that there exists a choice of sequence $\{k_r\}_{r=1}^{+\infty}$ for which

Goal: $\forall n \in \mathbb{N} \ : \ P_n(K(\{k_r\}_{r=1}^{+\infty})) \geq 1 - \varepsilon.$

Notice that, no matter what sequence we choose, we get:

$$\forall n \in \mathbb{N} \ : \ 1 - P_n(K(\{k_r\}_{r=1}^{+\infty})) = P_n\left(\bigcup_{r=1}^{+\infty} (F_{k_r}^{(r)})^c\right) \leq \sum_{r=1}^{+\infty} P_n((F_{k_r}^{(r)})^c).$$

So it suffices to show that for each $r \in \mathbb{N}$ we can choose $k_r \in \mathbb{N}$ with $\sup_n P_n((K_{k_r}^{(r)})^c) \leq \epsilon/2^r$.

To do this, fix some r. By our assumption that $\lim_{k \to +\infty} \liminf_{n \in \mathbb{N}} P_n(F_k^{(r)}) = 1$, we can find $k_r^* \in \mathbb{N}$, $n_0 \in \mathbb{N}$ such that

$$\forall n \geq n_0 \ : \ P_n(F_{k_r^*}^{(r)}) \geq 1 - \epsilon/2^r.$$

On the other hand, for $n < n_0$ we still have:

$$P_n(F_k^{(r)}) \nearrow 1 \text{ as } k \to +\infty \text{ due to assumption 1.}$$

Since there are only finitely many $n < n_0$, we can find k_r^{**} such that

$$\forall n < n_0 \ : \ P_n(F_{k_r^{**}}^{(r)}) \geq 1 - \frac{\epsilon}{2^r}.$$

We may now choose $k_r := \max\{k_r^*, k_r^{**}\}$. Since the sets $F_k^{(r)}$ are nested, we see at once that

$$\forall n \in \mathbb{N} \ : \ P_n(F_{k_r}^{(r)}) \geq 1 - \frac{\epsilon}{2^r},$$

which is the property we needed. $\qquad\square$

B.1.2 Criterion for convergence

Proposition B.2. *Assume $\{P_n\}_{n \in \mathbb{N}}$ is a tight sequence of probability measures, with a sequence of compact sets $\{K_k\}_{k \in \mathbb{N}}$ attesting tightness in the sense that.*

$$\lim_k \{\liminf_n P_n(K_k)\} = 1.$$

Then there exists a countable family $\mathcal{A}$ of bounded Lipschitz functions from M to $\mathbb{R}$, which depends only on $\{K_k\}_{k \in \mathbb{N}}$, such that, if $\lim_n P_n f$ exists for all $f \in \mathcal{A}$, then $P_n \Rightarrow P$ for some probability measure P. In that case, P is uniquely characterized by the fact that $Pf = \lim_n P_n f$ for all $f \in \mathcal{A}$.

Proof of Proposition B.2. It is known that a tight sequence $\{P_n\}_{n \in \mathbb{N}}$ converges weakly if and only $\lim_n P_n f$ exists for each function in the set

$$\mathcal{A}_0 := \{f : M \to \mathbb{R} \ : \ \|f\|_\infty \leq 1 \text{ and } \|f\|_{\text{Lip}} \leq 1\}.$$

In that case, Pf is uniquely defined by the fact that $Pf = \lim_n P_n f$ for all $f \in \mathcal{A}_0$.

Now consider a positive rational $\epsilon > 0$. The fact that $\{P_n\}_{n \in \mathbb{N}}$ is tight implies that there exists a compact set $K_k \subset M$, with $k = k(\epsilon)$, so that with $P_n(K_\epsilon) \geq 1 - \epsilon$ for all large enough n. For each such set, the functions $\mathcal{A}_0$ restricted to $K_{k(\epsilon)}$ form a bounded equicontinuous family. Thus the Ascoli-Arzèla Theorem implies that there exists a countable subset $A_\epsilon \subset \mathcal{A}_0$ with

$$\forall \delta > 0 \, \forall f \in \mathcal{A}_0 \, \exists f_{\epsilon,\delta} \in A_\epsilon \ : \ \sup_{x \in k(\epsilon)} |f(x) - f_{\epsilon,\delta}(x)| \leq \delta. \tag{B.1}$$

We *claim*

$$\mathcal{A} := \bigcup_{\epsilon \in \mathbb{Q}_+} A_\epsilon,$$

is the set we are looking for. For assume that $\lim_n P_n f$ exists for each $f \in \mathcal{A}$. For each $f \in \mathcal{A}_0$, one can choose $f_{\epsilon,\delta} \in \mathcal{A}$ as in (B.1) and obtain:

$$|P_n(f - f_{\epsilon,\delta})| \leq |P_n \mathbb{1}_{K_{k(\epsilon)}}(f - f_{\epsilon,\delta})| + P_n(K_{k(\epsilon)}^c) \leq \delta + \epsilon \text{ for large enough } n.$$

With this, one can show that $\{P_n f\}_{n \in \mathbb{N}}$ is Cauchy. Indeed,

$$\limsup_{m,n \to +\infty} |(P_n - P_m) f| \leq 2(\epsilon + \delta) + \limsup_{m,n \to +\infty} |(P_n - P_m) f_{\epsilon,\delta}| = 2(\epsilon + \delta),$$

since we are assuming $\{P_n f_{\epsilon,\delta}\}_n$ is Cauchy. Letting $\epsilon, \delta \to 0$, we obtain that $\{P_n f\}_n$ is Cauchy, as desired. A similar reasoning shows that:

$$\lim_n P_n f = \lim_{\epsilon,\delta \to 0} \lim_n P_n f_{\epsilon,\delta},$$

exists for all $f \in \mathcal{A}_0$. So $P_n \Rightarrow P$. Passing to limits, we obtain from the above estimates that

$$|P(f - f_{\epsilon,\delta})| \le \epsilon + \delta.$$

In particular, P is uniquely defined by the values of $Pf = \lim_n P_n f$ for $f \in \mathcal{A}_0$, which are uniquely specified by the numbers $Pf = \lim_n P_n f$ for $f \in \mathcal{A}$. □

B.2 Application to local topology over rooted graphs

The general theory of the previous section will now be applied to the space $\mathcal{T}_*$ of rooted, locally finite trees.

Proof of Proposition 7.1. We apply the criteria for tightness and convergence in Propositions B.1 and B.2. For each $r, k \in \mathbb{N}$ the set:

$$F_k^{(r)} := \{[T, o] \in \mathcal{T}_* \,:\, D_r([T, o]) \le k\},$$

is closed and $\cup_k F_k^{(r)} = \mathcal{T}_*$ for each fixed r. Lemma 3.5 in [4] implies that all sets of the form $K(\{k_r\}_{r \in \mathbb{N}}) = \cap_{r=1}^{+\infty} F_{k_r}^{(r)}$ are compact. Moreover, Assumption 1 implies that

$$\forall r \in \mathbb{N} \,:\, \lim_k \liminf_n P_n(F_k^{(r)}) = 1.$$

So $\{P_n\}_{n \in \mathbb{N}}$ is tight by virtue of Proposition B.1.

We now use Proposition B.2 to show convergence. That proposition gives us a countable family of functions such that if $P_n f$ converges for all $f \in \mathcal{A}$, then P_n converges. In the present setting, we need a countable family of r-local functions. To do this, define for each $n \in \mathbb{N}$ the projection map $\Pi_m : \mathcal{T}_* \to \mathcal{T}_*$ that takes $[G, o]$ to the m-neighborhood $[G, o]_m$. This map is a contraction of $\mathcal{T}_*$ and is therefore continuous, and moreover

$$\forall [G, o] \in \mathcal{G}_* \,:\, d([G, o], \Pi_m([G, o])) = d([G, o], [G, o]_m) \le \frac{1}{m+1}.$$

In particular, this implies that for any bounded Lipschitz $f : \mathcal{T}_* \to \mathbb{R}$,

$$\sup\{|f([G, o]) - f \circ \Pi_m([G, o])| \,:\, [G, o] \in K\} \le \frac{\|f\|_{\mathrm{Lip}}}{m+1}.$$

It transpires that one can replace the family $\mathcal{A}$ with

$$\mathcal{S} := \{f \circ \Pi_m \,:\, f \in \mathcal{A}, m \in \mathbb{N}\},$$

which is still countable, and obtain the desired result. □

References

[1] G. Amir, I. Benjamini, O. Gurel-Gurevich, and G. Kozma. Random Walk in Changing Environment. *ArXiv e-prints*, April 2015. MR-4167212

[2] Bernardo Amorim, Daniel R. Figueiredo, Giulio Iacobelli, and Giovanni Neglia. Growing networks through random walks without restarts. In *Proc 7th Workshop on Complex Networks (CompleNet'16)*, pages 199–211, 2016.

[3] Luca Avena, Hakan Güldaş, Remco van der Hofstad, and Frank den Hollander. Mixing times of random walks on dynamic configuration models. *The Annals of Applied Probability*, 28(4):1977–2002, 2018. MR-3843821

[4] Charles Bordenave. Lecture notes on random graphs and probabilistic combinatorial optimization. https://www.math.univ-toulouse.fr/~bordenave/coursRG.pdf, 2016.

[5] Chris Cannings and Jonathan Jordan. Random walk attachment graphs. *Electronic Communications in Probability*, 18:1–5, 2013. MR-3109632

[6] Andrea Collevecchio. On the transience of processes defined on galton-watson trees. *The Annals of Probability*, 34(3):870–878, 2006. MR-2243872

[7] Andrea Collevecchio, Mark Holmes, and Daniel Kious. On the speed of once-reinforced biased random walk on trees. *Electronic Journal of Probability*, 23, 2018. MR-3858914

[8] Codina Cotar and Debleena Thacker. Edge-and vertex-reinforced random walks with super-linear reinforcement on infinite graphs. *The Annals of Probability*, 45(4):2655–2706, 2017. MR-3693972

[9] Amir Dembo, Ruojun Huang, Ben Morris, and Yuval Peres. Transience in growing subgraphs via evolving sets. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(3):1164–1180, 2017. MR-3689964

[10] Margherita Disertori, Christophe Sabot, and Pierre Tarrès. Transience of edge-reinforced random walk. *Communications in Mathematical Physics*, 339(1):121–148, 2015. MR-3366053

[11] Rick Durrett, Harry Kesten, and Vlada Limic. Once edge-reinforced random walk on a tree. *Probability Theory and Related Fields*, 122(4):567–592, 2002. MR-1902191

[12] Daniel Figueiredo, Philippe Nain, Bruno Ribeiro, Edmundo de Souza e Silva, and Don Towsley. Characterizing continuous time random walks on time varying graphs. In *ACM SIGMETRICS Performance Evaluation Review*, volume 40, pages 307–318. ACM, 2012.

[13] Giulio Iacobelli, Daniel R. Figueiredo, and Giovanni Neglia. Transient and slim versus recurrent and fat: random walks and the trees they grow. *Journal of Applied Probability*, 56(3):769–786, 2019. MR-4015636

[14] Giulio Iacobelli and Daniel Ratton Figueiredo. Edge-attractor random walks on dynamic networks. *Journal of Complex Networks*, 5(1):84–110, 2016. MR-3614918

[15] Daniel Kious and Vladas Sidoravicius. Phase transition for the once-reinforced random walk on $\mathbb{Z}^d$-like trees. *The Annals of Probability*, 46(4):2121–2133, 2018. MR-3813987

[16] Russell Lyons, Robin Pemantle, and Yuval Peres. Ergodic theory on galton-watson trees: speed of random walk and dimension of harmonic measure. *Ergodic Theory and Dynamical Systems*, 15(3):593–619, 1995. MR-1336708

[17] Yuval Peres, Alexandre Stauffer, and Jeffrey E. Steif. Random walks on dynamical percolation: mixing times, mean squared displacement and hitting times. *Probability Theory and Related Fields*, 162(3-4):487–530, 2015. MR-3383336

[18] Firas Rassoul-Agha. The point of view of the particle on the law of large numbers for random walks in a mixing random environment. *The Annals of Probability*, 31(3):1441–1463, 2003. MR-1989439

[19] Jari Saramäki and Kimmo Kaski. Scale-free networks generated by random walkers. *Physica A: Statistical Mechanics and its Applications*, 341:80–86, 2004. MR-2092677