COMPARING MIXING TIMES ON SPARSE RANDOM GRAPHS

ANNA BEN-HAMOU, EYAL LUBETZKY, AND YUVAL PERES

ABSTRACT. It is natural to expect that nonbacktracking random walk will mix faster than simple random walks, but so far this has only been proved in regular graphs. To analyze typical irregular graphs, let G be a random graph on n vertices with minimum degree 3 and a degree distribution that has exponential tails. We determine the precise worst-case mixing time for simple random walk on G, and show that, with high probability, it exhibits cutoff at time $\mathbf{h}^{-1} \log n$, where $\mathbf{h}$ is the asymptotic entropy for simple random walk on a Galton–Watson tree that approximates G locally. (Previously this was only known for typical starting points.) Furthermore, we show this asymptotic mixing time is strictly larger than the mixing time of nonbacktracking walk, via a delicate comparison of entropies on the Galton–Watson tree.

1. INTRODUCTION

We study the mixing time of simple random walk (SRW) vs. the nonbacktracking random walk (NBRW) on sparse random graphs. It is natural to expect, as highlighted in [3] for the case of regular expander graphs, that forbidding the walk to backtrack (traverse an edge twice in a row) would allow the walk to mix faster. It was thereafter shown in [13] that, on a typical random d-regular graph on n vertices, both walks exhibit the *cutoff phenomenon*—a sharp transition in the total-variation distance from equilibrium, dropping from near its maximum to near 0 over a negligible time period referred to as the cutoff window: SRW mixes at time $\frac{d}{d-2} \log_{d-1} n + O(\sqrt{\log n})$, as conjectured by Durrett [10], whereas the NBRW mixes faster, having cutoff already at time $\log_{d-1} n + O(1)$. Our goal here is to obtain analogous results for the irregular case.

The (worst-case) total-variation distance of a Markov chain with transition kernel P and state space Ω from its stationary distribution π is defined as

$$d_{\text{TV}}(t) = \max_{x_0 \in \Omega} \left\| P^t(x_0, \cdot) - \pi \right\|_{\text{TV}} = \max_{x_0 \in \Omega} \sup_{A \subset \Omega} \left| P^t(x_0, A) - \pi(A) \right| ,$$

and, for $0 < \varepsilon < 1$, its corresponding mixing time to within distance ε is given by

$$t_{\text{MIX}}(\varepsilon) = \min \left\{ t : d_{\text{TV}}(t) \le \varepsilon \right\} .$$

(When addressing a prescribed initial state $x_0 \in \Omega$ rather than the worst one, this definition is replaced by $t_{\text{MIX}}^{(x_0)}(\varepsilon) = \min\{t : d_{\text{TV}}^{(x_0)}(t) \le \varepsilon\}$ where $d_{\text{TV}}^{(x_0)} = \|P^t(x_0, \cdot) - \pi\|_{\text{TV}}$.) The notion of cutoff, due to Aldous, Diaconis and Shahshahani [1, 2, 9], captures the dependence of t_{MIX} on the parameter ε: a sequence of chains is said to exhibit cutoff if $\lim_{|\Omega| \to \infty}[t_{\text{MIX}}(\varepsilon)/t_{\text{MIX}}(\varepsilon')] = 1$ for every fixed $0 < \varepsilon, \varepsilon' < 1$.

Roughly put, the locally-treelike geometry of sparse random graphs makes the mixing behavior of random walk resemble that on trees. In the regular case, a walk on a regular tree, conditioned to be at a given distance from the root, is uniform on that level, and the analogous mixing time on a random d-regular graph coincides with the time t at which this distance is about $\log_{d-1} n$ (so as to contain almost all vertices in its range). This corresponds to $t = \log_{d-1} n + O(1)$ for the NBRW, and a slowdown of the SRW by a factor of $d/(d-2)$ due to the reduced speed of random walk on a tree (with a coarser $O(\sqrt{\log n})$-window due to the normal fluctuations of its height), as in the results of [13].

In the irregular case, however, the walk can have vastly different probabilities of traversing different paths, and mixing occurs when almost all vertices are—not only reachable by paths—but reachable by ones that are sufficiently probable. It might then be the case that the NBRW, allbeit faster to reach the leaves of a tree, could potentially be "trapped" in a set of lower-probability paths, compared with the SRW that retains a decent probability of backtracking and exploring more favorable parts of the tree. It turns out that the effect of such traps is not strong enough to compensate for the slowdown of the SRW, and, even in the irregular case, the NBRW mixes faster.

Let $G = (V, E)$ be a random graph on n vertices with vertex set V and degree distribution $(p_k)_{k \geq 1}$; that is, the degree D_x of each vertex $x \in V$ is independently sampled with $\mathbb{P}(D_x = k) = p_k$, conditioned on $\sum_x D_x$ being even, and G is thereafter generated by the configuration model. Let Z be a random variable with distribution

$$\mathbb{P}(Z = k - 1) \quad = \quad \frac{k p_k}{\sum_{\ell \geq 1} \ell p_\ell},$$

consider a rooted Galton–Watson tree (T, ρ) with offspring variable Z, and let (X_t) and (Y_t) be SRW and NBRW, respectively, on T started at ρ. The mixing times on G can be expressed in terms of the asymptotic entropy of these walks on T as follows. Let

$$\mathbf{h}_X \overset{\text{a.s.}}{=} \lim_{t \to \infty} \frac{1}{t} H\Big(\mathbb{P}_\rho(X_t \in \cdot \mid T)\Big) \quad \text{and} \quad \mathbf{h}_Y \overset{\text{a.s.}}{=} \lim_{t \to \infty} \frac{1}{t} H\Big(\mathbb{P}_\rho(Y_t \in \cdot \mid T)\Big), \quad (1.1)$$

where the entropy $H(\mu_T)$ of a probability measure μ_T on the vertices of T is given by

$$H\left(\mu_T\right) = -\sum_{x \in T} \mu_T(x) \log \mu_T(x).$$

It was shown in [7] that, when the initial vertex v_1 is *fixed* (independently of the graph) and the degree distribution $(p_k)_{k \geq 1}$ satisfies suitable moment assumptions, with high probability (w.h.p.) the SRW has cutoff at time $\mathbf{h}_X^{-1} \log n$ whereas the NBRW has cutoff at time $\mathbf{h}_Y^{-1} \log n$. Comparing these two mixing times was left open.

Comparing this with the regular case, observe that, for the NBRW, $\mathbf{h}_Y = \mathbb{E} \log Z$, which satisfies $\mathbf{h}_Y < \log \mathbb{E} Z$ whenever Z is not a constant by Jensen's inequality. Hence, the NBRW mixes well after the time at which its range covers most vertices (local neighborhoods are approximately Galton–Watson trees with offspring variable Z), unlike the regular setting. The same phenomenon occurs for SRW: denoting by ν the limiting

speed of SRW on T, then $\mathbf{h}_X/\nu < \log \mathbb{E}Z$ whenever Z is not a constant (the "dimension drop" of harmonic measure), as shown in [14].

From a *worst-case* initial vertex, it was showed independently in [4,7] that w.h.p. the NBRW also exhibits cutoff at $\mathbf{h}_Y^{-1} \log n$ under similar moment assumptions (in [4] the Gaussian tail of the distance profile within the cutoff window was further established). Here we extend the arguments of [7] to provide the analogous cutoff result for the SRW from a worst-case starting point, as well as compare these cutoff locations.

Theorem 1. *Let G be a random graph with degree distribution $(p_k)_{k=1}^\infty$, such that*

$$p_1 = p_2 = 0\,, \qquad \sum k^2 p_k < \infty\,, \tag{1.2}$$

and for some fixed $\delta > 0$, the random variable Z given by $\mathbb{P}(Z = k-1) \propto kp_k$ satisfies

$$\mathbb{P}(Z > \Delta_n) = o(1/n) \quad for \quad \Delta_n := \exp\left[(\log n)^{1/2-\delta}\right]. \tag{1.3}$$

Then w.h.p., SRW from a worst-case vertex has cutoff at $\mathbf{h}_X^{-1} \log n$ with window $\sqrt{\log n}$, with $\mathbf{h}_X$ as defined in (1.1) for a Galton–Watson tree T with offspring variable Z. Moreover, $\mathbf{h}_X < \mathbf{h}_Y$ and the NBRW mixes faster than SRW.

Condition (1.3) is weaker than requiring Z to have an exponential tail.

The two statements of Theorem 1 are proved separately. In Section 2 we prove Proposition 1.1, which extends the result of [7] to the worst-case starting point. In Section 3 we establish Proposition 1.2, which shows that $\mathbf{h}_X < \mathbf{h}_Y$ when the degrees are at least 3. Together, this implies that, as in the regular case, the backtracking moves of the SRW on sparse irregular random graphs delay its mixing compared to the NBRW.

Proposition 1.1. *Under the assumptions (1.2)–(1.3), w.h.p. the worst-case mixing time of the SRW on G exhibits cutoff at time $\mathbf{h}_X^{-1} \log n$ with window $O(\sqrt{\log n})$.*

Proposition 1.2. *Let Z be a random variable taking nonnegative integer values such that $\mathbb{E}Z < \infty$ and $Z \geq 2$ a.s. Then $\mathbf{h}_X < \mathbf{h}_Y$ on the GW-tree with offspring variable Z.*

2. SIMPLE RANDOM WALK FROM THE WORST STARTING POINT

In this section we prove Proposition 1.1, establishing cutoff and its location for SRW on a sparse random graph with degree distribution $(p_k)_{k=0}^\infty$. Recalling that the minimum degree is 3 by assumption ($p_1 = p_2 = 0$), let $\Delta = \Delta(n)$ denote the maximum degree in G, which we may assume is at most $\exp[(\log n)^{1/2-\delta}]$ for some $\delta > 0$ fixed. The lower bound on the cutoff window from a worst-case starting point follows immediately from [7, Theorem 2] (where it was established for a uniformly chosen initial vertex, and in particular carries to the worst one), and it remains to show a matching upper bound.

The first step in the proof is to reduce the analysis of worst-case starting points, as was done in [13], to vertices whose neighborhood up to distance $c \log \log n$, for an appropriate constant $c > 0$, is a tree. For $x \in V$ and $k > 0$, let $B_k(x) = \{y : \mathrm{dist}(x, y) \leq k\}$ be the k-radius neighborhood of x, and let $\partial B_k(x) = \{y : \mathrm{dist}(x, y) = k\}$. Further define the

tree excess of $B_k(x)$, denoted $\mathtt{tx}(B_k(x))$, to be the maximum number of edges that can be deleted from the induced subgraph on $B_k(x)$ while keeping it connected.

Definition 2.1. Let $K > 0$. A vertex $x \in V$ is said to be a K-root if $\mathtt{tx}(B_K(x)) = 0$.

The next two straightforward lemmas follow [13, Lemmas 2.1 and 3.2].

Lemma 2.2. *Let $K = O(\sqrt{\log n})$. W.h.p., every vertex $x \in V$ satisfies $\mathtt{tx}(B_{5K}(x)) \leq 1$.*

Proof. Fix $x \in V$, and condition on the degrees $\{D_u : u \in V\}$, noticing $\sum D_u \geq 3n$ since $p_1 = p_2 = 0$. We generate the ball of radius $5K$ around x sequentially, by the standard breadth-first search process of the configuration model (*cf.*, *e.g.*, [8]): identifying each vertex u with $\deg(u)$ "half-edges," denoted $(u, *)$, we start by inserting all half-edges of x to a first-in-first-out queue. Upon extracting a half-edge $(u, *)$ from the queue it is matched to a uniformly chosen unmatched half-edge $(v, *)$ in G, to form the edge (x, y), and all other half-edges of the vertex v are inserted to the queue (if not already there).

If M is the total number of half-edges encountered until first observing a vertex at distance $5K + 1$ from x (whence $B_{5K}(x)$ is fully exposed), then clearly we will perform at most M steps of extracting a half-edge from the queue and matching it, and in each such step, the number of possible matches in the queue (adding a cycle in G and increasing $\mathtt{tx}(B_{5K}(x))$) is at most M vs. at least $3n - M$ half-edges outside the queue. Since $M \leq \Delta^{5K+1}$, it follows that $\mathtt{tx}(B_{5K}(x))$ is stochastically dominated by a binomial random variable $\mathrm{Bin}(\Delta^{5K+1}, \Delta^{5K+1}/(3n - \Delta^{5K+1}))$, and noting that $\Delta^{5K+1} \leq \exp\left[O((\log n)^{1-\delta})\right] = n^{o(1)}$ by our assumptions on K and Δ, we find that

$$\mathbb{P}\Big(\mathtt{tx}(B_{5K}(x)) \geq 2\Big) \leq \binom{\Delta^{5K+1}}{2} \left(\frac{\Delta^{5K+1}}{(3 - o(1))n}\right)^2 = n^{-2+o(1)}.$$

A union bound over x concludes the proof. ∎

Lemma 2.3. *Let $K > 0$, and let x be a vertex in a graph H with minimum degree 3, such that $\mathtt{tx}(B_{5K}(x)) \leq 1$. Then SRW of length $4K$ started at x ends at a K-root with probability at least $1 - e^{-K/128}$.*

Proof. Let (X_t) be SRW on H started at x. If x is a $5K$-root, then the claim is trivial. Otherwise, the induced subgraph on $B_{5K}(x)$ has exactly one cycle $\mathcal{C}$, by assumption. We claim that in this situation,

$$\mathbb{P}\big(\mathrm{dist}(X_{4K}, \mathcal{C}) < K\big) \leq \exp(-K/128). \tag{2.1}$$

Let $\rho_t = \mathrm{dist}(X_t, \mathcal{C})$. By the minimum degree assumption, $\mathbb{P}(\rho_{t+1} = 0 \mid \rho_t = 0, X_t) \leq \frac{2}{3}$, whereas on the event $\rho_t > 0$ one has $|\rho_{t+1} - \rho_t| = 1$ and $\mathbb{P}(\rho_{t+1} < \rho_t \mid \rho_t > 0, X_t) \leq \frac{1}{3}$. Combining both situations, $\mathbb{E}[\rho_{t+1} - \rho_t \mid X_t] \geq \frac{1}{3}$, so $M_t = \rho_t - \frac{1}{3}t$ is a submartingale with $|M_{t+1} - M_t| \leq \frac{4}{3}$. By the Hoeffding–Azuma inequality,

$$\mathbb{P}\big(\rho_{4K} \leq \rho_0 + K\big) = \mathbb{P}\left(M_{4K} \leq M_0 - \frac{K}{3}\right) \leq \exp\left(-\frac{(K/3)^2}{2(\frac{4}{3})^2 \cdot 4K}\right) = e^{-K/128},$$

establishing (2.1) and concluding the proof. ∎

Observe that Lemma 2.2 gives, w.h.p., the hypothesis of Lemma 2.3 simultaneously for all $x \in V$ provided that $K = O(\sqrt{\log n})$, implying then that

$$d_{\mathrm{TV}}(t + 4K) \le \max \left\{ d_{\mathrm{TV}}^{(x_0)}(t) \; : \; x_0 \text{ is a } K\text{-root} \right\} + e^{-K/128} \tag{2.2}$$

holds w.h.p. By choosing a suitably large K, this will allow us to reduce the problem to initial vertices that are K-roots. To bound the mixing time from a K-root x_0, let

$$\tau_K = \tau_K(x_0) := \inf \left\{ t > 0 \; : \; X_t \in \partial B_K(x_0) \right\},$$

where X_t is the SRW on G started at x_0, and further define

$$\Lambda_t = \Lambda_t(x_0, K, \varepsilon) := \left\{ x \in \partial B_K(x_0) \; : \; d_{\mathrm{TV}}^{(x)}(t) > \varepsilon \right\}.$$

The next simple lemma bounds $d_{\mathrm{TV}}^{(x_0)}$ via the probability that $X_{\tau_K} \in \Lambda_t$ started from x_0.

Lemma 2.4. *For every $x_0 \in V$ and every $\varepsilon, K, s, t > 0$,*

$$d_{\mathrm{TV}}^{(x_0)}(t + s) \le \tfrac{1}{2} \mathbb{P}_{x_0}^G (\tau_K > s) + \mathbb{P}_{x_0}^G (X_{\tau_K} \in \Lambda_t) + \varepsilon.$$

Proof. Write $S_K = \partial B_K(x_0)$ for brevity. By the triangle inequality,

$$d_{\mathrm{TV}}^{(x_0)}(t + s) \le \tfrac{1}{2} \mathbb{P}_{x_0}^G (\tau_K > s) + \tfrac{1}{2} \left\| \mathbb{P}_{x_0}^G (X_{t+s} \in \cdot, \tau_K \le s) - \pi \right\|_1.$$

Using the strong Markov property, $\mathbb{P}_{x_0}^G (X_{t+s} \in \cdot, \tau_K \le s)$ is equal to

$$\sum_{\ell=1}^s \sum_{z \in S_K} \mathbb{P}_{x_0}^G (\tau_K = \ell, X_{\tau_K} = z) \, \mathbb{P}_{x_0}^G \left(X_{t+s} \in \cdot \mid \tau_K = \ell, X_{\tau_K} = z \right)$$

$$= \sum_{\ell=1}^s \sum_{z \in S_K} \mathbb{P}_{x_0}^G (\tau_K = \ell, X_{\tau_K} = z) \, \mathbb{P}_z^G (X_{t+s-\ell} \in \cdot).$$

Combining these two statements, we find that

$$d_{\mathrm{TV}}^{(x_0)}(t + s) \le \tfrac{1}{2} \mathbb{P}_{x_0}^G (\tau_K > s) + \sum_{\ell=1}^s \sum_{z \in S_K} \mathbb{P}_{x_0}^G (\tau_K = \ell, X_{\tau_K} = z) \, d_{\mathrm{TV}}^{(z)}(t + s - \ell)$$

$$\le \tfrac{1}{2} \mathbb{P}_{x_0}^G (\tau_K > s) + \sum_{z \in S_K} \mathbb{P}_{x_0}^G (X_{\tau_K} = z) \, d_{\mathrm{TV}}^{(z)}(t),$$

where we used that $t \mapsto d_{\mathrm{TV}}^{(z)}(t)$ is non-decreasing. The proof is concluded by breaking the summation over z to $S_K \cap \Lambda_t$ and $S_K \cap \Lambda_t^c$ and using the bound $d_{\mathrm{TV}}^{(z)}(t) \le \varepsilon$ in the latter case (by definition of Λ_t) and $d_{\mathrm{TV}}^{(z)}(t) \le 1$ in the former. ∎

For every K-root $x_0 \in V$, one has

$$\mathbb{P}_{x_0}^G (\tau_K > 4K) \le e^{-K/128}$$

by the exact same proof of (2.1) (the root vertex x_0 plays the role of $\mathcal{C}$ in that argument). Together with (2.2) and Lemma 2.4, this implies the following.
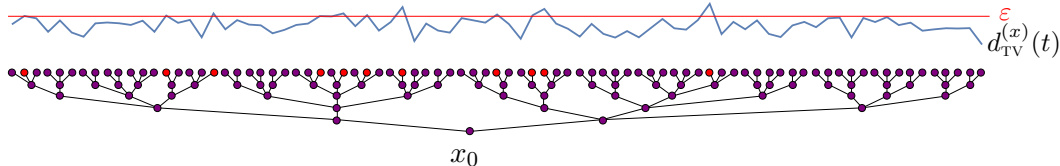
FIGURE 1. The set $\Lambda_t \subset \partial B_K(x_0)$ marked in red, above which are the corresponding values of $d_{\mathrm{TV}}^{(x)}(t)$ for $x \in \partial B_K(x_0)$, with a threshold at ε.

Corollary 2.5. *For all $\varepsilon, t > 0$ (that may depend on n) and $K = O(\sqrt{\log n})$, w.h.p.*

$$d_{\mathrm{TV}}(t + 8K) \leq \max\left\{\mu_{x_0}(\Lambda_t) \;:\; x_0 \text{ is a } K\text{-root}\right\} + \varepsilon + \tfrac{3}{2}e^{-K/128}, \qquad (2.3)$$

where $\mu_{x_0} = \mathbb{P}_{x_0}^G(X_{\tau_K} \in \cdot)$ is the hitting measure of the SRW from x_0 on $\partial B_K(x_0)$.

Combining this corollary with the following theorem will conclude the proof.

Theorem 2.6. *For every $\varepsilon > 0$ there exists some $\gamma_\star > 0$ such that, if*

$$t_1 = \left\lceil \mathbf{h}_X^{-1} \log n + \gamma_\star \sqrt{\log n} \right\rceil \qquad \text{and} \qquad K = \lceil \gamma_\star \log \log n \rceil, \qquad (2.4)$$

then for every tree T_0 such that $\mathbb{P}(B_K(x_0) = T_0) > 0$ and sufficiently large n,

$$\mathbb{P}\left(\mu_{x_0}(\Lambda_{t_1}) > \varepsilon \;\big|\; B_K(x_0) = T_0\right) < n^{-2}.$$

Indeed, modulo Theorem 2.6, for every fixed $\varepsilon > 0$, we may take t_1 and K as in (2.4), and deduce from the lemma (via a union bound) that, w.h.p., every K-root vertex x_0 satisfies $\mu_{x_0}(\Lambda_{t_1}) \leq \varepsilon$. Therefore, $d_{\mathrm{TV}}(t_1 + 8K) \leq 2\varepsilon + o(1)$ w.h.p. by (2.3), yielding the desired upper bound in Proposition 1.1.

Proof of Theorem 2.6. Throughout the proof of this theorem, let $\mathbf{P}^{T_0}$ denote the conditional probability $\mathbb{P}(\cdot \mid B_K(x_0) = T_0)$, and similarly let $\mathbf{E}^{T_0}$ denote the analogous conditional expectation.

For a given vertex $x \in V$, consider a breadth-first-search exploration of G from x, as described in the proof of Lemma 2.2, with the following addition: upon matching a half-edge $(u, *)$ from the queue to some (uniformly chosen) unmatched half-edge $(v, *)$, if the vertex v had already been encountered (*i.e.*, $(v, *)$ is in the queue), we stop the exploration at u and at v, and mark u and v as EXIT vertices. Also, when u is at distance L from x, we stop the exploration at this vertex and mark it as an EXIT vertex. Let $\Gamma_L(x) \subset B_L(x)$ denote the tree, rooted at x, that is obtained by running this process until no further exploration is allowed.

We first recall some key facts established in [7] in the analysis of SRW on sparse irregular random graphs. For a given graph $G = (V, E)$ on n vertices and vertex $x \in V$, an integer $t > 0$ and constants $\gamma, \varepsilon > 0$, define the event

$$\mathcal{A}_x^G(t, \gamma, \varepsilon) = \left\{ \sum_{y \in V} \left(\mathbb{P}_x^G(X_t = y) \wedge n^{-1} e^{\gamma\sqrt{\log n}} \right) \geq 1 - \varepsilon \right\}. \qquad (2.5)$$

Note that, if gap is the spectral-gap of SRW on a graph G, then for large enough n,

$$\mathcal{A}_x^G(t_0, \gamma, \varepsilon) \qquad \Rightarrow \qquad d_{\mathrm{TV}}^{(x)}(t_0 + s) < 2\varepsilon \quad \text{for} \quad s = 2\gamma \mathsf{gap}^{-1}\sqrt{\log n}. \qquad (2.6)$$

Indeed, on the event $\mathcal{A}_x^G$, the sub-measure $\nu(y) := P^{t_0}(x, y) \wedge n^{-1}\exp(\gamma\sqrt{\log n})$ satisfies $\nu(G) \geq 1 - \varepsilon$, and using that $1/(\Delta n) \leq \pi(y) \leq \Delta/n$ for all y (together with the assumption $\Delta = \exp(o(\sqrt{\log n}))$) we get $\|\nu/\pi - \nu(G)\|_{L^2(\pi)} \leq \exp((\gamma + o(1))\sqrt{\log n})$, so

$$d_{\mathrm{TV}}^{(x)}(t_0 + s) \leq \varepsilon + \|\nu P^s/\pi - 1\|_{L^2(\pi)} \leq \varepsilon + \exp(-(\gamma - o(1))\sqrt{\log n})) = \varepsilon + o(1).$$

It was shown in [7] that, if G is a random graph as specified in Theorem 1 and x is a uniformly chosen vertex in G, then for every $\varepsilon > 0$ there exist $\gamma_1, \gamma_2 > 0$ such that

$$\mathbb{P}\Big(\mathcal{A}_x^G(t_0, \gamma_1, \varepsilon)\Big) \geq 1 - \varepsilon \qquad \text{for} \qquad t_0 := \left\lceil \mathbf{h}_X^{-1}\left(\log n - \gamma_2\sqrt{\log n}\right)\right\rceil. \qquad (2.7)$$

(See [7, Eq. (3.19)], recalling $\mathbf{h}_X = \nu\mathbf{d}$ and the definitions in [7, Eqs. (3.1)–(3.3)]: our ε here replaces $\sqrt{5}\varepsilon$ from that equation, and γ_1 and γ_2 replace $\frac{4}{3}\gamma$ and $\frac{7}{8}\gamma$, respectively.) Moreover, by introducing an object referred to as *truncated* GW-trees [7, §3.1], it was shown there that for all $\varepsilon > 0$, there exist $c_0, \gamma_3 > 0$, depending only on ε and on the law of Z, such that, if

$$R := \lceil c \log\log n \rceil, \qquad L = \left\lceil \nu\big(t_0 + \mathbf{h}_X^{-1}\gamma_3\sqrt{\log n}\big)\right\rceil, \qquad (2.8)$$

for some fixed $c \geq c_0$, and $x \in V$ is fixed independently of G, then for large n there exists a subtree $\Gamma_L'(x) \subset \Gamma_L(x)$ (the *truncation* of $\Gamma_L(x)$, in which the exploration is also stopped at vertices which the SRW is less likely to visit, with those vertices being marked as EXIT vertices as well) with the following properties:

  (i) The first R levels of $\Gamma_L'(x)$ and $\Gamma_L(x)$ are equal, and can be coupled w.h.p. to a standard GW-tree[1] with variable Z (so every vertex there has at least 2 offspring). (See the definition of truncation in [7, §3.1] and the coupling in [7, §3.2].)
 (ii) The total size of the tree Γ_L' is at most $n\exp(-\frac{1}{4}\gamma_1\sqrt{\log n})$. (See [7, Eq. (3.11)], where ℓ_1 equals L from (2.8) for $\gamma_3 = \frac{1}{8}\gamma$ and γ_1, γ_2 as before.)
(iii) The SRW $(X_t)_{t=1}^{t_0}$ on G from x does not hit any EXIT vertex before time t_0 with probability $1 - O(\varepsilon)$ and can be coupled to the SRW on a *truncated* GW-tree with probability $1 - O(\varepsilon)$. Denote the event of such a successful coupling by Π_{t_0}. Furthermore, $\Gamma_L'(x)$ can be constructed while revealing at most $n\exp(-\frac{1}{4}\gamma_1\sqrt{\log n})$ vertices of $B_L(x)$. (See the description of this coupling in [7, §3.2] and the event Π_k defined there.)
(iv) There exists an event $\mathcal{E}_x^G \in \sigma(\Gamma_L'(x))$ such that

$$\mathcal{E}_x^G \subset \mathcal{A}_x^G(t_0, \gamma_1, \varepsilon) \qquad \text{and} \qquad \mathbb{P}(\mathcal{E}_x^G) \geq 1 - \varepsilon. \qquad (2.9)$$

(See [7, §3.4.1] and the event $\{\mathbb{P}_G(\Upsilon) \geq 1 - \sqrt{5}\varepsilon\}$ there, which is our event $\mathcal{E}_x^G$.)

---

[1]To be precise, the offspring distribution of the root is different — being given by (p_k) vs. all other vertices where it is given by (q_k) for $q_k \propto (k+1)p_{k+1}$ — but this does not essentially change the proofs.
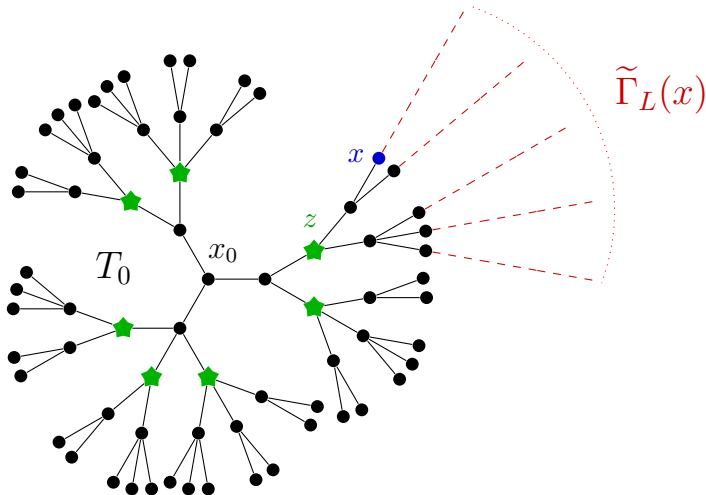
FIGURE 2. The tree $T_0 = \partial B_K(x_0)$ with $K = 4$.

We claim that this analysis extends to the case where x, instead of being a fixed vertex (chosen independently of the random graph G), is taken to be a leaf of $B_K(x_0)$.

Consider the following exploration: initially, $B_K(x_0)$ is fully exposed, and we assume $B_K(x_0) = T_0$ for some tree T_0. Now, for $x \in \partial T_0$, we may expose the truncated tree $\Gamma'_L(x)$ just as in [7], the only difference being that some part of this tree, namely T_0, is assumed to be exposed already at the beginning (no truncation will occur there). We will also consider a variant of this exposure process: given $B_K(x_0) = T_0$, for $x \in \partial T_0$ with ancestor $z \in \partial B_{\lfloor K/2 \rfloor}(x_0)$, we sequentially expose the neighborhood of x, stopping the exploration at u when we are about to create a cycle (*i.e.*, a half-edge $(u, *)$ is matched to a previously encountered vertex) or when u meets the truncation criterion [7, §3.1], just as for $\Gamma'_L(x)$, but in addition, we also stop the exploration at every $u \in \partial T_0$ which is not a descendant of z (and, just as before, mark every such vertex as an EXIT vertex). Let $\widetilde{\Gamma}_L(x)$ be the subtree obtained this way. An illustration of this exposure process is depicted in Figure 2. In what follows, when referring to the σ-fields generated by $\Gamma'_L(x)$ and $\widetilde{\Gamma}_L(x)$ we include the information of which vertices are marked EXIT (in our setting of $p_1 = 0$ per (1.2), these are but the leaves in these trees).

Lemma 2.7. *Let $\varepsilon > 0$, let t_0 and γ_1 as given in* (2.7), *and let T_0 be a tree such that $\mathbb{P}(B_K(x_0) = T_0) > 0$. For every $x \in \partial T_0$ there exists an event $\mathcal{E}_x^G \in \sigma\left(\Gamma'_L(x)\right)$ such that*

$$\mathcal{E}_x^G \subset \mathcal{A}_x^G(t_0, \gamma_1, \varepsilon) \qquad and \qquad \mathbf{P}^{T_0}\left(\mathcal{E}_x^G\right) \geq 1 - \varepsilon. \qquad (2.10)$$

Moreover, there exists an event $\widetilde{\mathcal{E}}_x^G \in \sigma\left(\widetilde{\Gamma}_L(x)\right)$ such that

$$\widetilde{\mathcal{E}}_x^G \subset \mathcal{A}_x^G(t_0, \gamma_1, 2\varepsilon) \qquad and \qquad \mathbf{P}^{T_0}\left(\widetilde{\mathcal{E}}_x^G\right) \geq 1 - 2\varepsilon. \qquad (2.11)$$

Proof. To prove (2.10), we describe in more detail the arguments used in [7] to derive the aforementioned facts (i)–(iv), and how these are affected by taking $x \in \partial B_K(x_0)$.

(i) The truncated GW-tree from [7] is a subtree of the standard GW-tree with variable Z, where the exploration of each level in the tree may skip certain vertices (where we say a truncation occurs), depending on a certain criterion (see [7, Eq. (3.8)]), which is measurable w.r.t. to the currently exposed subtree. No truncation occurs in the first R levels, immediately implying Property (i).

Using $c = c_0 \vee 2\gamma_\star$ for the definition of R in (2.8), with $\gamma_\star$ from the definition of K in (2.4), indeed $B_K(x_0) \subset B_R(x)$ in G, which is not truncated by definition.

(ii) Property (ii) was a consequence of the truncation criterion, which censors the parts of the tree where the random walk is less likely to visit; as such, the more likely subset of each level of the tree cannot be too large by conservation of mass.

Having merely adjusted R to be at least $2K$, this bound remains valid unchanged.

(iii) The coupling—now to SRW on a tree rooted at x, containing $B_K(x_0)$ with truncated GW-trees at x and all its leaves—is successful unless one of the following occurs:
   (a) *cycles*: SRW on G encounters a vertex at which exploration has been stopped because of a cycle (thereby hitting a vertex already marked EXIT in $\Gamma_L(x)$): this is controlled by the size of $\Gamma'_L(x)$, and is thus unchanged (for the new R);
   (b) *truncation* — SRW on G visits a truncated vertex (thereby hitting an EXIT vertex of $\Gamma'_L(x)$): this is controlled by the truncation criterion, thus unchanged (for the new R);
   (c) *degrees*: SRW on G hits a vertex whose degree is inconsistent with its analog in the GW-tree: again, this is controlled by the size of $\Gamma'_L(x)$, thus unchanged.

Note that encountering the truncation event (a) along the random walk up to time t_0 had probability $O(\varepsilon)$, whereas each of the other two events had probability $o(1)$ (in every given time step the events (b) and (c) had probability $\exp(-c\sqrt{\log n})$, outweighing a union bound over $t_0 = O(\log n)$ steps).

(iv) The event $\mathcal{E}_x^G$ was given by $\{\mathbb{P}_x^G(\Upsilon) \geq 1 - \varepsilon\}$ where Υ was the intersection of Π_{t_0} with two events, each occurring w.h.p.:
   • Υ_1 said that the distance of SRW on the standard GW-tree, at the target time t_0, is within a given number of standard deviations from its mean;
   • Υ_3 said that the (loop-erasure of) SRW on $\Gamma'_L(x)$ at time t_0 will belong to the subtree of a given set of vertices (denoted S''_{ℓ_0} in that paper), where the probability that SRW visits any given vertex in it is at most $n^{-1} \exp(\gamma_1 \sqrt{\log n})$.

The event Π_{t_0} was addressed in the previous item. As for the two events Υ_1, Υ_3, each occurs w.h.p. also in the situation where the truncated GW-tree rooted at x is further attached to a tree given by $B_K(x_0)$ (with additional truncated GW-trees rooted at all other vertices of $\partial B_K(x_0)$), as a consequence of the fact that each vertex of $B_K(x_0)$ has at least 2 offspring. This implies that SRW will encounter a regeneration point below level K by a time period that is $O(K)$ in expectation. In particular, for each of the above events, the estimates of SRW on the (standard) truncated GW-tree apply, once we ignore an additive time shift of, say, K^2 steps.

Altogether, we have verified (2.10). Now to prove (2.11), we have to verify an additional property allowing us to consider an event that is measurable with respect to $\widetilde{\Gamma}_L(x)$. To this end, let us define

$$\tilde{\tau} = \inf \left\{ t > 0 : \ X_t \in B_{\lfloor K/2 \rfloor}(x_0) \right\} ,$$

and consider the event

$$\widetilde{\mathcal{E}}_x^G = \left\{ \mathbb{P}_x^G \left(\Upsilon \cap \{\tilde{\tau} > t_0\} \right) > 1 - 2\varepsilon \right\} ,$$

where the event Υ is from item (iv). On the event $\Upsilon \cap \{\tilde{\tau} > t_0\}$, the walk does not hit any EXIT vertex of $\Gamma'_L(x)$ nor does it visit $B_{\lfloor K/2 \rfloor}(x_0)$ before time t_0. In particular, it does not hit any EXIT vertex $\widetilde{\Gamma}_L(x)$ before time t_0 and $\widetilde{\mathcal{E}}_x^G \in \sigma\left(\widetilde{\Gamma}_L(x)\right)$. Moreover,

$$\widetilde{\mathcal{E}}_x^G \subset \left\{ \mathbb{P}_x^G\left(\Upsilon\right) > 1 - 2\varepsilon \right\} \subset \mathcal{A}_x^G(t_0, \gamma_1, 2\varepsilon) .$$

Let us now show that

$$\mathbf{P}^{T_0}(\widetilde{\mathcal{E}}_x^G) \geq 1 - 2\varepsilon . \tag{2.12}$$

Using two successive union bounds (one on $\mathbb{P}_x^G$, the other on $\mathbf{P}^{T_0}$), we have

$$\mathbf{P}^{T_0}\left(\widetilde{\mathcal{E}}_x^G\right) \geq \mathbf{P}^{T_0}\left(\mathbb{P}_x^G(\Upsilon) \geq 1 - \varepsilon\right) - \mathbf{P}^{T_0}\left(\mathbb{P}_x^G(\tilde{\tau} \leq t_0) > \varepsilon\right) .$$

On the one hand, by (2.10), we know that $\mathbf{P}^{T_0}\left(\mathbb{P}_x^G(\Upsilon) \geq 1 - \varepsilon\right) = \mathbf{P}^{T_0}\left(\mathcal{E}_x^G\right) \geq 1 - \varepsilon$. On the other hand, by Markov's Inequality,

$$\mathbf{P}^{T_0}\left(\mathbb{P}_x^G(\widetilde{\tau} \leq t_0) > \varepsilon)\right) \leq \frac{\mathbf{P}^{T_0}(\widetilde{\tau} \leq t_0)}{\varepsilon} .$$

Since each vertex in the tree T_0 has at least 2 offspring, $\rho_t := \operatorname{dist}(X_t, x_0)$ dominates a one-dimensional biased random walk with increment ξ given by $\mathbb{P}(\xi = 1) = \frac{2}{3}$ and $\mathbb{P}(\xi = -1) = \frac{1}{3}$. It then follows (*e.g.*, via the exponential martingale $2^{-\rho_t}$) that

$$\sup_{y \in \partial B_{\lceil 3K/4 \rceil}(x_0)} \mathbf{P}_y^{T_0}\left(\tau_{\lfloor K/2 \rfloor} < \tau_K\right) \leq 2^{-K/4}$$

for

$$\tau_\ell = \min\{t : \rho_t = \ell\} ,$$

with which we can afford a union bound over at most $C \log n$ time-points serving as the potentially first visit to $\partial B_{\lceil 3K/4 \rceil}(x_0)$ before encountering the event $\tau_{\lfloor K/2 \rfloor} < \tau_K$, yielding $\mathbf{P}^{T_0}(\tau_{\lfloor K/2 \rfloor} \leq t_0) = o(1)$, and thus, for n large enough,

$$\mathbf{P}^{T_0}\left(\mathbb{P}_x^G(\widetilde{\tau} \leq t_0) > \varepsilon)\right) \leq \varepsilon ,$$

establishing (2.12) and concluding the proof of Lemma 2.7. ∎

Our next goal is showing that $\mu_{x_0}(\Lambda_{t_1})$ is concentrated around its mean with an exponential tail. (Recall that $\Lambda_{t_1} \subset \partial B_K(x_0)$ and $|\partial B_K(x_0)| \geq \log^c n$ for some $c > 0$, whereas our target error probability in Theorem 2.6 is $O(n^{-2})$.)
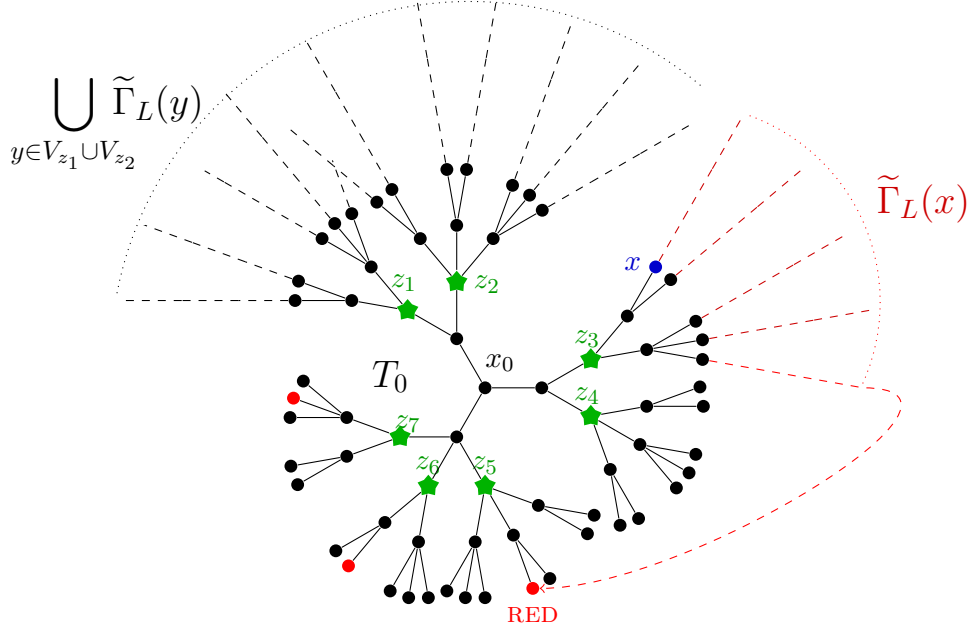
FIGURE 3. The ith stage of the sequential exposure process ($i = 3$) with the RED coloring of vertices.

To this end, let $\mathcal{S} = \partial B_{\lfloor K/2 \rfloor}(x_0)$ and $z_1, \ldots, z_{|\mathcal{S}|}$ an ordering of the elements of $\mathcal{S}$ and let

$$V_{z_i} = \{x \in \partial T_0, \ x \text{ is a descendant of } z_i\}.$$

Sequentially for i going from 1 to $|\mathcal{S}|$ and for $x \in V_{z_i}$, we expose the trees $\widetilde{\Gamma}_L(x)$ (in particular, if upon matching half edge $(u, *)$ during the exposure of $\widetilde{\Gamma}_L(x)$, we happen to reach a vertex in the boundary of the previously exposed trees, then we stop the exploration at u). While doing so, if at some stage i, we attempt to match half-edge $(u, *)$ with a half-edge $(v, *)$ where $v \in V_{z_j}$ for some $j > i$, then we mark v as RED .

Lemma 2.8. *Let $\mathbf{R}_L$ be the number of* RED *vertices in ∂T_0 at the end of the exploration process. Then, for γ_1 as given in* (2.7) *and every sufficiently large n,*

$$\mathbf{P}^{T_0}\left(\mathbf{R}_L > 15\gamma_1^{-1}\sqrt{\log n}\right) < n^{-2}.$$

Proof. By our assumption on the maximum degree, at any stage of the exploration process, the number of unmatched half-edges attached to a vertex of ∂T_0 is smaller than $\Delta^{K+1} \leq \exp\left((\log n)^{\frac{1-\delta}{2}}\right)$ for large enough n. Now recall that the truncation ensures that each of the trees $\widetilde{\Gamma}_L(x)$ has size at most $n \exp\left(-\frac{1}{4}\gamma_1\sqrt{\log n}\right)$, implying that the total number of pairs formed during the exploration process is at most $n \exp\left(-\frac{1}{5}\gamma_1\sqrt{\log n}\right)$.

All in all, the probability that at least $C\sqrt{\log n}$ times an unmatched half-edge attached to a vertex of ∂T_0 is chosen as target during the process is smaller than

$$\binom{n\exp\left(-\frac{1}{5}\gamma_1\sqrt{\log n}\right)}{C\sqrt{\log n}}\left(\frac{\exp\left[(\log n)^{\frac{1-\delta}{2}}\right]}{(3-o(1))n}\right)^{C\sqrt{\log n}}.$$

Using that $\binom{a}{b} \leq a^b$ for all $a, b \geq 0$, we see that

$$\mathbf{P}^{T_0}\left(\mathbf{R}_L > C\sqrt{\log n}\right) \leq \exp\left(-\frac{C\gamma_1}{5}\log n + C(\log n)^{1-\frac{\delta}{2}}\right),$$

which is smaller than n^{-2} for $C \geq 15/\gamma_1$ and large enough n. ∎

Let us denote by $\mathcal{F}_0$ the σ-field generated by T_0, and for $i = 1, \ldots, |\mathcal{S}|$, let $\mathcal{F}_i$ be the σ-field generated by T_0 and $\bigcup_{j=1}^{i}\bigcup_{x\in V_{z_j}}\widetilde{\Gamma}_L(x)$, together with the RED coloring of vertices of ∂T_0 up to stage i. We say that a vertex $z \in \mathcal{S}$ is NICE if none of its descendants in ∂T_0 is colored RED. Note that the event $\{z_i$ is NICE$\}$ belongs to $\mathcal{F}_{i-1}$.

Lemma 2.9. *For $i \geq 1$, consider running the exploration process up to stage $i-1$, and let $x \in \partial T_0$ be a descendant of z_i. Then*

$$\mathbb{P}\left(\widetilde{\mathcal{E}}_x^G \mid \mathcal{F}_{i-1}\right) \geq \mathbb{1}_{\{z_i \text{ is NICE}\}}(1-2\varepsilon).$$

Proof. The proof of Lemma 2.9 closely parallels the proof of Lemma 2.7. Indeed, all that remains to show is that having exposed the trees $\widetilde{\Gamma}_L(y)$ for $y \in V_{z_1} \cup \cdots \cup V_{z_{i-1}}$ does not essentially change the probability of $\widetilde{\mathcal{E}}_x$ for $x \in V_{z_i}$, at least when z_i is NICE. First note that $\widetilde{\mathcal{E}}_x \in \sigma\left(\widetilde{\Gamma}_L(x)\right)$ and that the tree $\widetilde{\Gamma}_L(x)$ does not intersect any the trees $\widetilde{\Gamma}_L(y)$ for $y \in V_{z_1} \cup \cdots \cup V_{z_{i-1}}$ (except for at T_0). Now, the only difference with the setting of Lemma 2.7 is that the number of exposed half-edges may now be larger. However, by the truncation criterion, we know that for all $y \in V_{z_1} \cup \cdots \cup V_{z_{i-1}}$, the size of $\widetilde{\Gamma}_L(y)$ is smaller than $n\exp\left(-\frac{1}{4}\gamma_1\sqrt{\log n}\right)$, so that the total number of exposed half-edges at the end of stage $i-1$ is at most $\Delta^K n\exp\left(-\frac{1}{4}\gamma_1\sqrt{\log n}\right) \leq n\exp\left(-\frac{1}{5}\gamma_1\sqrt{\log n}\right)$. Recalling that the the three possible impediments to a successful coupling with a truncated GW-tree (*truncation, cycles* and *degrees*) were controlled either by the truncation criterion itself (thus unchanged), or by the upper bound on the size of the exposed subgraph (which has merely increased from $n\exp\left(-\frac{1}{4}\gamma_1\sqrt{\log n}\right)$ to $n\exp\left(-\frac{1}{5}\gamma_1\sqrt{\log n}\right)$) and noticing that the event of a visit to $\partial B_{\lfloor K/2 \rfloor}(x_0)$ before time t_0 is not affected by this extra-exposure, this is enough to ensures that, conditionally on $\mathcal{F}_{i-1}$, for all $x \in V_{z_i}$ with z_i NICE, the coupling of SRW on G started at x with SRW on a tree rooted at x (containing $B_K(x_0)$ with truncated GW-trees on those leaves which are descendants of z_i) is successful with large probability. Also, the events Υ_1 and Υ_3 can be handled exactly as in the proof of Lemma 2.7, a regeneration point below level K being quickly found by the walk. ∎

Lemma 2.10. *Let $\varepsilon > 0$, take t_0 and γ_1 as in (2.7), and let $\mu_{x_0} = \mathbb{P}_{x_0}^G(X_{\tau_K} \in \cdot)$ and*

$$U = \left\{ x : \ \widetilde{\mathcal{E}}_x^G \ \text{does not hold} \right\}. \tag{2.13}$$

If T_0 is a tree with $\mathbb{P}(B_K(x_0) = T_0) > 0$ then for every sufficiently large n,

$$\mathbf{P}^{T_0}\left(\mu_{x_0}(U) \geq 4\varepsilon\right) < 2n^{-2}.$$

Proof. Note that

$$\mu_{x_0}(U) \leq \sum_{i=1}^{|\mathcal{S}|} \mu_{x_0}(V_{z_i}) \mathbb{1}_{\{z_i \text{ is not NICE}\}} + \sum_{i=1}^{|\mathcal{S}|} \mu_{x_0}(V_{z_i} \cap U) \mathbb{1}_{\{z_i \text{ is NICE}\}}. \tag{2.14}$$

As for the first term in the right-hand side of (2.14), we have

$$\sum_{i=1}^{|\mathcal{S}|} \mu_{x_0}(V_{z_i}) \mathbb{1}_{\{z_i \text{ is not NICE}\}} \leq \max_{z \in \mathcal{S}} \mu_{x_0}(V_z) \mathbf{R}_L.$$

Observe that

$$\max_{z \in \mathcal{S}} \mu_{x_0}(V_z) \leq \mathbf{P}_{x_0}^{T_0}(\tau_z < \tau_{\partial T_0}) \leq 3 \cdot 2^{-\lfloor K/2 \rfloor}, \tag{2.15}$$

as the probability of ever visiting a vertex at level k in the infinite binary tree is at most $3 \cdot 2^{-k}$ (at each point along the path from the root to this vertex, the random walk has a probability of $1/3$ of escaping to infinity through an alternative branch), and adding edges can only decrease the probability of visiting a vertex. And by Lemma 2.8, the number of RED vertices is smaller than $\frac{15}{\gamma_1}\sqrt{\log n}$ with probability at least $1 - n^{-2}$. Choosing $\gamma_\star$ large enough in the definition of K then ensures that

$$\max_{z \in \mathcal{S}} \mu_{x_0}(V_z) \mathbf{R}_L \leq \varepsilon,$$

with probability at least $1 - n^{-2}$.

Moving on to the second term in the right-hand side of (2.14), let

$$W = \sum_{i=1}^{|\mathcal{S}|} W_i \quad \text{for} \quad W_i = \mu_{x_0}(V_{z_i} \cap U) \mathbb{1}_{\{z_i \text{ is NICE}\}},$$

and

$$M_t := \sum_{i \leq t} (W_i - \mathbb{E}[W_i \mid \mathcal{F}_{i-1}]) \qquad (t = 1, \ldots, |\mathcal{S}|).$$

Note that, for all $1 \leq i \leq |\mathcal{S}|$, the variable $\sum_{j<i} W_j$ is $\mathcal{F}_{i-1}$-measurable and

$$\mathbb{E}\left[W_i \mid \mathcal{F}_{i-1}\right] = \mathbb{1}_{\{z_i \text{ is NICE}\}} \sum_{x \in V_{z_i}} \mu_{x_0}(x) \mathbb{P}\left(\left(\widetilde{\mathcal{E}}_x^G\right)^c \mid \mathcal{F}_{i-1}\right) \leq 2\varepsilon \mu_{x_0}(V_{z_i}),$$

with the last inequality by Lemma 2.9. In particular, $W \leq M_{|\mathcal{S}|} + 2\varepsilon$. Also, by (2.15),

$$\sum_{z \in \mathcal{S}} \mu_{x_0}(V_z)^2 \leq \max_{z \in \mathcal{S}} \mu_{x_0}(V_z) \leq 3 \cdot 2^{-\lfloor K/2 \rfloor},$$

and so $\sum_t \left| M_t - M_{t-1} \right|^2 \leq 3 \cdot 2^{-\lfloor K/2 \rfloor}$ with probability 1. Thus, we can infer from the Hoeffding–Azuma inequality for the martingale (M_t) that

$$\mathbf{P}^{T_0} \left(W \geq 3\varepsilon \right) \leq \mathbf{P}^{T_0} \left(M_{|\mathcal{S}|} \geq \varepsilon \right) \leq \exp \left(-\frac{\varepsilon^2}{6 \cdot 2^{-\lfloor K/2 \rfloor}} \right) < n^{-2} \,,$$

provided that $\gamma_\star$ in the definition (2.4) is chosen to be sufficiently large. ∎

Together with (2.6) this completes the proofs of Theorem 2.6 and Proposition 1.1. ∎

3. Entropy comparison of walks on Galton–Watson trees

In this section we prove Proposition 1.2, showing that the ratio $\mathbf{h}_X/\mathbf{h}_Y$ (which, as established in Proposition 1.1, is the ratio between the cutoff locations for SRW and NBRW on our sparse random graphs) is at most some $c(Z) < 1$. Assume w.l.o.g. that Z is non-constant (i.e., $\mathbb{P}(Z \neq \mathbb{E}Z) > 0$), otherwise this ratio is $\frac{\mathbb{E}Z - 1}{\mathbb{E}Z + 1}$ as mentioned above.

Let (T, ρ) be a rooted *Augmented* Galton–Watson tree (i.e., the tree formed by joining the roots (one of which being ρ) of two i.i.d. Galton–Watson trees by an edge) with offspring variable Z. As before, let (X_t) and (Y_t) be SRW and NBRW on T, resp.; as first observed in [14], (T, ρ, SRW) is a *stationary environment*, i.e., $(T, \rho) \stackrel{\mathrm{d}}{=} (T, \mathbb{P}_\rho(X_1 \in \cdot))$. Conditioned on (T, ρ), let $H_t(T, \rho)$ be the entropy of SRW after t steps:

$$H_t(T, \rho) = H\Big(\mathbb{P}_\rho(X_t \in \cdot \mid T) \Big) \quad \text{and} \quad h_t = \mathbb{E}[H_t(T, \rho)] \,,$$

and similarly defined $L_t(T, \rho)$ for the NBRW by

$$L_t(T, \rho) = H\Big(\mathbb{P}_\rho(Y_t \in \cdot \mid T) \Big) \quad \text{and} \quad \ell_t = \mathbb{E}[L_t(T, \rho)] \,.$$

With these notations, we have

$$\mathbf{h}_X = \lim_{t \to \infty} \frac{h_t}{t} \,, \quad \text{and } \mathbf{h}_Y = \lim_{t \to \infty} \frac{\ell_t}{t} \,, \tag{3.1}$$

where the identity for $\mathbf{h}_X$ is by [14, Theorem 9.7], and ℓ_t ($t \geq 1$) is explicitly given by

$$\ell_t = \mathbb{E}[\log(Z + 1)] + (t - 1)\mathbb{E}[\log Z] = \mathbb{E}[\log(Z + 1)] + (t - 1)\mathbf{h}_Y \,.$$

Since X_1 and Y_1 have the same distribution, we further have

$$h_1 = \ell_1 = \mathbb{E}[\log(Z + 1)] \,.$$

We need the following result (cf., e.g., the proof of Theorem 3.2 in [5] and Corollary 10 in [6]), which was first observed in the case of random walks on groups by [12]. (Entropy of random walks on random stationary environments were thereafter studied in [11], and it was recently shown in [5] that on stationary environments $t \mapsto h_t$ is sub-additive.) We include the short proof for completeness.

Claim 3.1. *The map $t \mapsto (h_t - h_{t-1})$ is non-increasing.*

Proof. Consider the joint entropy of X_1 and X_t given T:

$$H_{1,t}(T, \rho) := H\left(\mathbb{P}_\rho\big((X_1, X_t) \in \cdot \mid T\big)\right) \quad \text{and} \quad h_{1,t} := \mathbb{E}[H_{1,t}(T, \rho)].$$

Factoring out $\mathbb{P}_\rho(X_1 = x \mid T)$ from $\mathbb{P}_\rho(X_1 = x, X_t = y \mid T)$, one sees that

$$H_{1,t}(T, \rho) = H_1(T, \rho) - \sum_{x \in T} \mathbb{P}_\rho(X_1 = x \mid T) \sum_{y \in T} \mathbb{P}_x(X_{t-1} = y \mid T) \log \mathbb{P}_x(X_{t-1} = y \mid T),$$

and taking expectation gives

$$h_{1,t} = h_1 + \mathbb{E}[H_{t-1}(T, \mathbb{P}_\rho(X_1 \in \cdot))] = h_1 + h_{t-1},$$

where the last equality is due to the stationarity of the environment. Therefore,

$$h_t - h_{t-1} = h_t - h_{1,t} + h_1 = \mathbb{E}[H_t(T, \rho) - H_{1,t}(T, \rho)] + h_1.$$

Conditioned on T, the term $H_{1,t}(T, \rho) - H_t(T, \rho)$ is the conditional entropy $H(X_1 \mid X_t)$, which satisfies $H(X_1 \mid X_t) = H(X_1 \mid X_t, X_{t+1}) \leq H(X_1 \mid X_{t+1})$, since X_1, X_{t+1} are conditionally independent given X_t, and extra information cannot increase entropy. So,

$$h_t - h_{t-1} \geq h_{t+1} - h_{1,t+1} + h_1 = h_{t+1} - h_t. \qquad \blacksquare$$

The fact that $(h_t - h_{t-1})$ is non-increasing in t implies that, for every t,

$$h_t - h_2 \leq \lceil \tfrac{t-2}{2} \rceil (h_3 - h_1),$$

from which we see (recalling (3.1)) that it suffices to show that $h_3 - h_1 < 2\mathbf{h}_Y$ in order to conclude that $\mathbf{h}_X = \lim_{t \to \infty} h_t / t < \mathbf{h}_Y$. To see this, consider $H_3(T, \rho)$, the entropy of SRW after 3 steps on the tree T; using the notation $T_k = \{z \in T : \text{dist}(\rho, z) = k\}$, one has that $H_3(T, \rho) = R + S$ where

$$R = -\sum_{z \in T_3} \mathbb{P}_\rho(X_3 = z) \log \mathbb{P}_\rho(X_3 = z), \qquad (3.2)$$

$$S = -\sum_{x \in T_1} \mathbb{P}_\rho(X_3 = x) \log \mathbb{P}_\rho(X_3 = x).$$

Using the notation $y \prec x$ to denote that y is a child of x and D_y to be the number of children of y in T, one has

$$\mathbb{E}[R \mid T_1, T_2, T_3] = \sum_{y \prec x \prec \rho} D_y \frac{\log D_\rho + \log(D_x + 1) + \log(D_y + 1)}{D_\rho (D_x + 1)(D_y + 1)}$$

(where each child of y played the role of z in (3.2), hence the factor D_y above), thus

$$\mathbb{E}[R \mid T_1, T_2] = \frac{1}{D_\rho} \sum_{y \prec x \prec \rho} \frac{1}{D_x + 1} \left(\mathbb{E}\Big[\frac{Z+1}{Z}\Big] (\log D_\rho + \log(D_x + 1)) + \mathbb{E}\Big[\frac{Z \log(Z+1)}{Z+1}\Big] \right)$$

$$= \frac{1}{D_\rho} \sum_{x \prec \rho} \frac{D_x}{D_x + 1} \left(\mathbb{E}\Big[\frac{Z}{Z+1}\Big] (\log D_\rho + \log(D_x + 1)) + \mathbb{E}\Big[\frac{Z \log(Z+1)}{Z+1}\Big] \right).$$

Continuing in the same manner, setting $\beta := \mathbb{E}\left[\frac{Z}{Z+1}\right]$ to simplify the notation, we have

$$\mathbb{E}[R] = \mathbb{E}\left[\mathbb{E}[R \mid T_1]\right] = \beta^2 \mathbb{E}\left[\log D_\rho\right] + 2\beta \mathbb{E}\left[\frac{Z\log(Z+1)}{Z+1}\right]$$

$$= \beta^2 \mathbb{E}\left[\log(Z+1)\right] + 2\beta \mathbb{E}\left[\frac{Z\log(Z+1)}{Z+1}\right] \tag{3.3}$$

where the factor $1/D_\rho$ cancelled the number of i.i.d. choices corresponding to $x \prec \rho$, and of the three summands corresponding to $x \prec \rho$ in the preceding display, each of the two that did not involve D_ρ yielded the same expression after averaging over T_2.

Turning our attention to S, by the convexity of $x \mapsto x \log x$ and Jensen's inequality for conditional expectation,

$$\mathbb{E}[S \mid T_1] \leq - \sum_{x \in T_1} \mathbb{E}\left[\mathbb{P}_\rho(X_3 = x) \mid T_1\right] \log \mathbb{E}\left[\mathbb{P}_\rho(X_3 = x) \mid T_1\right]. \tag{3.4}$$

For every $x \in T_1$, accounting for whether $X_2 = \rho$ or $X_2 = y$ for some $y \prec x$ shows that

$$\mathbb{P}_\rho(X_3 = x) = \frac{1}{D_\rho}\left(\sum_{x' \in T_1} \frac{1}{D_\rho D_{x'}} + \sum_{y \prec x} \frac{1}{D_x D_y}\right);$$

thus, again using that $\mathbb{E}[\sum_{y \prec x} \frac{1}{D_x D_y} \mid T_1, T_2] = \mathbb{E}[\frac{1}{Z+1}]\frac{D_x - 1}{D_x}$ as reasoned above,

$$\mathbb{E}\left[\mathbb{P}_\rho(X_3 = x) \mid T_1\right] = \frac{\mathbb{E}\left[\frac{1}{Z+1}\right] + \mathbb{E}\left[\frac{1}{Z+1}\right]\mathbb{E}\left[\frac{Z}{Z+1}\right]}{D_\rho} = \frac{1 - \beta + (1-\beta)\beta}{D_\rho} = \frac{1 - \beta^2}{D_\rho}.$$

Plugging this in (3.4) shows that

$$\mathbb{E}[S \mid T_1] \leq (1 - \beta^2) \log \frac{D_\rho}{1 - \beta^2},$$

which, after taking an average over T_1 and combining it with (3.3), yields

$$h_3 = \mathbb{E}[R + S] \leq \mathbb{E}[\log(Z+1)] + 2\beta \mathbb{E}\left[\frac{Z\log(Z+1)}{Z+1}\right] - (1 - \beta^2)\log(1 - \beta^2). \tag{3.5}$$

By Jensen's inequality and the fact that Z is non-constant (thus the same holds also for $\frac{Z}{Z+1}$), we have $\beta^2 < \mathbb{E}[(\frac{Z}{Z+1})^2]$, hence $\log(1 - \beta^2) > \log(\mathbb{E}[1 - (\frac{Z}{Z+1})^2])$, and another application of Jensen's inequality to $x \mapsto \log x$ implies that

$$(1 - \beta^2)\log(1 - \beta^2) > (1 - \beta)\mathbb{E}\left[\frac{2Z+1}{Z+1}\right]\mathbb{E}\left[\log\left(\frac{2Z+1}{(Z+1)^2}\right)\right]$$

$$\geq (1 - \beta)\mathbb{E}\left[\frac{2Z+1}{Z+1}\log\left(\frac{2Z+1}{(Z+1)^2}\right)\right],$$

where the second inequality used that $f_1(z) = \frac{2z+1}{z+1} = 2 - \frac{1}{z+1}$ is increasing whereas $f_2(z) = \log\left(\frac{2z+1}{(z+1)^2}\right) = \log(1 - (\frac{z}{z+1})^2)$ is decreasing, implying—noting $\mathbb{E}[f_1(Z)^2] < \infty$

and $\mathbb{E}[f_2(Z)^2] < \infty$ since $Z \geq 1$—that $\mathrm{Cov}(f_1(Z), f_2(Z)) \leq 0$. Revisiting (3.5), while recalling that $h_1 = \mathbb{E}[\log(Z + 1)]$, we now infer that

$$
\begin{aligned}
h_3 - h_1 &< 2\beta\mathbb{E}\left[\frac{Z\log(Z+1)}{Z+1}\right] - (1-\beta)\mathbb{E}\left[\frac{2Z+1}{Z+1}\log\left(\frac{2Z+1}{(Z+1)^2}\right)\right] \\
&= 2\mathbb{E}\left[\frac{Z\log(Z+1)}{Z+1}\right] - (1-\beta)\mathbb{E}\left[\frac{2Z+1}{Z+1}\log\left(\frac{2Z+1}{(Z+1)^2}\right) + \frac{2Z\log(Z+1)}{Z+1}\right] \\
&= 2\mathbb{E}\left[\frac{Z\log(Z+1)}{Z+1}\right] - (1-\beta)\mathbb{E}\left[f_3(Z)\right],
\end{aligned}
\tag{3.6}
$$

where

$$
f_3(z) := \frac{2z+1}{z+1}\log(2z+1) - 2\log(z+1).
$$

It is easy to verify that $f_3'(z) = (z+1)^{-2}\log(2z+1) > 0$ for every $z > 0$, thus $f_3(z)$ is increasing and $\mathbb{E}[f_3(Z)^2] < \infty$ thanks to the facts $Z \geq 1$ and $\mathbb{E}Z < \infty$. Therefore, when considered with the function $f_4(z) = \frac{1}{z+1}$ which is decreasing and has a finite second moment, we have $\mathrm{Cov}(f_3(Z), f_4(Z)) \leq 0$, or equivalently,

$$
(1-\beta)\mathbb{E}\left[f_3(Z)\right] \geq \mathbb{E}\left[\frac{f_3(Z)}{Z+1}\right] = \mathbb{E}\left[\frac{2Z+1}{(Z+1)^2}\log(2Z+1) - \frac{2}{Z+1}\log(Z+1)\right].
$$

Combining these with (3.6) yields

$$
h_3 - h_1 < 2\mathbb{E}\left[\log(Z+1) - \frac{2Z+1}{2(Z+1)^2}\log(2Z+1)\right].
$$

The proof will thus be concluded by showing that the function

$$
g(x) := \log\left(\frac{x+1}{x}\right) - \frac{2x+1}{2(x+1)^2}\log(2x+1).
$$

satisfies $g(x) < 0$ for all $x \geq 2$, which would then imply that

$$
h_3 - h_1 < 2\mathbb{E}[\log Z] = 2\mathbf{h}_Y = \ell_3 - \ell_1.
$$

To see this, first observe that

$$
g'(x) = (2x+1)(x+1) - \frac{x^2\log(2x+1)}{x(x+1)^3}.
$$

Now, along the interval $[1, \infty)$, the function $x \mapsto \log(2x+1)$ increases from 0 to ∞, whereas $x \mapsto 2 + 3x^{-1} + x^{-2}$ decreases from 6 to 2, thus $g'(x)$ has a unique root x_0 and is negative on $[1, x_0)$ and positive on (x_0, ∞). Hence, it suffices to show that $g(x)$ is negative at $x = 2$ and near ∞, which is indeed the case: $g(2) = \log(3/2) - \frac{5}{18}\log 5 < 0$ and $g(x)\frac{x}{\log(x)} \to -1$ as $x \to \infty$, as claimed. ∎

REFERENCES

[1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.

[2] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly*, 93(5):333–348, 1986.

[3] N. Alon, I. Benjamini, E. Lubetzky, and S. Sodin. Non-backtracking random walks mix faster. *Comm. Contemp. Math.*, 9(4):585–603, 2007.

[4] A. Ben-Hamou and J. Salez. Cutoff for non-backtracking random walks on sparse random graphs. *Ann. Probab.*, 45(3):1752–1770, 2017.

[5] I. Benjamini and N. Curien. Ergodic theory on stationary random graphs. *Electron. J. Probab*, 17(93):1–20, 2012.

[6] I. Benjamini, H. Duminil-Copin, G. Kozma, and A. Yadin. Disorder, entropy and harmonic functions. *Ann. Probab.*, 43(5):2332–2373, 2015.

[7] N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly. Random walks on the random graph. *Ann. Probab.* To appear.

[8] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[9] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[10] R. Durrett. *Random graph dynamics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.

[11] V. Kaimanovich. Boundary and entropy of random walks in random environment. *Prob. Theory and Math. Stat*, 1:573–579, 1990.

[12] V. A. Kaimanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, pages 457–490, 1983.

[13] E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.*, 153(3):475–510, 2010.

[14] R. Lyons, R. Pemantle, and Y. Peres. Ergodic theory on Galton-Watson trees: speed of random walk and dimension of harmonic measure. *Ergodic Theory Dynam. Systems*, 15(3):593–619, 1995.

A. BEN-HAMOU

IMPA, ESTRADA DONA CASTORINA, 110, RIO DE JANEIRO 22460-320, BRAZIL.
   E-mail address: benhamou@impa.br

E. LUBETZKY

COURANT INSTITUTE, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012, USA.
   E-mail address: eyal@courant.nyu.edu

Y. PERES

MICROSOFT RESEARCH, ONE MICROSOFT WAY, REDMOND, WA 98052, USA.
   E-mail address: peres@microsoft.com