RANDOM WALKS ON THE RANDOM GRAPH

By Nathanaël Berestycki[1], Eyal Lubetzky[2], Yuval Peres and
Allan Sly

*University of Cambridge, New York University, One Microsoft Way and
University of California, Berkeley*

We study random walks on the giant component of the Erdős–Rényi
random graph $G(n, p)$ where $p = \lambda/n$ for $\lambda > 1$ fixed. The mixing time
from a worst starting point was shown by Fountoulakis and Reed, and in-
dependently by Benjamini, Kozma and Wormald, to have order $\log^2 n$. We
prove that starting from a uniform vertex (equivalently, from a fixed ver-
tex conditioned to belong to the giant) both accelerates mixing to $O(\log n)$
and concentrates it (the cutoff phenomenon occurs): the typical mixing is at
$(\nu \mathbf{d})^{-1} \log n \pm (\log n)^{1/2+o(1)}$, where ν and $\mathbf{d}$ are the speed of random walk
and dimension of harmonic measure on a Poisson(λ)-Galton–Watson tree.
Analogous results are given for graphs with prescribed degree sequences,
where cutoff is shown both for the simple and for the nonbacktracking ran-
dom walk.

1. Introduction. The time it takes random walk to approach its stationary
distribution on a graph is a gauge for an array of properties of the underlying ge-
ometry: it reflects the histogram of distances between vertices, both typical and
extremal (radius and diameter); it is affected by local traps (e.g., escaping from a
bad starting position) as well as by global bottlenecks (sparse cuts between large
sets); and it is closely related to the Cheeger constant and expansion of the graph.
In this work, we study random walk on the giant component $\mathcal{C}_1$ of the classical
Erdős–Rényi random graph $\mathcal{G}(n, p)$, and build on recent advances in our under-
standing of its geometry on one hand, and random walks on trees and on random
regular graphs on the other, to provide sharp results on mixing on $\mathcal{C}_1$ and on the
related model of a random graph with a prescribed degree sequences.

The Erdős–Rényi random graph $\mathcal{G}(n, p)$ is the graph on n vertices where each of
the $\binom{n}{2}$ possible edges appears independently with probability p. In their celebrated
papers from the 1960s, Erdős and Rényi discovered the so-called "double jump"
in the size of $\mathcal{C}_1$, the largest component in this graph: taking $p = \lambda/n$ with λ fixed,
at $\lambda < 1$ it is with high probability (w.h.p.) logarithmic in n; at $\lambda = 1$ it is of order

$n^{2/3}$ in expectation; and at $\lambda > 1$, it is w.h.p. linear (a "giant component"). Of these facts, the critical behavior was fully established only much later by Bollobás [7] and Łuczak [23], and extends to the *critical window* $p = (1 \pm \varepsilon)/n$ for $\varepsilon = O(n^{-1/3})$ as discovered in [7].

An important notion for the rate of convergence of a Markov chain to stationarity is its (worst-case) total-variation mixing time: for a transition kernel P on a state-space Ω with a stationary distribution π, recall $\|\mu - \nu\|_{\mathrm{TV}} = \sup_{A \subset \Omega}[\mu(A) - \nu(A)]$ and write

$$t_{\mathrm{MIX}}(\varepsilon) = \min\{t : d_{\mathrm{TV}}(t) < \varepsilon\} \qquad \text{where } d_{\mathrm{TV}}(t) = \max_x \|P^t(x, \cdot) - \pi\|_{\mathrm{TV}};$$

let $t_{\mathrm{MIX}}^{(x)}(\varepsilon)$ and $d_{\mathrm{TV}}^{(x)}(t)$ be the analogs from a fixed (rather than worst-case) initial state x. The effect of the threshold parameter $0 < \varepsilon < 1$ is addressed by the *cutoff phenomenon*, a sharp transition from $d_{\mathrm{TV}}(t) \approx 1$ to $d_{\mathrm{TV}}(t) \approx 0$, whereby $t_{\mathrm{MIX}}(\varepsilon) = (1 + o(1))t_{\mathrm{MIX}}(\varepsilon')$ for any fixed $0 < \varepsilon, \varepsilon' < 1$ [making $t_{\mathrm{MIX}}(\varepsilon)$ asymptotically independent of this ε].

In recent years, the understanding of the geometry of $\mathcal{C}_1$ and techniques for Markov chain analysis became sufficiently developed to give the typical order of the mixing time of random walk, which transitions from n in the critical window ([26]) to $\log^2 n$ at $p = \frac{\lambda}{n}$ on for $\lambda > 1$ fixed ([5, 16]) through the interpolating order $\varepsilon^{-3} \log^2(\varepsilon^3 n)$ when $p = \frac{1+\varepsilon}{n}$ for $n^{-1/3} \ll \varepsilon \ll 1$ ([11]). Of these facts, the lower bound of $\log^2 n$ on the mixing time in the supercritical regime ($\lambda > 1$) is easy to see, as $\mathcal{C}_1$ w.h.p. contains a path of $c \log n$ degree-2 vertices for a suitable fixed $c > 0$ [escaping this path when started at its midpoint would require $(\frac{c}{2} \log n)^2$ steps in expectation]. The two works that independently obtained a matching $\log^2 n$ upper bound used very different approaches: Fountoulakis and Reed [16] relied on a powerful estimate from [15] on mixing in terms of the conductance profile (following Lovász–Kannan [21]), while Benjamini, Kozma and Wormald [5] used a novel decomposition theorem of $\mathcal{C}_1$ as a "decorated expander."

As the lower bound was based on having the initial vertex v_1 be part of a rare bottleneck (a path of length $c \log n$), one may ask what $t_{\mathrm{MIX}}^{(v_1)}$ is for a typical v_1. Fountoulakis and Reed [16] conjectured that if v_1 is not part of a bottleneck, the mixing time is accelerated to $O(\log n)$. This is indeed the case for *almost every* initial vertex v_1, and moreover $t_{\mathrm{MIX}}^{(v_1)}(\varepsilon)$ then concentrates on $c_0 \log n$ for c_0 independent of ε (cutoff occurs).

THEOREM 1. *Let $\mathcal{C}_1$ denote the giant component of the random graph $\mathcal{G}(n, p = \lambda/n)$ for $\lambda > 1$ fixed, and let v and $\mathbf{d}$ denote the speed of random walk and the dimension of harmonic measure on a Poisson(λ)-Galton–Watson tree, respectively. For any $0 < \varepsilon < 1$ fixed, w.h.p. the random walk from a uniformly chosen vertex $v_1 \in \mathcal{C}_1$ satisfies*

$$(1.1) \qquad \left| t_{\mathrm{MIX}}^{(v_1)}(\varepsilon) - (v\mathbf{d})^{-1} \log n \right| \leq \log^{1/2 + o(1)} n.$$
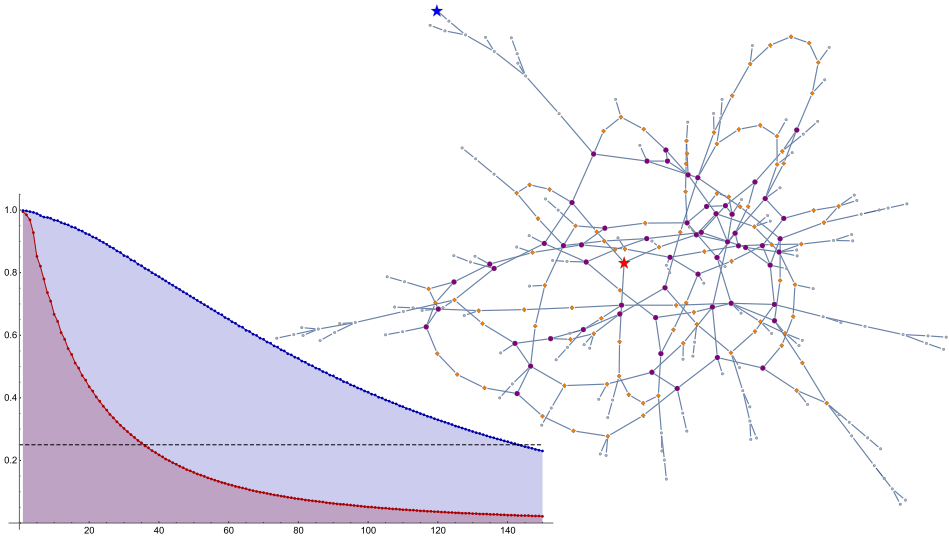
FIG. 1. *Total-variation distance to stationarity from typical/worst initial vertices (marked red/blue) on the giant component of* $\mathcal{G}(n, p = \frac{2}{n})$.

In particular, w.h.p. the random walk from v_1 has cutoff with a $\log^{1/2+o(1)} n$ window.

Before we examine the roles of ν and **d** in this result, it is helpful to place it in the context of the structure theorem for $\mathcal{C}_1$, recently given in [11] (see Theorem 3.3 below): a contiguous model for $\mathcal{C}_1$ is given by (i) choosing a *kernel* uniformly over graphs on degrees i.i.d. Poisson truncated to be at least 3; (ii) subdividing every edges via i.i.d. geometric variables; and (iii) hanging i.i.d. Poisson Galton–Watson (GW) trees on every vertex. Observe that steps (ii) and (iii) introduce i.i.d. delays with an exponential tail for the random walk; thus, starting the walk from a uniform vertex (rather than on a long path or a tall tree) would essentially eliminate all but the typical $O(1)$-delays, and it should rapidly mix on the kernel (w.h.p. an expander) in time $O(\log n)$; see Figure 1.

It is well known (see [8, 13, 18, 28]) that $\bar{D}$, the average distance between two vertices in $\mathcal{C}_1$, is $\log_\lambda n + O_P(1)$, analogous to the fact that $\bar{D} = \log_{d-1} n + O_P(1)$ in $\mathcal{G}(n, d)$, the uniform d-regular graph on n vertices[3] [both are locally-tree-like: $\mathcal{G}(n, d)$ resembles a d-regular tree while $\mathcal{G}(n, p)$ resembles a Poisson(λ)-GW tree]. It is then natural to expect that $t_{\mathrm{MIX}}^{(v_1)}$ coincides with the time it takes the walk to reach this typical distance from its origin v_1, which would be $\nu^{-1}\bar{D}$ for a random walk on a Poisson(λ)-GW tree.

---

[3] Here, $O_P(1)$ denotes a random variable that is bounded in probability.

Supporting evidence for this on the random 3-regular graph $\mathcal{G}(n, 3)$ was given in [6], where it was shown that the distance of the walk from v_1 after $t = c \log n$ steps is w.h.p. $(1 + o(1))(\nu t \wedge \bar{D})$, with $\nu = \frac{1}{3}$ being the speed of random walk on a binary tree. Durrett [13], Section 6, conjectured that reaching the correct distance from v_1 indicates mixing, namely that $t_{\text{MIX}}(\frac{1}{4}) \sim 2\nu^{-1}\bar{D}$ for the lazy (hence the extra factor of 2) random walk on $\mathcal{G}(n, 3)$. This was confirmed in [22], and indeed on $\mathcal{G}(n, d)$ the simple random walk has $t_{\text{MIX}}(\varepsilon) = \frac{d}{d-2} \log_{d-1} n + O(\sqrt{\log n})$, that is, there is cutoff at $\nu^{-1}\bar{D}$ with an $O(\sqrt{\log n})$ window (in particular, random walk has cutoff on almost every d-regular; prior to the work [22] cutoff was confirmed almost exclusively on graphs with unbounded degree).

However, Theorem 1 shows that $t_{\text{MIX}} \sim (\nu \mathbf{d})^{-1} \log n$ versus the $\nu^{-1} \log_\lambda n$ steps needed for the distance from v_1 to reach its typical value (and stabilize there; see Corollary 3.4). As it turns out, the "dimension drop" of harmonic measure (whereby $\mathbf{d} < \log \mathbb{E}Z$ unless the offspring distribution Z is a constant), discovered in [24], plays a crucial role here, and stands behind this slowdown factor of $(\mathbf{d}/\log \lambda)^{-1} > 1$. Indeed, while generation k of the GW-tree has size about $(\mathbb{E}Z)^k$, random walk at distance k from the root concentrates on an exponentially small subset of size about $\exp(\mathbf{d}k)$ (see Figures 2 and 3). Hence, $(\nu \mathbf{d})^{-1} \log n$ is certainly a lower bound on t_{MIX} (the factor ν^{-1} translates time to the distance k from the root), and Theorem 1 shows this bound is tight on $\mathcal{G}(n, p)$.

The same phenomenon occurs more generally in a random walk on a random graph with a degree distribution $(p_k)_{k=1}^\infty$, generated by first sampling the degree of each vertex v via an i.i.d. random variable D_v with $\mathbb{P}(D_v = k) = p_k$ conditioned on $\sum_v D_v$ being even, then choosing the graph uniformly over all graphs with these prescribed degrees.

THEOREM 2. *Let G be a random graph with degree distribution $(p_k)_{k=1}^\infty$, such that for some fixed $\delta > 0$, the random variable Z given by $\mathbb{P}(Z = k - 1) \propto kp_k$*
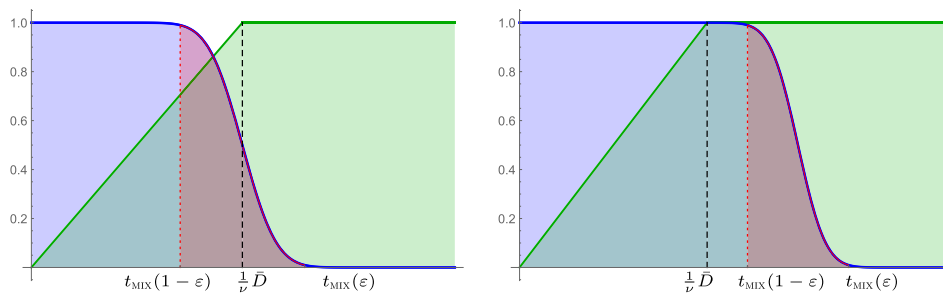

FIG. 2. *Mixing in total-variation versus distance from the origin (rescaled to [0, 1]). On left: random walk on a random regular graph mixes once its distance from the origin reaches the typical distance $\bar{D}$. On right: the analog on a random nonregular graph, where mixing is further delayed due to the dimension drop of harmonic measure.*
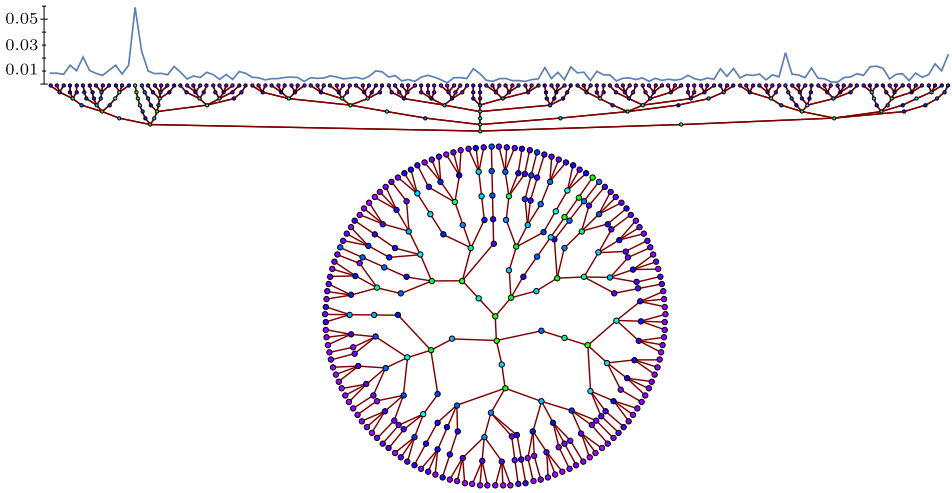
FIG. 3. *Harmonic measure on the first 7 generations of a GW-tree with offspring distribution* $\mathbb{P}(Z=1) = \mathbb{P}(Z=3) = \frac{1}{2}$.

satisfies

$$(1.2) \qquad \mathbb{P}(Z > \Delta_n) = o(1/n) \qquad for \ \Delta_n := \exp[(\log n)^{1/2 - \delta}],$$

and let $t_\star = (\nu \mathbf{d})^{-1} \log n$, *where* ν *and* $\mathbf{d}$ *are the speed of random walk and dimension of harmonic measure on a Galton–Watson tree with offspring distribution* Z:

(i) *Set* $w_n = \sqrt{\log n}(\log \log n)^3$. *If* $p_2 < 1 - \delta$ *and* $1 + \delta < \mathbb{E}Z < K$ *for an absolute constant* K, *then w.h.p. on the event that* v_1 *is in the largest component of* G,

$$d_{\mathrm{TV}}^{(v_1)}(t_\star - w_n) > 1 - \varepsilon, \qquad whereas \ d_{\mathrm{TV}}^{(v_1)}(t_\star + w_n) < \varepsilon.$$

(ii) *Set* $w_n = \sqrt{\log n}$. *If* $Z \geq 2$ *and* $\mathbb{E}Z < K$ *for some absolute constants* K, *then for any* $\varepsilon > 0$ *there exists some* $\gamma > 0$ *such that, with probability at least* $1 - \varepsilon - o(1)$,

$$d_{\mathrm{TV}}^{(v_1)}(t_\star - \gamma w_n) > 1 - \varepsilon, \qquad whereas \ d_{\mathrm{TV}}^{(v_1)}(t_\star + \gamma w_n) < \varepsilon.$$

Condition (1.2) is weaker than requiring Z to have finite exponential moments.

The intuition behind these results is better seen for the *nonbacktracking* (as opposed to simple) random walk (NBRW), which, upon arriving to a vertex v from some other vertex u, moves to a uniformly chosen neighbor $w \neq u$ (formally, this is a Markov chain whose state-space is the set of directed edges in the graph). This walk has speed $\nu = 1$ on a GW-tree (as it never backtracks toward the root), and on $\mathcal{G}(n, 3)$ it was shown in [22] to satisfy $|t_{\mathrm{MIX}}(\varepsilon) - \log_2 n| < C(\varepsilon)$ for any fixed $0 < \varepsilon < 1$ w.h.p.—indeed, cutoff [with an $O(1)$-window] occurs

once the distance from the origin reaches the average graph distance. If we instead take a random graph on $2n/3$ vertices of degree 2 and $n/3$ vertices of degree 4, this corresponds to a GW-tree with an offspring distribution $\mathbb{P}(Z = 1) = \mathbb{P}(Z = 3) = \frac{1}{2}$; since its kth generation grows as 2^k, the distance between two typical vertices is again asymptotically $\log_2 n$. However, the probability that the NBRW follows a given path $v_1, v_2, \ldots, v_k$ is $\prod 1/Z_i$ (with Z_i denoting the number of children of v_i); setting $\mathbf{d} = \mathbb{E} \log Z = \log \sqrt{3} < \log 2$, observe that $\sum_{i < k} \log Z_i$ concentrates around $k\mathbf{d}$ by CLT, and hence $k \sim \mathbf{d}^{-1} \log n$ is a lower bound on mixing. (More generally, by Jensen's inequality $\mathbf{d} = \mathbb{E} \log Z < \log \mathbb{E} Z$ unless Z is constant.)

This straightforward description of harmonic measure for the NBRW allows one to directly control the location of this walk in the random graph. Consequently, by adding a few ingredients to the approach originally used in [22], we were able to extend the NBRW analysis of that work to the nonregular setting (Theorems 4.1–4.2; similar results for the NBRW were independently obtained in [4]; see Section 4 for further details).

However, the harmonic measure for the simple random walk (SRW) remains mysterious, and there is no explicit formula for $\mathbf{d}$ even for very simple offspring distributions (see Figure 3). Formally, let T be an infinite GW-tree rooted at ρ with offspring distribution Z. Under our assumptions, the random walk (X_t) is transient, so its loop-erasure defines a unique ray ξ. Denoting graph distance by $\mathrm{dist}(\cdot, \cdot)$, let ν be the asymptotic speed of the walk (well defined for almost every tree; see [24], Section 3) given by

$$(1.3) \qquad \nu \overset{\text{a.s.}}{=} \lim_{t \to \infty} \frac{1}{t} \mathrm{dist}(X_t, \rho).$$

Let $\mathbb{P}_T$ denote the conditional probability given T, and set

$$(1.4) \qquad W_T(v) = -\log \mathbb{P}_T \qquad (v \in \xi).$$

Consider the metric $d(\xi, \eta) = \exp(-|\xi \wedge \eta|)$ on all rays ∂T, where $\xi \wedge \eta$ is the longest common prefix of ξ and η. It was shown in [24] that in the joint probability space of a GW-tree and SRW on it,

$$(1.5) \qquad W_T(\xi_t)/t \overset{\text{a.s.}}{\longrightarrow} \mathbf{d},$$

where $\mathbf{d}$ is nonrandom and is the Hausdorff dimension of harmonic measure in the above metric. Building on the results of [9, 24, 25], we establish the following refinement of (1.5), which plays a central role in our proof and seems to be of independent interest.

PROPOSITION 3. *Let T be a Galton–Watson tree, conditioned to survive, whose offspring variable Z satisfies $1 < \mathbb{E} Z < \infty$. For every $\varepsilon > 0$, there exists $\delta > 0$ so that, for all t,*

$$(1.6) \qquad \mathbb{P}\big(\big|W_T(\xi_t) - \mathbf{d}t\big| > \delta\sqrt{t}\big) < \varepsilon.$$

Indeed, the upper bound on t_{MIX} will hinge on showing that w.h.p. at a suitable time $t = (\nu\mathbf{d})^{-1}\log n + O(\sqrt{\log n})$ the L^2-distance of SRW from equilibrium is $\exp(O(\sqrt{\log n}))$—a term that originates from the $O(\sqrt{t})$-fluctuations of $W_T(\xi_t)$ as per Proposition 3.

Organization and notation. In Section 2, we will establish Proposition 3 along with several other estimates for random walk on GW-trees, building on the works of [9, 24, 25]. Section 3 studies SRW on random graphs and contains the proofs of Theorem 1 and 2, while Section 4 is devoted to the analysis of the NBRW.

Throughout the paper, a sequence of events A_n is said to hold with high probability (w.h.p.) if $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. We use the notation $f \ll g$ and $f \lesssim g$ to abbreviate $f = o(g)$ and $f = O(g)$, respectively (as well as their converse forms), and $f \asymp g$ to denote $f \lesssim g \lesssim f$. Finally, in the context of an offspring distribution Z with $\mathbb{E}Z > 1$, we say that T is a GW*-tree to refer to the corresponding GW-tree conditioned on survival.

2. Random walk estimates on Galton–Watson trees.

Let T be an infinite tree, rooted at some vertex ρ, on which random walk is transient. In what follows, we will always use (X_t) to denote SRW on T and let the random variable $\xi \in \partial T$ denote its limit, that is, the unique ray that (X_t) visits i.o., or equivalently, the loop-erased trace of $(X_t)_{t=0}^{\infty}$. For any vertex $v \in T$ other than the root, we let v^- denote its parent in T, and let $\theta_T(v)$ denote the incoming flow at v relative to its parent corresponding to harmonic measure:

$$(2.1) \qquad \theta_T(v) = \mathbb{P}_T\big(v \in \xi \mid v^- \in \xi\big),$$

so that if $(v_0 = \rho, v_1, \ldots, v_k = v)$ is a shortest path in T then $\mathbb{P}_T(v \in \xi) = \prod_{i=1}^k \theta_T(v_i)$.

Our goal in this section is to establish Proposition 3 as well as the next two estimates, in each of which the underlying offspring distribution Z is assumed to have $\mathbb{E}Z > 1$.

DEFINITION (Hitting measure). For a tree T rooted at ρ and an integer R, let T_R denote the tree induced on $\{v : \text{dist}(\rho, v) \leq R\}$. For $v \in T_R$, let $\tilde{\theta}_{T_R}(v)$ denote the probability that SRW from ρ first hits level R of T at a descendant of v.

LEMMA 2.1. *Let T be a GW*-tree rooted at ρ. There exists some $c > 0$ (depending only on the law of Z) so that for every $R > 0$ the following holds. With GW*-probability at least $1 - \exp(-cR)$, every child v of ρ satisfies*

$$(2.2) \qquad \big|\theta_T(v) - \tilde{\theta}_{T_R}(v)\big| \leq \exp(-cR),$$

$$(2.3) \qquad \big|\log\theta_T(v) - \log\tilde{\theta}_{T_R}(v)\big| \leq \exp(-cR).$$

LEMMA 2.2. *Let T be a GW*-tree. There exists $c > 0$ such that, for all $R, t > 0$,*

$$\mathbb{P}\big(\text{dist}(X_t, \xi) > R\big) \leq \exp(-cR).$$

2.1. *Proof of Lemma* 2.1. Let Z_1 be the degree of ρ, and denote the children of ρ as $v_1, v_2, \ldots, v_{Z_1}$. We will show that (2.2)–(2.3) hold for $v = v_1$ except with probability $\exp(-cR)$, and the desired statement will follow from a union over the children of ρ and averaging over Z_1 [multiplying said probability by the expectation of Z_1 w.r.t. GW*, which is $O(1)$].

Let L_i be the set of leaves of the depth-$(R-1)$ subtree of v_i [i.e., descendants of v_i at distance $(R-1)$ from it] and let $\tau = \min\{t : X_t \in \bigcup L_i\}$ be the hitting time of SRW to level R of T, so that $\tilde{\theta}_{T_R}(v) = \mathbb{P}_T(X_\tau \in L_1)$. Now let $B = \bigcup_{k>0}\{X_{\tau+k} = \rho\}$ denote the event that X_t revisits $\rho = v^-$ after time τ. Clearly, on the event B^c, we have $\xi_1 = v$ iff $X_\tau \in L_1$, and so

$$(2.4) \qquad\qquad \big|\theta_T(v) - \tilde{\theta}_{T_R}(v)\big| \leq \mathbb{P}_T(B).$$

Estimating $\mathbb{P}_T(B)$ follows from the following result in [9], p. 21, which was obtained as a corollary of a powerful lemma of Grimmett and Kesten [17] (cf. [9], Lemma 2.2).

LEMMA 2.3. *There exist $c, \alpha > 0$ such that the following holds. Let T be a GW*-tree and let $R \geq 1$. With probability at least $1 - \exp(-cR)$, every depth-R vertex x satisfies that the ray P from the root to x has at least αR vertices v such that $\mathbb{P}_v(\tau_P^+ = \infty) > c$, where τ_P^+ denotes the return time of RW to the ray P.*

By this lemma, there is a measurable set $\mathcal{A}$ of GW*-trees such that $\mathbb{P}(\mathcal{A}) \leq \exp(-cR)$ and for every $T \notin \mathcal{A}$, every $z \in \bigcup L_i$ satisfies the following: there are at least αR vertices on its path to ρ so that the SRW from such a vertex y has a probability $q > 0$ of escaping to ∞ never again visiting its parent y^-. In particular,

$$\mathbb{P}_T(B)\mathbf{1}_{T \notin A} \leq (1-q)^{\alpha R} \leq \exp(-c'R)$$

for some $c' > 0$, which we take to be at most $c/2$. This establishes (2.2).

It remains to show (2.3). Let γ_T be the probability that SRW on T does not revisit the root beyond time 0, and notice that by definition, for every i,

$$\theta_T(v_i) \wedge \tilde{\theta}_{T_R}(v_i) \geq \gamma_i/Z_1,$$

where γ_i is the copy of γ_T corresponding to the subtree rooted at v_i. By [24], Lemma 9.1,

$$\mathbb{E}[1/\gamma] \leq \frac{1}{1 - \mathbb{E}[1/Z]} = O(1).$$

Define the event

$$\tilde{\mathcal{A}} = \mathcal{A} \cup \{\gamma_1 < e^{-\frac{1}{2}c'R}\} \cup \{Z_1 > e^{\frac{1}{4}c'R}\},$$

and observe that $\mathbb{P}(\tilde{\mathcal{A}}) = O(e^{-\frac{1}{4}c'R})$ by Markov's inequality.

On the event $\tilde{\mathcal{A}}^c$,

$$\left| \log \theta_T(v) - \log \tilde{\theta}_{T_R}(v) \right| \leq \frac{|\theta_T(v) - \tilde{\theta}_{T_R}(v)|}{\theta_T(v) \wedge \tilde{\theta}_{T_R}(v)} \leq \frac{\mathbb{P}_T(B)}{\gamma_1/Z_1} \leq \frac{e^{-c'R}}{e^{-\frac{3}{4}c'R}} = e^{-\frac{1}{4}c'R},$$

establishing (2.3).

2.2. *Proof of Lemma* 2.2. Following [24], consider the Augmented GW-tree (AGW), which is obtained by joining the roots of two i.i.d. GW-trees by an edge. A highly useful observation of [24] is that the process in which SRW acts on (T, ρ) by moving the root to one of its neighbors is stationary when started at an AGW-tree T and one of its roots. Clearly, the probability of the event under consideration here in the GW-tree is, up to constant factors from below and above, the same as the analogous probability in the AGW-tree (with positive probability the walk never traverses the edge joining the two copies; conversely, a union bound can be applied to the two GW-tree instances). For brevity, an AGW*-tree will denote an AGW-tree conditioned to be infinite, for which the above stationarity property w.r.t. SRW is still maintained.

The event $\text{dist}(X_t, \xi) > R$ implies that, if u is the vertex at distance R from X_t on the loop-erased trace of $(X_0, X_1, \ldots, X_t)$ (the result of progressively erasing cycles as they appear) then the SRW must revisit u at some time $s > t$.

The stationarity reduces our goal to showing that $\mathbb{P}(\bigcup_{s>0}\{X_s = u\}) \leq \exp(-cR)$, where u is the vertex at distance R from X_0 on the loop-erased trace of $(X_0, \ldots, X_{-t})$. Condition on the past of the walk. By Lemma 2.3, with probability $1 - O(\exp(-cR))$ the AGW*-tree T satisfies that the path from ρ to u contains some αR vertices from which the walk would escape and never return to this path with probability bounded away from 0, whence the probability of visiting u at a positive time is at most $\exp(-cR)$ for some $c > 0$.

2.3. *Proof of Proposition* 3. A *regeneration point* for the SRW on the GW*-tree is a time $\tau \geq 1$ such that the edge $(X_{\tau-1}, X_\tau)$ is traversed exactly once along the trace of the random walk. Set $\tau_0 = 0$ and let $\tau_1 < \tau_2 < \cdots$ denote the sequence of regeneration points on a random GW*-tree T. The following remarkable result due to Kesten was reproduced in [27] (see also [24, 25] where the first part was observed).

LEMMA 2.4. *Let T be a GW*-tree rooted at ρ, let $(\tau_i)_{i \geq 0}$ be the regeneration points of SRW on it, and let $\varphi_i = \text{dist}(\rho, X_{\tau_i})$ denote their depths in the tree:*

(a) *Let T_i ($i \geq 1$) denote the depth-$(\varphi_i - \varphi_{i-1})$ tree that is rooted at $X_{\tau_{i-1}}$. Then $\{(T_i, (X_t)_{\tau_{i-1} \leq t < \tau_i})\}_{i \geq 1}$ are mutually independent, and for $i \geq 2$ they are i.i.d.*
(b) *There exist some $\alpha > 0$ such that $\mathbb{E} \exp[\alpha(\varphi_i - \varphi_{i-1})] < \infty$.*

By the standard decomposition of a GW*-tree into a GW-tree with no leaves (the *skeleton* of the tree) and critical bushes (see, e.g., [2], Section I.12), together with the fact that the harmonic measure is supported on the skeleton, we may assume throughout this proof that $\mathbb{P}(Z = 0) = 0$.

Let $\varphi_i = \mathrm{dist}(\rho, X_{\tau_i})$ ($i \geq 1$) be as in the above lemma (noting that necessarily $X_{\tau_i} = \xi_{\varphi_i}$ as regeneration points are by definition part of the loop-erased trace ξ) and set

$$Y_i = -\log \theta_T(\xi_i),$$

so that our goal is to show the concentration of $\sum_{i=1}^t Y_i = W_T(\xi_t)$. Next, we show that Y_i has exponential moments. If $v_1, \ldots, v_{Z_1}$ are the children of the root (recall $Z_1 > 0$), then $\theta_T^{-1}(\xi_1) = \sum_{i=1}^{Z_1} \mathbf{1}_{\xi_1 = v_i} \theta_T^{-1}(v_i)$ with the summands being i.i.d. given Z_1. Hence,

$$\begin{aligned}
\mathbb{E}[\theta_T^{-1}(\xi_1) \mid Z_1] &= Z_1 \mathbb{E}[\mathbf{1}_{\xi_1 = v_1} \theta_T^{-1}(v_1) \mid Z_1] \\
&= Z_1 \mathbb{E}[\theta_T^{-1}(v_1) \mathbb{E}[\mathbf{1}_{\xi_1 = v_1} \mid T] \mid Z_1] = Z_1,
\end{aligned}$$

and it follows that

$$\mathbb{E} \exp(Y_1) = \mathbb{E}[\theta_T(\xi_1)^{-1}] = \mathbb{E} Z_1 = O(1).$$

By [24], Theorem 8.1, there exists a probability measure $\mathrm{GW}_{\mathrm{HARM}}$ on GW-trees, which satisfies the following. It is stationary for the harmonic flow θ_T, mutually absolutely continuous w.r.t. GW (moreover, $\mathrm{GW}_{\mathrm{HARM}}/\mathrm{GW}$ is uniformly bounded from above), and under $\mathrm{GW}_{\mathrm{HARM}}$ one has $\mathbb{E}[W_T(\xi_t)] = \mathbf{d}t$. We will show that

(2.5) $$\mathrm{Var}(W_T(\xi_t)) \leq C_z t \qquad \text{w.r.t. } \mathrm{GW}_{\mathrm{HARM}} \ltimes \mathrm{SRW}$$

for some constant C_z that depends only on the law of the offspring variable Z, where $\mathrm{GW}_{\mathrm{HARM}} \ltimes \mathrm{SRW}$ is the joint distribution of a $\mathrm{GW}_{\mathrm{HARM}}$-tree and SRW on it. A direct application of Chebyshev's inequality will then imply (1.6) under $\mathrm{GW}_{\mathrm{HARM}} \ltimes \mathrm{SRW}$ (for $\delta = \sqrt{C_z/\varepsilon}$). The analogous statement for $\mathrm{GW} \ltimes \mathrm{SRW}$ will then follow by the absolute continuity: for every $\varepsilon > 0$ there exists ε' such that if the event in (1.6) has probability at most ε' under $\mathrm{GW}_{\mathrm{HARM}} \times \mathrm{SRW}$ then it has probability at most ε under $\mathrm{GW} \ltimes \mathrm{SRW}$.

It remains to show (2.5). Note that the bounds established above on $\mathbb{E} Y_1^2$ under GW remain valid under $\mathrm{GW}_{\mathrm{HARM}}$ thanks to the fact that $d\mathrm{GW}_{\mathrm{HARM}}/d\mathrm{GW} = O(1)$.

Since the sequence $(Y_i)_{i \geq 1}$ is stationary under $\mathrm{GW}_{\mathrm{HARM}} \ltimes \mathrm{SRW}$, it suffices to show that there exist some constants $c, c' > 0$ such that $\mathrm{Cov}(Y_1, Y_k) \leq c \exp(-c'k)$ for all k. Let $T_{\lfloor k/2 \rfloor}$ be the depth-$\lfloor k/2 \rfloor$ subtree of T (sharing the same root), and

write $\theta(v) = \theta_T(v)$ and $\tilde{\theta}(v) = \tilde{\theta}_{T_{\lfloor k/2 \rfloor}}(v)$ for brevity. Denoting the children of the root by $v_1, \ldots, v_{Z_1}$, recall from Lemma 2.1 (with $R = \lfloor k/2 \rfloor$) and Lemma 2.4 that there exist some $c_1, c_2 > 0$ (depending only on the offspring distribution Z) such that the events $\mathcal{E}_1, \mathcal{E}_2$ given by

$$\mathcal{E}_1 = \bigcup_{1 \le i \le Z_1} \{|\tilde{\theta}(v_i) - \theta(v_i)| > e^{-c_1 k}\}, \qquad \mathcal{E}_2 = \{\varphi_1 \ge k/2\}$$

satisfy $\mathbb{P}(\mathcal{E}_1 \cup \mathcal{E}_2) \le \exp(-c_2 k)$. It follows from Hölder's inequality that

$$\operatorname{Cov}(Y_1 \mathbf{1}_{E_1}, Y_k) \le [\mathbb{E} Y_1^2]^{\frac{1}{2}} [\mathbb{E} Y_k^4]^{\frac{1}{4}} \mathbb{P}(\mathcal{E}_1)^{\frac{1}{4}} \lesssim \exp(-c_2 k)$$

(recall Y_i has finite moments of every order), and similarly for $\operatorname{Cov}(Y_1 \mathbf{1}_{E_1^c}, Y_k \mathbf{1}_{E_2})$; hence, to estimate $\operatorname{Cov}(Y_1, Y_k)$ it remains to bound $\operatorname{Cov}(Y_1 \mathbf{1}_{E_1^c}, Y_k \mathbf{1}_{E_2^c})$.

On the event $\mathcal{E}_1^c$, either (i) we have $\theta(v_i) \le \exp(-\frac{3}{4} c_1 k)$, whence

$$\theta(v_i) \log^2 \theta(v_i) \lesssim e^{-\frac{1}{2} c_1 k} \quad \text{and} \quad \tilde{\theta}(v_i) \log^2 \tilde{\theta}(v_i) \lesssim e^{-\frac{1}{2} c_1 k},$$

or (ii) we have $\theta(v_i) > \exp(-\frac{3}{4} c_1 k)$ and infer from the bound on $|\tilde{\theta}(v_i) - \theta(v_i)|$ that

$$\left| \log \theta(v_i) - \log \tilde{\theta}(v_i) \right|^2 \le \left(\frac{e^{-c_1 k}}{\theta(v_i) \wedge \tilde{\theta}(v_i)} \right)^2 \lesssim e^{-\frac{1}{2} c_1 k},$$

where in both cases the implicit constants are independent of k. Hence, if we let Y_1' assume the value $\log(1/\tilde{\theta}(v_i))$ with probability $\tilde{\theta}(v_i)$ for each child v_i of the root of T,

$$\mathbb{E} |Y_1 \mathbf{1}_{E_1^c} - Y_1'|^2 \lesssim e^{-\frac{1}{2} c_1 k} \mathbb{E} Z \lesssim e^{-\frac{1}{2} c_1 k};$$

thus [recalling that $\mathbb{E} Y_k^2 = O(1)$],

$$\operatorname{Cov}(Y_1 \mathbf{1}_{E_1^c} - Y_1', Y_k \mathbf{1}_{E_2^c}) \le \sqrt{\mathbb{E}(Y_1 \mathbf{1}_{E_1^c} - Y_1')^2 \mathbb{E} Y_k^2} \lesssim e^{-\frac{1}{4} c_1 k}.$$

Thanks to the discrete renewal theorem—crucially using that the renewal intervals φ_i have exponential moments (see [19] as well as [20], Section II.4)—we can couple $Y_k \mathbf{1}_{E_2^c}$ with an i.i.d. copy of Y_1 with probability $1 - O(\exp(-c_3 k))$, and so $\operatorname{Cov}(Y_1', Y_k \mathbf{1}_{E_2^c}) \lesssim \exp(-c_3 k)$. This completes the desired bound on $\operatorname{Cov}(Y_1, Y_k)$ and concludes the proof.

3. Random walk from a typical vertex in a random graph. We begin by studying SRW and LERW on an infinite GW-tree along what would be the cutoff window. This analysis will then carry over to SRW on G via coupling arguments, and provide initial control over the distribution of the walk within the cutoff window. The final step will be to boost mixing in that window via the graph spectrum.

3.1. *Quantitative estimates on the infinite tree.* Fix $\varepsilon > 0$ and set

$$(3.1) \qquad m = n \exp(-\gamma \sqrt{\log n}),$$

$$(3.2) \qquad \ell_0 = \frac{1}{\mathbf{d}} \log m = \frac{1}{\mathbf{d}} \log n - O(\sqrt{\log n}),$$

$$(3.3) \qquad \ell_1 = \ell_0 + \frac{1}{4\mathbf{d}} \gamma \sqrt{\log n},$$

where $\gamma > 0$ is a suitably large constant with the following two properties:

(i) By Proposition 3, for large enough γ (in terms of the law of Z and ε),

$$(3.4) \qquad \mathbb{P}\left(W_T(\xi_{\ell_0}) > \log m - \frac{1}{3}\gamma\sqrt{\log n} \right) > 1 - \varepsilon,$$

$$(3.5) \qquad \mathbb{P}\left(W_T(\xi_{\ell_1}) < \log m + \frac{1}{3}\gamma\sqrt{\log n} \right) > 1 - \varepsilon.$$

(ii) It is known (see [27]) that the distance of SRW (X_t) from the root (its origin) of GW-a.e. tree has mean $vt + O(1)$ and variance at most $C_z t$ for some fixed $C_z > 0$ (which depends only on the offspring variable Z). Thus,

$$\mathbb{P}\left(|\mathrm{dist}(\rho, X_t) - vt| \leq \frac{1}{16\mathbf{d}}\gamma\sqrt{\log n} \right) > 1 - \varepsilon \qquad \text{for any } t \leq v^{-1}\ell_1$$

provided γ is large enough in terms of $\mathbf{d}$, C_z and ε. Combined with the exponential tails of the regeneration times (recall Section 2.3), taking γ large enough further gives

$$(3.6) \qquad \mathbb{P}\left(\max_{t \leq v^{-1}\ell_1} |\mathrm{dist}(\rho, X_t) - vt| \leq \frac{1}{16\mathbf{d}}\gamma\sqrt{\log n} \right) > 1 - \varepsilon.$$

Let $c_0 > 0$ be some constant (to be specified later), and set

$$(3.7) \qquad R = c_0 \log \log n.$$

DEFINITION. For a tree T and integer $R > 0$, if $u_0 = \rho, u_1, \ldots, u_k = u$ is the shortest path from the root to a vertex u and $\tilde{\theta}$ is as defined in Section 2, define

$$\widetilde{W}_T(u) = -\sum_{i=1}^{k} \log \tilde{\theta}_{T_R(u_i^-)}(u_i),$$

where $T_R(u_i^-)$ is the depth-R subtree of T rooted at the parent of u_i.

Consider the exploration process where at step k—corresponding to $\mathcal{F}_k$ in the filtration—we expose the depth-$(k+1)$ descendants of the root, as in the standard

breadth-first-search process, with one modification: we explore the children of a
vertex v (which is at depth k) only if its distance-R ancestor, denoted u, satisfies

$$(3.8) \qquad\qquad \widetilde{W}_T(u) \leq \log m + \frac{1}{2}\gamma\sqrt{\log n}.$$

Instead, if (3.8) is violated, we declare that the exploration process is *truncated* at
all the depth-R descendants of u, and denote this $\mathcal{F}_k$-measurable event by $\mathcal{B}_{\mathrm{tr}}(u)$.

We stress that, when determining whether to truncate the exploration at the ver-
tex v as described above, the variable $\widetilde{W}_T(u)$ is $\mathcal{F}_k$-measurable [as the event $\mathcal{B}_{\mathrm{tr}}(u_i)$
has not occurred for any of the ancestors of u, thus the subtree of depth R of each
of them has been fully explored].

LEMMA 3.1. *Let T be a GW*-tree, and define ℓ_1, γ and $\mathcal{B}_{\mathrm{tr}}$ as above. If c_0
from* (3.7) *is sufficiently large, then*

$$\mathbb{P}\bigg(\bigcup_{k\leq\ell_1}\mathcal{B}_{\mathrm{tr}}(\xi_k)\bigg) \leq \mathbb{P}\bigg(W_T(\xi_{\ell_1}) > \log m + \frac{1}{3}\gamma\sqrt{\log n}\bigg) + o(1) < \varepsilon + o(1).$$

To prove this result, we need the following simple claim.

CLAIM 3.2. *Let T be a GW*-tree. There exist $c_1, c_2 > 0$ such that the follow-
ing holds. For every measurable set $\mathcal{A}$ of trees and integer ℓ,*

$$\mathbb{P}\bigg(\bigcup_{i\leq\ell}\{T_\infty(\xi_i) \in \mathcal{A}\}\bigg) \leq c_1\ell\mathbb{P}(\mathcal{A}) + \exp(-c_2\ell^{1/3}).$$

PROOF.   It suffices to prove this for the AGW*-tree (as argued in the proof of
Lemma 2.2). Let $(X_t)_{t=0}^\infty$ be simple random walk on T whose loop-erased trace
is ξ. Then

$$\mathbb{P}\bigg(\bigcup_{i\leq\ell}T_\infty(\xi_i) \in \mathcal{A}\bigg) \leq \mathbb{P}\bigg(\bigcup_{t\leq 2\ell/v}\{T_\infty(X_t) \in \mathcal{A}\}\bigg) + \mathbb{P}\bigg(\bigcup_{t\geq 2\ell/v}\{\mathrm{dist}(X_t, \rho) < \ell\}\bigg)$$

$$\leq \frac{2\ell}{v}\mathbb{P}(\mathcal{A}) + \sum_{t\geq 2\ell/v}\exp(-c_3 t^{1/3}),$$

using the stationarity of the AGW*-tree and the large deviation estimate from [27]
on the speed of random walk on a GW-tree (see also [9]).   □

PROOF OF LEMMA 3.1.   Note that the right-hand inequality follows from the
choice of γ as per (3.5), and it remains to prove the left-hand inequality.

Let $\mathcal{A}$ denote the set of trees T for which there exists a child v of the root such
that

$$\big|\log\tilde{\theta}_{T_R}(v) - \log\theta_T(v)\big| > \exp(-cR)$$

for $c > 0$ the constant from Lemma 2.1. Note that $\mathbb{P}(\mathcal{A}) < \exp(-cR)$ by (2.3) in that lemma. Furthermore, on the event

$$\mathcal{E}_k := \bigcap_{i=0}^{k-1} \left(\mathcal{B}_{\mathrm{tr}}(\xi_i)^c \cap \{ T_\infty(\xi_i) \notin \mathcal{A} \} \right) \cap \mathcal{B}_{\mathrm{tr}}(\xi_k)$$

we have

$$\left| \log \tilde{\theta}_{T_R}(\xi_i) - \log \theta_T(\xi_i) \right| \leq \exp(-cR) \qquad \text{for all } i = 1, \ldots, k.$$

Thus, by the triangle inequality, the event $\mathcal{E}_k$ implies that

$$\left| \widetilde{W}_T(\xi_k) - W_T(\xi_k) \right| \leq k e^{-cR} \lesssim e^{-cR} \log n = O(\log^{-2} n),$$

provided that $k = O(\log n)$ and $c_0 \geq 3/c$ in the definition (3.7) of R. Therefore, by the truncation criterion (3.8), we find that $\mathcal{E}_k$ implies that

$$W_T(\xi_k) \geq \log m + \frac{1}{2} \gamma \sqrt{\log n} - O(\log^{-2} n) > \log m + \frac{1}{3} \gamma \sqrt{\log n}$$

for sufficiently large n, whence [using that $W_T(\xi_k) \leq W_T(\xi_{\ell_1})$ for all $k \leq \ell_1$],

$$(3.9) \qquad \bigcup_{k \leq \ell_1} \mathcal{E}_k \subset \left\{ W_T(\xi_{\ell_1}) \geq \log m + \frac{1}{3} \gamma \sqrt{\log n} \right\}.$$

On the other hand, Claim 3.2 shows that

$$(3.10) \qquad \mathbb{P}\left(\bigcup_{i \leq \ell_1} \{ T_\infty(\xi_i) \in \mathcal{A} \} \right) \leq c_1 \ell_1 e^{-cR} + e^{-c_2 \ell_1^{1/3}} = O(\log^{-2} n).$$

Since, by the definition of $\mathcal{E}_k$,

$$\bigcup_{k \leq \ell_1} \mathcal{B}_{\mathrm{tr}}(\xi_k) \subset \left(\bigcup_{k \leq \ell_1} \mathcal{E}_k \right) \cup \left(\bigcup_{k \leq \ell_1} \{ T_\infty(\xi_i) \in \mathcal{A} \} \right),$$

combining (3.9)–(3.10) completes the proof. $\quad\square$

Let S_k ($k \leq \ell_1$) denote the set of explored vertices at distance k from the root (so that S_k is $\mathcal{F}_k$-measurable), and define the $\mathcal{F}_{k+R}$-measurable S'_k by

$$S'_k = \{ u \in S_k : \mathcal{B}_{\mathrm{tr}}(u)^c \}.$$

By the definition of the truncation criterion, for every $u \in S'_k$,

$$-\widetilde{W}_T(u) \geq -\log m - \frac{1}{2} \gamma \sqrt{\log n}.$$

It is easy to see (e.g., by induction on k) that

$$\sum_{u \in S'_k} \exp\left(-\widetilde{W}_T(u) \right) \leq 1.$$

In particular,

$$1 \geq |S_k'| m^{-1} e^{-\frac{1}{2}\gamma\sqrt{\log n}},$$

thus it follows from (3.1) that

$$|S_k'| \leq m \exp\left(\frac{1}{2}\gamma\sqrt{\log n}\right) = n \exp\left(-\frac{1}{2}\gamma\sqrt{\log n}\right).$$

Therefore, as the maximum degree is at most $\Delta = \exp[(\log n)^{1/2-\delta}]$, the number of vertices reached in the first ℓ_1 exploration rounds (i.e., those revealed in $\mathcal{F}_{\ell_1}$) is at most

(3.11)
$$\begin{aligned}
\sum_{k \leq \ell_1} |S_k| &\leq \sum_{k \leq \ell_1} |S_k'| \Delta^R \\
&\leq \ell_1 n \exp\left(-\frac{1}{2}\gamma\sqrt{\log n}\right)\Delta^R \\
&\leq n \exp\left(-\frac{1}{3}\gamma\sqrt{\log n}\right),
\end{aligned}$$

using the fact $\Delta^R = \exp[O((\log n)^{1/2-\delta} \log\log n)] = \exp[o(\sqrt{\log n})]$. [We stress that the estimate (3.11) holds with probability 1].

Finally, recall from (3.4) that $W_T(\xi_{\ell_0}) > \log m - \frac{1}{3}\gamma\sqrt{\log n}$ with probability at least $1 - \varepsilon$. Since Lemma 3.1 implies that $\mathbb{P}(\xi_{\ell_0} \in S_{\ell_0}') \geq 1 - \varepsilon - o(1)$, it therefore follows [once we recall from (3.1) that $\frac{1}{m}\exp(\frac{1}{3}\gamma\sqrt{\log n}) = \frac{1}{n}\exp(\frac{4}{3}\gamma\sqrt{\log n})$] that

$$\mathbb{P}(\xi_{\ell_0} \in S_{\ell_0}'') \geq 1 - 2\varepsilon - o(1)$$

for $S_{\ell_0}'' \subset S_{\ell_0}'$ defined by

(3.12)
$$S_{\ell_0}'' = \left\{ u \in S_{\ell_0}' : \mathbb{P}_T(\xi_{\ell_0} = u) \leq \frac{1}{n}\exp\left(\frac{4}{3}\gamma\sqrt{\log n}\right)\right\}.$$

3.2. *Coupling the GW-tree to the random graph.* We describe a coupling of SRW (X_t) on a GW-tree[4] T, started at its root ρ, to SRW $(\mathscr{X}_t)$ on G, started at a fixed vertex v_1. Let $\tilde{T} \subset T$ be the truncated GW-tree from Section 3.1, and let τ be the time at which the SRW (X_t) on T reaches level ℓ_1. We will first construct a coupling of G and T, in which we will specify a tree $\tilde{\Gamma} \subset G$ rooted at v_1 and a one-to-one mapping $\phi : \tilde{\Gamma} \to \tilde{T}$, such that

(3.13)
$$\deg_G(x) = \deg_T(\phi(x)) \qquad \text{for all } x \in \tilde{\Gamma}$$

---

[4]More precisely, the degree of the root will be different—given by (p_k) as opposed to the size-biased law of Z—but this has no effect since this model is clearly contiguous to the standard GW-tree.

[where $\deg_G(u)$ accounts for multiple edges]. Moreover, defining, for $k \geq 1$, the event

$$\Pi_k := \bigcap_{t \leq k} \{X_t \in \phi(\tilde{\Gamma})\},$$

we will show that

(3.14) $$\mathbb{P}(\Pi_\tau) \geq 1 - \varepsilon - o(1).$$

In what follows, we refer to the event Π_k as a "successful coupling" of $(\mathscr{X}_t, X_t)_{t=1}^k$, for the following reason: denoting $\tilde{\tau} = \min\{t : X_t \notin \phi(\tilde{\Gamma})\}$, take $\mathscr{X}_t = \phi^{-1}(X_t)$ for $t < \tilde{\tau}$, let $\mathscr{X}_{\tilde{\tau}}$ choose a uniform neighbor of $\mathscr{X}_{\tilde{\tau}-1}$ among those that belong to $G \setminus \tilde{\Gamma}$ [recalling that $\deg_G(\mathscr{X}_{\tilde{\tau}-1}) = \deg_T(X_{\tilde{\tau}-1})$] and run $\mathscr{X}_t$ independently of X_t for $t > \tilde{\tau}$. This yields the correct marginal of SRW on G by (3.13). When using this coupling, on the event Π_τ one has $\mathscr{X}_t = \phi^{-1}(X_t)$ for all $t \leq \tau$.

Construct the coupling of (G, T) by exposing, in tandem, the truncated GW-tree $\tilde{T}$ and a subgraph of the exploration process of v_1 in G via the configuration model.[5] Initially, we let $\deg_G(v_1) = \deg_T(\rho)$ and set $\tilde{\Gamma} = \{v_1\}$ and $\phi(v_1) = \rho$. We proceed by exposing, vertex by vertex, the tree $\tilde{T}$, and attempting to find $\tilde{\Gamma}$ in G such that $\tilde{\Gamma}$ is isomorphic to a subtree of $\tilde{T}$ rooted at ρ, while maintaining that the queue of active vertices (ones that await exploration) in G will always be a subset of $\tilde{\Gamma}$.

Suppose we are now exploring the offspring of $z \in \phi(\tilde{\Gamma})$. Note one can map the offspring $y_1, \ldots, y_r$ of z one-to-one to the half-edges of $u = \phi^{-1}(z)$ in G, a currently active vertex in the exploration queue, with (3.13) in mind.

(a) If the offspring of z are to be truncated, remove u from the exploration queue of G (do not explore its half-edges). Otherwise, process $y_1, \ldots, y_r$ sequentially as follows.

(b) If the half-edge corresponding to y_i is matched to some previously encountered vertex v in the exploration of G [i.e., the edge (u, v) formed by this pairing closes a cycle in G], mark both u and v by saying that the events $\mathcal{B}_{\text{cyc}}(u)$ and $\mathcal{B}_{\text{cyc}}(v)$ hold, and remove u and its offspring from the exploration queue of G. In addition, we delete u and its offspring from $\tilde{\Gamma}$.

(c) If, on the other hand, the half-edge corresponding to y_i is to be matched to a vertex v that has not yet been encountered, we sample $\deg_T(y_i)$ and $\deg_G(v)$

---

[5]Every vertex u is associated with a number of "half-edges" corresponding to its degree, and the (multi)graph is generated by a uniform perfect matching of the half-edges (see, e.g., [8]). In our context, we reveal the connected component of v_1 by successively revealing the paired half-edges of vertices in a queue (breadth-first-search). If multiple edges or loops are found, or the component is exhausted with at most $\sqrt{n}$ vertices, we abort the analysis [rejection-sampling for having $v_1 \in \mathcal{C}_1(G)$].

via the maximal coupling between their distributions (the former has the law of Z, and the law of the latter is a function of the remaining vertex degrees in G). If this sample has $\deg_T(y_i) = \deg_G(v)$, then we add v to $\tilde{\Gamma}$ and set $\phi(v) = y_i$; otherwise, we say $\mathcal{B}_{\deg}(v)$ holds, and remove v from the exploration queue.

The above construction satisfies (3.13) by definition. To verify (3.14), we must estimate the probability that one of the following events occurs: (a) the walk (X_t) visits a truncated vertex [i.e., $\mathcal{B}_{tr}(X_t)$ occurs for some $t \leq \tau$], (b) its counterpart $(\mathcal{X}_t)$ visits a vertex with nontree-edges [i.e., $\mathcal{B}_{cyc}(\mathcal{X}_t)$ occurs for some $t \leq \tau$], or (c) the walk (X_t) visits some vertex $y \in \tilde{T}$ such that its counterpart in G had a mismatched degree [i.e., $\mathcal{B}_{\deg}(\mathcal{X}_t)$ occurs for some $t \leq \tau$].

The estimate of the event in (a) will follow directly from our results in Section 3.1. For the sake of estimating the events in (b)–(c), and in order to circumvent potentially problematic dependencies between the random walk and the random environment, we will estimate the probabilities of the following larger events: if w^R denotes the distance-R ancestor of a vertex w in a tree, then let

$$\widehat{\mathcal{B}}(w) = \bigcup \{\mathcal{B}_{cyc}(v) \cup \mathcal{B}_{\deg}(v) : v \in \tilde{\Gamma}_{R+1}(w^R)\},$$

that is, we test whether there exists a vertex with nontree-edges or a mismatched degree in the entire depth-R subtree of the distance-$(R-1)$ ancestor of w. We will estimate the probability of $\widehat{\mathcal{B}} = \bigcup_{i=0}^{\ell_1} \widehat{\mathcal{B}}(\mathcal{X}_{\varsigma_i})$, where ς_i is the first time that $\operatorname{dist}(\mathcal{X}_t, v_1) = R + i$.

We first reveal the first $R + 1$ levels of $\tilde{\Gamma}$ and $\tilde{T}$, and examine whether $\mathcal{B}_{cyc}(v)$ or $\mathcal{B}_{\deg}(v)$ hold for any one of their vertices. (If so, then $\widehat{\mathcal{B}}$ occurs due to $i = 0$.) For each $i \geq 1$, we reveal level $R + i + 1$ of $\tilde{\Gamma}$ and $\tilde{T}$ in the following manner: we first run $\mathcal{X}_t$ until ς_i, then examine whether $\mathcal{B}_{cyc}(v)$ or $\mathcal{B}_{\deg}(v)$ hold for some v that is a depth-$(R+1)$ descendant of $\mathcal{X}_{\varsigma_i}^R$ (if so, $\widehat{\mathcal{B}}$ occurs) and proceed to reveal the rest of this level.

Using this analysis, we guarantee that we examine whether $\mathcal{B}_{cyc}(v) \cup \mathcal{B}_{\deg}(v)$ occur for vertices v in a given level, only at the first time that the walk has reached that level (thus preserving the underlying independence in the configuration model). In addition, if $\mathcal{X}_t$ leaves $\tilde{\Gamma}$ for the first time on account of either $\mathcal{B}_{cyc}(\mathcal{X}_t)$ or $\mathcal{B}_{\deg}(\mathcal{X}_t)$, then either $\widehat{\mathcal{B}}$ holds, or $\mathcal{X}_t$ must have revisited some vertex u after reaching one of its depth-R descendants. The latter occurs by time t with probability at most $t \exp(-cR)$ by Lemma 2.2, which is $o(1)$ for $t \asymp \log n$ provided that the constant c_0 in the definition of R is sufficiently large.

(a) *Avoiding a truncation*: By definition, a truncated vertex v corresponds to a distance-R ancestor u for which the event $\mathcal{B}_{tr}(u)$ holds, and by Lemma 3.1 the probability that the LERW ξ visits such a vertex u by time ℓ_1 is at most $\varepsilon + o(1)$.

In the event that the LERW ξ does not visit such a vertex, the SRW (X_t) must escape from the ray ξ to distance at least R in order to hit a truncated

vertex, a scenario which has probability at most $\exp(-cR)$ by Lemma 2.2; thus, the overall probability of the SRW hitting a truncation by some time $t \asymp \log n$ is at most

$$te^{-cR} \lesssim e^{-cR} \log n = o(1)$$

provided the constant c_0 in the definition of R is chosen large enough.

(b) *Avoiding a cycle*: A set of vertices S of size s in the boundary of the exploration process of G has at most Δs half-edges that are to be matched in the next round. A nontree-edge at a given vertex $v \in S$ is generated by either:

1. matching a pair of its half-edges together—which has probability $O(\Delta^2/n)$ as long as the number of remaining unmatched half-edges has order n, or

2. pairing one of them to another unmatched half-edge of S (the more common scenario inducing cycles), which has probability $O(s\Delta^2/n)$.

By (3.11), we know that $s \leq n \exp(-\frac{1}{3}\gamma\sqrt{\log n}) = o(n/\Delta)$ and hence, indeed, the number of unmatched half-edges that remain at time ℓ_1 is of order n. Thus, to bound the contribution to $\mathbb{P}(\widehat{\mathcal{B}})$ due to some $\mathcal{B}_{\text{cyc}}(v)$, we estimate the probability of encountering such a vertex along some $\ell_1 \Delta^{R+1}$ trials. A union bound yields that this contribution has order at most

$$\ell_1 \Delta^{R+1} \frac{s\Delta^2}{n} \lesssim \exp\left(-\frac{1}{3}\gamma\sqrt{\log n} + o(\sqrt{\log n})\right)$$
$$= o(1).$$

(c) *Coupling vertex degrees*: While the tree uses a fixed offspring distribution Z, the exploration process in G dynamically changes as the half-edges are sampled without replacement. Let N_k denote the number of vertices of degree k in G, and note that

$$\mathbb{P}\big(|N_k - np_k| > 4(\sqrt{np_k \log n} \vee \log n)\big)$$
$$= O\big(n^{-3/2}\big)$$

by a standard large deviation estimate for the binomial variable $N_k \sim \text{Bin}(n, p_k)$, for example, [18], Theorem 2.1, where the $O(\cdot)$ includes an extra factor of 2 due to the conditioning that the sum of the degrees, $\sum k N_k$, should be even. We may thus condition on the degrees $(D_u)_{u \in V}$ (whence G becomes a random graph with a given degree sequence) and assume, by neglecting an event of probability $o(1)$, that

(3.15) $\qquad |N_k - np_k| \leq 4(\sqrt{np_k \log n} \vee \log n) \qquad$ for all $k = 1, \ldots, \Delta$

and that $N_k = 0$ for $k > \Delta$. Further observe that, by Cauchy–Schwarz,

$$(3.16) \qquad \sum_{k=1}^{\Delta}(\sqrt{np_k \log n} \vee \log n) \leq \sqrt{\Delta n \log n} \vee \Delta \log n \leq n^{1/2+o(1)}.$$

Letting $\tilde{N}_k$ denote the number of vertices with k unmatched half-edges at some point in time, a given half-edge gets matched to a vertex of degree j with probability $j\tilde{N}_j / \sum_k k\tilde{N}_k$. Denoting the last variable by $\tilde{Z}$, and note that

$$\sum_j j|\tilde{N}_j - np_j| \leq \Delta \sum_j (|\tilde{N}_j - N_j| + |N_j - np_j|)$$

$$(3.17) \qquad\qquad \leq \Delta \sum_{i \leq \ell_1} |S_i| + n^{1/2+o(1)}$$

$$\lesssim \Delta n \exp\left(-\frac{1}{3}\gamma\sqrt{\log n}\right),$$

where the inequality between the lines used (3.15)–(3.16), and the last inequality used (3.11). Thus,

$$\|\mathbb{P}(Z \in \cdot) - \mathbb{P}(\tilde{Z} \in \cdot)\|_{\mathrm{TV}} = \frac{1}{2}\sum_{j \leq \Delta}\left|\frac{jp_j}{\sum_k kp_k} - \frac{j\tilde{N}_j}{\sum_k k\tilde{N}_k}\right|$$

$$= \frac{1}{2}\sum_{j \leq \Delta} j\left|\frac{p_j \sum_k k\tilde{N}_k - \tilde{N}_j \sum_k kp_k}{\sum_k kp_k \sum_k k\tilde{N}_k}\right|,$$

which, if we write $\Xi = \sum_k k(\tilde{N}_k - np_k)$, is equal to

$$\frac{1}{2}\sum_{j \leq \Delta} j\left|\frac{np_j - \tilde{N}_j}{n\sum_k kp_k} + \frac{\tilde{N}_j \Xi}{n\sum_k kp_k \sum_k k\tilde{N}_k}\right|$$

$$\leq \frac{1}{2n}\sum_j j|\tilde{N}_j - np_j| + |\Xi|\sum_{j \leq \Delta}\frac{\tilde{N}_j}{n^2},$$

for large n, using $\sum_k k\tilde{N}_k \geq (1 - o(1))n$. By (3.17) and the hypothesis on Δ,

$$\|\mathbb{P}(Z \in \cdot) - \mathbb{P}(\tilde{Z} \in \cdot)\|_{\mathrm{TV}} \lesssim (\Delta + \Delta^2)\exp\left(-\frac{1}{3}\gamma\sqrt{\log n}\right).$$

A union bound over $\ell_1 \Delta^{R+1}$ trials of examining $\mathcal{B}_{\mathrm{deg}}(v)$ shows that this contribution to $\mathbb{P}(\widehat{\mathcal{B}})$ is also $o(1)$.

3.3. *Lower bound on mixing time in Theorem* 2.   Set

$$t^- = \frac{\ell_0}{\nu} = \nu^{-1}\left(\ell_1 - \frac{1}{4\mathbf{d}}\gamma\sqrt{\log n}\right),$$

and observe that (3.6) implies that $\mathrm{dist}(\rho, X_{t^-}) \leq \ell_1 - (8\mathbf{d})^{-1}\gamma\sqrt{\log n}$ with probability at least $1 - \varepsilon$. By Lemma 2.2, applied in a union bound over $t \leq t^- = O(\log n)$, w.h.p.

$$(3.18) \qquad\qquad \max\{\mathrm{dist}(X_t, \xi) : t \leq t^-\} \leq R$$

for a large enough $c_0 > 0$ (depending only on the distribution of Z) in the definition (3.7) of R. It follows that upon a successful coupling, $\mathscr{X}_{t^-}$ is within distance $R = O(\log\log n)$ from the set of vertices corresponding to $\bigcup_{i\leq\ell_1} S_i$, the vertices exposed in ℓ_1 rounds of our truncated exploration process of the GW-tree. Thus, by (3.11), the distribution of $\mathscr{X}_{t^-}$ given a successful coupling has a mass of $1 - \varepsilon$ on at most

$$\Delta^R \sum_{i\leq\ell_1} |S_i| = o(n)$$

vertices, yielding the required lower bound.

3.4. *Upper bound on mixing time in Theorem* 2.

3.4.1. *The case of minimum degree* 3 [*Theorem* 2, *Part* (ii)].   Set

$$t^+ = \frac{\ell_0 + \ell_1}{2\nu} = \nu^{-1}\left(\ell_1 - \frac{1}{8\mathbf{d}}\gamma\sqrt{\log n}\right).$$

We claim that, if P_G is the transition matrix of the SRW ($\mathscr{X}_t$), then for large n,

$$(3.19) \quad \mathbb{P}\left(\sum_{v\in V(G)}\left[P_G^{t^+}(v_1, v) \wedge \frac{1}{n}\exp\left(\frac{4}{3}\gamma\sqrt{\log n}\right)\right] \geq 1 - \sqrt{5\varepsilon}\right) \geq 1 - \sqrt{5\varepsilon}.$$

To see this, first define the event $\Upsilon = \Upsilon_1 \cap \Upsilon_2 \cap \Upsilon_3$, where

$$\Upsilon_1 = \left\{\ell_0 + \frac{1}{16\mathbf{d}}\gamma\sqrt{\log n} \leq \mathrm{dist}(\rho, X_{t^+}) \leq \ell_1 - \frac{1}{16\mathbf{d}}\gamma\sqrt{\log n}\right\},$$

$$\Upsilon_2 = \{(\mathscr{X}_t, X_t)_{t=1}^{t^+} \text{ are successfully coupled}\},$$

$$\Upsilon_3 = \{\xi_{\ell_0} = (\text{loop-erased trace of } (X_t)_{t\leq t^+})_{\ell_0}\} \cap \{\xi_{\ell_0} \in S''_{\ell_0}\} \qquad \text{for } (X_t) \text{ on } T,$$

where S''_{ℓ_0} is as defined in (3.12). Recall from (3.6) that $\mathbb{P}(\Upsilon_1^c) \leq \varepsilon$; we established in Section 3.2 that $\mathbb{P}(\Upsilon_2^c \cap \Upsilon_1) \leq \varepsilon + o(1)$ (utilizing the upper bound portion of the event Υ_1); and as for Υ_3, the first event in that intersection occurs w.h.p. on the event of the lower bound on $\mathrm{dist}(\rho, X_{t^+})$ from Υ_1 [otherwise there would be some

$t \leq t^+$ at which $\mathrm{dist}(X_t, \xi) \gtrsim \sqrt{\log n}$, which is unlikely as argued above (3.18)] and, therefore, $\mathbb{P}(\Upsilon_3^c \cap \Upsilon_1) \leq 2\varepsilon + o(1)$ by the bound above (3.12). Overall,

$$\mathbb{P}(\Upsilon^c) \leq 4\varepsilon + o(1),$$

and by Markov's inequality, for sufficiently large n we have

$$\mathbb{P}(\mathbb{P}_G(\Upsilon) \geq 1 - \sqrt{5\varepsilon}) \geq 1 - \sqrt{5\varepsilon}.$$

In particular, with the definition of S_{ℓ_0}'' in mind, with probability at least $1 - \sqrt{5\varepsilon}$,

$$\sum_{v \in V(G)} \left[\mathbb{P}_G(\mathscr{X}_{t^+} = v, \Upsilon \mid \mathscr{X}_0 = v_1) \wedge \frac{1}{n} \exp\left(\frac{4}{3}\gamma\sqrt{\log n}\right) \right]$$

$$\geq \sum_{v \in V(G)} \mathbb{P}_G(\mathscr{X}_{t^+} = v, \Upsilon \mid \mathscr{X}_0 = v_1) \geq \mathbb{P}_G(\Upsilon) \geq 1 - \sqrt{5\varepsilon},$$

and (3.19) follows.

We now reveal G, and can assume that the event in (3.19) occurs. That is, if we set $\mu(v) = P_G^{t^+}(v_1, v) \wedge \frac{1}{n}\exp(\frac{4}{3}\gamma\sqrt{\log n})$ then $1 - \mu(G) \leq \sqrt{5\varepsilon}$. With the fact $\pi(v) \geq 1/(\Delta n)$ in mind, we infer that

$$\left\| P_G^{t^++s}(v_1, \cdot) - \pi \right\|_{\mathrm{TV}} \leq \sqrt{5\varepsilon} + \left\| \mu P_G^s - \pi \right\|_{\mathrm{TV}} \leq \sqrt{5\varepsilon} + \left\| \mu P_G^s - \pi \right\|_{L^2(\pi)},$$

and

$$\left\| \mu/\pi - \mu(G) \right\|_{L^2(\pi)} \leq \exp\left(\left(\frac{4}{3} + o(1)\right)\gamma\sqrt{\log n}\right).$$

In the setting where $p_1 = p_2 = 0$, we can now conclude the proof, using the well-known fact that G is then w.h.p. an expander and by Cheeger's inequality the spectral gap of the SRW on G, denoted gap, is uniformly bounded away from 0 (see [1] on expanders and, *e.g.*, [10], Lemma 3.5, (treating the kernel of the giant component) for the standard argument showing expansion for a random graph with said degree sequence). Indeed, if we take

(3.20) $$s = c\mathrm{gap}^{-1}(\gamma\sqrt{\log n} + \log(1/\varepsilon))$$

for a large enough absolute constant $c > 0$, then we will have

$$\left\| P_G^{t^++s}(v_1, \cdot) - \pi \right\|_{\mathrm{TV}} \leq \sqrt{5\varepsilon} + \varepsilon,$$

as desired.

3.4.2. *Allowing degrees* 1 *and* 2 [*Theorem* 2, *Part* (i)]. The general case requires one final ingredient, since G is no longer an expander. Instead of G, consider the graph G' on $n' \leq n$ vertices where every path of degree-2 vertices whose length is larger than $R = c_0 \log\log n$ is contracted into a single edge, and similarly, every tree whose volume is at least R is replaced by a single vertex.

Let U be the set of vertices in $V(G) \setminus V(G')$ as well as their neighbors in G. We claim that, if v is a uniform vertex conditioned to belong to $\mathcal{C}_1(G)$ and

$$\mathcal{B}_{U,t} = \bigcup_{k \leq t} \{X_k \in U\},$$

where X_k is SRW from v, then $\mathbb{P}_v(\mathcal{B}_{U,\log^2 n}) = o(1)$. This will reduce the analysis to G', as clearly on the event $\mathcal{B}_{U,t}^c$ we can couple the walks on G and G' up to time t. Since $\mathcal{C}_1$ is linear w.h.p., it suffices to take X_0 to be uniform over V, that is, to show that

(3.21) $$\frac{1}{n} \sum_{v \in V} \mathbb{P}_v(\mathcal{B}_{U,\log^2 n}) = o(1).$$

The fact that p_2 is uniformly bounded away from 1 implies that

(3.22) $$\mathbb{P}(v \in U \mid (D_u)_{u \in V}) < D_v (\log n)^{-10}$$

for every vertex v provided we choose the constant c_0 to be sufficiently large [for the long paths this uses the decay of $\mathbb{P}(Y > R)$ where $Y \sim \mathrm{Geom}(p_2)$, whereas for the hanging trees this uses tail estimates for the size of subcritical GW-trees (see [14])].

Condition on the degree sequence $(D_u)_{u \in V}$, let N_k denote the number of vertices of degree k in G, and assume (3.15) holds, neglecting an event of probability $o(1)$ as described above that inequality. Since $\sum_k k p_k = O(1)$ thanks to our assumption on $\mathbb{E}Z$, it follows [see (3.15)–(3.16)] that $\sum_u D_u = (1 + o(1))n \sum_k k p_k = O(n)$. In particular, for every $v \in V$ we have $\pi(v)^{-1} = (\sum_u D_u)/D_v = O(n)$. Therefore, for every t,

$$\frac{1}{n} \sum_v \mathbb{P}_v(\mathcal{B}_{U,t} \mid (D_u)_{u \in V}) \lesssim \sum_v \pi(v) \mathbb{P}_v(\mathcal{B}_{U,t} \mid (D_u)_{u \in V})$$

$$= \mathbb{E}\left[\sum_v \pi(v) \mathbb{P}_v\left(\bigcup_{k \leq t} \{X_k \in U\} \mid G \right) \Big| (D_u)_{u \in V} \right]$$

$$\leq t \sum_v \pi(v) \mathbb{P}(v \in U \mid (D_u)_{u \in V}),$$

where the last step combined a union bound with the stationarity of the SRW. Let $A = \{v : D_v \geq \log^3 n\}$, and notice that $\sum_{k \geq \log^3 n} k p_k = O(1/\log^3 n)$ using $\mathbb{E}Z = O(1)$, which, by (3.15)–(3.16), yields $\pi(A) = O(1/\log^3 n)$. The above display then gives

$$\frac{1}{n} \sum_v \mathbb{P}_v(\mathcal{B}_{U,t} \mid (D_u)_{u \in V}) \lesssim t\pi(A) + t \frac{\log^3 n}{n} \sum_{v \in A^c} \mathbb{P}(v \in U \mid (D_u)_{u \in V})$$

$$\lesssim \frac{t}{\log^4 n},$$

where the last inequality used (3.22) to replace each summand in the sum over $v \in A^c$ by $D_v(\log n)^{-10} \le 1/\log^7 n$. This establishes (3.21).

The proof is concluded by noting that the kernel of G' is an expander w.h.p., and the effect of expanding some of its edges into paths of length at most R, or hanging trees of volume at most R on some of its vertices, can only decrease its Cheeger constant to c/R for some fixed $c > 0$. Hence, by Cheeger's inequality, the spectral gap of G' satisfies $\mathtt{gap} > c/R^2$ for a fixed $c > 0$ (depending only on the degrees of G'), and thereafter taking s as in (3.20) [noting that $s \asymp \sqrt{\log n}(\log \log n)^2$ now] establishes the following: for every $\varepsilon > 0$ there exists some $\gamma > 0$ such that, with probability at least $1 - \varepsilon - o(1)$,

$$d_{\mathrm{TV}}^{(v_1)}(t_\star - \gamma w_n') > 1 - \varepsilon, \qquad \text{whereas } d_{\mathrm{TV}}^{(v_1)}(t_\star + \gamma w_n') < \varepsilon,$$

where $w_n = \sqrt{\log n}(\log \log n)^2$. This immediately implies cutoff, w.h.p., for any choice of a sequence w_n' such that $w_n' = o(w_n)$.

3.5. *Random walk on the Erdős–Rényi random graph*: *Proof of Theorem* 1. The proof will follow from a modification of the argument used for a random graph on a prescribed degree sequence, with the help of the following structure theorem for $\mathcal{C}_1$.

THEOREM 3.3 ([12], Theorem 1).   *Let $\mathcal{C}_1$ be the largest component of $\mathcal{G}(n, p)$ for $p = \lambda/n$ where $\lambda > 1$ is fixed. Let $\mu < 1$ be the conjugate of λ, that is, $\mu e^{-\mu} = \lambda e^{-\lambda}$. Then $\mathcal{C}_1$ is contiguous to the following model $\tilde{\mathcal{C}}_1$:*

1. [*Kernel*] *Let $\Lambda \sim \mathcal{N}(\lambda - \mu, 1/n)$, take $D_u \sim \mathrm{Poisson}(\Lambda)$ for $u \in [n]$ i.i.d. and condition that $\sum D_u \mathbf{1}_{D_u \ge 3}$ is even. Let $N_k = \#\{u : D_u = k\}$ and $N = \sum_{k \ge 3} N_k$. Draw a uniform multigraph $\mathcal{K}$ on N vertices with N_k vertices of degree k for $k \ge 3$.*
2. [*2-Core*] *Replace the edges of $\mathcal{K}$ by paths of i.i.d. $\mathrm{Geom}(1 - \mu)$ lengths.*
3. [*Giant*] *Attach an independent $\mathrm{Poisson}(\mu)$-Galton–Watson tree to each vertex.*

That is, $\mathbb{P}(\tilde{\mathcal{C}}_1 \in \mathcal{A}) \to 0$ implies $\mathbb{P}(\mathcal{C}_1 \in \mathcal{A}) \to 0$ for any set of graphs $\mathcal{A}$.

The first point is that we develop the neighborhood of a fixed vertex of G (not of $\mathcal{C}_1$) as a GW-tree with offspring distribution $Z \sim \mathrm{Poisson}(\lambda)$ for $\lambda > 1$ fixed. Should its component turn out to be smaller than, say, $\log^2 n$, we abort the analysis (with nothing to prove).

The second point is that, when analyzing the probability of failing to couple the SRW on G versus the GW-tree, mismatched degrees are no longer related to sampling with/without replacement from a degree distribution, but rather to the total-variation distance between $\mathrm{Poisson}(\lambda)$ and $\mathrm{Bin}(n', \lambda/n)$ where n' is the number of remaining (unvisited) vertices. By (3.11), we know that $n' \ge n(1 - \exp(-\frac{1}{3}\gamma\sqrt{\log n}))$ and, therefore, by the Stein–Chen method (see, e.g., [3],

equation (1.23)), this total-variation distance is at most $\lambda/n + \|\operatorname{Poisson}(\lambda) - \operatorname{Poisson}(\lambda n'/n)\|_{\mathrm{TV}} \leq (\log n)^{-10}$ (with room to spare).

The final and most important point involves the modification of G into G', where the spectral gap is at least of order R^{-2} where $R = c_0 \log \log n$ for some suitably large constant $c_0 > 0$. We are entitled to do so, with the exact same estimates on the probability that SRW visits the set U of contracted vertices, by Theorem 3.3 (noting the traps introduced in Step 2 are i.i.d. with an exponential tail, and similarly for the traps from Step 3); the crucial fact that G' has a Cheeger constant of at least c/R follows immediately from that theorem using the expansion properties of the kernel.

3.6. *The typical distance from the origin.* A consequence of our arguments above is the following result on the typical distance of random walk from its origin at $t \asymp \log n$.

COROLLARY 3.4. *Consider random walk* $(\mathscr{X}_t)$ *from a uniform vertex* v_1 *in* $\mathcal{C}_1$ *either in the setting of Theorem* 1 *or of Theorem* 2, *let* ν *denote the speed of random walk on the corresponding Galton–Watson tree, and let* λ *be the mean of its offspring distribution. For every fixed* $a > 0$, *if* $t = a \log_\lambda n$ *then* $\operatorname{dist}(v_1, \mathscr{X}_t)/\log_\lambda n \to (\nu a \wedge 1)$ *in probability.*

PROOF. First, consider the case $a < \nu^{-1}$, and fix $\varepsilon > 0$ small enough such that $a\nu < 1 - \varepsilon$. Let T be a GW-tree with an offspring distribution that has the law of Z (this time without any truncation), and consider the coupling described in Section 3.2 of SRW (X_t) on T to SRW $(\mathscr{X}_t)$ on the subtree $\Gamma \subset T$ explored by the standard breadth-first-search of G from v_1. That is, we first reveal T and the walk $(X_1, \ldots, X_t)$, then explore the neighborhood of v_1 in G, and estimate the probability of $\bigcup_{k \leq t} \mathcal{B}_{\mathrm{cyc}}(X_k)$. By the definition of ν in (1.3), the assumption on a implies that $\max_{k \leq t} \operatorname{dist}(\rho, X_k) < (1 - \varepsilon/2) \log_\lambda n$ w.h.p. and, therefore, at all times $k \leq t$, the boundary of the exploration process of G (which is at most the size of T) contains at most $n^{1-\varepsilon/3}$ half-edges. Recalling the argument from Section 3.2 [part (b)], the probability of hitting a cycle at a given step X_k is therefore $O(\Delta n^{-\varepsilon/3}) \leq n^{-\varepsilon/4}$ for large enough n, and a union bound over k implies that w.h.p. $\bigcup_{k \leq t} \mathcal{B}_{\mathrm{cyc}}(X_k)$ does not occur and the coupling is successful. In this case, $\operatorname{dist}(v_1, \mathscr{X}_t) = \operatorname{dist}(\rho, X_t)$ and the desired result follows from the definition (1.3) of ν.

Next, consider $a \geq \nu^{-1}$, and suppose first that $p_1 = p_2 = 0$. As stated in the Introduction, it is then the case that the diameter of the graph is $(1 + o(1))\bar{D}$ where $\bar{D} = \log_\lambda n$, and it remains to provide a matching lower bound on $\operatorname{dist}(v_1, \mathscr{X}_t)$. Let

$$R = \lfloor (\log n)^{1/4} \rfloor$$

and

$$\ell_1 = \lfloor (1 - \varepsilon) \log_\lambda n \rfloor, \qquad \ell_0 = \ell_1 - R.$$

Consider the exploration tree Γ, as defined above, up to level ℓ_1.

CLAIM 3.5. *Let H be the auxiliary graph on level ℓ_0 of Γ, where $(u, v) \in E(H)$ if there is an edge of G connecting $\Gamma_R(u)$ and $\Gamma_R(v)$. With high probability, the largest connected component of H has size at most $10/\varepsilon$.*

PROOF. As stated above, since the boundary of Γ contains at most $n^{1-\varepsilon/3}$ half-edges, the analysis of Section 3.2 shows that the probability of $\mathcal{B}_{\mathrm{cyc}}(z)$ is at most $n^{-\varepsilon/3+o(1)}$ for every vertex z in Γ at distance at most ℓ_1 from v_1. Therefore, the degree of a vertex u in H is dominated by a binomial random variable $\zeta \sim \mathrm{Bin}(n^{\varepsilon/8}, n^{-\varepsilon/4})$, where we used that $|\Gamma_R(u)| \leq \Delta^R = n^{o(1)}$. Consequently, the size of the component of u in H is dominated by the total progeny Λ of a branching process with offspring random variables ζ. By exploring this branching process via breadth-first-search (or alternatively, through the Dwass–Otter formula; see, e.g., [14]), one finds that

$$\mathbb{P}(\Lambda > \ell) \leq \mathbb{P}\left(\sum_{i=1}^{\ell} (\zeta_i - 1) \geq 0\right)$$

$$\leq \mathbb{P}\left(\mathrm{Bin}\left(\ell n^{\varepsilon/8}, n^{-\varepsilon/4}\right) \geq \ell\right)$$

$$\leq \sum_{k \geq \ell} \binom{\ell n^{\varepsilon/8}}{k} n^{-k\varepsilon/4}$$

$$\leq \sum_{k \geq \ell} \left(\frac{e\ell n^{\varepsilon/8}}{k}\right)^k n^{-k\varepsilon/4}$$

$$\leq \sum_{k \geq \ell} (e n^{-\varepsilon/8})^k = O(n^{-\ell\varepsilon/8}).$$

Taking $\ell > 10/\varepsilon$ supports a union bound over u and completes the proof. □

COROLLARY 3.6. *Let H be as defined in Claim 3.5. With high probability, for every connected component C of H there exists an interval $I \subset [\ell_0, \ell_1]$ of length at least $\varepsilon R/10$ so that the induced subgraph of G on levels in I of $\bigcup_{u \in C} \Gamma_R(u)$ is a forest.*

Suppose that G satisfies the property of the above corollary. We claim that if $\mathrm{dist}(v_1, \mathscr{X}_k) = \ell_1$ for some k, then w.h.p. (w.r.t. the walk)

$$\min_{1 \leq j \leq \log^2 n} \mathrm{dist}(v_1, \mathscr{X}_{k+j}) \geq \ell_0.$$

Indeed, in light of the above corollary, in order for the walk to reach level ℓ_0 starting from level ℓ_1, it must cross an interval of length $\varepsilon R/10$ where every vertex has at least two children in a lower level. The probability that the biased random walk (S_t), where $S_t - S_{t-1}$ is 1 with probability $\frac{1}{3}$ and -1 with probability $\frac{2}{3}$, reaches

$\varepsilon R/10$ before returning to the origin is at most $2^{-\varepsilon R/10} < (\log n)^{-5}$, and taking a union bound over the $\log^2 n$ possible time steps now completes the treatment of the case $p_1 = p_2 = 0$.

Finally, when we allow degrees 1 and 2, recall from Section 3.4.2 that one can modify the random graph G to a graph G' where the long paths and large hanging trees all have size $O(\log\log n)$, and couple the SRW on G to the SRW on G' w.h.p. for all $t \leq \log^2 n$. The only change in the above argument is in the analysis of the biased random walk (S_t) along the interval of length $\varepsilon R/10$ that corresponds to a forest in G. There the problem reduces to an interval shorter by a factor of $c\log\log n$, and the fact that $2^{-\varepsilon R/(c\log\log n)}$ is again at most $(\log n)^{-5}$ for large enough n concludes the proof. $\square$

4. Nonbacktracking random walk on random graphs.
In this section, we consider the NBRW on a random graph G with i.i.d. degrees with distribution $(p_i)_{i\geq 0}$ (see the paragraph preceding Theorem 2) where $p_0 = p_1 = 0$. Recall that the NBRW is a Markov chain on the set $\vec{E}$ of directed edges, which moves from $e = (u,v)$ to a uniformly chosen edge among $\{e' = (v,w) : w \neq u\}$. Since $\deg(v) \geq 2$ (as $p_0 = p_1 = 0$) this Markov chain is well defined for all t, and that the uniform measure π on $\vec{E}$ is stationary. Let $P_G^t(e,e')$ be the t-step transition probability of the NBRW.

Our assumption on the degree distribution is that

$$(4.1) \qquad M_0 := \sum_k k(\log k)^2 p_k < \infty.$$

Let $\lambda = \sum_k k p_k$ be the mean degree [noting that $2 \leq \lambda < \infty$ by (4.1)], and let

$$q_{k-1} := k p_k/\lambda \qquad (k \geq 1)$$

be the shifted size-biased distribution. We further assume that the random variable Z given by $\mathbb{P}(Z = k) = q_k$ satisfies

$$(4.2) \qquad \mathbb{P}(Z > \Delta_n) = o(1/n) \qquad \text{for } \Delta_n := \exp[(\log n)^{1/2-\delta}].$$

Recall that the dimension of harmonic measure for NBRW in the GW-tree with offspring distribution (q_k) is

$$(4.3) \qquad \tilde{\mathbf{d}} = \sum_k q_k \log k,$$

which is finite under (4.1). To state our next result, we let

$$t_{\mathrm{MIX}}^{(\pi)}(\varepsilon) = \min\{t : d_{\mathrm{TV}}^{(\pi)}(t) < \varepsilon\} \qquad \text{where } d_{\mathrm{TV}}^{(\pi)}(t) = \frac{1}{|\vec{E}|}\sum_{e\in\vec{E}}\|P_G^t(e,\cdot) - \pi\|_{\mathrm{TV}}.$$

THEOREM 4.1 (Typical starting point).   *Let G be a random graph on n vertices with i.i.d. degrees with distribution $(p_k)_{k \geq 0}$, where $p_0 = p_1 = 0$ and (4.1)–(4.2) hold. For every fixed $0 < \varepsilon < 1$, with probability at least $1 - \varepsilon$,*

$$t_{\mathrm{MIX}}^{(\pi)}(\varepsilon) = \tilde{\mathbf{d}}^{-1} \log n + O(\sqrt{\log n}).$$

THEOREM 4.2 (Worst starting point).   *In the setting of Theorem 4.1 with the extra assumption $p_2 = 0$, for every fixed $0 < \varepsilon < 1$, with probability at least $1 - \varepsilon$,*

$$t_{\mathrm{MIX}}(\varepsilon) = \tilde{\mathbf{d}}^{-1} \log n + O(\sqrt{\log n}).$$

REMARK 4.3.   After completing this work, we learned that another proof of Theorem 4.2 was obtained independently by Ben-Hamou and Salez [4] with a weaker assumption on Δ and an asymptotic estimate of the total-variation distance within the cutoff window.

4.1. *Mixing from a typical starting point.*

4.1.1. *The truncated exploration process.*   We first condition on the degree sequence of the random graph, $\{D_v : v \in V\}$, and may assume, as in Section 3.2, that (3.15) holds by neglecting an event of probability $o(1)$. Let (Y_t) denote the NBRW, and for every directed edge $e = (u, v)$, let $D_e = D_v - 1$ (the number of directed edges that the NBRW can transition to from e).

Throughout this proof, let $t_\star$ be an odd integer given by

$$(4.4) \qquad t_\star = 2L + 1 \qquad \text{where } L = \lceil (2\tilde{\mathbf{d}})^{-1} \log n + 2K \tilde{\mathbf{d}}^{-3/2} \sqrt{\log n} \rceil,$$

for

$$(4.5) \qquad\qquad K := 8\sqrt{M_0}/\varepsilon \qquad \text{with } M_0 \text{ from (4.1)}.$$

Let $e = (u, v) \in \vec{E}$. Let T_e be the tree of depth L constructed as follows. Explore the neighborhood of e via breadth-first-search (as explained in Section 3.2). Exclude every half-edge that is matched to a previously encountered half-edge from T_e. Moreover, do not explore an edge e' (truncate its subtree) if it satisfies

$$(4.6) \qquad\qquad \left| \sum_{e_0 \in \Psi(e, e')} (\log D_{e_0} - \tilde{\mathbf{d}}) \right| \geq K \sqrt{L},$$

where $\Psi(e, e')$ denotes the directed path from e to e' in T_e. Let ∂T_e denote the half-edges at level L that had been encountered (and not yet matched) in this process.

Next, consider a second edge $f = (x, y) \in \vec{E}$ such that $x, y \notin T_e$. Define T_f^e to be the tree of depth L formed by exploring T_{f^*} corresponding to the *reversed* edge $f^* = (y, x)$, as described above, with the following modification: if a half-edge is matched to ∂T_e, we do not include it in the tree. Set

$$\mathcal{F}_e = \sigma(T_e), \qquad \mathcal{F}_f = \sigma(T_f^e), \qquad \mathcal{F}_{e,f} = \sigma(\mathcal{F}_e, \mathcal{F}_f).$$

Consider the sequence (a_k) defined by

(4.7) $$a_k := (2M_0)^{1/2}/k \qquad \text{with } M_0 \text{ from (4.1)},$$

and let $\mathcal{E}$ denote the set of all edges $e \in \vec{E}$ such that the NBRW from e remains in T_e except with probability a_K, for K the constant in (4.5); that is, define

$$\mathcal{E} = \{e \in \vec{E} : Q_e \leq a_K\} \qquad \text{where } Q_e := \mathbb{P}_G\Big(\bigcup_{t \leq L}\{Y_t \notin \vec{E}(T_e)\} \mid Y_0 = e\Big).$$

Analogously, let

$$\mathcal{E}_e = \{f \in \vec{E} : Q_{e,f} \leq a_K\} \qquad \text{where } Q_{e,f} := \mathbb{P}_G\Big(\bigcup_{t \leq L}\{Y_t \notin \vec{E}(T_f^e)\} \mid Y_0 = f^*\Big).$$

Note that, by labeling the half-edges of each vertex, we may refer to a directed edge $e = (u, v)$ also as $e = (i, v)$ where $i \in \{1, \ldots, D_v\}$ is the label of a half-edge matched to a half-edge of the vertex u. Similarly, we may refer to $f = (x, y)$ as $f = (x, j)$ where j corresponds to a half-edge $j \in \{1, \ldots, D_x\}$ matched to a half-edge of y.

With this notation, each half-edge in ∂T_e of the form (u', i) for some $u' \in V$ and $i \in \{1, \ldots, D_{u'}\}$ corresponds to a unique edge in $\vec{E}$, which we include in $\vec{E}(T_e)$, the set of directed edges induced on T_e. We claim that, for each $1 \leq k \leq L$,

(4.8) $$\#\{e' \in \vec{E}(T_e) : \text{dist}(e, e') = k\} \leq n^{1/2+o(1)}.$$

Indeed, by condition (4.6), and using its notation as well as (4.4), for every such e',

(4.9)
$$\mathbb{P}_e(Y_k = e') = \prod_{e_0 \in \Psi(e,e')} \frac{1}{D_{e_0}}$$
$$\geq e^{-k\tilde{\mathbf{d}}-K\sqrt{L}}$$
$$\geq e^{-L\tilde{\mathbf{d}}-K\sqrt{L}} = n^{-1/2-o(1)}.$$

This implies (4.8) since the events $\{Y_k = e'\}$ are disjoint for all edges e'. In particular, using that $\Delta_n = n^{o(1)}$ and $L = O(\log n)$, the total number of half-edges encountered in the exploration process of T_e is at most $n^{1/2+o(1)}$, and, similarly, this also holds for T_f^e.

We next estimate the probability that $e \in \mathcal{E}$ and $f \in \mathcal{E}_e$ for some fixed e, f.

LEMMA 4.4. *For every two edges $e = (i, v)$, $f = (x, j)$ (with $1 \leq i \leq D_v$, $1 \leq j \leq D_x$),*

$$\mathbb{P}(e \notin \mathcal{E}) \leq a_K, \qquad \mathbb{P}(f \notin \mathcal{E}_e \mid T_e) \leq a_K.$$

PROOF. The argument in Section 3.2 shows that the NBRW $(Y_t)_{t\geq 0}$ from $Y_0 = e$ on the random graph G can be coupled to a NBRW $(\xi_t)_{t\geq 0}$ on a (non-truncated) GW-tree T w.h.p., as only the events $\mathcal{B}_{\mathrm{cyc}}$ and $\mathcal{B}_{\mathrm{deg}}$ are to be avoided (*avoiding a cycle* and *coupling vertex degrees*, resp.) when no truncation is performed on the tree.

On the GW-tree, the random variables $(D_{\xi_i})_{i\geq 1}$, recording the number of offspring of each site visited by the NBRW, are i.i.d. with distribution given by $(q_k)_{k\geq 0}$, and in particular, $\mathbb{E}[\log D_{\xi_i}] = \tilde{\mathbf{d}}$ and $\mathbb{E}[\log^2(D_{\xi_i})] \leq M_0$ for M_0 from (4.1). Hence,

$$\mathbb{P}\left(\max_{1\leq k\leq L}\left|\sum_{i=1}^{k}(\log D_{\xi_i} - \tilde{\mathbf{d}})\right| \geq K\sqrt{L}\right) \leq \frac{M_0}{K^2}$$

by Kolmogorov's maximal inequality. Hence, the above coupling of (ξ_i) and (Y_i), which succeeds with probability $1 - o(1)$, implies that

$$\mathbb{E}[Q_e] \leq M_0/K^2 + o(1) = \left(\frac{1}{2} + o(1)\right)a_K^2.$$

It thus follows from Markov's inequality that

$$\mathbb{P}(Q_e \geq a_K) \leq \frac{\mathbb{E}[Q_e]}{a_K} \leq \left(\frac{1}{2} + o(1)\right)a_K < a_K,$$

provided that n is large enough. The statement concerning the NBRW from f^* follows in the exact same manner. $\square$

4.1.2. *Estimating the distribution of the walk via two truncated trees.* We wish to bound the quantity $P_G^t(e, f)$ from below. Note that, since $t_\star = 2L + 1$,

$$P_G^{t_\star}(e, f) \geq \sum_{e_0 \in \partial T_e}\sum_{f_0 \in \partial T_f^e} P_G^L(e, e_0)\mathbf{1}_{\{e_0 = f_0\}}P_G^L(f_0, f),$$

where the event $\{e_0 = f_0\}$ for $e_0 = (u_0, i) \in \partial T_e$ and $f_0 = (x_0, j) \in \partial T_f^e$ is equivalent to having the ith half-edge incident to x be matched with the jth edge incident to y. Further note that $P_G^L((j, x_0), (v, v')) = P_G^L((v', v), (x_0, j))$. Hence,

$$P_G^{t_\star}(e, f) \geq \sum_{e_0 \in \partial T_e}\sum_{f_0 \in \partial T_f^e} \bar{P}_{T_e}^L(e, e_0)\mathbf{1}_{\{e_0 = f_0\}}\bar{P}_{T_f^e}^L(f^*, f_0^*) =: Z_{e,f},$$

where $\bar{P}_T(\cdot, \cdot)$ corresponds to the walk restricted (not conditioned) to remain in $\vec{E}(T)$. We will therefore estimate $\mathbb{E}[Z_{e,f}]$ and then its variance.

LEMMA 4.5. *In the setting of Lemma 4.4, let* $A = \{e \in \mathcal{E}\} \cap \{f \in \mathcal{E}_f\}$*, which is* $\mathcal{F}_{e,f}$*-measurable. Then for every sufficiently large* n*,*

$$(4.10) \qquad \mathbb{E}[Z_{e,f} \mid \mathcal{F}_{e,f}] \geq \frac{1 - 2a_K}{|\vec{E}|}\mathbf{1}_A,$$

$$(4.11) \qquad \mathrm{Var}(Z_{e,f} \mid \mathcal{F}_{e,f}) \leq n^{-2}\exp\left(-K\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}\right).$$

PROOF. Recall that $\mathbb{P}(e_0 = f_0 \mid \mathcal{F}_{e,f}) = 1/(|\vec{E}| - m_{e,f})$, where the $\mathcal{F}_{e,f}$-measurable random variable $m_{e,f}$ denotes the number of half-edges matched during the exploration process in $\mathcal{F}_{e,f}$, which satisfies $m_{e,f} \leq n^{1/2+o(1)}$ by (4.8) and the discussion below it. Since $\bar{P}_{T_e}^L(e, e_0)$ and $\bar{P}_{T_f^e}^L(f, f_0)$ are both $\mathcal{F}_{e,f}$-measurable, we find that

$$\mathbb{E}[Z_{e,f} \mid \mathcal{F}_{e,f}] = \frac{1}{|\vec{E}| - m_{e,f}} \left(\sum_{e_0 \in \partial T_e} \bar{P}_{T_e}^L(e, e_0) \right) \left(\sum_{f_0 \in T_f^e} \bar{P}_{T_f^e}^L(f^*, f_0^*) \right)$$

$$\geq \frac{1}{|\vec{E}|}(1 - Q_e)(1 - Q_{e,f}) \geq \frac{1}{|\vec{E}|}(1 - 2a_K)\mathbf{1}_A,$$

using the definitions of $e \in \mathcal{E}$ and $f \in \mathcal{E}_e$ and that $(1 - a_K)^2 \geq 1 - 2a_K$. This yields (4.10).

To establish (4.11), note that by definition

$$\text{Var}(Z_{e,f} \mid \mathcal{F}) \leq \sum_{e_0, e_1 \in \partial T_e} \sum_{f_0, f_1 \in \partial T_f^e} \bar{P}_{T_e}^L(e, e_0) \bar{P}_{T_e}^L(e, e_1) \bar{P}_{T_f^e}^L(f^*, f_0^*) \bar{P}_{T_f^e}^L(f^*, f_1^*)$$

$$\times \text{Cov}(\mathbf{1}_{\{e_0 = f_0\}}, \mathbf{1}_{\{e_1 = f_1\}} \mid \mathcal{F}).$$

Clearly, if $e_0 \neq e_1$ or $f_0 \neq f_1$ then the term $\text{Cov}(\mathbf{1}_{\{e_0 = f_0\}}, \mathbf{1}_{\{e_1 = f_1\}} \mid \mathcal{F})$ equals

$$\mathbb{P}(e_0 = f_0 \mid \mathcal{F})\big(\mathbb{P}(e_1 = f_1 \mid \mathcal{F}, e_0 = f_0) - \mathbb{P}(e_1 = f_1 \mid \mathcal{F})\big)$$

$$\leq \frac{1}{|\vec{E}| - m_{e,f}} \left(\frac{1}{|\vec{E}| - m_{e,f} - 2} - \frac{1}{|\vec{E}| - m_{e,f}} \right) = \frac{2 + o(1)}{|\vec{E}|^3},$$

where in the last inequality we used that $m_{e,f} = o(n)$ while $|\vec{E}| \geq 2n$. On the other hand, for the diagonal terms where $e_0 = e_1$ and $f_0 = f_1$, we have

$$\text{Var}(\mathbf{1}_{\{e_0 = f_0\}} \mid \mathcal{F}_{e,f}) \leq \mathbb{P}(e_0 = f_0 \mid \mathcal{F}_{e,f}) = \frac{1 + o(1)}{|\vec{E}|}.$$

Overall, using $|\vec{E}| \geq 2n$, we conclude that

$$\text{Var}(Z_{e,f} \mid \mathcal{F}_{e,f})$$

$$\leq \frac{1 + o(1)}{4n^3} + \frac{1 + o(1)}{2n} \left(\max_{e_0 \in \partial T_e} \bar{P}_{T_e}^L(e, e_0) \right) \left(\max_{f_0 \in \partial T_f^e} \bar{P}_{T_f^e}^L(f^*, f_0^*) \right).$$

Now, if $e_0 \in \partial T_e$, then the criterion (4.6) implies that

$$\bar{P}_{T_e}^L(e, e_0) \leq e^{-L\tilde{\mathbf{d}} + K\sqrt{L}}$$

$$= \frac{1}{\sqrt{n}} \exp\left[-\left(2 - \frac{1}{\sqrt{2}} + o(1) \right) K \tilde{\mathbf{d}}^{-1/2} \sqrt{\log n} \right]$$

$$\leq \frac{1}{\sqrt{n}} e^{-K\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}}$$

for n sufficiently large, and the same holds for $\bar{P}^L_{T^e_f}(f^*, f^*_0)$. Consequently,

$$\mathrm{Var}(Z_{e,f} \mid \mathcal{F}_{e,f}) \leq \frac{1 + o(1)}{2n^2} \exp\left(-K\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}\right),$$

and the result follows. $\square$

REMARK 4.6.   The expectation estimate (4.10) remains valid for all smaller L; it is for the variance estimate (4.11) to hold that we must choose L as in (4.4).

COROLLARY 4.7.   *In the setting of Lemma* 4.4, *for every sufficiently large n,*

$$\mathbb{P}\left(Z_{e,f} < (1 - 3a_K)/|\vec{E}|, f \in \mathcal{E}_e \mid T_e\right)\mathbf{1}_{\{e\in E\}} \lesssim \exp\left(-K\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}\right).$$

PROOF.   Since $|\vec{E}| = O(n)$, we deduce from Lemma 4.5 via Chebyshev's inequality that

$$
\begin{aligned}
(4.12) \qquad \mathbb{P}\left(Z_{e,f} < \frac{1 - 3a_K}{|\vec{E}|} \mid \mathcal{F}_{e,f}\right)\mathbf{1}_A &\leq \frac{e^{-K\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}}}{n^2(a_K/|\vec{E}|)^2} \\
&\lesssim e^{-K\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}}.
\end{aligned}
$$

Taking expectation given T_e yields the desired result. $\square$

4.1.3. *Upper bound on the mixing time in Theorem* 4.1.   Let $N = |\vec{E}|$ and $e = (i, v) \in \vec{E}$. Recalling that $P^t_G(e, f) \geq Z_{e,f}$, it follows that

$$
\begin{aligned}
&\mathbb{E}\left[d^{(e)}_{\mathrm{TV}}(t_\star) \mid T_e\right]\mathbf{1}_{\{e\in E\}} \\
&\quad = \mathbb{E}\left[\sum_f \left|\frac{1}{N} - P^{t_\star}_G(e, f)\right|^+\right]\mathbf{1}_{\{e\in E\}} \\
&\quad \leq \frac{\mathbf{1}_{\{e\in E\}}}{N}\mathbb{E}\left[\#\{f : f \notin \mathcal{E}_e\} \mid T_e\right] \\
&\qquad + \frac{\mathbf{1}_{\{e\in E\}}}{N}\mathbb{E}\left[\#\left\{f \in \mathcal{E}_e : Z_{e,f} < \frac{1 - 3a_K}{N}\right\} \mid T_e\right] + 3a_K.
\end{aligned}
$$

The first term in the right-hand side is at most a_K by Lemma 4.4, and the second is $o(1)$ by Corollary 4.7. Therefore,

$$(4.13) \qquad \mathbb{E}\left[d^{(e)}_{\mathrm{TV}}(t_\star) \mid T_e\right]\mathbf{1}_{\{e\in E\}} \leq 4a_K + o(1).$$

Taking expectation over a uniform $e \in \vec{E}$ and using Lemma 4.4 once again, we get

$$\mathbb{E}\left[d^{(\pi)}_{\mathrm{TV}}(t_\star)\right] \leq 5a_K + o(1) < \varepsilon$$

by the choices of a_K and K.

4.1.4. *Lower bound on the mixing time in Theorem* 4.1.   Now consider

$$L' = \lfloor \tilde{\mathbf{d}}^{-1} \log n - 2K'\tilde{\mathbf{d}}^{-3/2}\sqrt{\log n} \rfloor, \qquad \text{where } K' = \sqrt{M_0/\varepsilon}.$$

Define the truncated tree T_e as in Section 4.1.1 (with K', L' replacing K, L). As in (4.9),

$$\mathbb{P}_e(Y_k = e') \geq e^{-k\tilde{\mathbf{d}} - K'\sqrt{L'}} \geq \frac{1}{n} e^{(K' - o(1))\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}}$$

for every $1 \leq k \leq L'$ and, therefore,

$$\#\{e' \in \vec{E}(T_e) : \text{dist}(e, e') = k\} \leq n e^{-(K' - o(1))\tilde{\mathbf{d}}^{-1/2}\sqrt{\log n}} \qquad \text{for each } 1 \leq k \leq L'.$$

Since the coupling argument of Section 3.2 was valid as long as the size of the truncated tree did not exceed $n \exp(-c\sqrt{\log n})$ for some fixed $c > 0$, we can repeat the analysis in the proof of Lemma 4.4 to find that

$$\mathbb{P}\left(\bigcup_{t \leq L'} \{Y_t \notin \vec{E}(T_e)\} \mid Y_0 = e \right) \leq \frac{M_0}{K'^2} + o(1) = \varepsilon + o(1),$$

where the probability is over G, (Y_t). The fact $|\vec{E}(T_e)| = o(n)$ completes the proof.

4.2. *Mixing from a worst starting point with minimum degree* 3.   We now describe how to modify the proof of Theorem 4.1 and derive the proof of Theorem 4.2. Since the lower bound applies also to the worst starting point, it remains to adapt the proof of the upper bound.

Recalling (4.13), we have in fact established not only an upper bound on $\mathbb{E}[d_{\text{TV}}^{(\pi)}(t)]$ a uniform starting edge e, but rather from every edge $e \in \mathcal{E}$, conditional on T_e. Denote by T_e^k the depth-k truncated tree from Section 4.1.1, take $t_\star$ and L as defined in (4.4) (so that T_e^L is T_e from that section), take K as in (4.5) and $R = 2\log_2 \log n$. Further denote by $T_e^{R;L}$ the tree obtained by first exposing T_e^R, then exposing $T_{e_0}^L$ sequentially for each $e_0 \in \partial T_e^R$, while (as before) excluding edges that form cycles/loops.

We will argue that, w.h.p., for *every* $e \in \vec{E}$, the truncated tree T_e^{L+R} satisfies

(4.14) $$P_G^R(e, \tilde{\mathcal{E}}_e) > 1 - 2a_K,$$

where (a_k) is as defined in (4.7) (so that $a_K \asymp \varepsilon$), and

$$\tilde{\mathcal{E}}_e = \{e_0 \in \partial T_e^R : \tilde{Q}_{e_0} \leq a_K\} \qquad \text{for } \tilde{Q}_{e_0} := \mathbb{P}_G\left(\bigcup_{t \leq L} \{Y_t \notin \vec{E}(T_e^{R;L})\} \mid Y_0 = e_0 \right).$$

(That is, the NBRW started at e_0 does not exit $T_e^{R;L}$, which, as opposed to the original definition of $\mathcal{E}$, also precludes it from visiting the subtree $T_{e_1}^L$ for another $e_1 \in \partial T_e^R$.) The proof of (4.13) also yields that

$$\mathbb{E}[d_{\text{TV}}^{(e_0)}(t_\star) \mid T_e^{R;L}]\mathbf{1}_{\{e_0 \in \tilde{E}_e\}} \leq 4a_K + o(1).$$

Hence, the upper bound will follow from (4.14) using the decomposition

$$P_G^{t_\star + R}(e, \cdot) = \sum_{e_0 \in \partial T_e^R} P_G^R(e, e_0) P_G^{t_\star}(e_0, \cdot).$$

To prove (4.14), note that the proof of Lemma 4.4 further implies that

$$\mathbb{P}(e_{j+1} \in \tilde{\mathcal{E}}_e \mid T_e^R, \{T_{e_i}\}_{i=1}^j)$$
$$\geq (1 - a_K) \mathbf{1}_{\{e_{j+1} \notin \bigcup_{i \leq j} T_{e_i}\}} \qquad \text{for all } e_1, \ldots, e_{j+1} \in \partial T_e^R.$$

The indicators of the events $\{e_{j+1} \in \bigcup_{i \leq j} T_{e_j}\}$ $(j = 1, 2, \ldots, |\partial T_e^R|)$ are stochastically dominated by i.i.d. Bernoulli random variables with parameter $n^{-1/2+o(1)}$.

Thus, appealing to Hoeffding–Azuma for the random variable

$$S_e := \sum_{e_j \in \partial T_e^R} P_G^R(e, e_j) \mathbf{1}_{\{e_j \in \tilde{\mathcal{E}}_e\}} = \mathbb{P}_G^R(e, \tilde{\mathcal{E}}_e)$$

[while noting that $P_G^R(e, e_0) \leq 2^{-R} \leq \log^{-2} n$ by the definition of R and the assumption on the minimum degree assumption], we get

$$\mathbb{P}(S_e \leq 1 - 2a_K) \leq \exp(-a_K^2 \log^2 n),$$

which, through a union bound on e, establishes (4.14) and completes the proof.

Acknowledgment. We thank the Theory Group of Microsoft Research, where much of this research was carried out, as well as the Newton Institute. We are grateful to Anna Ben-Hamou and the referees for a careful reading of the paper and many significant comments and corrections on an earlier version. Indeed, one of the referee's insightful comments lead us to redefine the notion of truncated tree, which plays a key role in the proofs.

REFERENCES

[1] ALON, N. and SPENCER, J. H. (2008). *The Probabilistic Method*, 3rd ed. Wiley, Hoboken, NJ. MR2437651

[2] ATHREYA, K. B. and NEY, P. E. (2004). *Branching Processes*. Dover, Mineola, NY. Reprint of the 1972 original [Springer, New York; MR0373040]. MR2047480

[3] BARBOUR, A. D., HOLST, L. and JANSON, S. (1992). *Poisson Approximation. Oxford Studies in Probability* **2**. Clarendon, Oxford. MR1163825

[4] BEN-HAMOU, A. and SALEZ, J. (2017). Cutoff for nonbacktracking random walks on sparse random graphs. *Ann. Probab.* **45** 1752–1770. MR3650414

[5] BENJAMINI, I., KOZMA, G. and WORMALD, N. (2014). The mixing time of the giant component of a random graph. *Random Structures Algorithms* **45** 383–407. MR3252922

[6] BERESTYCKI, N. and DURRETT, R. (2008). Limiting behavior for the distance of a random walk. *Electron. J. Probab.* **13** 374–395. MR2386737

[7] BOLLOBÁS, B. (1984). The evolution of random graphs. *Trans. Amer. Math. Soc.* **286** 257–274. MR0756039

[8] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge University Press, Cambridge. MR1864966

[9] DEMBO, A., GANTERT, N., PERES, Y. and ZEITOUNI, O. (2002). Large deviations for random walks on Galton–Watson trees: Averaging and uncertainty. *Probab. Theory Related Fields* **122** 241–288. MR1894069

[10] DING, J., KIM, J. H., LUBETZKY, E. and PERES, Y. (2011). Anatomy of a young giant component in the random graph. *Random Structures Algorithms* **39** 139–178. MR2850267

[11] DING, J., LUBETZKY, E. and PERES, Y. (2012). Mixing time of near-critical random graphs. *Ann. Probab.* **40** 979–1008. MR2962084

[12] DING, J., LUBETZKY, E. and PERES, Y. (2014). Anatomy of the giant component: The strictly supercritical regime. *European J. Combin.* **35** 155–168. MR3090494

[13] DURRETT, R. (2010). *Random Graph Dynamics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **20**. Cambridge University Press, Cambridge. MR2656427

[14] DWASS, M. (1969). The total progeny in a branching process and a related random walk. *J. Appl. Probab.* **6** 682–686. MR0253433

[15] FOUNTOULAKIS, N. and REED, B. A. (2007). Faster mixing and small bottlenecks. *Probab. Theory Related Fields* **137** 475–486. MR2278465

[16] FOUNTOULAKIS, N. and REED, B. A. (2008). The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures Algorithms* **33** 68–86. MR2428978

[17] GRIMMETT, G. and KESTEN, H. (1984). Random electrical networks on complete graphs II: Proofs. Unpublished manuscript. Available at arXiv:math/0107068.

[18] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs*. Wiley, New York. MR1782847

[19] KENDALL, D. G. (1959). Unitary dilations of Markov transition operators, and the corresponding integral representations for transition-probability matrices. In *Probability and Statistics*: *The Harald Cramér Volume* (U. Grenander, ed.). 139–161. Almqvist & Wiksell, Stockholm; Wiley, New York. MR0116389

[20] LINDVALL, T. (1992). *Lectures on the Coupling Method*. Wiley, New York. A Wiley-Interscience Publication. MR1180522

[21] LOVÁSZ, L. and KANNAN, R. (1999). Faster mixing via average conductance. In *Annual ACM Symposium on Theory of Computing* (*Atlanta*, *GA*, 1999) 282–287. ACM, New York. MR1798047

[22] LUBETZKY, E. and SLY, A. (2010). Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.* **153** 475–510. MR2667423

[23] ŁUCZAK, T. (1990). Component behavior near the critical point of the random graph process. *Random Structures Algorithms* **1** 287–310. MR1099794

[24] LYONS, R., PEMANTLE, R. and PERES, Y. (1995). Ergodic theory on Galton–Watson trees: Speed of random walk and dimension of harmonic measure. *Ergodic Theory Dynam. Systems* **15** 593–619. MR1336708

[25] LYONS, R., PEMANTLE, R. and PERES, Y. (1996). Biased random walks on Galton–Watson trees. *Probab. Theory Related Fields* **106** 249–264. MR1410689

[26] NACHMIAS, A. and PERES, Y. (2008). Critical random graphs: Diameter and mixing time. *Ann. Probab.* **36** 1267–1286. MR2435849

[27] PIAU, D. (1996). Functional limit theorems for the simple random walk on a supercritical Galton–Watson tree. In *Trees* (*Versailles*, 1995) (B. Chauvin, S. Cohen and A. Rouault, eds.). *Progress in Probability* **40** 95–106. Birkhäuser, Basel. MR1439974

[28] RIORDAN, O. and WORMALD, N. (2010). The diameter of sparse random graphs. *Combin. Probab. Comput.* **19** 835–926. MR2726083

N. BERESTYCKI
UNIVERSITY OF CAMBRIDGE
DPMMS
WILBERFORCE RD.
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: n.berestycki@statslab.cam.ac.uk

Y. PERES
MICROSOFT RESEARCH
ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052
USA
E-MAIL: peres@microsoft.com

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

A. SLY
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: sly@stat.berkeley.edu