The Annals of Probability
2017, Vol. 45, No. 3, 1752–1770
DOI: 10.1214/16-AOP1100

CUTOFF FOR NONBACKTRACKING RANDOM WALKS ON SPARSE RANDOM GRAPHS

BY ANNA BEN-HAMOU AND JUSTIN SALEZ

University of Paris Diderot, LPMA

A finite ergodic Markov chain exhibits *cutoff* if its distance to stationarity remains close to 1 over a certain number of iterations and then abruptly drops to near 0 on a much shorter time scale. Discovered in the context of card shuffling (Aldous–Diaconis, 1986), this phenomenon is now believed to be rather typical among fast mixing Markov chains. Yet, establishing it rigorously often requires a challengingly detailed understanding of the underlying chain. Here, we consider nonbacktracking random walks on random graphs with a given degree sequence. Under a general sparsity condition, we establish the cutoff phenomenon, determine its precise window and prove that the cutoff profile approaches a remarkably simple, universal shape.

1. Introduction.

1.1. *Setting*. Given a finite set V and a function $\deg\colon V \to \{2, 3, \dots\}$ such that

$$N := \sum_{v \in V} \deg(v) \tag{1.1}$$

is even, we construct a graph G with vertex set V and degrees $(\deg(v))_{v\in V}$ as follows. We form a set $\mathcal{X}$ by "attaching" $\deg(v)$ *half-edges* to each vertex $v \in V$:

$$\mathcal{X} := \{(v, i)\colon v \in V, 1 \le i \le \deg(v)\}.$$

We then simply choose a *pairing* π on $\mathcal{X}$ (i.e., an involution without fixed points), and interpret every pair of matched half-edges $\{x, \pi(x)\}$ as an edge between the corresponding vertices; loops and multiple edges are allowed (see Figure 1).

The *nonbacktracking random walk* (NBRW) on the graph $G = G(\pi)$ is a discrete-time Markov chain with state space $\mathcal{X}$ and transition matrix

$$P(x, y) = \begin{cases} \dfrac{1}{\deg(\pi(x))}, & \text{if y is a neighbour of } \pi(x), \\ 0, & \text{otherwise.} \end{cases}$$

In this definition and throughout the paper, two half-edges $x = (u, i)$ and $y = (v, j)$ are called *neighbours* if $u = v$ and $i \ne j$, and we let $\deg(x) := \deg(u) - 1$

Received April 2015; revised January 2016.
MSC2010 subject classifications. 60J10, 60G50, 05C80, 05C81.
Key words and phrases. Cutoff phenomenon, nonbacktracking random walks, random graphs.

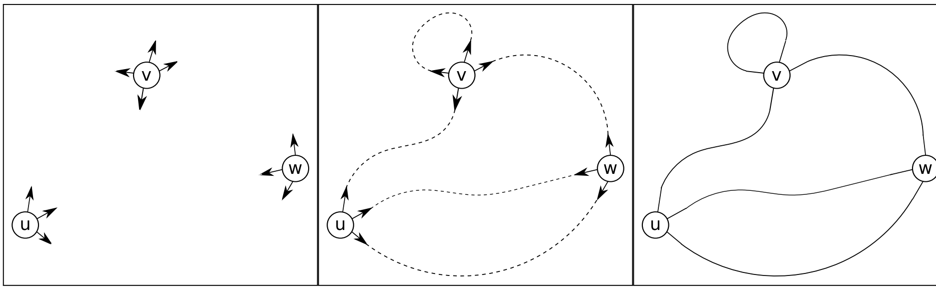

FIG. 1. *A set of half-edges $\mathcal{X}$, a pairing π and the resulting graph G.*

denote the number of neighbours of the half-edge $x = (u, i)$. In words, the chain moves at every step from the current state x to a uniformly chosen neighbour of $\pi(x)$ (see Figure 2).

Note that the matrix P is symmetric with respect to π: for all $x, y \in \mathcal{X}$,

$$P\big(\pi(y), \pi(x)\big) = P(x, y). \tag{1.2}$$

In particular, P is doubly stochastic: the uniform law on $\mathcal{X}$ is invariant for the chain. The *worst-case total-variation distance to equilibrium* at time $t \in \mathbb{N}$ is

$$\mathcal{D}(t) := \max_{x\in\mathcal{X}} \mathcal{D}_x(t) \qquad \text{where } \mathcal{D}_x(t) := \frac{1}{2}\sum_{y\in\mathcal{X}} \left| P^t(x, y) - \frac{1}{N} \right|. \tag{1.3}$$

This quantity is nonincreasing in t, and the number of transitions that have to be made before it falls below a given threshold $0 < \varepsilon < 1$ is the *mixing time*:

$$t_{\mathrm{MIX}}(\varepsilon) := \inf\{t \in \mathbb{N} : \mathcal{D}(t) < \varepsilon\}.$$

1.2. *Result.* The present paper is concerned with the typical profile of the function $t \mapsto \mathcal{D}(t)$ under the so-called *configuration model* (see, e.g., [35]), that is, when the pairing π is chosen uniformly at random among the $(N - 1)!!$ possible pairings on $\mathcal{X}$. In order to study large-size asymptotics, we let the vertex set V and degree function $\deg : V \to \mathbb{N}$ depend on an implicit parameter $n \in \mathbb{N}$, which we omit from the notation for convenience. The same convention applies to all related quantities, such as N or $\mathcal{X}$. All asymptotic statements are understood as

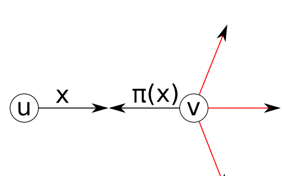

FIG. 2. *The nonbacktracking moves from x (in red).*

$n \to \infty$. Our interest is in the *sparse regime*, where the number N of half-edges diverges at a much faster rate than the maximum degree. Specifically, we assume that

$$\Delta := \max_{v \in V} \deg(v) = N^{o(1)}. \tag{1.4}$$

As the behaviour of the NBRW at degree-2 vertices is deterministic, we assume that

$$\min_{v \in V} \deg(v) \geq 3. \tag{1.5}$$

Remarkably enough, the asymptotics in this regime depends on the degrees through two simple statistics: the mean logarithmic degree of an half-edge,

$$\mu := \frac{1}{N} \sum_{v \in V} \deg(v) \log\bigl(\deg(v) - 1\bigr), \tag{1.6}$$

and the corresponding variance

$$\sigma^2 := \frac{1}{N} \sum_{v \in V} \deg(v) \bigl\{\log\bigl(\deg(v) - 1\bigr) - \mu\bigr\}^2. \tag{1.7}$$

We will also need some control on the third absolute moment:

$$\varrho := \frac{1}{N} \sum_{v \in V} \deg(v) \bigl|\log\bigl(\deg(v) - 1\bigr) - \mu\bigr|^3. \tag{1.8}$$

It might help the reader to think of μ, σ and ϱ as being fixed, or bounded away from 0 and ∞. However, we only impose the following (much weaker) condition:

$$\frac{\sigma^2}{\mu^3} \gg \frac{(\log\log N)^2}{\log N} \quad \text{and} \quad \frac{\sigma^3}{\varrho\sqrt{\mu}} \gg \frac{1}{\sqrt{\log N}}. \tag{1.9}$$

Our main result states that on most graphs with degrees $(\deg(v))_{v \in V}$, the NBRW exhibits a remarkable behaviour, visible on Figure 3 and known as a *cutoff*: the distance to equilibrium remains close to 1 for a rather long time, roughly

$$t_\star := \frac{\log N}{\mu}, \tag{1.10}$$

and then abruptly drops to nearly 0 over a much shorter time scale,[1] of order

$$\omega_\star := \sqrt{\frac{\sigma^2 \log N}{\mu^3}}. \tag{1.11}$$

Moreover, the cutoff shape inside this window approaches a surprisingly simple function $\Phi \colon \mathbb{R} \to [0, 1]$, namely the tail distribution of the standard normal:

$$\Phi(\lambda) := \frac{1}{2\pi} \int_\lambda^\infty e^{-\frac{u^2}{2}}\, \mathrm{d}u.$$

It is remarkable that this limit shape does not depend at all on the precise degrees.

[1]The fact that $\omega_\star \ll t_\star$ follows from condition (1.4).

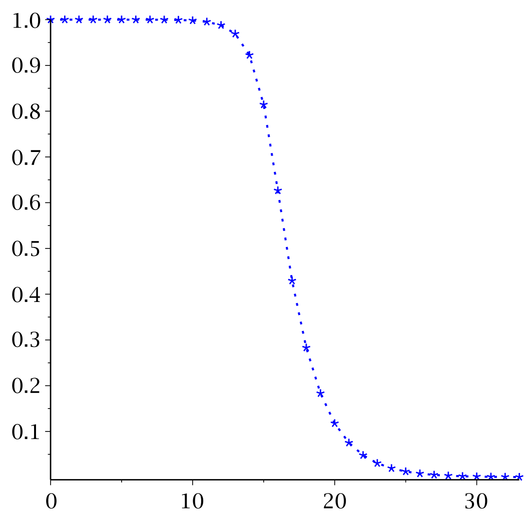

FIG. 3. *Distance to stationarity along time for the* NBRW *on a random graph with* 10^6 *degree* 3-*vertices and* 10^6 *degree* 4-*vertices.*

THEOREM 1.1 (Cutoff for the NBRW on sparse graphs). *For every* $0 < \varepsilon < 1$,

$$\frac{t_{\mathrm{MIX}}(\varepsilon) - t_\star}{\omega_\star} \xrightarrow{\mathbb{P}} \Phi^{-1}(\varepsilon).$$

Equivalently, for $t = t_\star + \lambda\omega_\star + o(w_\star)$ *with* $\lambda \in \mathbb{R}$ *fixed, we have* $\mathcal{D}(t) \xrightarrow{\mathbb{P}} \Phi(\lambda)$.

1.3. *Comments.* It is interesting to compare this with the d-regular case (i.e., $\deg\colon V \to \mathbb{N}$ constant equal to d) studied by Lubetzky and Sly [27]: by a remarkably precise path counting argument, they establish cutoff within constantly many steps around $t_\star = \log N/\log(d-1)$. To appreciate the effect of heterogeneous degrees, recall that μ and σ are the mean and variance of $\log D$, where D is the degree of a uniformly sampled half-edge. Now, by Jensen's inequality,

$$t_\star \geq \frac{\log N}{\log \mathbb{E}[D]},$$

and the less concentrated D, the larger the gap. The right-hand side is a well-known characteristic length in G, namely the typical inter-point distance (see, e.g., [36]). One notable effect of heterogeneous degrees is thus that the mixing time becomes significantly larger than the natural graph distance. A heuristic explanation is as follows: in the regular case, all paths of length t between two points are equally likely for the NBRW, and mixing occurs as soon as t is large enough for many such paths to exist. In the nonregular case, however, different paths have very different weights, and most of them actually have a negligible chance of being seen by the walk. Consequently, one has to make t larger in order to see paths with a "reasonable" weight. Even more remarkable is the impact of heterogeneous degrees on the cutoff width $\omega_\star$, which satisfies $\omega_\star \gg \log\log N$ against $\omega_\star = \Theta(1)$ in the regular case. Finally, the Gaussian limit shape Φ itself is specific to the nonregular case and is directly related to the fluctuations of degrees along a typical trajectory of the NBRW.

REMARK 1.1 (Simple graphs). A classical result by Janson [23] asserts that the graph produced by the configuration model is simple (no loops or multiple edges) with probability asymptotically bounded away from 0, as long as

$$\sum_{v\in V} \deg(v)^2 = O(N). \tag{1.12}$$

Moreover, conditionally on being simple, it is uniformly distributed over all simple graphs with degrees $(\deg(v))_{v\in V}$. Thus, every property which holds w.h.p. under the configuration model also holds w.h.p. for the uniform simple graph model. In particular, under (1.12), the conclusion of Theorem 1.1 extends to simple graphs.

REMARK 1.2 (i.i.d. degrees). A common setting consists in generating an infinite i.i.d. degree sequence $(\deg(v))_{v\in\mathbb{N}}$ from some fixed degree distribution Q and then restricting it to the index set $V = \{1,\dots,n\}$ for each $n \geq 1$. Let D denote a random integer with distribution Q. Assuming that

$$\mathbb{P}(D \leq 2) = 0, \qquad \operatorname{Var}(D) > 0 \quad \text{and} \quad \mathbb{E}\big[e^{\theta D}\big] < \infty \qquad \text{for some } \theta > 0,$$

ensures that the conditions (1.4), (1.5) and (1.9) hold almost surely. Thus, Theorem 1.1 applies with the parameters μ, σ and N now being random. But the latter clearly concentrate around their deterministic counterparts, in the following sense:

$$\begin{aligned}
N &= n\mathbb{E}[D] + O_{\mathbb{P}}\big(n^{\frac{1}{2}}\big),\\
\mu &= \mu_\star + O_{\mathbb{P}}\big(n^{-\frac{1}{2}}\big) \qquad \text{with } \mu_\star = \mathbb{E}\big[D\log(D-1)\big]/\mathbb{E}[D],\\
\sigma &= \sigma_\star + O_{\mathbb{P}}\big(n^{-\frac{1}{2}}\big) \qquad \text{with } \sigma_\star^2 = \mathbb{E}\big[D\big\{\log(D-1) - \mu_\star\big\}^2\big]/\mathbb{E}[D].
\end{aligned}$$

Those error terms are small enough to allow one to substitute $n, \mu_\star, \sigma_\star$ for N, μ, σ without affecting the convergence stated in Theorem 1.1.

1.4. *Related work.* The first instances of the cutoff phenomenon were discovered in the early 1980s by Diaconis and Shahshahani [15] and Aldous [2], in the context of card shuffling: given a certain procedure for shuffling a deck of cards, there exists a quite precise number of shuffles slightly below which the deck is far from being mixed, and slightly above which it is almost completely mixed. The term *cutoff* and the general formalization appeared shortly after, in the seminal paper by Aldous and Diaconis [3]. Since then, this remarkable behaviour has been identified in a variety of other contexts; see, for example, Diaconis [14], Chen and

Saloff-Coste [10], or the survey by Saloff-Coste [34] for random walks on finite groups.

Interacting particle systems in statistical mechanics provide a rich class of dynamics displaying cutoff. One emblematic example is the *stochastic Ising model* at high enough temperature, for which the cutoff phenomenon has been established successively on the complete graph (Levin, Luczak and Peres [25]), on lattices (Ding, Lubetzky and Peres [16], Lubetzky and Sly [28]), and finally on any sequence of graphs (Lubetzky and Sly [29]). Other examples include the *Potts model* (Cuff *et al.* [13]), the *East process* (Ganguly, Lubetzky and Martinelli [20]) or the *Simple Exclusion process* on the cycle (Lacoin [24]).

The problem of singling out abstract conditions under which the cutoff phenomenon occurs, without necessarily pinpointing its precise location, has drawn considerable attention. In 2004, Peres [31] proposed a simple spectral criterion for reversible chains, known as the *product condition*. Although counter-examples have quickly been constructed (see Levin, Peres and Wilmer [26], Chapter 18, and Chen and Saloff-Coste [10], Section 6), the condition is widely believed to be sufficient for "most" chains. This has already been verified for birth-and-death chains (Ding, Lubetzky and Peres [17]) and, more generally, for random walks on trees (Basu, Hermon and Peres [5]). The latter result relies on a promising characterization of cutoff in terms of the concentration of hitting times of "worst" (in some sense) sets. See also Oliveira [30], Peres and Sousi [32], Griffiths *et al.* [21] and Hermon [22].

Many natural families of Markov chains are now believed to exhibit cutoff. Yet, establishing this phenomenon rigorously requires a very detailed understanding of the underlying chain, and often constitutes a challenging task even in situations with a high degree of symmetry. The historical case of random walks on the symmetric group, for example, is still far from being completely understood: see Saloff-Coste [34] for a list of open problems, and Berestycki and Sengul [7] for a recent proof of one of them.

Understanding the mixing properties of random walks on sparse random graphs constitutes an important theoretical problem, with applications in a wide variety of contexts (see, e.g., the survey by Cooper [11]). A classical result of Broder and Shamir [8] states that random d-regular graphs with d fixed are *expanders* with high probability (see also Friedman [19]). In particular, the simple random walk (SRW) on such graphs satisfies the product condition, and should therefore exhibit cutoff. This long-standing conjecture was confirmed only recently in an impressive work by Lubetzky and Sly [27], who also determined the precise cutoff window and profile. Their result is actually derived from the analysis of the NBRW itself, via a clever transfer argument. Interestingly, the mixing time of the SRW is $d/(d-2)$ times larger than that of the NBRW. This confirms the practical advantage of NBRW over SRW for efficient network sampling and exploration, and complements a well-known spectral comparison for regular expanders due to Alon *et al.* [4], as well as a recent result by Cooper and Frieze [12] on the cover time of random regular graphs. For other ways of speeding up random walks, see Cooper [11].

In the nonregular case, however, the tight correspondence between the SRW and the NBRW breaks down, and there seems to be no direct way of transferring our main result to the SRW. We note that the latter should exhibit cutoff since the product condition holds, as can be seen from the fact that the *conductance* of sparse random graphs with a given degree sequence is bounded away from 0 (see Abdullah, Cooper and Frieze [1]). Confirming this is a challenging open problem. In particular, it would be interesting to see whether the NBRW still mixes faster than the SRW.

REMARK 1.3. After completing this work, we learned that the cutoff phenomenon for the SRW on random graphs with given degrees (starting from a typical vertex) has finally been established by Berestycki *et al.* [6]. Their paper also contains a weaker version of our result on the NBRW, namely that $t_{\mathrm{MIX}}(\varepsilon) = t_\star + O_{\mathbb{P}}(\sqrt{\log N})$ (under a more restrictive assumption on Δ).

2. Proof outline. The proof of Theorem 1.1 is divided into two (unequal) halves: for

$$t = t_\star + \lambda w_\star + o(w_\star), \tag{2.1}$$

with $\lambda \in \mathbb{R}$ fixed, we show that

$$\min_{x\in\mathcal{X}} \mathbb{E}\big[\mathcal{D}_x(t)\big] \ge \Phi(\lambda) - o(1), \tag{2.2}$$

$$\max_{x\in\mathcal{X}} \mathcal{D}_x(t) \le \Phi(\lambda) + o_{\mathbb{P}}(1). \tag{2.3}$$

Note that this actually shows that the maximization over all possible states in (1.3) is irrelevant. The lower bound (2.2) is proved in Section 3. The difficult part is the upper bound (2.3), due to the worst-case maximization: our approximations for a given initial state $x \in \mathcal{X}$ need to be valid with probability $1 - o(1/N)$, so that we may then take union bound. Our starting point is the key identity

$$P^t\big(x, \pi(y)\big) = \sum_{(u,v)\in\mathcal{X}\times\mathcal{X}} P^{t/2}(x,u)P^{t/2}(y,v)\mathbb{1}_{\{\pi(u)=v\}}, \tag{2.4}$$

which follows from the symmetry (1.2). As a first approximation, let us assume that the balls of radius $t/2$ around x and y consist of disjoint trees, as in Figure 4.

This is made rigorous by a particular exposure process described in Section 4. Then the *weight* $\mathbf{w}(u) := P^{t/2}(x,u)$ [resp., $\mathbf{w}(v) := P^{t/2}(y,v)$] can be unambiguously written as the inverse product of degrees along the unique path from x to u (resp., y to v). A second approximation consists in eliminating those paths whose weight exceeds some threshold $\theta > 0$ (the correct choice turns out to be $\theta \approx \frac{1}{N}$):

$$P^t\big(x, \pi(y)\big) \approx \sum_{u,v} \mathbf{w}(u)\mathbf{w}(v)\mathbb{1}_{(\mathbf{w}(u)\mathbf{w}(v)\le\theta)}\mathbb{1}_{(\pi(u)=v)}.$$

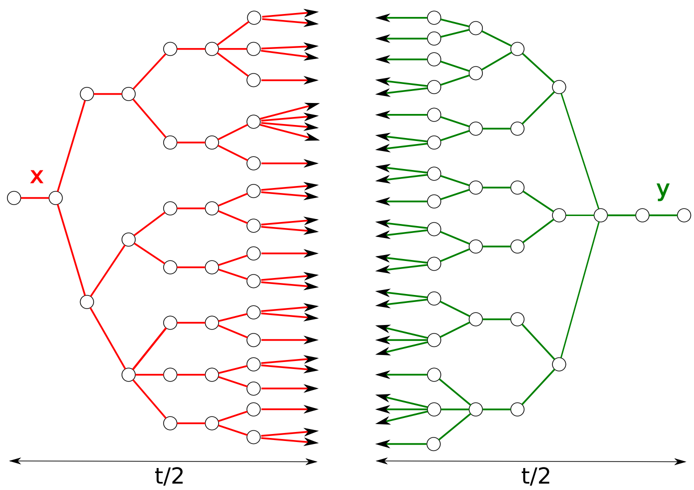

FIG. 4. *The tree-approximation.*

Conditionally on the two trees of height $t/2$, this is a weighted sum of weakly dependent Bernoulli variables, and the large-weight truncation should prevent it from deviating largely from its expectation. We make this argument rigorous in Section 5, using Stein's method of exchangeable pairs. Provided the exposure process did not reveal too many pairs of matched half-edges, the conditional expectation of $\mathbb{1}_{(\pi(u)=v)}$ remains close to $1/N$, and we obtain the new approximation

$$NP^t\big(x,\pi(y)\big) \approx \sum_{u,v} \mathbf{w}(u)\mathbf{w}(v)\mathbb{1}_{(\mathbf{w}(u)\mathbf{w}(v)\leq\theta)}.$$

Now, the right-hand side corresponds to the quenched probability that the product of weights seen by two independent NBRWs of length $t/2$, one starting from x and the other from y, does not exceed θ. The last step consists in approximating those trajectories by independent uniform samples $X_1^\star, \dots, X_t^\star$ from $\mathcal{X}$:

$$\begin{aligned}\sum_{u,v} \mathbf{w}(u)\mathbf{w}(v)\mathbb{1}_{\mathbf{w}(u)\mathbf{w}(v)\leq\theta} &\approx \mathbb{P}\left[\frac{1}{\deg(X_1^\star)}\cdots\frac{1}{\deg(X_t^\star)} \leq \theta\right] \\ &\approx \mathbb{P}\left[\frac{\sum_{k=1}^t(\mu - \log\deg(X_k^\star))}{\sigma\sqrt{t}} \leq \frac{\mu t + \log\theta}{\sigma\sqrt{t}}\right] \\ &\approx 1 - \Phi(\lambda),\end{aligned}$$

by the central limit theorem (recall that $\theta \approx 1/N$ and $t \approx t_\star + \lambda\omega_\star$). Consequently,

$$\mathcal{D}_x(t) = \sum_y \left(\frac{1}{N} - P^t\big(x,\pi(y)\big)\right)_+ \approx \Phi(\lambda),$$

as desired. This argument is made rigorous in Sections 6, 7 and 8.

3. The lower bound. Fix $t \geq 1$, two states $x, y \in \mathcal{X}$, and a parameter $\theta \in (0, 1)$. Let $P_\theta^t(x, y)$ denote the contribution to $P^t(x, y)$ from paths having weight less than θ. Note that $P_\theta^t(x, y) < P^t(x, y)$ if and only if some path of length t from x to y has weight larger than θ, implying in particular that $P^t(x, y) > \theta$. Thus,

$$\frac{1}{N} - P_\theta^t(x, y) \leq \left(\frac{1}{N} - P^t(x, y)\right)_+ + \frac{1}{N}\mathbf{1}_{P^t(x,y)>\theta}.$$

Summing over all $y \in \mathcal{X}$ and observing that there cannot be more than $1/\theta$ half-edges $y \in \mathcal{X}$ satisfying $P^t(x, y) > \theta$, we obtain

$$1 - \sum_{y \in \mathcal{X}} P_\theta^t(x, y) \leq \mathcal{D}_x(t) + \frac{1}{\theta N}.$$

Now, the left-hand side is the quenched probability (i.e., conditional on the pairing π) that a NBRW $\{X_k\}_{0 \leq k \leq t}$ on $G(\pi)$ starting at x satisfies $\prod_{k=1}^t \frac{1}{\deg(X_k)} > \theta$. Taking expectation w.r.t. the pairing, we arrive at

$$\mathbb{P}\left(\prod_{k=1}^t \frac{1}{\deg(X_k)} > \theta\right) \leq \mathbb{E}[\mathcal{D}_x(t)] + \frac{1}{\theta N}, \tag{3.1}$$

where the average is now taken over both the NBRW and the pairing (annealed law). A useful property of the uniform pairing is that it can be constructed sequentially, the pairs being revealed along the way, as we need them. We exploit this degree of freedom to generate the walk $\{X_k\}_{k \geq 0}$ and the pairing simultaneously, as follows. Initially, all half-edges are unpaired and $X_0 = x$; then at each time $k \geq 1$:

(i) if X_{k-1} is unpaired, we pair it with a uniformly chosen other unpaired half-edge; otherwise, $\pi(X_{k-1})$ is already defined and no new pair is formed.
(ii) in both cases, we let X_k be a uniformly chosen neighbour of $\pi(X_{k-1})$.

The sequence $\{X_k\}_{k \geq 0}$ is then exactly distributed according to the annealed law. Now, if we sample uniformly from $\mathcal{X}$ instead of restricting the random choice made at (i) to unpaired half-edges, then the uniform neighbour chosen at step (ii) also has the uniform law on $\mathcal{X}$. This creates a coupling between the process $\{X_k\}_{k \geq 1}$ and a sequence $\{X_k^\star\}_{k \geq 1}$ of i.i.d. samples from $\mathcal{X}$, valid until the first time T where the uniformly chosen half-edge or its uniformly chosen neighbour is already paired. As there are less than $2k$ paired half-edges by step k, a crude union-bound yields

$$\mathbb{P}(T \leq t) \leq \frac{2t^2}{N}.$$

Consequently,

$$\left|\mathbb{P}\left(\prod_{k=1}^t \frac{1}{\deg(X_k)} > \theta\right) - \mathbb{P}\left(\prod_{k=1}^t \frac{1}{\deg(X_k^\star)} > \theta\right)\right| \leq \frac{2t^2}{N}. \tag{3.2}$$

On the other hand, since $\{X_1^\star, \dots, X_t^\star\}$ are i.i.d., Berry–Esseen's inequality implies

$$\left|\mathbb{P}\left(\prod_{k=1}^{t} \frac{1}{\deg(X_k^\star)} > \theta\right) - \Phi\left(\frac{\mu t + \log\theta}{\sigma\sqrt{t}}\right)\right| \le \frac{\varrho}{\sigma^3\sqrt{t}}. \tag{3.3}$$

We may now combine (3.1), (3.2), (3.3) to obtain

$$\mathbb{E}[\mathcal{D}_x(t)] \ge \Phi\left(\frac{\mu t + \log\theta}{\sigma\sqrt{t}}\right) - \frac{1}{\theta N} - \frac{2t^2}{N} - \frac{\varrho}{\sigma^3\sqrt{t}}.$$

With t as in (2.1) and $\theta = (\log N)/N$, the right-hand side is $\Phi(\lambda) + o(1)$, thanks to our assumptions on μ, σ, ϱ. This establishes the lower bound (2.2).

4. The upper-bound. Following Lubetzky and Sly [27], we call $x \in \mathcal{X}$ a *root* (written $x \in \mathcal{R}$) if the (directed) ball of radius r centered at x (denoted by $\mathcal{B}_x$) is a tree, where

$$r := \left\lfloor \frac{\log N}{10\log\Delta} \wedge \log\log N \right\rfloor. \tag{4.1}$$

Note that $1 \ll r \ll \omega_\star$ by assumptions (1.4) and (1.9). The first proposition below shows that we may restrict our attention to paths between roots. The second proposition provides a good control on such paths.

PROPOSITION 4.1 (Roots are quickly reached).

$$\max_{x\in\mathcal{X}} P^r(x, \mathcal{X}\setminus\mathcal{R}) \xrightarrow{\mathbb{P}} 0.$$

PROPOSITION 4.2 (Roots are well inter-connected). *For t as in* (2.1),

$$\min_{x\in\mathcal{R}} \min_{y\in\mathcal{R}\setminus\mathcal{B}_x} P^t\big(x, \pi(y)\big) \ge \frac{1 - \Phi(\lambda) - o_{\mathbb{P}}(1)}{N}.$$

Let us first see how those results imply the upper-bound (2.3). Observe that

$$\mathcal{D}(t + r) \le \max_{x\in\mathcal{X}} P^r(x, \mathcal{X}\setminus\mathcal{R}) + \max_{x\in\mathcal{R}} \mathcal{D}_x(t).$$

The first term is $o_{\mathbb{P}}(1)$ by Proposition 4.1. For the second one, we write

$$\mathcal{D}_x(t) = \sum_{y\in\mathcal{R}\setminus\mathcal{B}_x} \left(\frac{1}{N} - P^t\big(x, \pi(y)\big)\right)_+ + \sum_{y\in\mathcal{B}_x\cup(\mathcal{X}\setminus\mathcal{R})} \left(\frac{1}{N} - P^t\big(x, \pi(y)\big)\right)_+.$$

Proposition 4.2 ensures that the first sum is bounded by $\Phi(\lambda) + o_{\mathbb{P}}(1)$ uniformly in $x \in \mathcal{R}$. To see that the second sum is $o_{\mathbb{P}}(1)$ uniformly in $x \in \mathcal{R}$, it suffices to bound its summands by $1/N$ and observe that $|\mathcal{B}_x| \le \Delta^r = o(N)$ by (4.1), while

$$|\mathcal{X}\setminus\mathcal{R}| = \sum_{x\in\mathcal{X}} P^r(x, \mathcal{X}\setminus\mathcal{R}) = o_{\mathbb{P}}(N).$$

(The first equality because P is doubly stochastic, the second by Proposition 4.1.)

PROOF OF PROPOSITION 4.1. Define $R := \lfloor \frac{\log N}{5 \log \Delta} \rfloor$ and fix $x \in \mathcal{X}$. The ball of radius R around x can be generated sequentially, its half-edges being paired one after the other with uniformly chosen other unpaired half-edges, until the whole ball has been paired. Observe that at most $k = \frac{\Delta((\Delta-1)^R-1)}{\Delta-2}$ pairs are formed. Moreover, for each of them, the number of unpaired half-edges having an already paired neighbour is at most $\Delta(\Delta-1)^R$ and hence the conditional chance of hitting such a half-edge (thereby creating a cycle) is at most $p = \frac{\Delta(\Delta-1)^R-1}{N-2k-1}$. Thus, the probability that more than one cycle is found is at most

$$(kp)^2 = O\left(\frac{(\Delta-1)^{4R}}{N^2}\right) = o\left(\frac{1}{N}\right).$$

Summing over all $x \in \mathcal{X}$ (union bound), we obtain that with high probability, no ball of radius R in $G(\pi)$ contains more than one cycle.

To conclude the proof, we now fix a pairing π with the above property, and we prove that the NBRW on $G(\pi)$ starting from any $x \in \mathcal{X}$ satisfies

$$\mathbb{P}(X_t \text{ is not a root}) \le 2^{1-t}, \tag{4.2}$$

for all $t \le R - r$. The claim is trivial if the ball of radius R around x is acyclic. Otherwise, it contains a single cycle $\mathcal{C}$, by assumption. Write $d(x, \mathcal{C})$ for the minimum length of a nonbacktracking path from x to some $z \in \mathcal{C}$. The nonbacktracking property ensures that if $d(X_t, \mathcal{C}) < d(X_{t+1}, \mathcal{C})$ for some $t < R - r$, then $X_{t+1}, X_{t+2}, \dots, X_{R-r}$ are all roots. Indeed, as soon as the NBRW makes a step away from $\mathcal{C}$ on one of the disjoint trees rooted to $\mathcal{C}$, it can only go further away from it. By (1.5), the conditional chance that $d(X_{t+1}, \mathcal{C}) = d(X_t, \mathcal{C}) + 1$ given the past is at least $1/2$ [unless $d(X_t, \mathcal{C}) = 1$, which can only happen once]. This shows (4.2). We then specialize to $t = r$. $\square$

The remainder of the paper is devoted to the proof of Proposition 4.2. By union bound, it is enough to fix two distinct half-edges $x, y \in \mathcal{X}$ and establish that, for every $\varepsilon > 0$,

$$\mathbb{P}\left(x \in \mathcal{R}, y \in \mathcal{R} \setminus \mathcal{B}_x, P^t(x, \pi(y)) \le \frac{1 - \Phi(\lambda) - \varepsilon}{N}\right) = o\left(\frac{1}{N^2}\right). \tag{4.3}$$

To do so, we shall analyse a special procedure that generates a uniform pairing on $\mathcal{X}$ together with a two-tree forest $\mathfrak{F}$ keeping track of certain paths from x and from y. Initially, all half-edges are unpaired and $\mathfrak{F}$ is reduced to its two ancestors, x and y. We then iterate the following three steps:

1. An unpaired half-edge $z \in \mathfrak{F}$ is selected according to some rule (see below).
2. z is paired with a uniformly chosen other unpaired half-edge z'.

3. If neither z' nor any of its neighbours was already in $\mathfrak{F}$, then all neighbours of z' become children of z in the forest $\mathfrak{F}$.

The *exploration stage* stops when no unpaired half-edge is compatible with the selection rule. We then complete the pairing π by matching all the remaining unpaired half-edges uniformly at random: this is the *completion stage*.

The condition in step 3 ensures that $\mathfrak{F}$ remains a forest: any $z \in \mathfrak{F}$ determines a unique sequence $(z_0, \dots, z_h)$ in $\mathfrak{F}$ such that $z_0 \in \{x, y\}$, z_i is a child of z_{i-1} for each $1 \le i \le h$, and $z_h = z$. We shall naturally refer to h and z_0 as the *height* and *ancestor* of z, respectively. We also define the *weight* of z as

$$\mathbf{w}(z) := \prod_{i=1}^{h} \frac{1}{\deg(z_i)}.$$

Note that this quantity is the quenched probability that the sequence $(z_0, \dots, z_h)$ is realized by a NBRW on G starting from z_0. In particular,

$$\mathbf{w}(z) \le P^h(z_0, z). \tag{4.4}$$

Our rule for step 1 consists in selecting a half-edge with maximal weight[2] among all unpaired $z \in \mathfrak{F}$ with height $\mathbf{h}(z) < t/2$ and weight $\mathbf{w}(z) > w_{\mathrm{MIN}}$, where

$$w_{\mathrm{MIN}} := N^{-\frac{2}{3}}.$$

The role of this parameter is to limit the number of pairs formed at the exploration stage; see (8.1) below. As outlined in Section 2, we shall be interested in

$$\mathfrak{W} := \sum_{(u,v) \in \mathcal{H}_x \times \mathcal{H}_y} \mathbf{w}(u)\mathbf{w}(v) \mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) \le \theta\}},$$

where $\mathcal{H}_x$ (resp., $\mathcal{H}_y$) denotes the set of unpaired half-edges with height $\frac{t}{2}$ and ancestor x (resp., y) in $\mathfrak{F}$ at the end of the exploration stage, and where

$$\theta := \frac{1}{N(\log N)^2}. \tag{4.5}$$

Write $\overline{\mathfrak{W}}$ for the quantity obtained by replacing $\le$ with $>$ in $\mathfrak{W}$, so that

$$\mathfrak{W} + \overline{\mathfrak{W}} = \sum_{(u,v) \in \mathcal{H}_x \times \mathcal{H}_y} \mathbf{w}(u)\mathbf{w}(v) \ge \sum_{z \in \mathcal{H}_x \cup \mathcal{H}_y} \mathbf{w}(z) - 1,$$

thanks to the inequality $ab \ge a + b - 1$ for $a, b \in [0, 1]$. Now, let $\mathfrak{U}$ denote the set of unpaired half-edges in $\mathfrak{F}$. By construction, at the end of the exploration stage, each $z \in \mathfrak{U}$ must have height $t/2$ or weight less than w_{MIN}, so that

$$\sum_{z \in \mathcal{H}_x \cup \mathcal{H}_y} \mathbf{w}(z) \ge \sum_{z \in \mathfrak{U}} \mathbf{w}(z) - \sum_{z \in \mathfrak{F}} \mathbf{w}(z) \mathbb{1}_{\{\mathbf{w}(z) < w_{\mathrm{MIN}}\}}.$$

Therefore, (4.3) is a consequence of the following four technical lemmas.

---

[2]For definiteness, let us say that we use the lexicographic order on $\mathcal{X}$ to break ties.

LEMMA 4.3. *For every* $\varepsilon > 0$,

$$\mathbb{P}\left(P^t(x, \pi(y)) \leq \frac{\mathfrak{W} - \varepsilon}{N}\right) = o\left(\frac{1}{N^2}\right).$$

LEMMA 4.4. *For every* $\varepsilon > 0$,

$$\mathbb{P}\left(\sum_{z \in \mathfrak{F}} \mathbf{w}(z) \mathbb{1}_{\{\mathbf{w}(z) < w_{\mathrm{MIN}}\}} > \varepsilon\right) = o\left(\frac{1}{N^2}\right).$$

LEMMA 4.5. *For every* $\varepsilon > 0$,

$$\mathbb{P}(\overline{\mathfrak{W}} > \Phi(\lambda) + \varepsilon) = o\left(\frac{1}{N^2}\right).$$

LEMMA 4.6. *For every* $\varepsilon > 0$,

$$\mathbb{P}\left(x \in \mathcal{R}, y \in \mathcal{R} \setminus \mathcal{B}_x, \sum_{z \in \mathfrak{U}} \mathbf{w}(z) < 2 - \varepsilon\right) = o\left(\frac{1}{N^2}\right).$$

5. Proof of Lemma 4.3. Combining the representation (2.4) with the observation (4.4) yields

$$P^t(x, \pi(y)) \geq \sum_{(u,v) \in \mathcal{H}_x \times \mathcal{H}_y} \mathbf{w}(u)\mathbf{w}(v) \mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) \leq \theta\}} \mathbb{1}_{\{\pi(u) = v\}}.$$

The right-hand side can be interpreted as the weight of the uniform pairing chosen at the completion stage, provided we define the weight of a pair (u, v) as

$$\mathbf{w}(u)\mathbf{w}(v) \mathbb{1}_{\{u \in \mathcal{H}_x\}} \mathbb{1}_{\{v \in \mathcal{H}_y\}} \mathbb{1}_{\{\mathbf{w}(u)\mathbf{w}(v) \leq \theta\}}. \tag{5.1}$$

With this interpretation, Lemma 4.3 becomes a special case of the following concentration inequality [which we apply conditionally on the exploration stage, with $\mathcal{I}$ being the set of half-edges that did not get paired, and weights given by (5.1)].

LEMMA 5.1. *Let* $\mathcal{I}$ *be an even set*, $\{w_{i,j}\}_{(i,j) \in \mathcal{I} \times \mathcal{I}}$ *an array of nonnegative weights*, *and* π *a uniform random pairing on* $\mathcal{I}$. *Then for all* $a > 0$,

$$\mathbb{P}\left(\sum_{i \in \mathcal{I}} w_{i,\pi(i)} \leq m - a\right) \leq \exp\left\{-\frac{a^2}{4\theta m}\right\},$$

where $m = \frac{1}{|\mathcal{I}|-1} \sum_{i \in \mathcal{I}} \sum_{j \neq i} w_{i,j}$ *and* $\theta = \max_{i \neq j}(w_{i,j} + w_{j,i})$.

Note that in our case, $m = \frac{\mathfrak{W}}{|\mathcal{I}|-1}$. Lemma 4.3 follows easily by taking $a = \frac{\varepsilon}{|\mathcal{I}|-1}$ and observing that $|\mathcal{I}| - 1 \leq N$ and $\mathfrak{W} \leq 1$.

PROOF OF LEMMA 5.1. We exploit the following concentration result for Stein pairs due to Chatterjee [9] (see also Ross [33], Theorem 7.4): let Y, Y' be bounded variables satisfying:

(i) $(Y, Y') \stackrel{d}{=} (Y', Y)$;
(ii) $\mathbb{E}[Y' - Y|Y] = -\lambda Y$;
(iii) $\mathbb{E}[(Y' - Y)^2|Y] \le \lambda(bY + c)$,

for some constants $\lambda \in (0, 1)$ and $b, c \ge 0$. Then for all $a > 0$,

$$\mathbb{P}(Y \le -a) \le \exp\left\{-\frac{a^2}{c}\right\} \quad \text{and} \quad \mathbb{P}(Y \ge a) \le \exp\left\{-\frac{a^2}{ab + c}\right\}.$$

We shall only use the first inequality. Consider the centered variable

$$Y := \sum_{i \in \mathcal{I}} w_{i,\pi(i)} - m,$$

and let Y' be the corresponding quantity for the pairing π' obtained from π by performing a random switch: two indices i, j are sampled uniformly at random from $\mathcal{I}$ without replacement, and the pairs $\{i, \pi(i)\}$, $\{j, \pi(j)\}$ are replaced with the pairs $\{i, j\}$, $\{\pi(i), \pi(j)\}$. This changes the weight by exactly

$$\begin{aligned}\Delta_{i,j} := {} & w_{i,j} + w_{j,i} + w_{\pi(i),\pi(j)} + w_{\pi(j),\pi(i)} - w_{i,\pi(i)} \\ & - w_{\pi(i),i} - w_{j,\pi(j)} - w_{\pi(j),j}.\end{aligned} \tag{5.2}$$

It is not hard to see that $(\pi, \pi') \stackrel{d}{=} (\pi', \pi)$, so that (i) holds. Moreover,

$$\begin{aligned}\mathbb{E}[Y' - Y|\pi] &= \frac{1}{|\mathcal{I}|(|\mathcal{I}| - 1)} \sum_{i \in \mathcal{I}} \sum_{j \ne i} \Delta_{i,j} \\ &= \frac{4}{|\mathcal{I}|(|\mathcal{I}| - 1)} \sum_{i \in \mathcal{I}} \sum_{j \ne i} w_{i,j} - \frac{4}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} w_{i,\pi(i)} \\ &= -\frac{4}{|\mathcal{I}|} Y.\end{aligned}$$

Regarding the square $\Delta_{i,j}^2 = |\Delta_{i,j}||\Delta_{i,j}|$, we may bound the first copy of $|\Delta_{i,j}|$ by 2θ and the second by changing all minus signs to plus signs in (5.2), yielding

$$\begin{aligned}\mathbb{E}[(Y' - Y)^2|\pi] &= \frac{1}{|\mathcal{I}|(|\mathcal{I}| - 1)} \sum_{i \in \mathcal{I}} \sum_{j \ne i} \Delta_{i,j}^2 \\ &\le \frac{8\theta}{|\mathcal{I}|(|\mathcal{I}| - 1)} \sum_{i \in \mathcal{I}} \sum_{j \ne i} w_{i,j} + \frac{8\theta}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} w_{i,\pi(i)} \\ &= \frac{8\theta}{|\mathcal{I}|}(2m + Y).\end{aligned}$$

Note that taking conditional expectation with respect to Y does not affect the right-hand side. Thus, (ii) and (iii) hold with $\lambda = \frac{4}{|\mathcal{I}|}$, $b = 2\theta$ and $c = 4m\theta$. $\square$

6. Proof of Lemma 4.4. We may fix $z_0 \in \{x, y\}$ and restrict our attention to the halved sum

$$Z := \sum_{z \in \mathfrak{F}} \mathbf{w}(z) \mathbb{1}_{\{\mathbf{w}(z) < w_{\mathrm{MIN}}\}} \mathbb{1}_{\{z \text{ has ancestor } z_0\}}.$$

Consider $m = \lfloor \log N \rfloor$ independent NBRWs on $G(\pi)$ starting at z_0, each being killed as soon as its weight falls below w_{MIN}, and write A for the event that their trajectories form a tree of height less than $t/2$. Clearly, $\mathbb{P}(A|\pi) \geq Z^m$. Taking expectation and using Markov inequality, we deduce that

$$\mathbb{P}(Z > \varepsilon) \leq \frac{\mathbb{P}(A)}{\varepsilon^m},$$

where the average is now taken over both the walks and the pairing. Recalling that $m = \lceil \log N \rceil$, it is more than enough to establish that $\mathbb{P}(A) = (o(1))^m$. To do so, we generate the m killed NBRWs one after the other, revealing the underlying pairs along the way, as described in Section 3. Given that the first $\ell - 1$ walks form a tree of height less than $t/2$, the conditional chance that the ℓth walk also fulfils the requirement is $o(1)$, uniformly in $1 \leq \ell \leq m$. Indeed,

- either its weight falls below $\eta = (1/\log N)^2$ before it ever leaves the graph spanned by the first $\ell - 1$ trajectories and reaches an unpaired half-edge: thanks to the tree structure, there are at most $\ell - 1 < m$ possible trajectories to follow, each having weight at most η, so the chance is less than

$$m\eta = o(1).$$

- or the remainder of its trajectory after the first unpaired half-edge has weight less than $\Delta w_{\mathrm{MIN}}/\eta$: this part consists of at most $t/2$ half-edges which can be coupled with uniform samples from $\mathcal{X}$ for a total-variation cost of mt^2/N, as in Section 3. Thus, the conditional chance is at most

$$\frac{mt^2}{N} + \mathbb{P}\left(\prod_{k=1}^{t/2} \deg(X_k^\star) \geq \frac{\eta}{\Delta w_{\mathrm{MIN}}}\right) = o(1),$$

  by Chebychev's inequality, since $\log(\frac{\eta}{\Delta w_{\mathrm{MIN}}}) - \frac{\mu t}{2} \gg \sigma \sqrt{\frac{t}{2}}$.

7. Proof of Lemma 4.5. Set $m = \lceil (\log N)^2 \rceil$. On $G(\pi)$, let $X^{(1)}, \ldots, X^{(m)}$ and $Y^{(1)}, \ldots, Y^{(m)}$ be $2m$ independent NBRWs of length $t/2$ starting at x and y, respectively. Let B denote the event that their trajectories form two disjoint trees and that for all $1 \leq k \leq m$,

$$\prod_{\ell=1}^{t/2} \frac{1}{\deg(X_\ell^{(k)})} \prod_{\ell=1}^{t/2} \frac{1}{\deg(Y_\ell^{(k)})} > \theta.$$

Then clearly, $\mathbb{P}(B|\pi) \geq \overline{\mathfrak{W}}^m$. Averaging w.r.t. the pairing π, we see that

$$\mathbb{P}\big(\overline{\mathfrak{W}} > \Phi(\lambda) + \varepsilon\big) \leq \frac{\mathbb{P}(B)}{(\Phi(\lambda) + \varepsilon)^m}.$$

Thus, it is enough to establish that $\mathbb{P}(B) \leq (\Phi(\lambda) + o(1))^m$. We do so by generating the $2m$ walks $X^{(1)}, Y^{(1)}, \ldots, X^{(m)}, Y^{(m)}$ one after the other along with the underlying pairing, as above. Given that $X^{(1)}, Y^{(1)}, \ldots, X^{(\ell-1)}, Y^{(\ell-1)}$ already satisfy the desired property, the conditional chance that $X^{(\ell)}, Y^{(\ell)}$ also does is at most $\Phi(\lambda) + o(1)$, uniformly in $1 \leq \ell \leq m$. Indeed,

- either one of the two walks attains length $s = \lceil 4 \log \log N \rceil$ before leaving the graph spanned by the first $2(\ell - 1)$ trajectories and reaching an unpaired half-edge: thanks to the tree structure, there are at most $\ell - 1 < m$ possible trajectories to follow for each walk, each having weight at most 2^{-s} by (1.5), so the conditional chance is at most

$$2m2^{-s} = o(1).$$

- or at least $t - 2s$ unpaired half-edges are encountered, and the product of their degrees falls below $\frac{1}{\theta}$ with conditional probability at most

$$\frac{4mt^2}{N} + \mathbb{P}\left(\prod_{k=1}^{t-2s} \deg\big(X_k^\star\big) < \frac{1}{\theta}\right) = \Phi(\lambda) + o(1),$$

  by the same coupling as above and Berry–Essen's inequality (3.3).

8. Proof of Lemma 4.6. Let τ denote the (random) number of pairs formed during the exploration stage. For $k \geq 0$, we let $\mathfrak{U}_k$ denote the set of unpaired half-edges in the forest after $k \wedge \tau$ pairs have been formed, and we consider the random variable

$$W_k := \sum_{z \in \mathfrak{U}_k} \mathbf{w}(z).$$

Initially $W_0 = 2$, and this quantity either stays constant or decreases at each stage, depending on whether the condition appearing in step 3 is satisfied or not. More precisely, denoting by z_k (resp., z_k') the half-edge selected at step 1 (resp., chosen at step 2) of the kth pair, we have for all $k \geq 1$,

$$W_k = W_{k-1} - \mathbb{1}_{\{k \leq \tau\}}\big(\mathbf{w}(z_k)\mathbb{1}_{\{z_k' \in \mathfrak{U}_{k-1}^+\}} + \mathbf{w}\big(z_k'\big)\mathbb{1}_{\{z_k' \in \mathfrak{U}_{k-1}\}}\big),$$

where $\mathfrak{U}_{k-1}^+$ is $\mathfrak{U}_{k-1}$ together with the unpaired neighbours of x and y. Now, let $\{\mathcal{G}_k\}_{k \geq 0}$ be the natural filtration associated with the exploration stage. Note that τ

is a stopping time, that $\mathbf{w}(z_k)$ is $\mathcal{G}_{k-1}$-measurable, and that the conditional law of z'_k given $\mathcal{G}_{k-1}$ is uniform on $\mathcal{X} \setminus \{z_1, \dots, z_k, z'_1, \dots, z'_{k-1}\}$. Thus,

$$\mathbb{E}[W_k - W_{k-1}|\mathcal{G}_{k-1}] = -\mathbb{1}_{\{k\le\tau\}} \frac{\mathbf{w}(z_k)(|\mathfrak{U}^+_{k-1}| - 2) + W_{k-1}}{N - 2k + 1},$$

$$\begin{aligned}&\mathbb{E}\big[(W_k - W_{k-1})^2|\mathcal{G}_{k-1}\big] \\ &\quad = \mathbb{1}_{\{k\le\tau\}} \frac{\mathbf{w}(z_k)^2(|\mathfrak{U}^+_{k-1}| - 4) + 2\mathbf{w}(z_k)W_{k-1} + \sum_{z\in\mathfrak{U}_{k-1}} \mathbf{w}(z)^2}{N - 2k + 1}.\end{aligned}$$

To bound those quantities, observe that each half-edge in $\mathfrak{U}_{k-1}$ has weight at least $\frac{\mathbf{w}(z_k)}{\Delta}$ because its parent has been selected at an earlier iteration and our selection rule ensures that the quantity $\mathbf{w}(z_k)$ is nonincreasing with k. Consequently,

$$|\mathfrak{U}_{k-1}| \frac{\mathbf{w}(z_k)}{\Delta} \le \sum_{z\in\mathfrak{U}_{k-1}} \mathbf{w}(z) \le 2.$$

Combining this with the bound $|\mathfrak{U}^+_{k-1}| \le |\mathfrak{U}_{k-1}| + 2\Delta$, we arrive at

$$\mathbb{E}[W_k - W_{k-1}|\mathcal{G}_{k-1}] \ge -\mathbb{1}_{\{k\le\tau\}} \frac{4\Delta}{N - 2k + 1},$$

$$\mathbb{E}\big[(W_k - W_{k-1})^2|\mathcal{G}_{k-1}\big] \le \mathbb{1}_{\{k\le\tau\}} \frac{4\Delta\mathbf{w}(z_k) + 2}{N - 2k + 1}.$$

Now recall that $\mathbf{w}(z_k) \ge w_{\mathrm{MIN}}$ and $\mathbf{h}(z_k) < \frac{t}{2}$ as per our selection rule, implying

$$w_{\mathrm{MIN}}\tau \le \sum_{k\ge1} \mathbf{w}(z_k)\mathbb{1}_{\{\tau\ge k\}} \le \sum_{z\in\mathfrak{F}} \mathbf{w}(z)\mathbb{1}_{\{\mathbf{h}(z)<\frac{t}{2}\}} \le t. \tag{8.1}$$

The right-most inequality follows from the fact that the total weight at a given height in $\mathfrak{F}$ is at most 2 (the total weight being preserved from a parent to its children, if any). We conclude that

$$\sum_{k=1}^{\tau} \mathbb{E}[W_k - W_{k-1}|\mathcal{G}_{k-1}] \ge -\frac{4\Delta t}{w_{\mathrm{MIN}}N - 2t} := -m,$$

$$\sum_{k=1}^{\tau} \mathbb{E}\big[(W_k - W_{k-1})^2|\mathcal{G}_{k-1}\big] \le \frac{4\Delta t w_{\mathrm{MIN}} + 2t}{N w_{\mathrm{MIN}} - 2t} := v.$$

Now, fix $\varepsilon > 0$ and consider the martingale $\{M_k\}_{k\ge0}$ defined by $M_0 = 0$ and

$$M_k := \sum_{i=1}^{k} \big\{(W_{i-1} - W_i) \wedge \varepsilon - \mathbb{E}\big[(W_{i-1} - W_i) \wedge \varepsilon|\mathcal{G}_{i-1}\big]\big\}.$$

Then the increments of $\{M_k\}_{k\ge0}$ are bounded by ε by construction, and the above computation guarantees that almost-surely,

$$\sum_{k=1}^{\tau} \mathbb{E}\big[(M_k - M_{k-1})^2|\mathcal{G}_{k-1}\big] \le v = N^{-\frac{1}{3}+o(1)}.$$

Thus, the martingale version of Bennett's inequality due to Freedman [18] yields

$$\mathbb{P}(M_\tau > 7\varepsilon) \le \left(\frac{ev}{v+7\varepsilon^2}\right)^7 = N^{-\frac{7}{3}+o(1)}. \tag{8.2}$$

But on the event $\{x \in \mathcal{R}, y \in \mathcal{R} \setminus \mathcal{B}_x\}$, all paths from the set $\{x, y\}$ to itself must have length at least r, and since $r \to \infty$, we must have asymptotically

$$\begin{aligned}\{x \in \mathcal{R}, y \in \mathcal{R} \setminus \mathcal{B}_x\} &\subseteq \left\{\max_k (W_{k-1} - W_k) \le \varepsilon\right\} \\ &\subseteq \{W_0 - W_\tau \le M_\tau + m\}.\end{aligned}$$

With (8.2), this proves Lemma 4.6 since $W_0 - W_\tau = 2 - \sum_{z\in\mathfrak{U}} \mathbf{w}(z)$ and $m = o(1)$.

REFERENCES

[1] ABDULLAH, M., COOPER, C. and FRIEZE, A. (2012). Cover time of a random graph with given degree sequence. *Discrete Math.* **312** 3146–3163. MR2957935

[2] ALDOUS, D. (1983). Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on Probability, XVII. Lecture Notes in Math.* **986** 243–297. Springer, Berlin. MR0770418

[3] ALDOUS, D. and DIACONIS, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly* **93** 333–348. MR0841111

[4] ALON, N., BENJAMINI, I., LUBETZKY, E. and SODIN, S. (2007). Non-backtracking random walks mix faster. *Commun. Contemp. Math.* **9** 585–603. MR2348845

[5] BASU, R., HERMON, J. and PERES, Y. (2017). Characterization of cutoff for reversible Markov chains. *Ann. Probab.* **45** 1448–1487.

[6] BERESTYCKI, N., LUBETZKY, E., PERES, Y. and SLY, A. (2015). Random walks on the random graph. Preprint. Available at arXiv:1504.01999.

[7] BERESTYCKI, N. and SENGUL, B. (2014). Cutoff for conjugacy-invariant random walks on the permutation group. Preprint. Available at arXiv:1410.4800.

[8] BRODER, A. and SHAMIR, E. (1987). On the second eigenvalue of random regular graphs. In 28*th Annual Symposium on Foundations of Computer Science*, 1987 286–294. IEEE Computer Society, Washington, DC.

[9] CHATTERJEE, S. (2007). Stein's method for concentration inequalities. *Probab. Theory Related Fields* **138** 305–321. MR2288072

[10] CHEN, G.-Y. and SALOFF-COSTE, L. (2008). The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.* **13** 26–78. MR2375599

[11] COOPER, C. (2011). Random walks, interacting particles, dynamic networks: Randomness can be helpful. In *Structural Information and Communication Complexity. Lecture Notes in Computer Science* **6796** 1–14. Springer, Heidelberg. MR2862823

[12] COOPER, C. and FRIEZE, A. (2014). Vacant sets and vacant nets: Component structures induced by a random walk. Preprint. Available at arXiv:1404.4403.

[13] CUFF, P., DING, J., LOUIDOR, O., LUBETZKY, E., PERES, Y. and SLY, A. (2012). Glauber dynamics for the mean-field Potts model. *J. Stat. Phys.* **149** 432–477. MR2992796

[14] DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* **93** 1659–1664. MR1374011

[15] DIACONIS, P. and SHAHSHAHANI, M. (1981). Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete* **57** 159–179. MR0626813

[16] DING, J., LUBETZKY, E. and PERES, Y. (2009). The mixing time evolution of Glauber dynamics for the mean-field Ising model. *Comm. Math. Phys.* **289** 725–764. MR2506768

[17] DING, J., LUBETZKY, E. and PERES, Y. (2010). Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields* **146** 61–85. MR2550359

[18] FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118. MR0380971

[19] FRIEDMAN, J. (2008). A proof of Alon's second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.* **195** viii+100. MR2437174

[20] GANGULY, S., LUBETZKY, E. and MARTINELLI, F. (2013). Cutoff for the East process. Preprint. Available at arXiv:1312.7863.

[21] GRIFFITHS, S., KANG, R. J., OLIVEIRA, R. I. and PATEL, V. (2014). Tight inequalities among set hitting times in Markov chains. *Proc. Amer. Math. Soc.* **142** 3285–3298. MR3223383

[22] HERMON, J. (2015). A technical report on hitting times, mixing and cutoff. Preprint. Available at arXiv:1501.01869.

[23] JANSON, S. (2009). The probability that a random multigraph is simple. *Combin. Probab. Comput.* **18** 205–225. MR2497380

[24] LACOIN, H. (2015). The cutoff profile for the simple-exclusion process on the cycle. Preprint. Available at arXiv:1502.00952.

[25] LEVIN, D. A., LUCZAK, M. J. and PERES, Y. (2010). Glauber dynamics for the mean-field Ising model: Cut-off, critical power law, and metastability. *Probab. Theory Related Fields* **146** 223–265. MR2550363

[26] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. AMS, Providence, RI. MR2466937

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

[28] LUBETZKY, E. and SLY, A. (2014). Cutoff for general spin systems with arbitrary boundary conditions. *Comm. Pure Appl. Math.* **67** 982–1027. MR3193965

[29] LUBETZKY, E. and SLY, A. (2014). Universality of cutoff for the Ising model. Preprint. Available at arXiv:1407.1761.

[30] OLIVEIRA, R. I. (2012). Mixing and hitting times for finite Markov chains. *Electron. J. Probab.* **17** 1–12. MR2968677

[31] PERES, Y. (2004). American Institute of Mathematics (AIM) research workshop "Sharp Thresholds for Mixing Times" (Palo Alto, December 2004). Summary. Available at http://www.aimath.org/WWN/mixingtimes.

[32] PERES, Y. and SOUSI, P. (2013). Mixing times are hitting times of large sets. *J. Theoret. Probab.* 1–32.

[33] ROSS, N. (2011). Fundamentals of Stein's method. *Probab. Surv.* **8** 210–293. MR2861132

[34] SALOFF-COSTE, L. (2004). Random walks on finite groups. In *Probability on Discrete Structures*. *Encyclopaedia Math. Sci.* **110** 263–346. Springer, Berlin. MR2023654

[35] VAN DER HOFSTAD, R. (2013). Random graphs and complex networks. Course notes. Available at http://www.win.tue.nl/~rhofstad/NotesRGCN.html.

[36] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and VAN MIEGHEM, P. (2005). Distances in random graphs with finite variance degrees. *Random Structures Algorithms* **27** 76–123. MR2150017

UFR DE MATHÉMATIQUES
UNIVERSITY OF PARIS DIDEROT, LPMA
BÂTIMENT SOPHIE GERMAIN, OFFICE 532
5 RUE THOMAS MANN
75205 PARIS CEDEX 13
FRANCE
E-MAIL: benhamou@math.univ-paris-diderot.fr
URL: http://www.proba.jussieu.fr/~benhamou/

UFR DE MATHÉMATIQUES
UNIVERSITY OF PARIS DIDEROT, LPMA
BÂTIMENT SOPHIE GERMAIN, OFFICE 516
5 RUE THOMAS MANN
75205 PARIS CEDEX 13
FRANCE
E-MAIL: salez@math.univ-paris-diderot.fr
URL: http://www.proba.jussieu.fr/~salez/