



HAL
open science

Cutoff for permuted markov chains

Anna Ben-Hamou, Yuval Peres

► **To cite this version:**

Anna Ben-Hamou, Yuval Peres. Cutoff for permuted markov chains. Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, In press, 10.1214/22-AIHP1248 . hal-03895653

HAL Id: hal-03895653

<https://hal.science/hal-03895653>

Submitted on 13 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CUTOFF FOR PERMUTED MARKOV CHAINS

ANNA BEN-HAMOU AND YUVAL PERES

ABSTRACT. Let P be a bistochastic matrix of size n , and Π be a permutation matrix of size n . In this paper, we are interested in the mixing time of the Markov chain whose transition matrix is given by $Q = P\Pi$. In other words, the chain alternates between random steps governed by P and deterministic steps governed by Π . We show that if the permutation Π is chosen uniformly at random, then under mild assumptions on P , with high probability, the chain Q exhibits cutoff at time $\frac{\log n}{\mathbf{h}}$, where \mathbf{h} is the entropic rate of P . Moreover, for deterministic permutations, we improve the upper bound on the mixing time obtained by Chatterjee and Diaconis [11].

RÉSUMÉ. Soit P une matrice bistochastique de taille n , et Π une matrice de permutation de taille n . Dans cet article, nous nous intéressons au temps de mélange de la chaîne de Markov dont la matrice de transition est donnée par $Q = P\Pi$. En d'autres termes, la chaîne alterne entre des sauts aléatoires gouvernés par P et des sauts déterministes gouvernés par Π . Nous montrons que si la permutation Π est choisie uniformément au hasard, alors, sous de légères hypothèses sur P , avec grande probabilité, la chaîne Q présente un cutoff au temps $\frac{\log n}{\mathbf{h}}$, où \mathbf{h} est le taux entropique de P . De plus, pour des permutations déterministes, nous améliorons la borne supérieure sur le temps de mélange obtenue par Chatterjee and Diaconis [11].

1. INTRODUCTION

1.1. **Setting and main results.** Let P be a bistochastic transition matrix on a finite state space Ω of size n , and let π be a permutation on Ω . Denote by Π the corresponding permutation matrix (defined by $\Pi_{i,j} = \mathbf{1}_{\{j=\pi(i)\}}$). We are interested in the mixing time of the Markov chain with transition matrix $Q = P\Pi$. By the bistochastic assumption, the uniform distribution on Ω is stationary for P , and it is stationary for Q as well. The total-variation distance to equilibrium at time $t \geq 1$ starting from state $x \in \Omega$ is then given by

$$\mathcal{D}_x(t) = \sum_{y \in \Omega} \left(\frac{1}{n} - Q^t(x, y) \right)_+.$$

For $\varepsilon \in (0, 1)$, the ε -mixing time is defined as

$$t_{\text{mix}}(\varepsilon) = \min\{t \geq 0, \mathcal{D}(t) \leq \varepsilon\},$$

where $\mathcal{D}(t) = \max_{x \in \Omega} \mathcal{D}_x(t)$. The original Markov chain P might have a very large mixing time (it might not even mix at all, if P is reducible or periodic), and it is natural to ask which permutations on the state space are able to speed up convergence to stationarity. Consider for instance the lazy random walk on the circle \mathbb{Z}_n given by $X_0 = 0$ and, for $k \geq 0$,

$$X_{k+1} = X_k + \varepsilon_{k+1} \pmod{n},$$

where $(\varepsilon_k)_{k \geq 1}$ is an i.i.d. sequence uniformly distributed on $\{-1, 0, 1\}$. It is well known that this chain mixes in order n^2 steps (and does not have cutoff). Now consider alternating random uniform jumps in $\{-1, 0, 1\}$ and deterministic jumps given by some bijection f on \mathbb{Z}_n . This corresponds to the permuted Markov chain

$$X_{k+1} = f(X_k) + \varepsilon_{k+1} \pmod{n}.$$

He, Pham, and Xu [18] recently proved that, for n prime and for $f(x) = x^{-1}$ when $x \neq 0$, and $f(0) = 0$, this chain takes only order $\log n$ steps to mix. For $f(x) = 2x$, it was shown by Eberhard and Varjú [16] that the chain has cutoff at time $c \log n$, for some absolute constant $c > 0$.

Key words and phrases. Markov chains, mixing times, random permutations.

In the general setting, Chatterjee and Diaconis [11] recently showed that, under some assumptions on P , if the permutation π satisfies some expansion condition with respect to P , then the mixing time of $Q = P\Pi$ is logarithmic, and that all but a vanishing fraction of permutations satisfy this expansion condition.

We refine this result by showing that, under mild assumptions on the matrix P , for almost every permutation π , the chain $Q = P\Pi$ has cutoff at a time $\frac{\log n}{\mathbf{h}}$, where \mathbf{h} is the entropic rate of P , that is

$$\mathbf{h} = \frac{1}{n} \sum_{x,y \in \Omega} P(x,y) \log \frac{1}{P(x,y)}. \quad (1)$$

To be more precise, this result holds for a sequence of chains $Q_n = P_n \Pi_n$ on Ω_n with $|\Omega_n| = n$, all asymptotic statements being understood as n tends to $+\infty$. To alleviate notation, we will later keep this dependence in n implicit.

Theorem 1. *Let P_n be a bistochastic transition matrix on Ω_n with $|\Omega_n| = n \geq 1$, and let \mathbf{h}_n be the corresponding entropy rate. Assume that*

$$\delta_n = \min \{P_n(x,y), x,y \in \Omega_n, P_n(x,y) > 0\} = \frac{1}{n^{o(1)}}, \quad (2)$$

$$\Delta = \sup_{n \geq 1} \max_{x,y \in \Omega_n} P_n(x,y) < 1, \quad (3)$$

$$\mathbf{h}_n = o\left(\frac{\log n}{\log \log n}\right). \quad (4)$$

Let π_n be a uniform random permutation on Ω_n , and Π_n be the corresponding permutation matrix. Then, for all $\varepsilon \in (0,1)$, the mixing time of the chain $Q_n = P_n \Pi_n$ satisfies

$$\frac{\mathbf{h}_n t_{\text{mix}}^{(n)}(\varepsilon)}{\log n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 1.$$

Note that assumption 2 means that δ_n satisfies $\frac{\log \delta_n}{\log n} \rightarrow 0$. In particular, it holds if $\inf_n \delta_n > 0$. Also note that assumption 3 implies that $\mathbf{h}_n \geq \log \frac{1}{\Delta} = \Omega(1)$. Hence the cutoff time $\frac{\log n}{\mathbf{h}_n}$ is always $O(\log n)$ but it may be much smaller than $\log n$: assumption 4 only requires that $\frac{\log n}{\mathbf{h}_n}$ grows faster than $\log \log n$.

An interesting consequence of Theorem 1 is the following.

Corollary 2. *Let $d \geq 2$ be a fixed integer and let $(G_n = (V_n, E_n))_{n \geq 1}$ be a sequence of d -regular digraphs (i.e. for all vertices, the in-degree and out-degree are both equal to d), with $|V_n| = n$. For each $n \geq 1$, let π_n be a uniformly chosen permutation on V_n , and k_n be a positive integer. Consider the Markov chain on V_n for which one step consists in k_n simple random walk steps on G_n followed by an application of π_n . This chain has transition matrix $Q_n = M_n^{k_n} \Pi_n$, where M_n is the transition matrix of the simple random walk on G_n , and Π_n is the matrix associated to π_n . Then, if $k_n = o\left(\frac{\log n}{\log \log n}\right)$, the sequence of chains $(Q_n)_{n \geq 1}$ has cutoff at time $\frac{\log n}{\mathbf{H}_n}$ where*

$$\mathbf{H}_n = \frac{1}{n} \sum_{u,v \in V(G_n)} M_n^{k_n}(u,v) \log \frac{1}{M_n^{k_n}(u,v)}.$$

Proof of Corollary 2. It suffices to observe that the bistochastic matrix $P_n = M_n^{k_n}$ satisfies the assumptions of Theorem 1. Indeed, since $d \geq 2$, the probabilities $M_n^{k_n}(u,v)$ are all at most $1/2$ and assumption 3 is satisfied. And since $k_n = o\left(\frac{\log n}{\log \log n}\right)$, for all $u,v \in V_n$ such that $M_n^{k_n}(u,v) > 0$, we have

$$M_n^{k_n}(u,v) \geq \left(\frac{1}{d}\right)^{k_n} = \frac{1}{n^{o(1)}},$$

hence assumption 2 is satisfied. Moreover, we have

$$\mathbf{H}_n \leq \log \left(d^{k_n} \right) = o \left(\frac{\log n}{\log \log n} \right),$$

and assumption 4 is satisfied as well. \blacksquare

The following proposition (whose proof is given in Section 2.4) shows that the threshold $k_n = o \left(\frac{\log n}{\log \log n} \right)$ is not far from being optimal.

Proposition 3. *In the setting of Corollary 2, for $k_n = \Omega(\log n)$, one may find examples of graphs for which the chain Q_n does not have cutoff.*

Our second result concerns deterministic permutations. Let us first recall the result of Chatterjee and Diaconis [11]. For $A \subset \Omega_n$, let $\mathcal{E}_n(A)$ be the subset of states attainable by P_n in one step starting from A . Assume that the sequence of permutations $(\pi_n)_{n \geq 1}$ is such that for all $n \geq 1$, there exists $\alpha_n \in (0, 1)$ such that for all $A \subset \Omega_n$ with $|A| \leq n/2$, we have

$$|\mathcal{E}_n \circ \pi_n \circ \mathcal{E}_n(A)| \geq (1 + \alpha_n)|A|. \quad (5)$$

Under some assumptions on P_n , Chatterjee and Diaconis [11] show that the total-variation distance of Q_n satisfies

$$\mathcal{D}^{(n)} \left(\frac{(1 + o(1))4 \log n}{\alpha_n^2 \delta_n^8} \right) \xrightarrow{n \rightarrow \infty} 0,$$

where $\delta_n = \min\{P_n(x, y), x, y \in \Omega_n, P_n(x, y) > 0\}$. We show that the dependence on δ_n can be improved from δ_n^{-8} to δ_n^{-4} , under weaker assumptions on P_n .

Theorem 4. *Let P_n be a bistochastic transition matrix on Ω_n with $|\Omega_n| = n$, and assume that there exists $\gamma > 0$ such that for all n and for all $x \in \Omega_n$, we have $P_n(x, x) \geq \gamma$. Let π_n be a permutation on Ω_n satisfying assumption (5). Then the total-variation distance of $Q_n = P_n \Pi_n$ satisfies*

$$\mathcal{D}^{(n)} \left(\frac{17 \log n}{\alpha_n^2 \delta_n^4} \right) \xrightarrow{n \rightarrow \infty} 0.$$

1.2. Related work. Theorem 1 describes a sharp transition in the convergence to equilibrium: the first order of the mixing time $t_{\text{mix}}(\varepsilon)$ does not depend on the target ε . In other words, the total-variation distance to the stationary measure abruptly falls from 1 to 0 over a short time-scale which is negligible with respect to the mixing time. This behavior is known as cutoff. Since its discovery in the early 1980's by Diaconis and Shahshahani [14] and Aldous [1], this phenomenon has been shown to occur for a wide variety of finite Markov chains. However, the question of characterizing the chains which have this property is still largely open, except in some specific settings such as birth-and-death chains [15], and more generally random walks on trees [2]. Recently, an impressive step was made by Salez [26], who showed that all non-negatively curved Markov chains satisfying a simple condition have cutoff.

In recent years, a lot of attention has been devoted to random instances: instead of being fixed, the sequence of transition matrices is drawn at random according to some probability distribution, and one looks for a result that holds for almost all realizations. A recurring observation in this framework is that random instances often exhibit cutoff, at a time that can be described in terms of entropy (the ‘‘entropic time’’). For random walks on random graphs, this was pioneered by Lubetzky and Sly [24], who established cutoff for the simple and non-backtracking random walks on random regular graphs. This was later extended to random walks on random graphs with given degrees [3, 4, 5], and to random directed graphs [7]. Other notable examples are random lifts of graphs [6, 13], random Cayley graphs for Abelian groups [19], and for upper triangular matrices [20].

Bordenave, Caputo, and Salez [8] considered random transition matrices obtained by permuting independently and uniformly at random each row of a fixed transition matrix P . They establish cutoff at the entropic time $\frac{\log n}{\mathbf{h}}$, with \mathbf{h} given in (1). Our setup is similar, except the perturbation is reduced from n to only one random permutation. We note however that our model is

significantly simplified by assuming that P has uniform stationary distribution. Also, their result holds under weaker versions of assumptions (2) and (3). On the other hand, they assume that \mathbf{h} is of order 1, whereas we only need $\mathbf{h} = o\left(\frac{\log n}{\log \log n}\right)$.

Another model that has some resemblance with our setup is the one considered by Hermon, Sly, and Sousi [21]. The authors consider a random graph formed by adding to a fixed bounded-degree graph the edges of a uniform random matching on the vertex set. The random walk is shown to have cutoff, at a time that can be described through the entropy of the walk on an auxiliary graph referred to as the quasi-tree. Similarly to our result, this gives a simple random perturbation that makes a large class of transition matrices have cutoff. Let us note that one main difference is that our model has a spontaneous non-backtracking tendency which simplifies the analysis of typical paths.

The spectral properties of the random matrix $Q = P\Pi$ were investigated by Bordenave, Qiu, Zhang, et al. [9]. In particular, they show that, under some sparsity and regularity assumptions on P , the spectral gap of Q is bounded away from 0, which implies that the mixing time is $O(\log n)$.

Finally, let us note that multiplying a transition matrix by a permutation fits into the general topic of speeding up Markov chains. For references on this topic, we refer the reader to the paper of Chatterjee and Diaconis [11] and to the survey of Hildebrand [22] for the case of random walks on groups.

1.3. Organization of the paper. Theorem 1 and Proposition 3 are proven in Section 2, and Theorem 4 in Section 3.

2. CUTOFF FOR RANDOM PERMUTATIONS

To alleviate notation, the dependence on n will often be implicit: for instance, we will write Ω instead of Ω_n and P instead of P_n . Also, δ will stand for δ_n as defined in (2), and \mathbf{h} for \mathbf{h}_n .

2.1. Coupling with i.i.d. samples. Before entering into the proof of Theorem 1, we describe a coupling for typical trajectories, which helps approximating the annealed law of the walk and will be crucially used later on. This coupling takes advantage of the fact that the walk and the permutation π along its trajectory can be generated simultaneously as follows: initially, $X_0 = x \in \Omega$ and $\text{Dom}(\pi) = \text{Ran}(\pi) = \emptyset$; then at each time $k \geq 0$,

- (1) choose an element $Y_k \in \Omega$ according to the distribution $P(X_k, \cdot)$;
- (2) if $Y_k \notin \text{Dom}(\pi)$, i.e. if the image of Y_k by π has not been chosen yet, choose $\pi(Y_k)$ uniformly at random in $\Omega \setminus \text{Ran}(\pi)$, and update $\text{Dom}(\pi)$ and $\text{Ran}(\pi)$ accordingly; otherwise, $\pi(Y_k)$ is already defined, and no new image is chosen;
- (3) in both cases, let $X_{k+1} = \pi(Y_k)$.

The sequence $(X_k)_{k \geq 0}$ is then exactly distributed according to the annealed law. Now, consider a sequence $(X_k^*, Y_k^*)_{k \geq 0}$ generated in the following way: initially $X_0^* = x \in \Omega$; then at each time $k \geq 0$,

- (1) choose an element $Y_k^* \in \Omega$ according to the distribution $P(X_k^*, \cdot)$;
- (2) let X_{k+1}^* be uniformly distributed on the entire set Ω .

Note that the process $(X_k^*, Y_k^*)_{k \geq 1}$ is simply an i.i.d. sequence and that both X_k^* and Y_k^* are uniformly distributed on Ω . The two processes $(X_k, Y_k)_{k \geq 0}$ and $(X_k^*, Y_k^*)_{k \geq 0}$ may be coupled until time

$$T = \inf\{k \geq 0, X_k^* \in \text{Ran}(\pi) \text{ or } Y_k^* \in \text{Dom}(\pi)\}.$$

Since at step k , both $|\text{Dom}(\pi)|$ and $|\text{Ran}(\pi)|$ are at most k , we have

$$\mathbb{P}(T \leq t) \leq \frac{2t^2}{n}. \tag{6}$$

2.2. Lower bound. Let $x \in \Omega$ be a fixed starting point and let

$$t = \left\lceil c \frac{\log n}{\mathbf{h}} \right\rceil,$$

for $0 < c < 1$. Let A be the set of $y \in \Omega$ such that there exists a path from x to y which has probability larger than $\frac{\log n}{n}$ to be taken by the chain in t steps. Since, for all $y \in A$, we have $Q^t(x, y) \geq \frac{\log n}{n}$, and since $Q^t(x, \cdot)$ is a probability, the set A has size less than $\frac{n}{\log n}$. Hence

$$\mathcal{D}_x(t) \geq Q^t(x, A) - \frac{|A|}{n} \geq Q^t(x, A) - \frac{1}{\log n}.$$

Taking expectation with respect to the permutation, we have

$$\mathbb{E}Q^t(x, A) \geq \mathbb{P}_x \left(\prod_{s=0}^{t-1} P(X_s, Y_s) > \frac{\log n}{n} \right) = \mathbb{P}_x \left(\prod_{s=0}^{t-1} P(X_s^*, Y_s^*) > \frac{\log n}{n} \right) + o(1),$$

where the last equality is by (6). Taking logarithms in the complementary event, singling out the particular first step (and using that $\log \frac{1}{P(X_0^*, Y_0^*)} \leq \log \frac{1}{\delta_n} = o(\log n)$), and applying Chebychev Inequality, we have

$$\begin{aligned} \mathbb{P}_x \left(\prod_{s=0}^{t-1} P(X_s^*, Y_s^*) \leq \frac{\log n}{n} \right) &= \mathbb{P}_x \left(\sum_{s=1}^{t-1} \log \frac{1}{P(X_s^*, Y_s^*)} \geq \log n + o(\log n) \right) \\ &= O \left(\frac{\sigma^2}{(1-c)^2 \mathbf{h} \log n} \right), \end{aligned}$$

where

$$\sigma^2 = \frac{1}{n} \sum_{x, y \in \Omega} P(x, y) \left(\log \frac{1}{P(x, y)} - \mathbf{h} \right)^2.$$

By assumption (2), we have

$$\sigma^2 = o(\mathbf{h} \log n). \quad (7)$$

Hence

$$\min_{x \in \Omega} \mathbb{E}[\mathcal{D}_x(t)] \xrightarrow{n \rightarrow +\infty} 1,$$

which yields $\mathcal{D}(t) \xrightarrow{\mathbb{P}} 1$.

2.3. Upper bound. Let

$$t = 2 \left\lceil \frac{c \log n}{2\mathbf{h}} \right\rceil,$$

for $c > 1$. Letting P_\star be the matrix defined by

$$\forall x, y \in \Omega, P_\star(x, y) = P(y, x),$$

and $Q_\star = P_\star \Pi^{-1}$, a key observation is that for all $x, y \in \Omega$,

$$Q^t(x, \pi(y)) = \sum_{u, v \in \Omega} Q^{t/2}(x, u) Q_\star^{t/2}(y, v) \mathbf{1}_{\{v = \pi^{-1}(u)\}}.$$

The proof of the upper bound then relies on a special exploration procedure, which generates the permutation π together with two disjoint weighted trees \mathcal{T}_x and \mathcal{T}_y , keeping track of only certain paths from x and from y . Initially at a time 0, $\text{Dom}_0(\pi) = \text{Ran}_0(\pi) = \emptyset$, the tree $\mathcal{T}_x(0)$ is reduced to x and the tree $\mathcal{T}_y(0)$ is reduced to y . Then, for $k = 1, 2, \dots$, we iterate the following steps to generate $\mathcal{T}_x(k)$ and $\mathcal{T}_y(k)$:

- (1) Each vertex $u \in \mathcal{T}_x(k-1) \cup \mathcal{T}_y(k-1)$ determines a unique sequence (u_0, \dots, u_d) where $u_0 \in \{x, y\}$ is the ancestor of u , for each $1 \leq i \leq d$, u_i is a child of u_{i-1} , and $u_d = u$. The integer d is called the height of u , denoted $h(u)$, and the weight of u is defined as

$$\mathbf{w}(u) = \begin{cases} \prod_{s=0}^{d-1} Q(u_s, u_{s+1}) & \text{if } u_0 = x, \\ \prod_{s=0}^{d-1} Q_\star(u_s, u_{s+1}) & \text{if } u_0 = y. \end{cases}$$

If $u_0 = x$ (resp. $u_0 = y$), we say that (u, i) is a stub if $P(u, i) > 0$ (resp. if $P_\star(u, i) > 0$). The weight of a stub is naturally defined as

$$\mathbf{w}(u, i) = \begin{cases} \mathbf{w}(u)P(u, i) & \text{if } u_0 = x, \\ \mathbf{w}(u)P_\star(u, i) & \text{if } u_0 = y, \end{cases}$$

and its height as $h(u, i) = h(u) + 1$. A stub (u, i) is said to be free at time $k - 1$, which is denoted $(u, i) \in \mathcal{F}_{k-1}$, if either $u_0 = x$ and $i \notin \text{Dom}_{k-1}(\pi)$ or $u_0 = y$ and $i \notin \text{Ran}_{k-1}(\pi)$. Pick a free stub $(u, i) \in \mathcal{F}_{k-1}$ such that

$$\mathbf{w}(u, i) \geq \mathbf{w}_{\min} = n^{-2/3} \quad \text{and} \quad h(u, i) < \frac{t}{2}. \quad (8)$$

- If $u_0 = x$, set $\text{Dom}_k(\pi) = \text{Dom}_{k-1}(\pi) \cup \{i\}$; if $u_0 = y$, set $\text{Ran}_k(\pi) = \text{Ran}_{k-1}(\pi) \cup \{i\}$.
- (2) If $u_0 = x$, then choose $i' = \pi(i)$ uniformly at random in $\Omega \setminus \text{Ran}_{k-1}(\pi)$ and set $\text{Ran}_k(\pi) = \text{Ran}_{k-1}(\pi) \cup \{i'\}$. If $u_0 = y$, then choose $i' = \pi^{-1}(i)$ uniformly at random in $\Omega \setminus \text{Dom}_{k-1}(\pi)$ and set $\text{Dom}_k(\pi) = \text{Dom}_{k-1}(\pi) \cup \{i'\}$.
 - (3) For $A \subset \Omega$, let

$$\mathcal{N}(A) = A \cup \{z \in \Omega, \exists a \in A, P(a, z) > 0 \text{ or } P_\star(a, z) > 0\}.$$

If $\mathcal{N}(\{i'\}) \cap \mathcal{N}(\mathcal{T}_x(k-1) \cup \mathcal{T}_y(k-1)) = \emptyset$, then add i' as a child of u in the corresponding tree and leave the other tree unchanged to form $\mathcal{T}_x(k)$ and $\mathcal{T}_y(k)$. Otherwise, both $\mathcal{T}_x(k) = \mathcal{T}_x(k-1)$ and $\mathcal{T}_y(k) = \mathcal{T}_y(k-1)$.

The exploration stage stops when the set of free stubs satisfying (8) is empty. Let τ be the number of times step (2) is performed and let $\mathcal{T}_x = \mathcal{T}_x(\tau)$ and $\mathcal{T}_y = \mathcal{T}_y(\tau)$. Note that, by condition (8), we have

$$\mathbf{w}_{\min}\tau \leq \sum_{u \in \mathcal{T}_x} \mathbf{w}(u) + \sum_{u \in \mathcal{T}_y} \mathbf{w}(u) \leq t. \quad (9)$$

Also note that the stub picked in step (1) can be chosen arbitrarily provided it satisfies condition (8). We choose to explore in the following order:

- (i) Reveal around x , choosing free stubs with minimal height, until no free stub has height strictly less than

$$t_0 := \left\lceil \frac{\log \log n}{\log(1/\Delta)} \right\rceil.$$

Let x_1, \dots, x_m be the vertices of \mathcal{T}_x at height t_0 at the end of this first stage.

- (ii) Reveal around y , choosing free stubs with minimal height, until no free stub has height strictly less than t_0 . Let $y_1, \dots, y_{m'}$ be the vertices of \mathcal{T}_y at height t_0 at the end of this first stage.
- (iii) Sequentially for $\ell = 1, \dots, m$, complete the exploration around x_ℓ , until height $t/2$.
- (iv) Sequentially for $\ell = 1, \dots, m'$, complete the exploration around y_ℓ , until height $t/2$.

Let

$$r := \left\lceil \frac{\log n}{12 \log(1/\delta)} \right\rceil.$$

By assumption (2), we have $r \rightarrow +\infty$. An oriented graph G (resp. G_\star) on n vertices may naturally be associated to the matrix Q (resp. Q_\star), by placing an oriented edge from u to v when $Q(u, v) > 0$ (resp. $Q_\star(u, v) > 0$). We say that x is a Q -root (resp. a Q_\star -root), which we denote $x \in \mathcal{R}$ (resp. $x \in \mathcal{R}_\star$), if the (oriented) neighborhood of x in G (resp. G_\star) up to level r is a tree.

Lemma 5. *For $s = t_0 \wedge r$, we have*

$$\max_{x \in \Omega} Q^s(x, \mathcal{R}^c) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \frac{|\mathcal{R}_\star^c|}{n} \xrightarrow{\mathbb{P}} 0.$$

Proof of Lemma 5. Let us first prove the first claim. Let $R = \left\lceil \frac{\log n}{5 \log(1/\delta)} \right\rceil$. When sequentially revealing the neighborhood of x in G up to level R , choosing free stubs with minimal height, the total number of steps is at most $(1/\delta)^R = n^{1/5+o(1)}$, and at each step, the probability of having a “bad event”, i.e. choosing a state u such that either $u = x$ or there exists v with $P(u, v) > 0$ and $P(u', v) > 0$ for some other already exposed u' , is at most $\frac{(1/\delta)^{R+2}}{n - (1/\delta)^R} = n^{-4/5+o(1)}$. Hence, the probability of having at least two such bad events up to level R is $n^{-6/5+o(1)}$. Hence, with high probability, all $x \in \Omega$ are such that there is at most one problematic vertex in their R -neighborhood. Let $x \in \Omega$. If there is no problematic vertex in its R -neighborhood, then the claim is clear. Otherwise, let u be this single problematic vertex. If by time s the walk has exited the only path from x to u , then it lies on a Q -root. By assumption (3), at each step, the probability to follow the path is at most $\Delta < 1$, except, in the case $u \neq x$, when the walk is at u , which can happen at most once in every two steps. Hence the probability not to have exited the path by time r is at most $\Delta^{\frac{r}{2}-1} = o(1)$, which proves the first claim. Since Q has uniform stationary distribution, this implies that $\frac{|\mathcal{R}^c|}{n} \xrightarrow{\mathbb{P}} 0$. Since P and P_* satisfy the same assumptions, the same can be said for $|\mathcal{R}_*|$, establishing the second claim. \blacksquare

Let \mathcal{B}_x be the set of vertices at (oriented) distance at most r from x in G (i.e. \mathcal{B}_x corresponds to the states that can be reached in at most r steps by the chain Q). We have $|\mathcal{B}_x| \leq (1/\delta)^r = o(n)$. Combined with Lemma 5, this yields

$$\mathcal{D}(t+s) \leq \max_{x \in \mathcal{R}} \mathcal{D}_x(t) + o_{\mathbb{P}}(1) \leq \max_{x \in \mathcal{R}} \sum_{y \in \mathcal{R}_* \setminus \mathcal{B}_x} \left(\frac{1}{n} - Q^t(x, \pi(y)) \right)_+ + o_{\mathbb{P}}(1).$$

The rest of the proof is devoted to establishing

$$\min_{x \in \mathcal{R}} \min_{y \in \mathcal{R}_* \setminus \mathcal{B}_x} Q^t(x, \pi(y)) \geq \frac{1 + o_{\mathbb{P}}(1)}{n}. \quad (10)$$

Let \mathcal{H}_x be the set of free stubs (u, i) with $u \in \mathcal{T}_x$ and $h(u, i) = \frac{t}{2}$ at the end of the exploration stage. Similarly, let \mathcal{H}_y be the set of free stubs (v, j) with $v \in \mathcal{T}_y$ and $h(v, j) = \frac{t}{2}$. Let also

$$\theta := \frac{1}{n(\log n)^2}.$$

We have

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

Note that, conditionally on the exploration stage, the sum on the right-hand side above can be written as $\sum_{i \notin \text{Dom}_\tau(\pi)} a_{i, \pi(i)}$, where the numbers $a_{i,j}$ satisfy

$$0 \leq a_{i,j} \leq \theta \quad \text{and} \quad \sum_{\substack{i \notin \text{Dom}_\tau(\pi) \\ j \notin \text{Ran}_\tau(\pi)}} a_{i,j} \leq 1.$$

Define

$$Z_\theta(x, y) = \sum_{(u,i) \in \mathcal{H}_x} \sum_{(v,j) \in \mathcal{H}_y} \mathbf{w}(u,i) \mathbf{w}(v,j) \mathbf{1}_{\mathbf{w}(u,i) \mathbf{w}(v,j) \leq \theta},$$

and note that $(n - \tau)^{-1} Z_\theta(x, y)$ corresponds to the conditional expectation of $\sum_{i \notin \text{Dom}_\tau(\pi)} a_{i, \pi(i)}$ given the exploration stage. Using Chatterjee [10, Proposition 1.1, or rather the refined bound for the left tail given in Theorem 1.5, (ii)], we have, for all $\varepsilon > 0$,

$$\begin{aligned} \mathbb{P}(nQ^t(x, \pi(y)) < Z_\theta(x, y) - \varepsilon) &\leq \mathbb{P}\left((n - \tau) \sum_{i \notin \text{Dom}_\tau(\pi)} a_{i, \pi(i)} < Z_\theta(x, y) - \varepsilon\right) \\ &\leq \exp\left(-\frac{\varepsilon^2}{4\theta n}\right) = o\left(\frac{1}{n^2}\right). \end{aligned}$$

To complete the proof of (10), we now have to show that $\min_{x \in \mathcal{R}} \min_{y \in \mathcal{R}_* \setminus \mathcal{B}_x} Z_\theta(x, y) = 1 + o_{\mathbb{P}}(1)$. To do so, we decompose $1 - Z_\theta(x, y)$ into three terms that are treated separately: the weight of free stubs at height $t/2$ but with product larger than θ , the weight of stubs that are not free and childless (i.e. those that did not satisfy the condition at step (3) of the exploration stage), and the weight of free stubs at height less than $t/2$ (i.e. those that have weight smaller than \mathbf{w}_{\min}).

Lemma 6. *For all $\varepsilon > 0$,*

$$\mathbb{P} \left(\sum_{(u,i) \in \mathcal{H}_x} \sum_{(v,j) \in \mathcal{H}_y} \mathbf{w}(u,i) \mathbf{w}(v,j) \mathbf{1}_{\{\mathbf{w}(u,i) \mathbf{w}(v,j) > \theta\}} > \varepsilon \right) = o \left(\frac{1}{n^2} \right).$$

Proof of Lemma 6. First observe that

$$\sum_{(u,i) \in \mathcal{H}_x} \sum_{(v,j) \in \mathcal{H}_y} \mathbf{w}(u,i) \mathbf{w}(v,j) \mathbf{1}_{\{\mathbf{w}(u,i) \mathbf{w}(v,j) > \theta\}} \leq \sum_{z \in \{x,y\}} \sum_{(u,i) \in \mathcal{H}_z} \mathbf{w}(u,i) \mathbf{1}_{\{\mathbf{w}(u,i) > \sqrt{\theta}\}}.$$

We will show that, for all $\varepsilon > 0$,

$$\mathbb{P} \left(\sum_{(u,i) \in \mathcal{H}_x} \mathbf{w}(u,i) \mathbf{1}_{\{\mathbf{w}(u,i) > \sqrt{\theta}\}} > \varepsilon \right) = o \left(\frac{1}{n^2} \right).$$

Since the argument for y is completely similar, we do not include it (the only difference is that we have to condition also on the whole tree \mathcal{T}_x but this does not change the argument).

Let \mathcal{J}_0 be the σ -field corresponding to steps (i) and (ii), with x_1, \dots, x_m the vertices at height t_0 . For $\ell = 1, \dots, m$, let \mathcal{J}_ℓ be the σ -field containing \mathcal{J}_0 plus the first ℓ steps of step (iii), i.e. the subtrees rooted at x_1, \dots, x_ℓ . Observe that

$$\sum_{(u,i) \in \mathcal{H}_x} \mathbf{w}(u,i) \mathbf{1}_{\{\mathbf{w}(u,i) > \sqrt{\theta}\}} \leq \sum_{\ell=1}^m \mathbf{w}(x_\ell) W_\ell,$$

where

$$W_\ell = \sum_{\substack{(u,i) \in \mathcal{H}_x \\ u_r = x_\ell}} \frac{\mathbf{w}(u,i)}{\mathbf{w}(x_\ell)} \mathbf{1}_{\{\frac{\mathbf{w}(u,i)}{\mathbf{w}(x_\ell)} > \sqrt{\theta}\}}.$$

Now consider the martingale sequence $(M_k)_{k=0}^m$ defined by $M_0 = 0$, and, for $k = 1, \dots, m$,

$$M_k = \sum_{\ell=1}^k \mathbf{w}(x_\ell) (W_\ell - \mathbb{E}[W_\ell \mid \mathcal{J}_{\ell-1}]).$$

Note that

$$\mathbb{E}[W_\ell \mid \mathcal{J}_{\ell-1}] \leq \mathbb{P}_{x_\ell} \left(\prod_{s=0}^{t/2-t_0-1} P(X_s, Y_s) > \sqrt{\theta} \mid \mathcal{J}_{\ell-1} \right).$$

Using the coupling of Section 2.1 and bound (9), we have

$$\mathbb{E}[W_\ell \mid \mathcal{J}_{\ell-1}] \leq O \left(\frac{t^2}{\mathbf{w}_{\min} n} \right) + \mathbb{P}_{x_\ell} \left(\prod_{s=0}^{t/2-t_0-1} P(X_s^*, Y_s^*) > \sqrt{\theta} \right).$$

Singling out the particular first step, taking logarithms, and using that $t_0 \mathbf{h} = o(\log n)$, we have

$$\mathbb{P}_{x_\ell} \left(\prod_{s=0}^{t/2-t_0-1} P(X_s^*, Y_s^*) > \sqrt{\theta} \right) \leq \mathbb{P} \left(\sum_{s=1}^{t/2-t_0-1} \left(\mathbf{h} - \log \frac{1}{P(X_s^*, Y_s^*)} \right) > u_n \right),$$

with $u_n = \frac{(c-1)}{2} \log n + o(\log n)$. Using Chebychev Inequality, we get $\mathbb{E}[W_\ell \mid \mathcal{J}_{\ell-1}] = o(1)$, uniformly in ℓ , which implies

$$\sum_{\ell=1}^m \mathbf{w}(x_\ell) W_\ell \leq M_m + o(1).$$

Now the increments of $(M_k)_{k=0}^m$ are bounded by

$$|M_k - M_{k-1}| \leq \mathbf{w}(x_k) \leq \Delta^{t_0} \leq \frac{1}{\log n},$$

and its quadratic variation satisfies

$$\sum_{k=1}^m \mathbb{E} [(M_k - M_{k-1})^2 \mid \mathcal{J}_{k-1}] \leq \sum_{k=1}^m \mathbf{w}(x_k)^2 \mathbb{E}[W_k \mid \mathcal{J}_{k-1}] = o\left(\frac{1}{\log n}\right).$$

By Freedman Inequality [17, Theorem 1.6] applied to the martingale $(\log(n)M_k)_{k=0}^m$ which has increments bounded by 1 and quadratic variation $v = o(\log n)$, we obtain

$$\mathbb{P}(M_m > \varepsilon) \leq \left(\frac{ve}{v + \varepsilon \log n}\right)^{\varepsilon \log n} = o\left(\frac{1}{n^2}\right). \quad \blacksquare$$

Let \mathcal{F} be the set of free stubs (u, i) with $u \in \mathcal{T}_x \cup \mathcal{T}_y$ at the end of the exploration stage.

Lemma 7. For all $\varepsilon > 0$,

$$\mathbb{P}\left(x \in \mathcal{R}, y \in \mathcal{R}_* \setminus \mathcal{B}_x, \sum_{(u,i) \in \mathcal{F}} \mathbf{w}(u, i) < 2 - \varepsilon\right) = o\left(\frac{1}{n^2}\right).$$

Proof of Lemma 7. For $k \geq 0$, let \mathcal{F}_k denote the set of free stubs (u, i) with $u \in \mathcal{T}_x(k \wedge \tau) \cup \mathcal{T}_y(k \wedge \tau)$ (i.e. after $k \wedge \tau$ rounds of the exploration process), and consider the random variable

$$W_k := \sum_{(u,i) \in \mathcal{F}_k} \mathbf{w}(u, i).$$

Initially, $W_0 = 2$, and, for $k \geq 1$, the difference $W_k - W_{k-1}$ is either 0 or negative, depending on whether the condition of step (3) is satisfied or not. More precisely, denoting by (u_k, i_k) the free stub selected at the k^{th} iteration of step (1) and letting i'_k be the state selected at step (2), we have for all $k \geq 1$,

$$W_k - W_{k-1} = -\mathbf{1}_{k \leq \tau} \mathbf{w}(u_k, i_k) \mathbf{1}_{\mathcal{N}(\{i'_k\}) \cap \mathcal{N}(\mathcal{T}_x(k-1) \cup \mathcal{T}_y(k-1)) \neq \emptyset},$$

where for $A \subset \Omega$,

$$\mathcal{N}(A) = A \cup \{z \in \Omega, \exists a \in A, P(a, z) > 0 \text{ or } P_*(a, z) > 0\}.$$

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 $\mathbf{w}(u_k, i_k)$ is \mathcal{G}_{k-1} -measurable and that

$$\begin{aligned} & \mathbb{P}(\mathcal{N}(\{i'_k\}) \cap \mathcal{N}(\mathcal{T}_x(k-1) \cup \mathcal{T}_y(k-1)) \neq \emptyset \mid \mathcal{G}_{k-1}) \\ & \leq \frac{(1 + 2/\delta)^2 |\mathcal{T}_x(k-1) \cup \mathcal{T}_y(k-1)|}{n - |\text{Ran}_{k-1}(\pi)|} \\ & \leq \frac{9(k+1)}{\delta^2(n-k+1)}, \end{aligned}$$

where we used that $|\text{Ran}_{k-1}(\pi)|$ is at most $k-1$ and $|\mathcal{T}_x(k-1) \cup \mathcal{T}_y(k-1)|$ at most $k+1$. Hence,

$$\mathbb{E}[W_k - W_{k-1} \mid \mathcal{G}_{k-1}] \geq -\mathbf{1}_{k \leq \tau} \mathbf{w}(u_k, i_k) \frac{9(k+1)}{\delta^2(n-k+1)},$$

and

$$\mathbb{E}[(W_k - W_{k-1})^2 \mid \mathcal{G}_{k-1}] \leq \mathbf{1}_{k \leq \tau} \mathbf{w}(u_k, i_k)^2 \frac{9(k+1)}{\delta^2(n-k+1)}.$$

Using that

$$\mathbf{w}_{\min} \tau \leq \sum_{k \leq \tau} \mathbf{w}(u_k, i_k) \leq t,$$

and assumption (2), we get

$$m := \sum_{k=1}^{\tau} \mathbb{E} [W_{k-1} - W_k \mid \mathcal{G}_{k-1}] \leq \frac{n^{o(1)}t^2}{n\mathbf{w}_{\min} - t},$$

and

$$v := \sum_{k=1}^{\tau} \mathbb{E} [(W_k - W_{k-1})^2 \mid \mathcal{G}_{k-1}] \leq \frac{n^{o(1)}t^2}{n\mathbf{w}_{\min} - t}.$$

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

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

The increments of $(M_k)_{k \geq 0}$ are bounded by ε by construction, and, recalling that $\mathbf{w}_{\min} = n^{-2/3}$ by (8) and that $t = O(\log n)$ by (3), the quadratic variation up to time τ satisfies

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

Thus, Freedman Inequality [17] yields

$$\mathbb{P}(M_{\tau} > 7\varepsilon) \leq \left(\frac{ev}{v + 7\varepsilon^2} \right)^7 = o\left(\frac{1}{n^2}\right). \quad (11)$$

Now on the event $x \in \mathcal{R}$ and $y \in \mathcal{R}_{\star} \setminus \mathcal{B}_x$, one must have $\max_k (W_{k-1} - W_k) \leq \varepsilon$ for n large enough (since $\Delta < 1$ and $r \rightarrow +\infty$), hence

$$\{x \in \mathcal{R}, y \in \mathcal{R}_{\star} \setminus \mathcal{B}_x\} \subseteq \{W_0 - W_{\tau} \leq M_{\tau} + m\}.$$

Since $W_0 - W_{\tau} = 2 - \sum_{(u,i) \in \mathcal{F}} \mathbf{w}(u,i)$ and $m = o(1)$, this concludes the proof. \blacksquare

Lemma 8. For all $\varepsilon > 0$,

$$\mathbb{P}\left(\sum_{(u,i) \in \mathcal{F}} \mathbf{w}(u,i) \mathbf{1}_{\{\mathbf{w}(u,i) < \mathbf{w}_{\min}\}} > \varepsilon\right) = o\left(\frac{1}{n^2}\right).$$

Proof of Lemma 8. Let \mathcal{F}_x be the set of free stubs (u,i) with $u \in \mathcal{T}_x$. Consider $m = \lfloor \log n \rfloor$ independent Markov chains with transition matrix Q all started at x , each being killed as soon as its weight falls below \mathbf{w}_{\min} , and write A for the event that their trajectories form a tree of height less than $t/2$. The conditional probability of A given the permutation π satisfies

$$\mathbb{P}(A \mid \pi) \geq \left(\sum_{(u,i) \in \mathcal{F}_x} \mathbf{w}(u,i) \mathbf{1}_{\{\mathbf{w}(u,i) < \mathbf{w}_{\min}\}} \right)^m.$$

Taking expectation and using Markov inequality, we deduce that

$$\mathbb{P}\left(\sum_{(u,i) \in \mathcal{F}_x} \mathbf{w}(u,i) \mathbf{1}_{\{\mathbf{w}(u,i) < \mathbf{w}_{\min}\}} > \varepsilon\right) \leq \frac{\mathbb{P}(A)}{\varepsilon^m},$$

where the average is now taken over both the Markov chains and the permutation. We will now show that $\mathbb{P}(A) = o(1)^m$, so that $\varepsilon^{-m} \mathbb{P}(A) = o(n^{-2})$. Generate the m killed trajectories one after the other, revealing the permutation along the way, as described in Section 2.1. Given that the first $\ell - 1$ trajectories form a tree of height less than $t/2$, we claim that the conditional chance that the ℓ^{th} one also fulfills 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 unvisited state: 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 $\mathbf{w}_{\min}/\eta\delta$ (where δ accounts for the step exiting the graph spanned by the previous trajectories): this part consists of at most $t/2$ states which can be coupled with uniform samples from Ω for a total-variation cost of $\frac{mt^2}{2n}$, as in Section 2.1. Thus, the conditional chance is at most

$$\frac{mt^2}{2n} + \mathbb{P}\left(\prod_{s=1}^{t/2} P(X_s^*, Y_s^*) < \frac{\mathbf{w}_{\min}}{\eta\delta}\right) = o(1),$$

by Chebychev Inequality.

Since P_* satisfies the same assumptions as P , the same argument may be repeated to show that

$$\mathbb{P}\left(\sum_{(u,i) \in \mathcal{F}_y} \mathbf{w}(u,i) \mathbf{1}_{\{\mathbf{w}(u,i) < \mathbf{w}_{\min}\}} > \varepsilon\right) = o\left(\frac{1}{n^2}\right). \quad \blacksquare$$

Combining Lemma 6, 7 and 8, we obtain $\min_{x \in \mathcal{R}} \min_{y \in \mathcal{R}_* \setminus \mathcal{B}_x} Z_\theta(x, y) = 1 + o_{\mathbb{P}}(1)$, which concludes the proof of the upper bound.

2.4. Proof of Proposition 3. In the setting of Corollary 2, let us show that, for $k_n = \Omega(\log n)$, there exist graphs for which the chain Q_n does not have cutoff. Consider a graph G_n formed by the disconnected union of one 3-regular Ramanujan graph R_n of size $r_n \sim \frac{n}{\sqrt{\log n}}$, and of $\frac{n-r_n}{4}$ copies of K_4 (the complete graph on four vertices). Take $k_n \geq a \log n$ with $a > \frac{3}{\log(2)}$, and consider the Markov chain with transition matrix $Q_n = M_n^{k_n} \Pi_n$ where M_n is the transition matrix of the simple random walk on G_n and Π_n is a uniform random permutation matrix. We claim that, starting from one copy of K_4 , the mixing time of Q_n is determined by the hitting time of R_n . Indeed, observe that, when restricted to $G_n \setminus R_n$, the chain Q_n is very similar to a random walk on a 4-regular directed random graph, which takes order $\log n$ steps to mix. Since the hitting time of R_n is approximately distributed as a Geometric random variable with parameter $\frac{1}{\sqrt{\log n}}$, the chain Q_n reaches R_n before it mixes on $G_n \setminus R_n$. Once it has reached R_n , then k_n simple random walk steps makes it basically uniformly distributed on R_n . This is because, by [23], the simple random walk on R_n has cutoff at time $\frac{3}{\log(2)} \log n$. Hence, applying the random permutation makes the chain approximately uniformly distributed over a set of r_n vertices, among which a vanishing proportion belong to R_n . Now, starting from the uniform distribution over a set of about r_n vertices, the chain on $G_n \setminus R_n$ only requires order $\log \log n$ steps to mix (this is because the range at time k is roughly of size $r_n 4^k$ and the walk mixes when the range has size about n). All in all, the mixing time is of order $\sqrt{\log n}$, and there is no cutoff.

3. DETERMINISTIC PERMUTATIONS

Let P be a bistochastic transition matrix on Ω , with $|\Omega| = n$, and let

$$\delta = \min \{P(x, y), x, y \in \Omega, P(x, y) > 0\}.$$

Assume that the matrix P has laziness parameter $\gamma > 0$, i.e.

$$\forall x \in \Omega, P(x, x) \geq \gamma.$$

Let π be a fixed permutation on Ω . For $A \subset \Omega$, we denote by $\mathcal{E}(A)$ the subset of elements attainable in one step of P starting from A , and we assume that there exists $\alpha \in (0, 1)$ such that for all $A \subset \Omega$ with $|A| \leq n/2$,

$$|\mathcal{E} \circ \pi \circ \mathcal{E}(A)| \geq (1 + \alpha)|A|. \quad (12)$$

Consider the matrix $Q = P\Pi$, where Π is the matrix associated to π . We proceed by bounding the mixing time of Q^2 . Note that by assumption (12), the matrix Q^2 is irreducible. For $x, y \in \Omega$, we have

$$Q^2(x, y) = \sum_{z \in \Omega} P(x, \pi^{-1}(z))P(z, \pi^{-1}(y)).$$

Following Morris and Peres [25], we define, for $S \subset \Omega$, $S \neq \emptyset$,

$$\varphi_S = \frac{1}{2|S|} \sum_{y \in \Omega} \min \left\{ \sum_{x \in S} Q^2(x, y), \sum_{x \in S^c} Q^2(x, y) \right\}.$$

Note that, if $y \in \pi \circ \pi(S)$, then for $x = \pi^{-1} \circ \pi^{-1}(y) \in S$, we have

$$Q^2(x, y) \geq P(x, x)P(\pi(x), \pi(x)) \geq \gamma^2.$$

Moreover, since Q^2 has uniform stationary distribution, $\sum_{x \in S^c} Q^2(x, y) \leq 1$. Hence,

$$\sum_{y \in \pi \circ \pi(S)} \min \left\{ \sum_{x \in S} Q^2(x, y), \sum_{x \in S^c} Q^2(x, y) \right\} \geq \gamma^2 \sum_{x \in S^c} Q^2(x, \pi \circ \pi(S)).$$

Similarly,

$$\sum_{y \in \pi \circ \pi(S^c)} \min \left\{ \sum_{x \in S} Q^2(x, y), \sum_{x \in S^c} Q^2(x, y) \right\} \geq \gamma^2 \sum_{x \in S} Q^2(x, \pi \circ \pi(S^c)).$$

Therefore,

$$\begin{aligned} \varphi_S &\geq \frac{\gamma^2}{2|S|} \left\{ \sum_{x \in S} Q^2(x, \pi \circ \pi(S^c)) + \sum_{x \in S^c} Q^2(x, \pi \circ \pi(S)) \right\} \\ &\geq \frac{\gamma^2 \delta^2}{2|S|} \left\{ |\pi^{-1} \circ \mathcal{E} \circ \pi \circ \mathcal{E}(S) \setminus S| + |\pi^{-1} \circ \mathcal{E} \circ \pi \circ \mathcal{E}(S^c) \setminus S^c| \right\}. \end{aligned}$$

Assume without loss of generality that $|S| \leq n/2$. Since P has some laziness, $S \subset \pi^{-1} \circ \mathcal{E} \circ \pi \circ \mathcal{E}(S)$, and, by assumption, we have

$$|\pi^{-1} \circ \mathcal{E} \circ \pi \circ \mathcal{E}(S)| = |\mathcal{E} \circ \pi \circ \mathcal{E}(S)| \geq (1 + \alpha)|S|.$$

We obtain

$$\varphi_\star := \min_{S \subset \Omega, S \neq \emptyset} \varphi_S \geq \frac{\gamma^2 \delta^2 \alpha}{2}.$$

By [25, Theorem 4 and Lemma 10], this implies that the mixing time of Q is bounded by

$$\begin{aligned} t_{\text{mix}}(\varepsilon) &\leq \int_{4/n}^{1/\varepsilon^2} \frac{4}{u \varphi_\star^2} du \\ &\leq \frac{16}{\gamma^4 \delta^4 \alpha^2} \left(\log \left(\frac{n}{4} \right) + 2 \log \left(\frac{1}{\varepsilon} \right) \right). \end{aligned}$$

Denoting by $\mathcal{D}^{(\gamma)}(\cdot)$ the worst case total-variation distance when P has laziness γ , we have in particular that for all γ such that $16 < 17\gamma^4$,

$$\mathcal{D}^{(\gamma)} \left(\frac{17}{\delta^4 \alpha^2} \log n \right) \xrightarrow{n \rightarrow \infty} 0.$$

Using Chen and Saloff-Coste [12, Proposition 3.1], we may then extend this bound to all positive laziness parameters: for all $\gamma \in (0, 1)$, we have

$$\mathcal{D}^{(\gamma)} \left(\frac{17}{\delta^4 \alpha^2} \log n \right) \xrightarrow{n \rightarrow \infty} 0.$$

Acknowledgments. We thank Charles Bordenave and Persi Diaconis for helpful discussions. We are also very grateful to Sam Olesker-Taylor for his careful reading and for pointing out to us confusing typos in a previous version.

REFERENCES

- [1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.
- [2] R. Basu, J. Hermon, and Y. Peres. Characterization of cutoff for reversible markov chains. *The Annals of Probability*, 45(3):1448–1487, 2017.
- [3] A. Ben-Hamou and J. Salez. Cutoff for non-backtracking random walks on sparse random graphs. *The Annals of Probability*, 45(3):1752–1770, 2017.
- [4] A. Ben-Hamou, E. Lubetzky, and Y. Peres. Comparing mixing times on sparse random graphs. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 55(2):1116 – 1130, 2019.
- [5] N. Berestycki, E. Lubetzky, Y. Peres, A. Sly, et al. Random walks on the random graph. *The Annals of Probability*, 46(1):456–490, 2018.
- [6] C. Bordenave and H. Lacoin. Cutoff at the entropic time for random walks on covered expander graphs. *Journal of the Institute of Mathematics of Jussieu*, pages 1–46, 2021. doi: 10.1017/S1474748020000663.
- [7] C. Bordenave, P. Caputo, and J. Salez. Random walk on sparse random digraphs. *Probability Theory and Related Fields*, 170(3-4):933–960, 2018.
- [8] C. Bordenave, P. Caputo, and J. Salez. Cutoff at the “entropic time” for sparse markov chains. *Probability Theory and Related Fields*, 173(1):261–292, 2019.
- [9] C. Bordenave, Y. Qiu, Y. Zhang, et al. Spectral gap of sparse bistochastic matrices with exchangeable rows. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(4): 2971–2995, 2020.
- [10] S. Chatterjee. Stein’s method for concentration inequalities. *Probability theory and related fields*, 138(1):305–321, 2007.
- [11] S. Chatterjee and P. Diaconis. Speeding up markov chains with deterministic jumps. *Probability Theory and Related Fields*, 178(3):1193–1214, 2020.
- [12] G.-Y. Chen and L. Saloff-Coste. Comparison of cutoffs between lazy walks and markovian semigroups. *Journal of Applied Probability*, 50(4):943–959, 2013.
- [13] G. Conchon. Cutoff for random lifts of weighted graphs. 2019.
- [14] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 57(2):159–179, 1981.
- [15] J. Ding, E. Lubetzky, and Y. Peres. Total variation cutoff in birth-and-death chains. *Probability theory and related fields*, 146(1-2):61–85, 2010.
- [16] S. Eberhard and P. P. Varjú. Mixing time of the Chung–Diaconis–Graham random process. *Probability Theory and Related Fields*, 179(1):317–344, 2021.
- [17] D. A. Freedman. On tail probabilities for martingales. *The Annals of Probability*, pages 100–118, 1975.
- [18] J. He, H. T. Pham, and M. W. Xu. Mixing time of fractional random walk on finite fields. *arXiv preprint arXiv:2102.02781*, 2021.
- [19] J. Hermon and S. Olesker-Taylor. Cutoff for almost all random walks on abelian groups. *arXiv preprint arXiv:2102.02809*, 2021.
- [20] J. Hermon and S. Olesker-Taylor. Cutoff for random walks on upper triangular matrices. *arXiv preprint arXiv:1911.02974*, 2021.
- [21] J. Hermon, A. Sly, and P. Sousi. Universality of cutoff for graphs with an added random matching. *arXiv preprint arXiv:2008.08564*, 2020.
- [22] M. Hildebrand. A survey of results on random random walks on finite groups. *Probability Surveys*, 2:33–63, 2005.
- [23] E. Lubetzky and Y. Peres. Cutoff on all ramanujan graphs. *Geometric and Functional Analysis*, 26(4):1190–1216, 2016.
- [24] E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010.

- [25] B. Morris and Y. Peres. Evolving sets, mixing and heat kernel bounds. *Probability Theory and Related Fields*, 133(2):245–266, 2005.
- [26] J. Salez. Cutoff for non-negatively curved markov chains. *arXiv preprint arXiv:2102.05597*, 2021.