CUTOFF AT THE ENTROPIC TIME FOR RANDOM WALKS ON COVERED EXPANDER GRAPHS

CHARLES BORDENAVE AND HUBERT LACOIN

ABSTRACT. It is a fact simple to establish that the mixing time of the simple random walk on a d-regular graph G_n with n vertices is asymptotically bounded from below by $\frac{d}{d-2}\frac{\log n}{\log(d-1)}$. Such a bound is obtained by comparing the walk on G_n to the walk on d-regular tree $\mathcal{T}_d$. If one can map another transitive graph $\mathcal{G}$ onto G_n, then we can improve the strategy by using a comparison with the random walk on $\mathcal{G}$ (instead of that of $\mathcal{T}_d$), and we obtain a lower bound of the form $\frac{1}{\mathfrak{h}}\log n$, where $\mathfrak{h}$ is the entropy rate associated with $\mathcal{G}$. We call this the entropic lower bound.

It was recently proved that in the case $\mathcal{G} = \mathcal{T}_d$, this entropic lower bound (in that case $\frac{d}{d-2}\frac{\log n}{\log(d-1)}$) is sharp when graphs have minimal spectral radius and thus that in that case the random walk exhibit cutoff at the entropic time. In this paper, we provide a generalization of the result by providing a sufficient condition on the spectra of the random walks on G_n under which the random walk exhibits cutoff at the entropic time. It applies notably to anisotropic random walks on random d-regular graphs and to random walks on random n-lifts of a base graph (including non-reversible walks).

1. INTRODUCTION

This paper is aimed at understanding the mixing properties of random walks on a finite regular graph. We are going to be focused on asymptotic properties when the number of vertices goes to infinity.

Minimal mixing time for the simple random walk. Let $3 \le d \le n-1$ be integers with nd even and let $G_n = (V_n, E_n)$ be a finite simple d-regular graph on a vertex set V_n of size $\#V_n = n$. Let $(X_t)_{t\ge 0}$ be the *simple random walk* on G_n, which is the Markov process taking values in V_n with transition matrix,

$$P_n(x,y) = \frac{\mathbf{1}_{\{\{x,y\}\in E_n\}}}{d} \qquad \text{for } x,y \in V_n.$$

The uniform measure on V_n denoted by π_n is reversible for the process. Furthermore if G_n is connected, then π_n is the unique invariant probability measure of P_n. If additionally G_n is not bipartite, then $P_n^t(x,\cdot)$ converges to π_n when t tends to infinity.

We are interested in estimating the time at which $P_n^t(x,\cdot)$ falls in a close neighborhood of π_n in terms of the *total variation* distance. More formally, the *total variation mixing time* associated

with threshold $\varepsilon \in (0, 1)$ and initial condition $x \in V_n$, is defined by

$$T_n^{\mathrm{mix}}(x, \varepsilon) := \inf \left\{ t \in \mathbb{N} \ : \ d_n(x, t) < \varepsilon \right\},$$

where $d_n(x, t)$ is the *total variation* distance to equilibrium

(1) $\qquad d_n(x, t) := \|P_n^t(x, \cdot) - \pi_n\|_{\mathrm{TV}} = \frac{1}{2} \sum_{y \in V_n} \left| P_n^t(x, y) - \pi_n(y) \right| = \max_{A \subset V_n} \left\{ P_n^t(x, A) - \pi_n(A) \right\}.$

The *worst-case mixing time* is classically defined as

$$T_n^{\mathrm{mix}}(\varepsilon) = \max_{x \in V_n} T_n^{\mathrm{mix}}(x, \varepsilon).$$

The mixing properties for the random walk are intimately related to the spectrum of P_n. An illustration of this is the classical computation based on the spectral decomposition of P_n (see [36, Theorem 12.4] for a proof in the reversible case) which allows to control the distance as a function of the *singular radius* of P_n. For all $x \in V_n$,

(2) $\qquad\qquad\qquad\qquad d_n(x, t) \leq \frac{\sqrt{n-1}}{2} \sigma_n^t.$

where the singular radius σ_n

$$\sigma_n = \|(P_n)_{|\mathbf{1}^\perp}\|_{2 \to 2}$$

is the ℓ_2 operator norm of P_n restricted to functions with zero sum. Since P_n is reversible, we have $\sigma_n = \varrho_n$ where ϱ_n is the spectral radius of P_n, that is the second largest eigenvalue of P_n in absolute value counting multiplicities. This yields in particular that

(3) $\qquad\qquad\qquad T_n^{\mathrm{mix}}(\varepsilon) \leq \frac{1}{|\log \varrho_n|} \left(\frac{1}{2} \log n - \log(2\varepsilon) \right).$

In particular, if we have $\varrho_n < 1 - \delta$ for some fixed $\delta \in (0, 1)$ along some sequence of integers going to infinity, then the upper bound in (3) is of order $\log n$ along that sequence.

On the other hand, a naive lower bound of the same order of $T_n^{\mathrm{mix}}(\varepsilon)$ can be obtaind by using the elementary fact that the graph distance $\mathrm{Dist}(x, X_t)$ between X_t and the initial condition x is stochastically dominated by a random walk on the set of non-negative integers, started at 0, with jump probabilities $1/d$ to the left and $(d-1)/d$ to the right, except at 0 where the probability to jump to the right is equal to 1. Thus when starting from $X_0 = x$, X_t remains within distance r from x at least during a random time of order $\frac{d}{d-2} r + O(\sqrt{r})$. Combining this with the fact that a ball of radius r contains at most $d(d-1)^{r-1}$ vertices, we obtain that for any $x \in V_n$

(4) $\qquad\qquad\qquad T_n^{\mathrm{mix}}(x, 1 - \varepsilon) \geq \frac{d}{(d-2)\log(d-1)} \left(\log n - C_\varepsilon \sqrt{\log n} \right).$

While the strategy might seem a bit rough, the above bound (4) can be sharp. This was first discovered for random d-regular graphs in [40].

However, an important observation is that the factor in front of $\log n$ in (3) and (4) cannot match. From Alon-Boppana lower bound [4, 42], we have for any sequence of d-regular graphs

$(G_n)_{n \geq 0}$ on n vertices, we have $\liminf_n \varrho_n \geq \varrho := \frac{2\sqrt{d-1}}{d}$. More precisely, there exists a constant $C = C(d)$ such that for every n and every d-regular graph on n vertices

$$(5) \qquad \varrho_n \geq \varrho - \frac{C}{(\log n)^2}.$$

The number $\varrho = 2\sqrt{d-1}/d$ is the spectral radius of the simple random walk on the infinite d-regular tree $\mathcal{T}_d$ (and incidentally also that of the biased random walk on integers used in the lower-bound strategy). A graph such that $\varrho_n \leq \varrho$ is called a (non-bipartite) *Ramanujan graph*. Hence a natural question is the following: *If a sequence of graphs on n vertices has an asymptotically minimal spectral radius in the sense that $\varrho_n = (1 + o(1))\varrho$, does it also have a minimal mixing time in the sense that $T_n^{\mathrm{mix}}(\varepsilon) = (1 + o(1))\frac{(d-2)}{d\log(d-1)} \log n$ for any fixed $\varepsilon \in (0,1)$?*

An affirmative answer was given to this question in [39] (see also [29]):

Theorem A ([39]). *Let $d \geq 3$ be an integer and let $(G_n)_{n \in \mathbb{N}}$ be a sequence of d-regular graphs on n vertices, for which the associated sequence of spectral radii satisfy $\lim_{n \to \infty} \varrho_n = \varrho = 2\sqrt{d-1}/d$. Then for any $\varepsilon \in (0,1)$, we have*

$$(6) \qquad \lim_{n \to \infty} \frac{T_n^{\mathrm{mix}}(\varepsilon)}{\log n} = \frac{d}{(d-2)\log(d-1)}.$$

Remark 1.1. *The result above remains of course valid if our sequence (G_n) is indexed by an infinite subset of $\mathbb{N}$ provided that ϱ_n converges to ϱ when $n \to \infty$ in this subset. In the remainder of the paper, with a some small abuse of notation, when using* $\lim$, *we always assume that the considered sequence may not be defined for every n.*

Theorem A is an illustration of the *cutoff phenomenon*. A sequence of finite Markov chains corresponding to the sequence of transition matrices (P_n) exhibits cutoff if up to first order in n, the mixing time $T_n^{\mathrm{mix}}(\varepsilon)$ does not depend on $\varepsilon \in (0,1)$, that is, for any $\varepsilon \in (0,1)$, $\lim_{n \to \infty} T_n^{\mathrm{mix}}(\varepsilon)/T_n^{\mathrm{mix}}(1 - \varepsilon) = 1$. Since its original discovery by Diaconis, Shashahani and Aldous in the context of card shuffling [2, 3, 21], this phenomenon has attracted much attention. We refer to [20, 36] for an introduction and to [9] for an alternative characterization of cutoff. For other recent contributions on cutoff for random walks on graphs with bounded degrees, see [10, 11, 14].

As a warmup, we provide a novel proof of Theorem A, which is simpler than those presented in [39] and [29] (independently observed by Lubetzky [38]). A more precise version of Theorem A will be proved in Proposition 11 below (it notably allows to obtain the second order term in the asymptotic development of $T_n^{\mathrm{mix}}(\varepsilon)$). With our approach we can also relax the assumption by allowing the presence of n^α eigenvalues at a positive distance from the interval $[-\varrho, \varrho]$, with $\alpha \in (0,1)$ small enough, at the cost of discarding a small proportion of possibly bad starting points (the methods in [29, 39] only allow for $n^{o(1)}$ outlying eigenvalues, see remark below [39, Corollary 5]). More precisely, given (G_n) a sequence of d-regular graphs on n vertices, we define the upper

3

semi-continuous function $I : [0,1] \to \{-\infty\} \cup [0,1]$, which can be interpreted as an asymptotic density of eigenvalues on log-log scale

$$(7) \qquad I(u) = \inf_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{\log \left(\sum_{\{\lambda \in \mathrm{Sp}(P_n) \,:\, ||\lambda| - u| < \varepsilon\}} \dim(E_n^\lambda) \right)}{\log n},$$

where $\dim(E_n^\lambda)$ denotes the dimension of the eigenspace corresponding to λ.

Theorem B. *Let $\delta \in (0,1)$, $d \geq 3$ an integer and let (G_n) be a sequence of d-regular graphs on n vertices whose spectral radii satisfy for all n, $\varrho_n \leq 1 - \delta$ and for all $u > \varrho$,*

$$(8) \qquad I(u) \leq 1 - 2 \frac{\log(u/\varrho + \sqrt{(u/\varrho)^2 - 1})}{\log(d-1)}.$$

Then, there exists $c = c(\delta, d) > 0$ such that for any $\varepsilon \in (0,1)$ and $\eta > 0$,

$$(9) \qquad \lim_{n \to \infty} \# \left\{ x \in V_n \; : \; \frac{T_n^{\mathrm{mix}}(x, \varepsilon)}{\log n} \geq (1 + \eta) \frac{d}{(d-2)\log(d-1)} \right\} / n^{1-c\eta} = 0.$$

We note that if the graph G_n is transitive (that is for any pair $x, y \in V_n$, there exists an automorphism of G_n which maps x to y) then $T_n^{\mathrm{mix}}(x, \varepsilon)$ does not depend on x, and (9) implies that $\lim_{n \to \infty} T_n^{\mathrm{mix}}(\varepsilon) / \log n = d/((d-2)\log(d-1))$. See Remark 3.1 for a variant of Theorem B which allows to control $T_n^{\mathrm{mix}}(\varepsilon)$ at the cost of modifying the definition of the function $I(u)$. The principal aim of this paper is to obtain a better understanding of this phenomenon via bringing the question to a larger setup.

Minimal mixing time for the anisotropic random walk. A first possible extension is to consider a random walk on G_n with non-uniform jump rates. For $d \in \mathbb{N}$, we set $[d] = \{1, \ldots, d\}$. One way to define a biaised random walk on a d-regular graph $G_n = (V_n, E_n)$ with $\#V_n = n$ is to assume that E_n can be partitioned into d sets of edges $(E_{n,i})_{i \in [d]}$ where each vertex of V_n is adjacent to exactly one edge of each type (this implies in particular that n is even), and to associate a transition rate p_i to each type of edge with $\sum_{i \in [d]} p_i = 1$. For more generality, we consider an involution $* : i \mapsto i^*$ of $[d] = \{1, \ldots, d\}$. We are going to make the weaker assumption that G_n is a *Schreier graph*. This means that its adjacency matrix P_n may be written as a sum of permutation matrices. That is to say that for all $x, y \in V_n$, we have

$$(10) \qquad P_n(x, y) = \sum_{i=1}^{d} S_i(x, y),$$

where, for every $i \in [d]$, $S_i(x, y) = \mathbf{1}(x = \alpha_i(y))$ for some permutation α_i on V_n and $S_{i*} = S_i^{-1}$. In full generality the expression (10) allows for both $P_n(x, y) \geq 2$, and $P_n(x, x) \geq 1$, so that we consider graphs which may include loops and/or multiple edges. For example, if the involution $*$ on $[d]$ is the identity, then the permutations α_i are involutions: for every $i \in [d]$, we have $\alpha_i^{-1} = \alpha_{i*} = \alpha_i$. In this case, the sets $(E_{n,i})_{i \in [d]}$ defined for every $i \in [d]$ by $E_{n,i} = \{\{x, \alpha_i(x)\} : x \in V_n\}$ is a partition of the edge set E_n. We thus recover the above setting.

4

If d is even, any finite d-regular graph is a Schreier graph for some collection of $d/2$ permutations and their inverses (another formulation of this result is: *any 2k-regular graph is 2-factorable* see [43], this is now a standard exercise in graph theory and can be proved using König's Theorem for bipartite graphs, see e.g. [37, Theorem 6.2.4]).

This definition of Schreier graphs can be extended to regular graphs on countably many vertices. Note that any *Cayley graph* of a finitely generated group with a symmetric set of generators of size d is a Schreier graph: the natural choice for the permutations S_i in (10) corresponds to the (left or right) multiplication by an element of the symmetric set of generators, the involution maps a generator to its inverse.

Now we consider G_n is a d-regular Schreier graph with $\#V_n = n$, given with an involution $*$ and a decompotion of the adjacency matrix into permutations (10), and $\mathbf{p} = (p_1, \ldots, p_d)$ a probability vector (that is, a vector whose coordinate are nonnegative and sum to one) we define the matrix

$$(11) \qquad\qquad P_{n,\mathbf{p}} = \sum_{i=1}^{d} p_i S_i.$$

Note that by construction $P_{n,\mathbf{p}}$ is a stochastic matrix. This is the transition kernel of a random walk on G_n which we refer to as the **p**-*anisotropic random walk*. Again, π_n, the uniform measure on V_n, is invariant for this process. We are going to assume that

$$(12) \qquad\qquad d \geq 3 \quad \text{and} \quad \forall i \in [d], \ p_i + p_{i^*} > 0.$$

The condition $p_i + p_{i^*} > 0$ is not really a restriction since it can be satisfied by just eliminating the coordinates for which $p_i + p_{i^*} = 0$. The condition $d \geq 3$ (which is not the same as asking that three coordinates of $\mathbf{p}$ are positive) is very natural and justified below Equation (19). The singular radius of $P_{n,\mathbf{p}}$ is defined as the ℓ_2 operator norm of $P_{n,\mathbf{p}}$ projected onto the orthogonal of constant functions

$$(13) \qquad\qquad \sigma_{n,\mathbf{p}} = \|(P_{n,\mathbf{p}})_{\mathbf{1}^\perp}\|_{2 \to 2}.$$

Recall that the *singular values* of a matrix T are the square of the eigenvalues of TT^*. By definition $\sigma_{n,\mathbf{p}}$ is the second largest singular value of $P_{n,\mathbf{p}}$ (we are counting eigenvalues with multiplicities, meaning that $\sigma_{n,\mathbf{p}} = 1$ for a non-connected graph). If $P_{n,\mathbf{p}}$ is reversible then $\sigma_{n,\mathbf{p}}$ coincides with the spectral radius $\varrho_{n,\mathbf{p}}$, that is the second largest eigenvalue of $P_{n,\mathbf{p}}$ in absolute value. Note that $P_{n,\mathbf{p}}$ is reversible if the following condition holds:

$$(14) \qquad\qquad \forall i \in [d], \quad p_{i^*} = p_i.$$

Our aim is to prove a result analogous to Theorem A for **p**-anisotropic walks on Schreier graphs. We fix the involution $*$ and $\mathbf{p}$ and then investigate the asymptotic behavior of the mixing time for **p**-anisotropic random walks on a sequence of Schreier graphs (G_n) associated with the involution $*$.

Instead of comparing the spectral radius of P_n with that of the simple random walk on the d-regular tree, we need here to compare it with that of a **p**-anisotropic walk on the tree $\mathcal{T}_d$ considered as a Cayley graph. There are several natural ways to endow $\mathcal{T}_d$ with a Cayley graph structure. For instance, we can consider k free copies of $\mathbb{Z}/2\mathbb{Z}$ and l free copies of $\mathbb{Z}$ with their natural generators, for any value of k and l satisfying $k + 2l = d$. We are going to choose k to be equal to the number of fixed points of $*$ so that the infinite object we consider has a structure which is analogous to our finite Schreier graphs (with Definition 2 below we will formalize this remark).

Using the Schreier graph structure of $\mathcal{T}_d$ considered as a Cayley graph, we define in a manner analogous to (11) the **p**-anisotropic random walk on $\mathcal{T}_d$. We denote by $\mathcal{P}_{\mathbf{p}}$ its transition kernel. These random walk have been extensively studied in the literature (see e.g. [22, 24, 34]).

In analogy with (5), in the reversible case where (14) holds, one can asymptotically compare the spectral radius of $P_{n,\mathbf{p}}$ with that of $\mathcal{P}_{\mathbf{p}}$. From [17, 26], the Alon-Bopanna lower bound for the spectral radius states that for any sequence of Schreier graphs we have

$$(15) \qquad \liminf_{n \to \infty} \varrho_{n,\mathbf{p}} \geq \varrho_{\mathbf{p}},$$

where $\varrho_{\mathbf{p}}$ is the spectral radius of $\mathcal{P}_{\mathbf{p}}$, given by the classical Akemann-Ostrand formula [1]. In the general case, a lower bound of this type holds for the singular radii of powers of $P_{n,\mathbf{p}}$. More precisely, for integer $t \geq 1$, we define the *t-th singular radius* as

$$(16) \qquad \sigma_{n,\mathbf{p}}(t) = \|(P_{n,\mathbf{p}}^t)_{\mathbf{1}^{\perp}}\|_{2 \to 2}^{1/t} \quad \text{and} \quad \sigma_{\mathbf{p}}(t) = \|(\mathcal{P}_{\mathbf{p}}^t)\|_{2 \to 2}^{1/t}.$$

We simply write $\sigma_{n,\mathbf{p}}$ and $\sigma_{\mathbf{p}}$ when $t = 1$. Moreover, Gelfand's formula asserts that the t-th singular radius converges to the spectral radius,

$$(17) \qquad \lim_{t \to \infty} \sigma_{n,\mathbf{p}}(t) = \varrho_{n,\mathbf{p}} \quad \text{and} \quad \lim_{t \to \infty} \sigma_{\mathbf{p}}(t) = \varrho_{\mathbf{p}}.$$

Note that if (14) holds then for any $t \geq 1$, $\sigma_{n,\mathbf{p}}(t) = \varrho_{n,\mathbf{p}}$ and $\sigma_{\mathbf{p}}(t) = \varrho_{\mathbf{p}}$. Beware here and througout this text that the spectral radius $\varrho_{\mathbf{p}}$ is the spectral radius of the bounded operator $\mathcal{P}_{\mathbf{p}}$ on $\ell^2(\mathcal{G})$. It can differ (in fact it is larger or equal to) from what is often called in the literature the spectral radius of the walk which is the asymptotic rate of decay the return probability, that is $\lim_{t \to \infty} \mathcal{P}_{\mathbf{p}}^{2t}(e, e)^{1/(2t)}$ (in the reversible case, the two notions coincide). From [26], the Alon-Bopanna lower bound claims that for any fixed t, for any sequence of Schreier graphs we have

$$(18) \qquad \liminf_{n \to \infty} \sigma_{n,\mathbf{p}}(t) \geq \sigma_{\mathbf{p}}(t),$$

In particular, from Gelfand's formula (17), we get

$$(19) \qquad \lim_{t \to \infty} \liminf_{n \to \infty} \sigma_{n,\mathbf{p}}(t) \geq \varrho_{\mathbf{p}}.$$

The latter formula can be thought as an extension of (15) to the non-reversible case. Note that our assumption (12) simply corresponds to assuming that $\varrho_{\mathbf{p}} < 1$ (in the discarded cases, the anisotropic random walk on $\mathcal{T}_d$ remains on a subset of the tree which is homemorphic to $\mathbb{Z}$).

Adapting the reasoning which yields (4) to the isotropic case, we can also obtain an asymptotic lower bound in n for the mixing time for the $\mathbf{p}$-anisotropic random walk on G_n. Consider $(\mathcal{X}_t)_{t\geq 0}$ an anisotropic random walk on $\mathcal{T}_d$ with transition kernel $\mathcal{P}_\mathbf{p}$ and starting from the root of $\mathcal{T}_d$ denoted by e. Introduced by Avez [8], the *entropy rate* $\mathfrak{h}(\mathbf{p})$ of $\mathcal{P}_\mathbf{p}$ is defined as

$$(20) \qquad \mathfrak{h}(\mathbf{p}) := \lim_{t\to\infty} -\frac{1}{t}\sum_{g\in\mathcal{T}_d}\mathcal{P}_\mathbf{p}^t(e,g)\log\mathcal{P}_\mathbf{p}^t(e,g).$$

We have $\mathfrak{h}(\mathbf{p}) > 0$ as soon as (12) holds. From Shannon-McMillan-Breiman Theorem [31, Theorem 2.1], we have almost surely

$$(21) \qquad \lim_{t\to\infty}\frac{\log\mathcal{P}_\mathbf{p}^t(e,\mathcal{X}_t)}{t} = -\mathfrak{h}(\mathbf{p}).$$

A way to interpret this convergence is to say that at large times t, the marginal distribution of $\mathcal{X}_t$ is roughly uniform on a (deterministic) set of size $e^{\mathfrak{h}(\mathbf{p})t(1+o(1))}$. We have chosen our setup so that we can construct the random walk on G_n by taking the image of $\mathcal{X}_t$ by some function $\mathcal{T}_d \to V_n$ (see Definition 2). Thus for any time $t > 0$ and $x \in V_n$, the entropy of $P_{n,\mathbf{p}}^t(x,\cdot)$ is at most the entropy of $\mathcal{P}_\mathbf{p}^t(e,\cdot)$ (details are in Proposition 6 below). As a consequence, for any fixed $\varepsilon \in (0,1)$, we have

$$(22) \qquad \liminf_{t\to\infty}\min_{x\in V_n}\frac{T_{n,\mathbf{p}}^{\mathrm{mix}}(x,1-\varepsilon)}{\log n} \geq \frac{1}{\mathfrak{h}(\mathbf{p})}.$$

In the spirit of Theorem A, for a given probability vector $\mathbf{p}$, a natural question is thus the following: *If a sequence of graphs on n vertices has a minimal asymptotic spectral radius in the sense that $\lim_{t\to\infty}\limsup_{n\to\infty}\sigma_{n,\mathbf{p}}(t) = \varrho_\mathbf{p}$, does it also have an asympotic minimal mixing time in the sense that $\lim_{n\to\infty}\frac{T_n^{\mathrm{mix}}(\varepsilon)}{\log n} = (\log n)/\mathfrak{h}(\mathbf{p})$ for any fixed $\varepsilon \in (0,1)$?*

It turns out that in the anisotropic setup, the relation between spectral gap and mixing time could be more subtle. We have an asymptotically minimal mixing time for the $\mathbf{p}$-anisotropic random walk if the spectral radius is asymptotically minimal for another anisotropy vector $\mathbf{p}'$.

Theorem 1. *Let $d \geq 3$ be an integer, $*$ an involution on $[d]$ and let $\mathbf{p}$ be a probability vector on $[d]$ which satisfies the condition (12). Then, there exists another probability vector $\mathbf{p}'$ with the same support than $\mathbf{p}$ such that the following holds. If a sequence of Schreier graphs G_n on n vertices as in (10) satisfies for all integers $t \geq 1$,*

$$(23) \qquad \lim_{n\to\infty}\sigma_{n,\mathbf{p}'}(t) = \sigma_{\mathbf{p}'}(t),$$

then for every $\varepsilon \in (0,1)$

$$(24) \qquad \lim_{n\to\infty}\frac{T_{n,\mathbf{p}}^{\mathrm{mix}}(\varepsilon)}{\log n} = \frac{1}{\mathfrak{h}(\mathbf{p})}.$$

Finally, if $\mathbf{p}$ satisfies (14) then $\mathbf{p}'$ also satisfies (14).

The condition (23) can be thought as a Ramanujan property for the anisotropic random walk with probability $\mathbf{p}'$. If (14) holds, condition (23) is equivalent to

$$\lim_{n\to\infty} \varrho_{n,\mathbf{p}'} = \varrho_{\mathbf{p}'}.$$

In the non-reversible case, since $\sigma_{n,\mathbf{p}}(t) \geq \varrho_{n,\mathbf{p}}$ for any $t \geq 1$, condition (23) implies that

$$\limsup_{n\to\infty} \varrho_{n,\mathbf{p}'} \leq \varrho_{\mathbf{p}'}.$$

Note also that in some cases, this condition (23) can be relaxed to allow for $n^{o(1)}$ singular values outside a neighborhood of the interval $[-\varrho_{\mathbf{p}'}, \varrho_{\mathbf{p}'}]$; see Remark 5.1 below. An explicit expression for the vector $\mathbf{p}'$ is provided in the proof. In particular we have that $\mathbf{p}' = \mathbf{p}$ in only two cases. The first one is the simple random walk: that is $\mathbf{p}$ is the uniform vector ($p_i = 1/d$ for all $i \in [d]$), and our result is thus a generalization of Theorem A. The other case is the totally asymetric isotropic walk. It corresponds to the case where $*$ has no fixed point and we have

$$\forall i \in [d], \quad p_i p_{i*} = 0 \quad \text{and} \quad p_i + p_{i*} = \frac{2}{d}.$$

In that case we have $\varrho_{\mathbf{p}} = \sqrt{2/d}$ (see [28, Example 5.5]) and $\mathfrak{h}(\mathbf{p}) = \log \frac{d}{2}$. From Poincaré inequality (32) below, Theorem 1 is extremely easy to prove in this case. For $\mathbf{p}$ different from the uniform vector, a source of example for Theorem 1 is in [15]. Up to the involution, we consider independent permutations σ_i on $[n]$ vertices: if $i \neq i^*$, σ_i is a uniform permutation on n elements and, if $i^* = i$, we take n even and σ_i is a uniform matching on n elements (where a matching is an involution without fixed point). Then, in probability, the condition (23) is true for any probability vector $\mathbf{p}'$ which satisfies the condition (14).

A couple of open questions concerning anisotropic random walks. Let us focus for simplicity on the reversible case (14). We emphasize again that as soon as $\mathbf{p}$ is not the uniform vector, the condition (23) differs from what would be the most natural generalization of the Ramanujan property in the anisotropic setup. We call this property $\mathfrak{R}(\mathbf{p})$:

$$(25) \qquad\qquad\qquad \lim_{n\to\infty} \varrho_{n,\mathbf{p}} = \varrho_{\mathbf{p}}.$$

We believe that this is not an artifact of our proof and that the result would be false if (23) is replaced by (25). More precisely we believe that for every $\mathbf{p}$ which is not the uniform vector there should exists sequences of graph satisfying (25) but such that (24) does not hold. We cannot however prove that $\mathfrak{R}(\mathbf{p})$ and $\mathfrak{R}(\mathbf{p}')$ are not equivalent. In fact, this question yields two natural open problems:

(A) With the setup described above can one find a sequence of Schreier graphs and two probability vectors $\mathbf{p}$ and $\mathbf{q}$ satisfying (14) such that $\mathfrak{R}(\mathbf{p})$ holds and $\mathfrak{R}(\mathbf{q})$ does not?

(B) Given $\mathbf{p} \neq \mathbf{q}$ satisfying (14) can one find a sequence of Schreier graphs such that $\mathfrak{R}(\mathbf{p})$ holds and $\mathfrak{R}(\mathbf{q})$ does not?

While we believe that the answer to (B) (and hence also to (A)) is positive, to our knowledge, all known examples of graphs satisfying $\mathfrak{R}(\mathbf{p})$ for one value of $\mathbf{p}$ satisfy the property for all values of $\mathbf{p}$.

Minimal mixing time for covered random walks. We now present another extension of Theorem A. We start with an extension of the notion of Schreier graph beyond the case of the free group.

Definition 2 (Group action, covering and Schreier graph). *Let $\mathcal{G}$ be a finitely generated group with unit e and V a finite set. A map $\varphi : \mathcal{G} \times V \to V$ is an action of $\mathcal{G}$ on V, if we have*

$$\forall x \in V, \ \forall g, h \in \mathcal{G}, \quad \varphi(e, x) = x \quad and \quad \varphi(gh, x) = \varphi(g, \varphi(h, x))$$

For any $g \in \mathcal{G}$, we denote by S_g the permutation matrix on V associated to the bijective map on $V: x \mapsto \varphi(g, x)$.

If $\mathcal{A}$ is a finite symmetric subset of $\mathcal{G}$ then the Schreier graph of $(\mathcal{G}, \mathcal{A}, \varphi)$ is the graph (with possible loops and multiple edges) on V whose adjacency matrix is $\sum_{g \in \mathcal{A}} S_g$. If $G = (V, E)$ is the Schreier graph of $(\mathcal{G}, \mathcal{A}, \varphi)$, we say that $(\mathcal{G}, \mathcal{A})$ is a covering of G.

Let us check that this definition of a Schreier graph is equivalent to that given earlier. If the adjacency matrix of a finite graph $G = (V, E)$ is of the form (10), then G is the Schreier graph of $(\mathcal{S}_V, \mathcal{A}, \varphi)$ where $\mathcal{S}_V$ is the symmetric group on V, $\mathcal{A} = (S_1, \ldots, S_d)$ and the covering map is $\varphi(\sigma, x) = \sigma(x)$. Conversely, if G is the Schreier graph of $(\mathcal{G}, \mathcal{A}, \varphi)$ as in Definition 2 with $\mathcal{A} = \{a_1, \ldots, a_d\}$ then its adjacency matrix is of the form (10) where the involution $* : i \mapsto i^*$ is defined as $i^* = j$ if and only if $a_j = a_i^{-1}$. Note that if the involution $*$ on $[d]$ has q_1 fixed points and $q_1 + q_2$ equivalence classes then G is d-regular with $d = q_1 + 2q_2$. As already pointed, the infinite tree $\mathcal{T}_d$ is the Cayley graph of the group $\mathcal{G}_{\text{free}}^{(q_1, q_2)}$ generated by q_1 free copies of $\mathbb{Z}$ and q_2 free copies of $\mathbb{Z}/2\mathbb{Z}$ with their natural generators denoted $\mathcal{A}_{\text{free}}$. By considering the group homomorphism from $\mathcal{G}_{\text{free}}^{q_1, q_2}$ to $\mathcal{G}$ which maps $\mathcal{A}_{\text{free}}$ to $\mathcal{A}$, we deduce that *all Schreier graphs are covered by $(\mathcal{G}_{\text{free}}^{q_1, q_2}, \mathcal{A}_{\text{free}})$* with the corresponding involution.

Remark 1.2. *Note that if $\mathbf{p}$ is fixed, the definition of the $\mathbf{p}$-anisotropic random walk $\mathcal{T}_d$ is the same (up to graph isomorphism) for all possible values of q_1 and q_2. However, the choice of the group structure we endow $\mathcal{T}_d$ with turns out to be of importance when considering coverings. As the groups corresponding to different values of q_1 and q_2 are not isomorphic, there are d-regular graphs G that can be expressed as the Schreier graph for $\mathcal{G}_{\text{free}}^{(q_1, q_2)}$ (with $q_1 + 2q_2 = d$) for some values of q_2 and not for others (more precisely it is harder to find a covering for smaller values of q_2).*

The standard example of an action on a finite set is the following. Let $\mathcal{G}$ be a finitely generated group and H a subgroup of $\mathcal{G}$ with a finite index. Then the set of left cosets $V = \{gH : g \in \mathcal{G}\}$ (with $gH = \{gh : h \in H\}$) is a finite set and φ defined by $\varphi(a, bH) = abH$ is an action of $\mathcal{G}$ on V.

9

We introduce now a notion of anisotropic walk on a (sequence of) Schreier graph. Fixing a group $\mathcal{G}$ we assume that we have a sequence of finite sets (V_n) with $\#V_n = n$ and (φ_n) a sequence of actions of $\mathcal{G}$ on V_n. We consider $\mathbf{p}$ to be a probability vector $\mathcal{G}$ with finite support. We are interested in the random walk on V_n with transition matrix

$$(26) \qquad P_{n,\mathbf{p}} = \sum_{g \in \mathcal{A}} p_g S_g,$$

where for $g \in \mathcal{G}$, S_g is as in Definition 2 the permutation matrix associated to the action φ_n. Note that if the support of $\mathbf{p}$ is contained in a finite symmetric set $\mathcal{A}$, then $P_{n,\mathbf{p}}$ is an anisotropic random walk on the Schreier graph of $(\mathcal{G}, \mathcal{A}, \varphi_n)$. This situation extends the previous setup in both directions: the underlying group is not necessarily the free group and the generating set is not necessarily the natural set of generators. Note that the uniform measure on V_n is always stationnary for this random walk. It is reversible if we assume

$$(27) \qquad \forall g \in \mathcal{G}, \ p_g = p_{g^{-1}}.$$

We are going to compare the random walk on V_n with a random walk on $\mathcal{G}$. To this end, we define $\mathcal{P}_{\mathbf{p}}$ the transition kernel of this random walk on $\mathcal{G}$ defined by

$$(28) \qquad \mathcal{P}_{\mathbf{p}} = \sum_{g \in \mathcal{A}} p_g \lambda(g),$$

where, for $g \in \mathcal{G}$, $\lambda(g)$ is the left multiplication operator in (or the left regular representation of g) defined on $\mathcal{G}$ by $\lambda(g)(h) = gh$.

Our aim is to provide an extension of Theorem 1 which gives a condition in terms of spectral properties for the mixing time to be minimal.

Let $\varrho_{\mathbf{p}}$ and $\sigma_{\mathbf{p}}$ be the spectral radius and the singular radius of $\mathcal{P}_{\mathbf{p}}$ and let $\sigma_{n,\mathbf{p}}$ be the singular radius of $P_{n,\mathbf{p}}$ defined in (13). For integer $t \geq 1$, we define $\sigma_{n,\mathbf{p}}(t)$ and $\sigma_{n,\mathbf{p}}(t)$ as in (16). From [26], the Alon-Boppana lower bound (19) is still valid.

We wish to focus on sequences of random walks whose spectral gap is uniformly bounded away from one. Hence, we will assume that $\varrho_{\mathbf{p}} < 1$. This is equivalent to assuming that the subgroup $\langle \mathcal{A}_{\mathbf{p}} \rangle$ of $\mathcal{G}$ generated by $\mathcal{A}_{\mathbf{p}} := \{g \ : \ p_g > 0\}$ is *non-amenable* (or simply that $\mathcal{G}$ is non-amenable if one takes as an assumption that $\mathcal{A}_{\mathbf{p}}$ generates $\mathcal{G}$). Recall that a group is said to be *amenable* if it admits a finitely-additive left-invariant probability measure. The equivalent between non-amenability of $\langle \mathcal{A}_{\mathbf{p}} \rangle$ and $\varrho_{\mathbf{p}} < 1$ for $\mathbf{p}$ satisfying (27) was established by Kesten [32, 33]. In the non-reversible case, see forthcoming Lemma 8.

As before, we can determine an asymptotic lower bound for the mixing time of the random walk with generator $P_{n,\mathbf{p}}$ (valid for any sequence of group actions) in terms of the entropy rate of $\mathcal{P}_{\mathbf{p}}$ denoted by $\mathfrak{h}(\mathbf{p})$ and defined by Equation (20). In Subsection 2.1 below, we will check that the Avez lower bound $\mathfrak{h}(\mathbf{p}) \geq -2 \log \varrho_{\mathbf{p}}$ holds and that the mixing time of the anisotropic random walk

on G_n is asymptotically larger than $(1 - o(1))(\log n)/\mathfrak{h}(\mathbf{p})$ in the sense that for any fixed $\varepsilon \in (0, 1)$, uniformly in $x \in V_n$, the inequality (22) holds.

For a given probability vector $\mathbf{p}$ supported by a generator, a natural question is thus the following: *Are there spectral conditions for a sequence of actions (φ_n) of $\mathcal{G}$ on V_n with $\#V_n = n$ to guarantee that the anisotropic random walk on V_n has an asymptotically minimal mixing time in the sense that $T_n^{\mathrm{mix}}(\varepsilon) = (1 + o(1))(\log n)/\mathfrak{h}(\mathbf{p})$ for any fixed $\varepsilon \in (0, 1)$?* Our answer to this question is based on the two following notions of group algebra.

Definition 3 (RD property). *A finitely generated group $\mathcal{G}$ has the RD property (for Markov operators) if for every finitely supported probability vector $\mathbf{p}$ the singular radius $\sigma_{\mathbf{p}} = \|\mathcal{P}_{\mathbf{p}}\|_{2 \to 2}$ of $\mathcal{P}_{\mathbf{p}}$ is well controlled by the ℓ^2-norm of $\mathbf{p}$ in the following sense : For any finite symmetric generating set $\mathcal{A}$ of $\mathcal{G}$, there exists a constant $C = C(\mathcal{G}, \mathcal{A})$ such that*

$$\sigma_{\mathbf{p}} \leq C R^C \|\mathbf{p}\|_2, \tag{29}$$

where R is the diameter of the support of $\mathbf{p}$ in the Cayley graph associated with $(\mathcal{G}, \mathcal{A})$.

We refer to [18] for an introduction to the RD property. Among non-amenable groups, we note that free groups and hyperbolic groups satisfy the RD property. Observe also that as the distance corresponding to different generating sets are comparable (if $d_{\mathcal{A}}$ and $d_{\mathcal{A}'}$ are the graph distance for the Caley graph associated with generators $\mathcal{A}$ and $\mathcal{A}'$ respectively we have $d_{\mathcal{A}'} \leq C_{\mathcal{A}, \mathcal{A}'} d_{\mathcal{A}}$ where $C_{\mathcal{A}, \mathcal{A}'} = \max_{y \in \mathcal{A}'} d_{\mathcal{A}}(e, y))$ it is sufficient to check (29) for a single finite symmetric generating set $\mathcal{A}$ of $\mathcal{G}$.

Recall that we automatically have $\sigma_{\mathbf{p}} \leq \|\mathbf{p}\|_1$, and hence (29) is trivially satisfied when $\|\mathbf{p}\|_2 \geq C R^{-C} \|\mathbf{p}\|_1$. Therefore, the condition (29) says something about $\sigma_{\mathbf{p}}$ for $\mathbf{p}$ which have large support and whose mass is well spread on that support. In fact, as we have $\sigma_{\mathbf{p}} \geq \|\mathbf{p}\|_2$, the property (29) for a non-amenable group tells us in particular that $\sigma_{\mathbf{p}}$ is close to this trivial lower bound (in the sense that $\sigma_{\mathbf{p}} = \|\mathbf{p}\|_2^{1 + o(1)}$) when $\mathbf{p}$ is reasonably spread-out on the ball of radius R for large values of R (recall that a non-amenable group has exponential growth).

Our second notion is the *strong convergence of operators algebras* which we define here in our specific Markovian setting. It can be thought as an analog of the Ramanujan property for a sequence of group actions on finite sets. It is a stronger assumption since the Ramanujan property only refers to one particular random walk on the free group and the property below must be valid for every random walk.

Definition 4 (Strong convergence). *Let $\mathcal{G}$ be a finitely generated group, (V_n) a sequence of finite sets and (φ_n) a sequence of covering maps of $\mathcal{G}$ on V_n. We say that the sequence of covering maps (φ_n) converges strongly (on Markov operators) if for every finitely supported probability vector $\mathbf{p}$, we have*

$$\lim_{n \to \infty} \sigma_{n, \mathbf{p}} = \sigma_{\mathbf{p}}.$$

11

where $\sigma_{\mathbf{p}}$ is the singular radius of $\mathcal{P}_{\mathbf{p}}$ defined in (28) and $\sigma_{n,\mathbf{p}}$ is the singular radius of $P_{n,\mathbf{p}}$ defined in (26) and (13).

From (17), the strong convergence of (φ_n) implies in particular that

$$(30) \qquad \lim_{t\to\infty}\lim_{n\to\infty}\sigma_{n,\mathbf{p}}(t) = \varrho_{\mathbf{p}}.$$

We are now ready to state the last result of this introduction.

Theorem 5. *Let $\mathcal{G}$ be a finitely generated non-amenable group with the property RD, (V_n) a sequence of finite sets with $\#V_n = n$ and (φ_n) a sequence of actions of $\mathcal{G}$ on V_n which converges strongly. Then for every finitely supported probability vector $\mathbf{p}$ on $\mathcal{G}$ such that $\varrho_{\mathbf{p}} < 1$, the mixing time of the random walk with transition matrix $P_{n,\mathbf{p}}$ satisfies, for every $\varepsilon \in (0,1)$,*

$$\lim_{n\to\infty}\frac{T_{n,\mathbf{p}}^{\mathrm{mix}}(\varepsilon)}{\log n} = \frac{1}{\mathfrak{h}(\mathbf{p})}.$$

The assumption that the group actions converge strongly is a strong assumption. Notably, Theorem 5 does not imply neither Theorem A nor Theorem 1. These two theorems rely on special properties of free groups. Nevertheless, in some cases, it is possible to relax the assumption that the group actions converge strongly by supposing instead that the strong convergence holds on some vector spaces of codimension $n^{o(1)}$ (we discuss this point further in Remark 4.2).

The paper [15] provides a source of examples for Theorem 5 by establishing that random actions of the free group are strongly convergent. We consider an involution $*$ in $[d]$ with q_1 fixed points, and $\mathcal{G}_{\mathrm{free}}^{(q_1,q_2)}$ the group generated by q_1 free copies of $\mathbb{Z}/2\mathbb{Z}$ and $q_2 = (d-q_1)/2$ free copies of $\mathbb{Z}$ with its natural set of generators. We consider permutations $\alpha_{n,i}$, $i \in [d]$ on $[n]$ vertices which are chosen such that:

(A) If $i \neq i^*$, $\alpha_{n,i}$ is a uniform permutation on n elements and $\alpha_{n,i^*} = \alpha_{n,i}^{-1}$.

(B) If $i = i^*$, $\alpha_{n,i}$ is a uniform involution on n elements without fixed point (the construction is made only for even n).

(C) The permutations are chosen independently for each equivalence class of the involution $*$.

We consider the action of $\mathcal{G}_{\mathrm{free}}^{(q_1,q_2)}$ on $V_n = [n]$ defined by $\varphi_n(a_i,x) = \alpha_{n,i}(x)$. Then, in probability, this sequence of actions is strongly convergent. These random actions on the free group are the only known examples of strongly convergent sequences of actions, but it could indicate that choosing the action at random among all possible choices might yield a strongly convergent sequence also for other choices of non-amenable groups.

Minimal mixing time for color covered random walks. Finally, we also consider yet another extension which allows in particular to consider random walks on n-lifts of a base graph (not necessarily regular). Since the work of Amit and Linial [6, 7] and Friedman [23], this model has attracted a substantial attention. In this context, we will give the analog of Theorem 5. To avoid any confusion on notation, we postpone the treatment of this model to Section 6.

Organization of the paper. In Section 2 we provide a short proof for the entropic lower bound (22) only stetched in this introduction, and provide a general result (Proposition 9) which allows to estimate the mixing time of a Markov chain in terms of the distribution of a stopping time at which the chain is close to equilibrium.

In Section 3, we provide a simple proof of Theorem A/B, proving and using a relation between the k-non-backtracking random walk on trees and Chebychev polynomials.

In Section 4, we prove Theorem 4 concerning cutoff in the more general setup under the assumption of Strong Convergence (Definition 4).

In Section 5, we prove Theorem 1 concerning anisotropic walks, by combining the ideas of Section 4 with an analysis of the resolvent of the anisotropic random walk on $\mathcal{T}_d$.

Finally, in Section 6 we deal with the model of color covered random walks.

Notation. If V is a countable set and M is a bounded operator in $\ell^2(V)$, we use the matrix notation $M(x, y) = \langle \mathbf{1}_x, M\mathbf{1}_y \rangle$ for $x, y \in V$ where $\mathbf{1}_x$ denotes the indicator function of x. The integer part of real number t is denoted by $\lfloor t \rfloor$.

2. Preliminaries

2.1. **The entropic time lower bound.** For the sake of completeness we provide a complete proof of the entropic lower bound (22) which is only sketched in the introduction. The result is stated in the more general setup of Theorem 5. Recall that $\mathcal{G}$ is a finitely generated group, $(V_n)_{n \geq 0}$ a sequence of finite sets with $\#V_n = n$, $(\varphi_n)_{n \geq 0}$ a sequence of actions of $\mathcal{G}$ on V_n, that $P_{n,\mathbf{p}}$, $\mathcal{P}_{\mathbf{p}}$ denote the transition matrices on V_n and $\mathcal{G}$ respectively defined by (26) and (28), and that $\mathfrak{h}(\mathbf{p})$ is the entropy rate associated with $\mathcal{P}_{\mathbf{p}}$.

Proposition 6. *Let $\mathbf{p}$ be a finitely supported probability vector on $\mathcal{G}$ such that $\mathfrak{h}(\mathbf{p}) > 0$. Given any sequence (V_n), $(\varphi_n)_{n \geq 0}$ as above, the mixing time associated with the random walk on V_n with transition $P_{n,\mathbf{p}}$ satisfies for any $\varepsilon \in (0, 1)$*

$$\liminf_n \min_{x \in V_n} \frac{T_{n,\mathbf{p}}^{\mathrm{mix}}(x, 1 - \varepsilon)}{\log n} \geq \frac{1}{\mathfrak{h}(\mathbf{p})}.$$

We consider $T_{n,\mathbf{p}}^{\mathrm{mix}}(x, 1 - \varepsilon)$ (rather than $T_{n,\mathbf{p}}^{\mathrm{mix}}(x, \varepsilon)$) when lower bounds on the mixing time are concerned so that for both the upper and the lower bound, it is sufficient to consider small values of ε.

Proof. Let $(\mathcal{X}_t)$ denote the random walk on $\mathcal{G}$ starting from the unit e and with transition $\mathcal{P}_{\mathbf{p}}$. Its distribution is denoted by $\mathbb{P}$. The result is an almost direct consequence of the Shannon-McMillan-Breiman Theorem [31, Theorem 2.1], which states that $\log \mathcal{P}_{\mathbf{p}}(e, \mathcal{X}_t)$ concentrates around its mean; see (21). In particular given $\varepsilon, \delta > 0$ we have for all t sufficiently large

$$\mathbb{P}\big[\log \mathcal{P}_{\mathbf{p}}^t(e, \mathcal{X}_t) < -(1+\delta)\mathfrak{h}(\mathbf{p})t\big] \leq \varepsilon/2.$$

Thus if one sets

$$\mathcal{V}_\delta(t) := \Big\{ g \in \mathcal{G} \ : \ \mathcal{P}_{\mathbf{p}}^t(e, g) \geq e^{-(1+\delta)\mathfrak{h}(\mathbf{p})t} \Big\}$$

we have

$$|\mathcal{V}_\delta(t)| \leq e^{(1+\delta)\mathfrak{h}(\mathbf{p})t} \quad \text{and} \quad \mathbb{P}[\mathcal{X}_t \notin \mathcal{V}_\delta(t)] \leq \varepsilon/2.$$

Now given $x \in V_n$ arbitrary, we consider $X_t := \varphi_n(\mathcal{X}_t, x)$, which is a random walk with transition matrix $P_{n,\mathbf{p}}$ started at x, and let $V_\delta(t) := \{\varphi_n(g, x) \ : \ g \in \mathcal{V}_\delta(t)\}$, the image of $\mathcal{V}_\delta(t)$ by the action. We have, for all t sufficiently large

$$\pi_n(V_\delta(t)) = \frac{|V_\delta(t)|}{n} \leq \frac{|\mathcal{V}_\delta(t)|}{n} \leq \frac{e^{(1+\delta)\mathfrak{h}(\mathbf{p})t}}{n},$$

and $P_{n,\mathbf{p}}^t(x, V_\delta(t)) = \mathbb{P}(X_t \in V_\delta(t)) \geq 1 - \varepsilon/2$. Thus we have

$$\|P_{n,\mathbf{p}}^t(x, \cdot) - \pi_n\|_{\mathrm{TV}} \geq 1 - \varepsilon/2 - e^{(1+\delta)\mathfrak{h}(\mathbf{p})t}/n.$$

Considering $t = \lfloor \log(n\varepsilon/2)/((1+\delta)\mathfrak{h}(\mathbf{p})) \rfloor$, we conclude that for any arbitrarily small and $\varepsilon, \delta > 0$, we have for n sufficiently large

$$T_{n,\mathbf{p}}^{\mathrm{mix}}(x, 1-\varepsilon) \geq \lfloor \log(n\varepsilon/2)/((1+\delta)\mathfrak{h}(\mathbf{p})) \rfloor \geq (1-\delta)\frac{(\log n)}{\mathfrak{h}(\mathbf{p})}.$$

It concludes the proof. $\qquad\square$

The next lemma is the classical Avez lower bound adapted to our definition of spectral radius. It implies notably that if $\varrho_{\mathbf{p}} < 1$ then $\mathfrak{h}(\mathbf{p}) > 0$.

Lemma 7. *If $\mathbf{p}$ is a finitely supported probability vector on $\mathcal{G}$ we have $\mathfrak{h}(\mathbf{p}) \geq -2\log \varrho_{\mathbf{p}}$.*

Proof. We may assume $\varrho_{\mathbf{p}} < 1$. Let $h \geq 0$ be such that $\varrho_{\mathbf{p}} < e^{-h}$. From (17), there exists an integer $t_0 \geq 1$ such that for $\|\mathcal{P}_{\mathbf{p}}^t\|^2 = \|\mathcal{P}_{\mathbf{p}}^t \mathcal{P}_{\mathbf{p}}^{t*}\| \leq e^{-2ht}$ for all $t \geq t_0$. In particular,

$$\sum_{g \in \mathcal{G}} \big(\mathcal{P}_{\mathbf{p}}^t(e, g)\big)^2 = \big(\mathcal{P}_{\mathbf{p}}^t(\mathcal{P}_{\mathbf{p}}^*)^t\big)(e, e) \leq e^{-2ht}.$$

From Jensen inequality, we deduce that

$$\sum_{g \in \mathcal{G}} \mathcal{P}_{\mathbf{p}}^t(e, g) \log(\mathcal{P}_{\mathbf{p}}^t(e, g)) \leq \log \left(\sum_{g \in \mathcal{G}} \big(\mathcal{P}_{\mathbf{p}}^t(e, g)\big)^2 \right) \leq -2ht.$$

It follows that $\mathfrak{h}(\mathbf{p}) \geq 2h$. $\qquad\square$

We conclude this subsection with a corollary of Kesten's criterion for non-amenability applicable to non-reversible walks.

Lemma 8. *Let* $\mathbf{p}$ *be a finitely supported probability vector on* $\mathcal{G}$ *and* $\mathcal{A}_\mathbf{p} = \{g : p_g > 0\}$. *The subgroup* $\langle \mathcal{A}_\mathbf{p} \rangle$ *generated by* $\mathcal{A}_\mathbf{p}$ *is non-amenable if and only if* $\varrho_\mathbf{p} < 1$.

Proof. It is convenient to introduce the lazy random walk, $\mathcal{L}_\mathbf{p} = (\mathcal{I} + \mathcal{P}_\mathbf{p})/2 = \mathcal{P}_{\delta_e/2 + \mathbf{p}/2}$ where $\mathcal{I}$ is the identity operator. Assume that $\langle \mathcal{A}_\mathbf{p} \rangle$ is non-amenable. Then $\mathcal{L}_\mathbf{p}\mathcal{L}_\mathbf{p}^*$ is of the form $\mathcal{P}_{\mathbf{p}'}$ for some $\mathbf{p}'$ which satisfies (27) and $\langle \mathcal{A}_\mathbf{p} \rangle = \langle \mathcal{A}_{\mathbf{p}'} \rangle$. Kesten's result then implies that the singular radius of $\mathcal{L}_\mathbf{p}$ is smaller than $1 - \delta$ for some $\delta > 0$. The operator norm of operators of the form $\mathcal{P}_{\mathbf{p}'}$ with $\mathbf{p}'$ a probability vector can be obtained by taking the supremum over non-negative functions. We can thus compare the norm of $\mathcal{L}_\mathbf{p}^{2t} = ((\mathcal{I} + \mathcal{P}_\mathbf{p})/2)^{2t}$ with that of any term appearing in its binomial expansion. In particular we have,

$$\|\mathcal{P}_\mathbf{p}^t\|_{2\to 2} \leq 2^{2t}\binom{2t}{t}^{-1}\|\mathcal{L}_\mathbf{p}^{2t}\|_{2\to 2} \leq 2\sqrt{t}\|\mathcal{L}_\mathbf{p}\|_{2\to 2}^{2t} \leq 2\sqrt{t}(1-\delta)^{2t},$$

(where we have used that $\binom{2t}{t} \geq 2^{2t}/(2\sqrt{t})$. From (17), it follows that $\varrho_\mathbf{p} \leq (1-\delta)^2 < 1$. Conversely, if $\varrho_\mathbf{p} < 1$ then, from (17), the singular radius of $\mathcal{L}_\mathbf{p}^t$ is strictly smaller than 1 for some t. From the definition, it follows that the spectral radius of $\mathcal{P}_{\mathbf{p}''} := \mathcal{L}_\mathbf{p}^t(\mathcal{L}_\mathbf{p}^t)^*$ is smaller than one, which, by Kesten's reciprocal implies the non-amenability of $\langle \mathcal{A}_{\mathbf{p}''} \rangle = \langle \mathcal{A}_\mathbf{p} \rangle$. $\qquad\square$

2.2. **Mixing time from stopping time.** We present here a result derived from [9], which allows to estimate the distance from equilibrium using arbitrary stopping times. In this subsection, (X_t) is an arbitrary Markov chain on a finite set V with transition matrix P and for $x \in V$, $\mathbb{P}_x$ denotes the distribution of this process with initial condition $X_0 = x$.

A classical way to obtain mixing time upper-bounds is via the use of strong stationary times (see [36, Chapter 6]). A strong stationary time is defined as a stopping time T for the chain X for which X_T is at equilibrium and X_T and T are independent. The standard bound [36, Lemma 6.11] says that if T is a strong stationary time for (1) then (the bound is in fact proved for the separation distance which is larger)

$$\|P^t(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \mathbb{P}_x[T > t].$$

A careful inspection of the proof in [36] reveals that one can allow X_T to admit another distribution provided an adequate error term is added. However the assumption that X_T and T are independent is crucial in the mechanism of proof. However using recent techniques developed in [30, 9] to compare mixing times with hitting times, we can by-pass this independence assumption if the chain is reversible and if the mixing time is much larger than the relaxation time, at the cost of a second error term. We will present a variant of this argument for general finite Markov chains (which in particular does not require reversibility).

We say that a filtration $(\mathcal{F}_t)$, is adapted to (X_t) if for any $t \geq 0$, $x, y \in V$.

(31)
$$\mathbb{E}[\mathbf{1}_{\{X_t=x, X_{t+1}=y\}} \mid \mathcal{F}_t] = P(x,y)\mathbf{1}_{\{X_t=x\}},$$

(in particular this implies that X_t is $\mathcal{F}_t$ measurable). We denote by $\ell^2(\pi)$ the vector space $\mathbb{R}^V$ endowed with the scalar product,

$$\langle f, g \rangle_\pi = \sum_{x \in V} \pi(x) f(x) g(x).$$

Let us recall the definition *singular radius* given in (16) for a finite Markov chain P with invariant measure π

$$\sigma = \|P_{|\mathbf{1}^\perp}\|_{\ell^2(\pi) \to \ell^2(\pi)} = \sup \left\{ \frac{\|P_{|\mathbf{1}^\perp} f\|_{\ell^2(\pi)}}{\|f\|_{\ell^2(\pi)}} : f \neq 0 \right\}.$$

For any integer $t \geq 1$, we define the *t-th singular radius* as

$$\sigma(t) = \|(P^t)_{|\mathbf{1}^\perp}\|_{\ell^2(\pi) \to \ell^2(\pi)}^{1/t}.$$

Note that $\sigma(t) \leq \sigma$. Moreover, in our context, the Poincaré inequality is the claim that for any vector $f \in \mathbb{R}^V$, with $\pi(f) = \langle \mathbf{1}, f \rangle_\pi$,

$$(32) \qquad \|P^t f - \pi(f)\mathbf{1}\|_{\ell^2(\pi)} = \|(P^t)_{|\mathbf{1}^\perp} f\|_{\ell^2(\pi)} \leq \sigma(t)^t \|f\|_{\ell^2(\pi)}.$$

This follows immediately from the definition of t-th singular radius. We control distance to equilibrium with the help of stopping time with the following result (in the present paper, the inequality (33) is sufficient for all purposes, but since the result might have other applications, we also include a reversible version which is significantly better when ϱ is close to 1).

Proposition 9. *Let (X_t) be a finite irreducible Markov chain with transition matrix P, equilibrium measure π and with t-th singular radius $\sigma(t)$. If T is a stopping time with respect to a filtration adapted to X and $\mathbb{P}_x(X_T \in \cdot) = \nu$, then we have for any positive integers t and s:*

$$(33) \qquad \|P^{t+s}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \|\nu - \pi\|_{\mathrm{TV}} + \mathbb{P}_x[T > t] + 2(1 - \sigma(s))^{-1/3} \sigma(s)^{2s/3}.$$

Moreover if (X_t) is reversible and $\varrho = \sigma(1)$ denotes the spectral radius of P, we have

$$(34) \qquad \|P^{t+s}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \|\nu - \pi\|_{\mathrm{TV}} + \mathbb{P}_x[T > t] + 3\varrho^{2s/3}.$$

Proof. In the reversible case, the main ingredient of our proof is [9, Corollary 2.4], which we reformulate as follows. Given a set $A \subset V$, $s \geq 0$ and $\varepsilon > 0$ we set

$$U(A, s, \varepsilon) := \{y \in V \; : \; \exists t \geq s, \; |P^t(y, A) - \pi(A)| > \varepsilon\}.$$

Then we have

$$\pi(U(A, s, \varepsilon)) \leq 2\varepsilon^{-2} \varrho^{2s}.$$

From the definition of total variation distance we deduce

$$(35) \qquad \nu(U(A, s, \varepsilon)) \leq 2\varepsilon^{-2} \varrho^{2s} + \|\nu - \pi\|_{\mathrm{TV}}.$$

For every x, t and s, using the triangle inequality and the fact that $X_T \sim \nu$ we obtain (using the short-hand notation U for $U(A, s, \varepsilon)$) that

$$P^{t+s}(x, A) - \pi(A) \leq \sum_{i=0}^{t} \sum_{y \notin U} \mathbb{P}_x(T = i \; ; X_T = y)(P^{s+t-i}(y, A) - \pi(A)) + \mathbb{P}_x[T > t] + \nu(U).$$

From the definition of $U(A, t, \varepsilon)$ the double sum above is smaller than ε. Thus, from (35), we obtain (maximizing over A)

$$\|P^{t+s}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq \mathbb{P}_x[T > t] + \|\nu - \pi\|_{\mathrm{TV}} + 2\varepsilon^{-2}\varrho^{2s} + \varepsilon,$$

and one can conclude by choosing $\varepsilon = \varrho^{2s/3}$. In the general case, we define for integer $s \geq 0$

$$V(A, s, \varepsilon) := \{y \in V \; : |P^s(y, A) - \pi(A)| > \varepsilon\} \quad \text{and} \quad U(A, s, \varepsilon) = \bigcup_{t=s}^{\infty} V(A, k, \varepsilon).$$

In particular, we recover the above definition for $U(A, s, \varepsilon)$. From Markov inequality and (32), for any integer $t \geq 0$, we have

$$\pi(V(A, t, \varepsilon)) \leq \varepsilon^{-2}\|P^t \mathbf{1}_A - \pi(A)\mathbf{1}\|_{\ell^2(\pi)}^2 \leq \varepsilon^{-2}\sigma(t)^{2t}\pi(A).$$

Since σ_t is non-increasing, we thus have for any s and A,

$$\pi(U(A, s, \varepsilon)) \leq \sum_{k=s}^{\infty} \pi(V(A, k, \varepsilon)) \leq \varepsilon^{-2}(1 - \sigma(s)^2)^{-1}\sigma(s)^{2s}.$$

We deduce a slightly modified form of Equation (35)

$$\nu(U(A, s, \varepsilon)) \leq \varepsilon^{-2}(1 - \sigma(s)^2)^{-1}\sigma(s)^{2s} + \|\nu - \pi\|_{\mathrm{TV}}.$$

We may thus reproduce the same argument. $\qquad\square$

3. Simple random walks on Ramanujan graphs revisited

3.1. **Sketch of proof of Theorem A and Theorem B.** In order to prove Theorem A and Theorem B, we apply Proposition 9 for a stopping time defined using a coupling between the random walk on G_n and that on $\mathcal{T}_d$, the infinite d-regular tree. This coupling is defined thanks to a covering map from $\mathcal{T}_d$ to G_n.

We denote by e the root vertex of $\mathcal{T}_d$. Let $\mathcal{X}$ be the simple random walk on $\mathcal{T}_d$ starting from e. Given $x \in V_n$, we fix a local graph homeomorphism $\varphi : \mathcal{T}_d \to G_n$ (each vertex v in $\mathcal{T}_d$ has its d neighbors mapped to the d neighbors of $\varphi(v)$ in G_n) such that $\varphi(e) = x$. We may construct the simple random walk on G_n by setting $X_t := \varphi(\mathcal{X}_t)$. For a well chosen integer $k \geq 1$, we define the stopping time τ as

$$(36) \qquad\qquad \tau = \inf\{t \geq 0 : \mathrm{Dist}(\mathcal{X}_t, e) = k\},$$

where $\mathrm{Dist}(v, e)$ is the distance of the vertex v in $\mathcal{T}_d$ to the root e. With $k = \frac{\log n}{\log(d-1)}(1 + o(1))$, we show that at the time τ, $X_\tau = \varphi(\mathcal{X}_\tau)$ is close to equilibrium. More precisely, we use that the distribution of X_τ can be expressed as an explicit polynomial of the transition matrix P_n (cf.

Lemma 10), and thus its ℓ_2-norm can be controlled in terms of the spectral radius of P_n (cf. Lemma 12). This spectral bound turns out to be optimal.

Then the proof is concluded easily by using Proposition 9 and the fact that the detailed behavior of τ, which is a hitting time for a biased random walk, is known. It is worth mentioning that this reasoning leads to a more quantitative result in Proposition 11 below (which can also be obtained using methods from [39]). We note also that the variables X_τ and τ are independent and Proposition 9 in its full strength is not really needed here.

3.2. **Non-backtracking walks and Chebyshev polynomials.**
In this subsection, let us consider $G = (V, E)$ an arbitrary simple d-regular graph. Given $k \geq 1$ integer, we let

$$W_k := \{(x_i)_{i=0}^k \in V^{k+1} \; : \; \forall i \in [k], \; \{x_{i-1}, x_i\} \in E\}$$

denote the set of paths of length k in G. Given $x, y \in V$, we define the set of non-backtracking paths of length k from x to y as (with the convention that $[0]$ is the empty set)

$$\mathbf{NB}_k(x, y) := \{\mathbf{x} \in W_k \; : \; x_0 = x, \; x_k = y \; , \; \forall i \in [k-1], x_{i-1} \neq x_{i+1}\},$$

and $\mathbf{NB}_k(x) := \bigcup_{y \in V} \mathbf{NB}_k(x, y)$. We define the non-backtracking operator of length k on G to be the following stochastic matrix on $V \times V$,

$$Q_k(x, y) := \frac{\#\mathbf{NB}_k(x, y)}{\#\mathbf{NB}_k(x)} = \frac{\#\mathbf{NB}_k(x, y)}{d(d-1)^{k-1}}.$$

We let P denote the transition matrix for the simple random walk on G (i.e. $P = Q_1$). The following well known result (see e.g. [5, 45] and [16] for an early reference) will help us to control the spectral radius of Q_k in terms of that of P.

Lemma 10. *For every integer k, there exists a polynomial p_k such that $Q_k = p_k(P)$ for every simple d-regular graph G. More precisely we have*

$$p_k(x) = \frac{1}{d(d-1)^{k/2}} \left((d-1) U_k \left(\frac{x}{\varrho} \right) - U_{k-2} \left(\frac{x}{\varrho} \right) \right),$$

where $\varrho := (2\sqrt{d-1})/d$ and (U_k), $k \geq -1$, are the Chebyshev polynomials of the second type defined rescursively by

$$U_{-1} = 0, \quad U_0 = 1, \quad and \quad U_{k+1}(x) = 2x U_k(x) - U_{k-1}(x).$$

Proof (sketch). For a more detailed proof, we refer to the above mentionned references [5, 16, 45]. It is sufficient to check that the identity $Q_k = p_k(P)$ is valid on the d-regular infinite tree $\mathcal{T}_d$ (it is the universal covering of G and the preimage by a covering map of the non-bactracking paths on G are the non-backtracking paths on $\mathcal{T}_d$). We set $\bar{Q}_k := d(d-1)^{k-1} Q_k$. Using that $\bar{Q}_k(x, y) = \mathbf{1}_{\{\text{Dist}(x,y)=k\}}$ it is simple to check that

$$(37) \qquad\qquad \bar{Q}_{k+1} = dP\bar{Q}_k - \bar{Q}_{k-1}.$$

The result follows then by induction on k. We find $p_1 = x$, $p_2 = x^2 d/(d-1) - 1/(d-1)$ and, from (37),

$$p_{k+1} = \frac{d}{d-1} x p_k - \frac{1}{d-1} p_{k-1}.$$

It is then immediate to check that this recursion coincides with the recursion satisfied by the polynomials $d^{-1}(d-1)^{-k/2}((d-1)U_k(x/\varrho) - U_{k-2}(x/\varrho))$. The conclusion follows. $\square$

The polynomials (p_k) are called the *Geronimus polynomials* (in reference to [25]) or the *non-backtracking polynomials*.

3.3. **Proof of Theorem A.** Recall that ϱ_n denotes the spectral radius for P_n restricted to non-constant function. We let

$$\eta_n := \max\left(0, \frac{d\varrho_n}{2\sqrt{d-1}} - 1\right)$$

quantify by how much G_n is far from being a Ramanujan graph. Theorem A is a consequence of this more quantitative statement.

Proposition 11. *Let (G_n) be a sequence of d-regular graphs on n vertices such that $\lim_{n\to\infty} \eta_n = 0$. There exists a constant C and a sequence δ_n tending to zero such that for all $\varepsilon \in (0,1)$ for all n sufficiently large (depending on ε)*

$$(38) \qquad T_n^{\mathrm{mix}}(\varepsilon) \leq \left(\frac{d}{(d-2)\log(d-1)} + C\sqrt{\eta_n}\right) \log n + (\Phi(\varepsilon) + \delta_n)\sqrt{\log n},$$

where, if Z is a standard normal distribution, the function $\Phi(\cdot)$ is defined as the inverse of

$$s \mapsto \mathbb{P}\left[Z \geq \frac{(d-2)^{3/2}}{2\sqrt{d(d-1)}} s\right].$$

In particular if $\lim_{n\to\infty} \eta_n \log n = 0$, then

$$(39) \qquad T_n^{\mathrm{mix}}(\varepsilon) = \frac{d}{(d-2)\log(d-1)} \log n + \Phi(\varepsilon)\sqrt{\log n} + o(\sqrt{\log n}).$$

Note that the upper-bound in (39) is an immediate consequence of (38), while the lower bound (displayed in [39, Fact 2.1]) - which is valid for any d-regular graph - follows from the argument sketched in the introduction. We note also that it follows from [13] that, if G_n is a uniform random d-regular graph on n vertices then $\eta_n\sqrt{\log n}$ converges to 0 in probability. Hence, we recover the main result of [40] from Proposition 11. The remainder of this subsection is devoted to the proof of Proposition 11. The proof includes a technical lemma whose proof is postponed to the end of the section.

Proof of Proposition 11. We apply the content of the previous subsection to our problem. Let x be in V_n and $\varphi : \mathcal{T}_d \to G_n$ be as in Section 3.1 be a local graph homeomorphism such that $\varphi(e) = x$, where e is the root of $\mathcal{T}_d$. Let $\mathcal{X}_t$ be the simple random walk on $\mathcal{T}_d$ started at the root vertex e. Then $X_t := \varphi(\mathcal{X}_t)$ is a simple random walk on G_n starting from x. For an integer k to be chosen later on, let τ be defined as in (36). Since non-backtracking paths in a tree are geodesic

19

paths, it is immediate to see that the distribution of X_τ is given by $Q_{k,n}(x,\cdot)$ where $Q_{k,n}(x,\cdot)$ is the non-backtracking operator on G_n. Hence in particular the standard ℓ_2 upper-bound on total variation distance (2) applied for $t = 1$ yields

$$\|Q_{k,n}(x,\cdot) - \pi_n\|_{\mathrm{TV}} \le r_{k,n}\sqrt{n}. \tag{40}$$

where, using Lemma 10

$$r_{k,n} := \max_{\lambda \in \mathrm{Sp}(Q_{k,n})\backslash 1} |\lambda| = \max_{\lambda \in \mathrm{Sp}(P_n)\backslash 1} p_k(\lambda).$$

Hence if one sets

$$k = k_n := \min\left\{k \ : \ r_{k,n} \le \frac{1}{\sqrt{n}\log n}\right\},$$

we deduce from (40) that $\|Q_{k_n,n}(x,\cdot) - \pi_n\|_{\mathrm{TV}} \le (\log n)^{-1}$. We now apply Proposition 9 for $T = \tau$. We obtain that, provided that $\varrho_n \le 1 - \delta$ (which is true for all n large enough if $\eta_n \to 0$ for e.g. $\delta = 1/20$), for all $t \ge 0$,

$$d_n(x, t+s) \le \|Q_{k_n,n}(x,\cdot) - \pi_n\|_{\mathrm{TV}} + \mathbb{P}[\tau \ge t] + 3(1-\delta)^{2s/3}. \tag{41}$$

The last term can be made smaller than $(\log n)^{-1}$ for all n large enough by choosing $s = s_n := (\log\log n)^2$. Hence, setting

$$t_n(\varepsilon) := \inf\left\{t \ : \ \mathbb{P}[\tau > t] \le \varepsilon - 2(\log n)^{-1}\right\},$$

we obtain

$$T_n^{\mathrm{mix}}(\varepsilon) \le t_n(\varepsilon) + s_n.$$

Now, the central limit theorem for the biased random walk on the line implies that

$$\frac{\mathrm{Dist}(\mathcal{X}_t, o) - t((d-2)/d)}{2\sqrt{d-1}/d}$$

converges weakly to a standard normal distribution. We may thus easily estimate t_n as a function of k_n. Hence, the only missing part is an estimate for k_n.

Lemma 12. *For any integer $d \ge 3$, there exists a constant C such that for all n sufficiently large we have*

$$k_n \le \begin{cases} \frac{\log n}{\log(d-1)} + C\,(\log\log n)\,, & \text{if } \eta \le (\log n)^{-2}(\log\log n)^2, \\ \frac{\log n}{\log(d-1)}(1 + C\sqrt{\eta}) & \text{if } \eta \ge (\log n)^{-2}(\log\log n)^2. \end{cases}$$

The above estimates combined with the use of the Central Limit Theorem (details are left to the reader) imply that

$$t_n(\varepsilon) \le \left(\frac{d}{(d-2)\log(d-1)} + C\sqrt{\eta_n}\right)\log n + (\Phi(\varepsilon) + \delta_n)\sqrt{\log n}.$$

This concludes the proof of Proposition 11 (provided that Lemma 12 has been established). $\qquad\square$

Proof of Lemma 12. We use the following classic identities, valid for all $\theta \in \mathbb{R}\backslash\{0\}$ and $k \in \mathbb{N}$,

$$(42) \qquad U_k(\cosh\theta) = \frac{\sinh((k+1)\theta)}{\sinh\theta} \quad \text{and} \quad U_k(\cos\theta) = \frac{\sin((k+1)\theta)}{\sin\theta}.$$

We note that U_k is either an even or an odd function (depending on the parity of k). We thus have for any λ,

$$|p_k(\lambda)| \le \frac{1}{d(d-1)^{k/2}}[(d-1)|U_k(|\lambda|/\varrho)| + |U_{k-2}(\lambda/\varrho)|]$$

Using the fact (it can be checked using (42)) that $|U_k(x)| \le k+1$ on $[0,1]$ and $U_k(x)$ is increasing on $[1,\infty)$ we obtain that

$$(43) \qquad \max_{\lambda \in \mathrm{Sp}(P_n)\backslash 1} |p_k(\lambda)| \le \frac{1}{d(d-1)^{k/2}}[(d-1)U_k(1+\eta_n) + U_{k-2}(1+\eta_n)].$$

and hence $r_{k,n} \le (d-1)^{-k/2}U_k(1+\eta_n)$. Using the identity (42) we obtain that there exists a constant C such that

$$r_{k,n} \le \frac{C}{(d-1)^{k/2}}\min(\eta^{-1/2}, k)e^{Ck\sqrt{\eta}}.$$

This is sufficient to obtain the desired estimate on k_n. $\qquad\square$

3.4. **Proof of Theorem B.**

Let $\eta > 0$. To prove Theorem B, we use (41) with k_n replaced by $k'_n = \frac{\log n}{\log(d-1)}(1+\eta/2)$. By the law of large numbers $\tau = \tau(k)$ is asymptotically equivalent to $\frac{kd}{d-2}$ when k goes to infinity. Hence, to prove Theorem B, it is sufficient to show that, there exists $c > 0$ such that for all n large enough, we have $\|Q_{k'_n,n}(x,\cdot) - \pi_n\|_{\mathrm{TV}} \le n^{-c\eta}$ for at least $n^{1-2c\eta}$ vertices x in V_n. It is thus sufficient to show that for all n sufficiently large,

$$(44) \qquad \sum_{x\in V_n} \|Q_{k'_n,n}(x,\cdot) - \pi_n\|_{\mathrm{TV}} \le n^{1-3c\eta}.$$

To take into account the information we have about the multiplicity of eigenvalues, we must be more precise than (40) in our decomposition. For $\lambda \in \mathrm{Sp}(P_n)\backslash\{1\}$ we let $\alpha_\lambda(x)$ be the square norm of the projection of the vector δ_x onto E_n^λ, the eigenspace of P_n corresponding to λ, that is

$$\alpha_\lambda(x) := \max_{f\in E_n^\lambda} \frac{f(x)^2}{\sum_{y\in V_n} f(y)^2}.$$

From the spectral theorem, we have $\sum_{x\in V_n} \alpha_\lambda(x) = \dim(E_n^\lambda)$. Using the Cauchy-Schwartz inequality and the decomposition on the eigenspaces of P_n we obtain

$$(2\|Q_{k,n}(x,\cdot) - \pi_n\|_{\mathrm{TV}})^2 \le n\sum_{y\in V_n}\left(Q_{k,n}(x,y) - \frac{1}{n}\right)^2 = \sum_{\lambda\in\mathrm{Sp}(P_n)\backslash\{1\}} np_k(\lambda)^2\alpha_\lambda(x).$$

Hence, averaging over x, we have

$$\frac{1}{n}\sum_{x\in V_n} \|Q_{k,n}(x,\cdot) - \pi_n\|_{\mathrm{TV}}^2 \le \sum_{\lambda\in\mathrm{Sp}(P_n)\backslash\{1\}} p_k(\lambda)^2\dim(E_n^\lambda).$$

Using the fact (recall (43)) that $p_k(\lambda) \le (d-1)^{-k/2}(k+1)$ when $\lambda \le \varrho$, we obtain that

$$(45) \qquad \sum_{\lambda \in \mathrm{Sp}(P_n) \cap [-\varrho,\varrho]} p_k(\lambda)^2 \dim(E_n^\lambda) \le (d-1)^{-k}(k+1)^2 n.$$

For $\lambda \notin [-\varrho, \varrho]$, as a consequence of (42) we have

$$\begin{aligned}
\limsup_{k \to \infty} \frac{1}{k} \log |p_k(\lambda)| &\le \lim_{k \to \infty} \frac{1}{k} \log U_k(|\lambda|/\varrho) - \frac{1}{2}\log(d-1) \\
&= \log\left(|\lambda|/\varrho - \sqrt{(\lambda/\varrho)^2 - 1}\right) - \frac{1}{2}\log(d-1).
\end{aligned}$$

Hence recalling the definition of $I(u)$ in (7) and the assumption $\varrho_n \le 1 - \delta$, we arrive at

$$(46) \quad \limsup_{n \to \infty} \frac{1}{\log n} \log \sum_{\lambda \in \mathrm{Sp}(P_n) \setminus [-\varrho,\varrho]} p_{k_n'}(\lambda)^2 \dim(E_n^\lambda)$$

$$\le \sup_{u \in [\varrho, 1-\delta]} \left[(1 + \eta/2) \left(\frac{2 \log\left(u/\varrho - \sqrt{(u/\varrho)^2 - 1}\right)}{\log(d-1)} - 1 \right) + I(u) \right],$$

(where we have used the upper semi-continuity of $u \mapsto I(u)$). Using the assumption (8), we obtain that the left-hand side of (46) is at most $c_0 \eta$ with

$$c_0 := \frac{1}{2} - \frac{\log\left((1-\delta)/\varrho - \sqrt{((1-\delta)/\varrho)^2 - 1}\right)}{\log(d-1)} > 0.$$

Together with (45), it concludes the proof of (44) with $c = c_0/4$. $\qquad \square$

Remark 3.1 (Variant of Theorem B). *If H is a vector space of $\mathbb{R}^{V_n}$ with $\#V_n = n$, we define the flat-dimension of H as $\dim_0(H) = n \max_{x \in V_n} \|P_H \mathbf{1}_x\|_2^2$ where P_H is the orthogonal projection onto H. This definition implies $\dim_0(H) \ge \dim(H)$, $\dim_0(\mathrm{span}(\pi_n)) = 1$ and $\dim_0(\mathrm{span}(\mathbf{1}_x)) = n$. If the graph G_n is a transitive graph and H is the invariant vector space generated by k eigenvalues of P_n, then we have $\dim_0(H) = \dim(H)$. Now, we define I_0 exactly as I in (7) except that we replace in (7) $\dim(E_n^\lambda)$ by $\dim_0(E_n^\lambda)$. The proof of Theorem B actually proves that (6) holds if $\varrho_n < 1 - \delta$ and for all $u > \varrho$, (8) holds with I_0 instead of I.*

4. Covered random walks: proof of Theorem 5

4.1. **Notation.** In this section, we fix a finitely supported probability vector $\mathbf{p}$ on $\mathcal{G}$ and we denote by $(\mathcal{X}_t)_{t \ge 0}$ the random walk on $\mathcal{G}$ with transition kernel $\mathcal{P}_\mathbf{p}$ started at $\mathcal{X}_0 = e$, the unit of $\mathcal{G}$. The underlying probability distribution of the process $(\mathcal{X}_t)_{t \ge 0}$ on $\mathcal{G}^\mathbb{N}$ will be denoted by $\mathbb{P}(\cdot)$. Finally, if φ_n is the action of $\mathcal{G}$ on V_n as in Theorem 5, given a fixed $x \in V_n$, we set $X_t = \varphi_n(\mathcal{X}_t, x)$. Then $(X_t)_{t \ge 0}$ is a trajectory of the Markov chain on V_n with transition kernel $P_{n,\mathbf{p}}$ and initial condition x.

4.2. **Proof strategy.** Our strategy shares some similarities with that adopted in the Ramanujan case: we try to build a *backbone walk* $(Y_s)_{s\geq 0}$ with $Y_s = X_{\tau_s}$ using stopping times which are defined in terms of the walk performed on the covering (our terms backbone comes from the fact that the complete walk can be recovered from the backbone by adding the missing pieces). The two important properties that our backbone walk must satisfy are the following :

  (i) At each step, one jumps more or less uniformly to one of k vertices for a large k.
  (ii) The spectral gap associated with the backbone walk is close to the Alon-Boppana bound.

The second property is obtained from our assumptions that the RD property on $\mathcal{G}$ holds and that the sequence of actions converges strongly. To obtain a backbone walk that jumps close to uniformly on large sets, we perform an explicit construction based on the Green's operator associated to $\mathcal{P}_{\mathbf{p}}$.

To conclude, we need to relate the mixing time of the backbone walk to that of the original one. This is done using the tools developed in Subsection 2.2 which relate mixing times and hitting times. Indeed hitting times of the backbone walk provide an upper bound for the hitting times of the original walk.

4.3. **Construction of the backbone walk from the Green's operator.** Given k a large integer, our task is to find a stopping time τ for the process $(\mathcal{X}_t)$ starting from $\mathcal{X}_0 = e$ such that $\mathcal{X}_\tau$ is close to be uniformly distributed on a set of k vertices. We denote by $\mathcal{A} = \{a_1, \ldots, a_d\}$ be the symmetric support of $\mathcal{P}_{\mathbf{p}}$. We define Γ as the Cayley graph of $\mathcal{G}$ generated by $\mathcal{A}$. By construction $(\mathcal{X}_t)$ is a random walk on Γ. We are going to choose our stopping time of the form

(47)
$$\tau := \inf\{t \geq 0 : \mathcal{X}_t \notin U\},$$

where U is finite and contains e. Notably, τ is almost surely finite and $\mathcal{X}_\tau$ is supported on the set ∂U defined by

$$\partial U := \{g \notin U : a_i^{-1}g \in U \text{ for some } i \in [d]\},$$

which satisfies $\#\partial U \leq (d-1)\#U$.

Now let us specify our choice for U. We let $\mathcal{R}_{\mathbf{p}} = (\mathcal{I} - \mathcal{P}_{\mathbf{p}})^{-1}$ be the Green's operator associated with $\mathcal{P}_{\mathbf{p}}$. The Green's operator is a well defined bounded operator, since $\varrho_{\mathbf{p}} < 1$, 1 is not in the spectrum of $\mathcal{P}_{\mathbf{p}}$. We define u to be the image of the coordinate vector at e by $\mathcal{R}_{\mathbf{p}}$. The scalar $u(g)$ corresponds to the expected number of visits at g starting from e:

(48)
$$u(g) := \mathcal{R}_{\mathbf{p}}(e,g) = \sum_{t=0}^{\infty} \mathcal{P}_{\mathbf{p}}^t(e,g).$$

Given $k \geq 1$ we define the set

(49)
$$U := \left\{g \in \mathcal{G} : u(g) > \frac{1}{k}\right\}.$$

In the reversible case, our set U can be interpreted as the complement of a ball around e in the Green metric on $\mathcal{G}$ associated to $\mathcal{P}_{\mathbf{p}}$. The Green metric is defined for all $g, h \in \mathcal{G}$ by $d(g,h) =$

$\log \mathcal{R}_{\mathbf{p}}(e, e) - \log \mathcal{R}_{\mathbf{p}}(g, h)$ and it is closely related to the entropy of the random walk $(\mathcal{X}_t)$, see [12]. Our backbone walk is the induced walk on the successive exit times from U. More precisely, we define $\tau_0 := 0$, $\tau_1 = \tau$ and, for integer $s \geq 1$, $\tau_{s+1} := \inf\{t \geq \tau_s : \mathcal{X}_t \mathcal{X}_{\tau_s}^{-1} \notin U\}$. We finally set $\mathcal{Y}_s := \mathcal{X}_{\tau_s}$. We denote by $\mathcal{Q}$ the transition kernel associated with the Markov chain $(\mathcal{Y}_s)$: for any $g, h \in \mathcal{G}$,

(50) $$\mathcal{Q}(g, h) = \mathbb{P}(\mathcal{X}_\tau = hg^{-1}).$$

By construction, we have $\mathcal{Q} = \mathcal{P}_{\mathbf{q}}$ where, for all $g \in \mathcal{G}$,

(51) $$q_g := \mathcal{Q}(e, g) = \mathbb{P}\left[\mathcal{Y}_1 = g\right] = \mathbb{P}\left[\mathcal{X}_\tau = g\right].$$

We let (Y_s) be the projection of the walk $(\mathcal{Y}_s)$ onto V_n, $Y_s := \varphi_n(\mathcal{Y}_s, x)$, and let Q_n denote the associated transition kernel. This Markov chain is our backbone walk. The following result establishes that U has the desired property.

Proposition 13. *Assume that $\varrho_{\mathbf{p}} < 1$. Then there exists a constant C such that for every integer $k \geq 2$, the set defined by Equation (49) satisfies $e \in U$, $\#U \leq Ck \log k$, $\mathrm{diam}(U) \leq C \log k$ (where diam denotes the diameter for the graph distance in Γ) and such that for $\mathbf{q}$ defined by (51),*

(52) $$\forall g \in \partial U, \quad q_g \leq \frac{1}{k}.$$

Proof. By definition of the function u, we have $e \in U$ and for any $g \in \partial U$, $q_g = \mathbb{P}\left[\mathcal{X}_\tau = g\right] \leq u(g) \leq 1/k$, as requested. We now check that the cardinality of U is controlled by $k \log k$. This is a simple consequence of the assumption that $\varrho_{\mathbf{p}} < 1$. We fix ϱ such that $\varrho_{\mathbf{p}} < \varrho < 1$. Then, from (17), there exists $s \geq 1$, such that $\|\mathcal{P}_{\mathbf{p}}^t\| \leq \varrho^t$ for all $t \geq s$. Hence, there exists a constant $C_0 \geq 1$, such that $\|\mathcal{P}_{\mathbf{p}}^t\| \leq C_0 \varrho^t$ for all $t \geq 1$. Notably, we deduce that for all $g, h \in \mathcal{G}$,

(53) $$\mathcal{P}_{\mathbf{p}}^t(g, h) \leq \|\mathcal{P}_{\mathbf{p}}^t\| \leq C_0 \varrho^t,$$

(see forthcoming Lemma 25 for an improvement of this inequality). Thus, if $\mathrm{Dist}(g, e)$ is the graph distance between g and e in Γ,

$$u(g) = \sum_{t \geq |g|} \mathcal{P}_{\mathbf{p}}^t(e, g) \leq C_0(1 - \varrho)^{-1} \varrho^{\mathrm{Dist}(g, e)}.$$

This implies that U is included in the ball B_r of radius $r = \lfloor C_1 \log k \rfloor$ around the unit e. For any integer $b \geq 1$, we find

$$\begin{aligned}
\frac{\#U}{k} &\leq \sum_{g \in U} u(g) \leq \sum_{g \in B_r} \sum_{t=1}^{\infty} \mathcal{P}_{\mathbf{p}}^t(e, g) \\
&\leq \sum_{t=1}^{br} \sum_{g \in B_r} \mathcal{P}_{\mathbf{p}}^t(e, g) + \sum_{t=br+1}^{\infty} \sum_{g \in B_r} \mathcal{P}_{\mathbf{p}}^t(e, g) \\
&\leq br + \sum_{t=br+1}^{\infty} (\#B_r) C_0 \varrho^t.
\end{aligned}$$

We choose $b > 0$ such that $(d-1)\varrho^b < 1$. Since $\#B_r \leq d(d-1)^{r-1}$, we thus find that $\#U/k$ is at most $C_2 \log k$ as requested (with $C_2 = 2bC_1$). $\qquad\square$

4.4. Deducing mixing times from RD property and the strong convergence.

To compare the original walk with the backbone walk, the first requirement is to control how much time each backbone step requires on average. This can be deduced from the definition of the entropy of $\mathcal{G}$. Recall the definition of τ in (47).

Lemma 14. *Assume that $\varrho_{\mathbf{p}} < 1$. For any $\varepsilon > 0$, there exists $k(\varepsilon) > 1$ such that for all integers $k \geq k(\varepsilon)$,*

$$\mathbb{E}[\tau] \leq (1+\varepsilon)\frac{\log k}{\mathfrak{h}(\mathbf{p})}.$$

Proof. Given $t_1 < t_2 < \infty$, we decompose the expectation in three contributions ($\tau \leq t_1$, $\tau \in (t_1, t_2]$, $\tau > t_2$) and obtain

$$(54) \qquad \mathbb{E}[\tau] \leq t_1 + t_2 \mathbb{P}(\tau > t_1) + \mathbb{E}[\tau \mathbf{1}_{\{\tau > t_2\}}].$$

We set

$$t_1 := (1+\varepsilon/2)\frac{\log k}{\mathfrak{h}(\mathbf{p})} \quad \text{and} \quad t_2 := C\log k$$

for some adequate constant C, and prove that the second and third term in (54) are smaller than $(\varepsilon/4)(\log k/\mathfrak{h}(\mathbf{p}))$. We start by bounding the tail probability of τ. Recall that $\varrho_{\mathbf{p}}$ is the spectral radius of $\mathcal{P}_{\mathbf{p}}$. We fix ϱ such that $\varrho_{\mathbf{p}} < \varrho < 1$. From (53),

$$\mathbb{P}(\tau > t) \leq \mathbb{P}(\mathcal{X}_t \in U) \leq \sum_{g \in U} \mathcal{P}_{\mathbf{p}}^t(e, g) \leq \#U C_0 \varrho^t.$$

Hence, for any $s > 0$,

$$\mathbb{P}\left(\tau > \frac{\log(C_0 \# U) + s}{\log(1/\varrho)}\right) \leq e^{-s}.$$

By Proposition 13, we deduce for some choice of constant $C > 0$, for any $s > 0$ and integer $k \geq 2$,

$$\mathbb{P}(\tau > (C/2)(\log k + s)) \leq e^{-s}.$$

It follows that for any $\varepsilon > 0$, for all k large enough

$$\mathbb{E}\left[\tau \mathbf{1}_{\{\tau > t_2\}}\right] \leq \frac{1}{\log k} \leq \frac{\varepsilon \log k}{4\mathfrak{h}(\mathbf{p})}.$$

Now to control the second term, we need to show that $\mathbf{P}(\tau > t_1) \leq \varepsilon/(4C\mathfrak{h}(\mathbf{p}))$. Set

$$H = \left\{g \in \mathcal{G} : \mathcal{P}_{\mathbf{p}}^{t_1}(e, g) \leq e^{-(1-\varepsilon/3)t_1 \mathfrak{h}(\mathbf{p})}\right\},$$

and arguing as above,

$$\mathbb{P}(\tau > t_1) \leq \mathbb{P}(\mathcal{X}_{t_1} \in U) \leq \sum_{g \in U \cap H} \mathcal{P}_{\mathbf{p}}^{t_1}(e, g) + \mathbb{P}(\mathcal{X}_t \notin H).$$

Now, from (21), if k is large enough, $\mathbb{P}(\mathcal{X}_t \notin H) \le \varepsilon/(8C\mathfrak{h}(\mathbf{p}))$ and, by Proposition 13,

$$\sum_{g \in U \cap H} \mathcal{P}_{\mathbf{p}}^{t_1}(e, g) \le \#U e^{-(1-\varepsilon/3)t_1 \mathfrak{h}(\mathbf{p})} \le (Ck \log k)k^{-(1+\varepsilon/10)} \le \varepsilon/(8C\mathfrak{h}(\mathbf{p})),$$

as requested. $\qquad\square$

Remark 4.1. *The above proof actually shows that the conclusion of Lemma 14 is true for any exit time from a set of cardinality $k^{1+o(1)}$. On the other hand (21) and the lower bound $u(\mathcal{X}_t) \ge \mathcal{P}_{\mathbf{p}}^t(e, \mathcal{X}_t)$ imply easily that $\mathbb{E}[\tau] \ge (1-\varepsilon)(\log k)/\mathfrak{h}(\mathbf{p})$ for all k large enough. Hence, our set U asymptotically maximizes the mean exit time (among all sets of cardinality $k^{1+o(1)}$).*

All ingredients are now gathered to conclude.

Proof of Theorem 5. We fix $\varepsilon \in (0, 1)$, $\delta > 0$ and $x \in V_n$ arbitrary and prove that for n sufficiently large

$$T_{n,\mathbf{p}}^{\mathrm{mix}}(x, \varepsilon) \le (1+\delta) \log n/\mathfrak{h}(\mathbf{p}).$$

We consider τ constructed with U from Proposition 13 for some large k which we are going to choose depending on δ but not on n, and we set $m := \lfloor(1+\delta/4)(\log n)/\log k\rfloor$. We use Proposition 9 for the walk $X_t := \varphi_n(\mathcal{X}_t, x)$ with

$$T = \tau_m, \quad t = t_n = \lfloor(1+\delta)\log n/\mathfrak{h}(\mathbf{p}) - \log \log n\rfloor \quad \text{and } s = s_n = \lfloor\log \log n\rfloor.$$

We have

$$(55) \qquad \|P_{n,\mathbf{p}}^{t_n+s_n}(x, \cdot) - \pi\|_{\mathrm{TV}} \le \|Q_n^m - \pi\|_{\mathrm{TV}} + \mathbb{P}[\tau_m > t_n] + 2(1 - \sigma_{n,\mathbf{p}}(s_n))^{-1/3}\sigma_{n,\mathbf{p}}(s_n)^{2s_n/3}.$$

We are going to show that for n sufficiently large each of the three terms in the r.h.s. are smaller than $\varepsilon/3$.

We start with the third term. To deal with it we prove the following statement

$$(56) \qquad \limsup_{\substack{n \to \infty \\ s \to \infty}} \sigma_{n,\mathbf{p}}(s) \le \varrho_{\mathbf{p}},$$

where the limit can be taken over arbitray sequences of n and s which both go to infinity (though it is sufficient for our purpose to know that the lim sup is < 1). If $s = as_0 + b$ with a, b, s_1 non-negative integers, we have

$$\|(P_{n,\mathbf{p}}^s)_{\mathbf{1}^\perp}\|_{2\to2} \le \|(P_{n,\mathbf{p}}^{s_0})_{\mathbf{1}^\perp}\|_{2\to2}^a\|(P_{n,\mathbf{p}}^b)_{\mathbf{1}^\perp}\|_{2\to2} \le \|(P_{n,\mathbf{p}}^{s_0})_{\mathbf{1}^\perp}\|_{2\to2}^a.$$

Applied to $a = \lfloor s/s_0\rfloor$, we deduce that for all $t \ge t_1$,

$$(57) \qquad\qquad\qquad \sigma_{n,\mathbf{p}}(s) \le \sigma_{n,\mathbf{p}}(s_0)^{1-\frac{s_0-1}{s}}.$$

From (30), we can choose $s_0(\delta)$ and $n_0(\delta)$ such that $\sigma_{n,\mathbf{p}}(s_0) \le \varrho_{\mathbf{p}} + \delta/2$ and we can deduce from (57) that $\sigma_{n,\mathbf{p}}(s) \le \varrho_{\mathbf{p}} + \delta$ for $n \ge n_0(\delta)$ and $s \ge s_1(\delta)$ sufficiently large. Let us now move the second and third term in (34)

The probability $\mathbb{P}[\tau_m > t_n]$ is small as a consequence of the law of large numbers. Indeed choosing $k(\delta)$ sufficiently large Lemma, 14, guaranties that $t_n \geq (1 + \delta/2)m\mathbb{E}[\tau]$. The smallness of $\|Q_n^m - \pi\|_{\mathrm{TV}}$ is obtained using spectral estimates for Q_n. Since $\mathcal{G}$ has the RD property (29), we deduce from Proposition 13 that for some constants C, C' (depending on $\mathbf{p}$)

$$\sigma_{\mathbf{q}} = \|\mathcal{P}_{\mathbf{q}}\| \leq C(\log k)^C \left(k^{-2}\#\partial U\right)^{1/2} \leq C'k^{-1/2}(\log k)^{C+1/2}.$$

Now, the assumption that (φ_n) converges strongly applied to $\mathbf{q}$ implies that for all n large enough (depending on k), the singular radius $\sigma_{n,\mathbf{q}}$ of Q_n satisfies.

$$\sigma_{n,\mathbf{q}} \leq 2C'k^{-1/2}(\log k)^{C+1/2}.$$

Then, we use the Cauchy-Schwarz inequality and the usual ℓ^2-distance bound (32). We obtain that for any $m \geq 1$,

$$(58) \qquad \|Q_n^m(x,\cdot) - \pi_n\|_{\mathrm{TV}} \leq \sqrt{n}\|Q_n^m(x,\cdot) - \pi_n\|_2 \leq \sqrt{n}\sigma_{n,\mathbf{q}}^m,$$

and we can conclude replacing m by its value. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 4.2 (Relaxation of the definition of the spectral radius.)**.** *We may relax a little bit the assumption of strong convergence. If H is a vector subspace of $\mathbb{R}^{V_n}$ which is invariant under $P_{n,\mathbf{p}}$, we set $\sigma_{n,\mathbf{p}}^H$ to be the operator norm of $P_{n,\mathbf{p}}$ on the orthogonal of H. Recall the definition of the flat-dimension $\dim_0$ in Remark 3.1.*

Now, we say that the sequence of actions (φ_n) converges relatively strongly if for any finitely supported probability vector $\mathbf{p} \in \ell^2(\mathcal{G})$, we have $\limsup_n \sigma_{n,\mathbf{p}} < 1$ and $\lim_n \sigma_{n,\mathbf{p}}^{H_n} = \sigma_{\mathbf{p}}$ for a sequence (H_n) of invariant vector spaces such that $\pi_n \in H_n$ and $\dim_0(H_n) \leq n^{\varepsilon_n}$ with $\lim_n \varepsilon_n = 0$. Then Theorem 5 also holds under this weaker assumption. Indeed, we simply replace the bound (58) by the bound valid for any invariant vector space H of Q_n which contains π_n:

$$(59) \qquad \|Q_n^m(x,\cdot) - \pi_n\|_{\mathrm{TV}} \leq \sqrt{n}\|Q_n^m(x,\cdot) - \pi_n\|_2 \leq \sqrt{n}\sigma_{n,\mathbf{q}}^m\sqrt{\dim_0(H)/n} + \sqrt{n}\left(\sigma_{n,\mathbf{q}}^H\right)^m,$$

which follows directly from the spectral theorem and the observation that, if P_H is the orthogonal projection onto a vector space H, then,

$$\|P_H f\|_2 \leq \sum_x |f(x)|\|P_H \mathbf{1}_x\|_2 \leq \|f\|_1 \sqrt{\dim_0(H)/n}.$$

Finally, we notice that if $\dim_0(H) = n^{o(1)}$ and $\limsup_n \sigma_{n,\mathbf{q}} < 1$ then the first term on the right-hand side of (59) goes to 0 as soon as m is of order $\log n$.

5. ANISOTROPIC RANDOM WALKS: PROOF OF THEOREM 1

5.1. **Notation.** In this section, we fix an involution as in Theorem 1. We define $\mathcal{G}$ as the group obtained by k free copies of $\mathbb{Z}$ and l free copies of $Z/2\mathbb{Z}$ where $k + l$ is the number of equivalence classes of the involution, as detailed below Definition 2. We denote by $\mathcal{A} = \{g_1, \ldots, g_d\}$ its natural set of generators. The probability vector $\mathbf{p} = (p_1, \ldots, p_d)$ as in Theorem 1 is identified with a

vector in $\ell^2(\mathcal{G})$ defined by, for all $i \in [d]$, $p_{g_i} = p_i$ and $p_g = 0$ otherwise. As in the previous section, we denote by $(\mathcal{X}_t)_{t \geq 0}$ the random walk on $\mathcal{G}$ with transition kernel $\mathcal{P}_{\mathbf{p}}$ started at $\mathcal{X}_0 = e$, the unit of $\mathcal{G}$. The underlying probability distribution of the process $(\mathcal{X}_t)_{t \geq 0}$ on $\mathcal{G}^{\mathbb{N}}$ will be denoted by $\mathbb{P}(\cdot)$. Finally, we define φ_n as the action of $\mathcal{G}$ on V_n such that for all $i \in [d]$, $S_{g_i} = S_i$ where S_i is as in (11) and S_g is the permutation matrix associated to $\varphi_n(g, \cdot)$. Finally, given $x \in V_n$ we set $X_t = \varphi_n(\mathcal{X}_t, x)$, that is $(X_t)_{t \geq 0}$ is a trajectory of the Markov chain on V_n with transition kernel $P_{n,\mathbf{p}}$ with initial condition x.

5.2. **Proof strategy and organization.** Our starting point is to use the same stopping time strategy that for the previous section. But instead of using the RD property to conclude, we are going to show that the generator of the backbone random walk can be reasonably approximated by a polynomial in $P_{n,\mathbf{p}'}$, the generator of the random walk with anisotropy given by $\mathbf{p}'$. Our first job is thus to identify the value of $\mathbf{p}'$ which is possible. We perform this approximation for the backbone walk on the covering graph $\mathcal{G}$ (it is then sufficient to use the covering to have an approximation for the walk on V_n). With the definition of the stopping set U in (49), a natural object to compare $\mathcal{Q}$ to is the Green's operator which is expressed as a series in $\mathcal{P}_{\mathbf{p}}$. After a suitable truncation, we can in fact obtain a polynomial in $\mathcal{P}_{\mathbf{p}}$ which is a good approximation of $\mathcal{Q}$. By *good approximation in the ℓ^1 sense* we mean that one can find a polynomial R which is such that

(60)
$$\mathcal{Q}(x, y) \leq R(\mathcal{P}_p)(x, y)$$

for all x and y, and also such that $\|R(\mathcal{P}_p)(e, \cdot)\|_{\ell_1(\mathcal{G})}$ is not much larger than $\|\mathcal{Q}(e, \cdot)\|_{\ell_1(\mathcal{G})}$ (which is equal to 1). However, for our spectral computations, we want an approximation of $\mathcal{Q}$ in the ℓ^2 sense and it turns out that the above one is not satisfactory. In the same way that the Green's operator helps to find a good approximation in ℓ^1, we want to use the operator $\mathcal{R}'_{\mathbf{p}}$ defined by

$$\mathcal{R}'_{\mathbf{p}}(x, y) := \sqrt{\mathcal{R}_{\mathbf{p}}(x, y)}$$

to find a good approximation of $\mathcal{Q} \in \ell^2(\mathcal{G})$. What makes this approach successful for anisotropic random walk on free groups is that $\mathcal{R}'_{\mathbf{p}}$ correspond to a point of the resolvent of *another anisotropic random walk $\mathcal{P}_{\mathbf{p}'}$ for a vector $\mathbf{p}'$ which has the same support as $\mathbf{p}$*. Again we can approximate the resolvent operator by a polynomial by an *ad-hoc* truncation procedure.

Our study of the resolvent of the random walk, presented in Section 5.3, allows us to derive an explicit relation between $\mathbf{p}$ and $\mathbf{p}'$. Then in Section 5.4 we show that this relation combined with a technical but somehow natural truncation proceedure yields relevant bound on the kernel of the backbone walk (Proposition 21). Combining this with a few ℓ_2 computations (Lemma 22) this allows us to prove Theorem 1 by adapting the approach used for Theorem 5.

5.3. **The relation between p and p′ via resolvent.** The resolvent of $\mathcal{P}_{\mathbf{p}}$ is defined for $z \notin \sigma(\mathcal{P}_{\mathbf{p}})$ by

$$\mathcal{R}^z_{\mathbf{p}} = (z\mathcal{I} - \mathcal{P}_{\mathbf{p}})^{-1}.$$

In the above expression, $\mathcal{I}$ is the identity operator on $\ell^2(\mathcal{G})$. Since we are particularly interested in the behavior of the operator $\mathcal{R}^z_{\mathbf{p}}$ as $|z| > \varrho_{\mathbf{p}}$ approaches $\varrho_{\mathbf{p}}$, we consider the following alternative definition of $\mathcal{R}^z_{\mathbf{p}}(x, y)$ (which coincides with one above for $|z| > \varrho_{\mathbf{p}}$),

$$(61) \qquad \mathcal{R}^z_{\mathbf{p}}(x, y) := \sum_{t \geq 0} z^{-(t+1)} \mathcal{P}^t_{\mathbf{p}}(x, y).$$

As shown in [24] (see Lemma 16 below), the above series converges for all x and y if and only if $|z| \geq \varrho'_{\mathbf{p}}$ where $(\varrho'_{\mathbf{p}})^{-1}$ is the radius of convergence of the series $\mathcal{P}^t_{\mathbf{p}}(e, e)$. It is given by the following generalization of the Akemann-Ostrand formula (see [24, Equation (2.6)]),

$$(62) \qquad \varrho'_{\mathbf{p}} = \min_{s > 0} \left\{ 2s + \sum_{i=1}^{d} \left(\sqrt{s^2 + p_i p_{i^*}} - s \right) \right\},$$

and satisfies $\varrho'_{\mathbf{p}} \leq \varrho_{\mathbf{p}}$ (with equality in the symmetric case $p_i = p_i^*$ for all $i \in [d]$).

As our group is non-amenable, the vector $(\mathcal{R}^1_{\mathbf{p}}(e, x))_{x \in \mathcal{G}}$ is very close to be integrable (it does not belong to $\ell^1(\mathcal{G})$ but $(\mathcal{R}^z_{\mathbf{p}}(e, x))_{x \in \mathcal{G}}$ is in $\ell^1(\mathcal{G})$ for all $z > 1$), while $(\mathcal{R}^{\varrho_{\mathbf{p}}}_{\mathbf{p}}(e, x))_{x \in \mathcal{G}}$ is close to be in $\ell^2(\mathcal{G})$ in the same sense. What we prove in this section (and which is made plausible by the observation above) is the following:

Proposition 15. *Given $\mathbf{p}$ a probability vector on $[d]$ such that (12) holds, there exists a unique probability vector $\mathbf{p}'$ and a real $C = C(\mathbf{p})$ such that for all $x, y \in \mathcal{G}$,*

$$\mathcal{R}^1_{\mathbf{p}}(x, y) = C \left(\mathcal{R}^{\varrho_{\mathbf{p}'}}_{\mathbf{p}'}(x, y) \right)^2.$$

To our knowledge, this quadratic identity has not been discovered before. It is of fundamental importance in what follows: it establishes a relation between the vector $(\mathcal{R}^1_{\mathbf{p}}(e, x))_{x \in \mathcal{G}}$ - which, as seen in Section 4.4 is intimately connected with the entropy $\mathfrak{h}(\mathbf{p})$ - and $(\mathcal{R}^{\varrho_{\mathbf{p}'}}_{\mathbf{p}'}(e, x))_{x \in \mathcal{G}}$, the resolvent of $\mathcal{P}_{\mathbf{p}'}$ at its spectral edge. As a consequence of the tree structure of the Cayley graph associated with $(\mathcal{G}, \mathcal{A})$, that can be identified with the regular tree $\mathcal{T}_d$, the resolvent admits a simple "multiplicative" expression (this is a well established result that can be found e.g. in [24] or [22]). Indeed, $\mathcal{R}^z_{\mathbf{p}}(e, x)$ can be obtained by multiplying $\mathcal{R}^z_{\mathbf{p}}(e, e)$ by a quantity $r^z_i(\mathbf{p})$ for each edge of type i which is crossed on the minimal path linking e to x. Hence to prove Proposition 15, we need to find a probability vector $\mathbf{p}'$ such that for all $i \in [d]$, $r^1_i(\mathbf{p}) = \left(r^{\varrho_{\mathbf{p}'}}_i(\mathbf{p}') \right)^2$.

We need some extra-notation to give an expression for the coefficients r_i. Let us denote by $\mathcal{R}^z_{\mathbf{p},i}$ the resolvent of the operator $\mathcal{P}_{\mathbf{p},i} = \mathcal{P}_{\mathbf{p}} - (p_i \delta_e \otimes \delta_{g_i} + p_{i^*} \delta_{g_i} \otimes \delta_e)$ (defined as in (61) for $|z| \geq \varrho'_{\mathbf{p}}$) obtained from $\mathcal{P}_{\mathbf{p}}$ by removing the transitions between e and g_i. Finally we let γ^z_i be the diagonal coefficient of $\mathcal{R}^z_{\mathbf{p},i}$:

$$(63) \qquad \gamma^z_i = \gamma^z_i(\mathbf{p}) := \mathcal{R}^z_{\mathbf{p},i}(e, e) = \sum_{t \geq 0} \sum_{t \geq 0} z^{-(t+1)} \mathcal{P}^t_{\mathbf{p},i}(e, e).$$

Note that $\mathcal{P}^t_{\mathbf{p}}(e, e)$ is a function of $p_i p_{i^*}$, $i \in [d]$ (since every transitions from e to e involves the same number of multiplication by g_i and g_{i^*}. It implies in particular that $\gamma^z_i = \gamma^z_{i^*}$. Note that the

29

reference in [24] only treats the case $q_2 = d/2$ and assumes that every coordinates are positive. The positivity assumption however is not used in the proof (save for the fact that the return probability to zero decays exponentially, which is ensured by (12)). The proof also adapts to arbitrary values of q_1 and q_2 without any change (cf. [22, Proposition 3.4] which only deals with the case $q_1 = d$).

Lemma 16 (see Lemma 2 and Lemma 3 in [24], [22]). *For any reduced word $x = g_{i_1} \ldots g_{i_n} \in \mathcal{G}$ written in reduced form (that is $g_{i_{k+1}} \neq g_{i_k^*}$ for all k) and $|z| \geq \varrho'_{\mathbf{p}}$ (recall (62))*

$$\mathcal{R}_{\mathbf{p}}^z(e, x) = \mathcal{R}_{\mathbf{p}}^z(e, e) \prod_{t=1}^{n} p_{i_t} \gamma_{i_t}^z.$$

Moreover,

(64) $$\mathcal{R}_{\mathbf{p}}^z(e, e) = \left(z - \sum_{j \in [d]} p_{j^*} p_j \gamma_j^z \right)^{-1} \quad and \quad \gamma_i^z = \left(z - \sum_{j \neq i^*} p_{j^*} p_j \gamma_j^z \right)^{-1}.$$

The above lemma allows to compute explicitly the resolvent operator.

Lemma 17. *We assume that (12) holds and that $z \in [\varrho'_{\mathbf{p}}, \infty)$. If $s = s_z = 1/(2\mathcal{R}_{\mathbf{p}}^z(e, e))$ and $r_i = p_i \gamma_i^z$, we have $r_i r_{i^*} < 1$,*

$$r_i = \frac{\sqrt{s^2 + p_i p_{i^*}} - s}{p_{i^*}} \quad when \ p_{i^*} > 0, \quad r_i = \frac{p_i}{2s} \quad when \ p_{i^*} = 0 \quad and \quad p_i = \frac{2s r_i}{1 - r_i r_{i^*}}.$$

Moreover, s_z is the largest real solution of the following equation in x

$$z = 2x + \sum_{j=1}^{d} \left(\sqrt{x^2 + p_j p_{j^*}} - x \right).$$

Proof. From (64), and the fact that $\gamma_i = \gamma_{i^*}$ (recall (63)) we have $2s = z - \sum_j p_{j^*} r_j$ and $p_i p_{i^*} \gamma_i^2 + 2s\gamma_i - 1 = 0$. The inequality $r_i r_{i^*} = p_i p_{i^*} \gamma_i^2 < 1$ and the formulas follow (also in the case $p_i p_{i^*} = 0$). It remains to prove that s is the largest solution of $f(x) = z$ with $f(x) = 2x + \sum_j \left(\sqrt{x^2 + p_j p_{j^*}} - x \right)$. Since f is strictly convex and has a unique minimizer $x_{\min} \geq 0$ such that $f(x_{\min}) = \varrho'_{\mathbf{p}}$. The equation $f(x) = z$ has either zero, one or two solutions according to whether $z < \varrho'_{\mathbf{p}}$, $z = \varrho'_{\mathbf{p}}$, $z > \varrho'_{\mathbf{p}}$. In the latter case, we let $x_-(z) < x_{\min} < x_+(z)$ denote the two solutions. As s_z is an increasing function of z we have $s_z > x_{\min}$ for $z > \varrho'_{\mathbf{p}}$ and thus $s(z) = x_+(z)$. $\qquad\square$

Lemma 18. *Let $z \in [\varrho_{\mathbf{p}}, \infty)$ and $r_i = p_i \gamma_i^z$. We have*

$$\sum_{i=1}^{d} \frac{r_i(1 - r_{i^*})}{1 - r_i r_{i^*}} = 1 \Leftrightarrow z = 1 \quad and \quad \sum_{i=1}^{d} \frac{r_i^2(1 - r_{i^*}^2)}{1 - (r_i r_{i^*})^2} = 1 \Leftrightarrow z = \varrho_{\mathbf{p}}.$$

The result is a direct consequence of the following combinatorial statement (whose proof we include in the appendix for completeness).

Lemma 19. *For any $(\alpha_i)_{i=1}^d$ in $[0,1)^d$, the function defined on $\mathcal{G}$ by $F(x) := \prod_{t=1}^n \alpha_{i_t}$ if $x = g_{i_1} \dots g_{i_n}$ in reduced form, then F is integrable for the uniform counting measure on $\mathcal{G}$ if and only if*

$$\sum_{i=1}^d \frac{\alpha_i(1 - \alpha_{i^*})}{1 - \alpha_i \alpha_{i^*}} < 1.$$

Proof of Lemma 18. From (61), $\sum_{x \in \mathcal{G}} \mathcal{R}_{\mathbf{p}}^z(e, x) < \infty$ if and only if $z > 1$. On the other hand, from spectral considerations, $\|\mathcal{R}_{\mathbf{p}}^z \delta_e\|_2$ is finite for $z > \varrho_{\mathbf{p}}$ and diverges as z goes to $\varrho_{\mathbf{p}}$. Hence, recalling definition (61), we have $\sum_{x \in \mathcal{G}} \left(\mathcal{R}_{\mathbf{p}}^z(e, x)\right)^2 < \infty$ if and only if $z > \varrho_{\mathbf{p}}$. Lemma 19 implies that

$$\sum_{i=1}^d \frac{r_i(1 - r_{i^*})}{1 - r_i r_{i^*}} < 1 \Leftrightarrow z > 1 \qquad \text{and} \qquad \sum_{i=1}^d \frac{r_i^2(1 - (r_{i^*})^2)}{1 - (r_i r_{i^*})^2} < 1 \Leftrightarrow z > \varrho_{\mathbf{p}}.$$

We can conclude using the fact that (cf. (63)) the γ_i^zs are continuous functions of z. $\qquad \square$

Now we are ready to identify the value of $\mathbf{p}'$ which is such that (15) holds. We set, for $i \in [d]$, $r_i^z(\mathbf{p}) = p_i \gamma_i^z(\mathbf{p})$ and we introduce the vectors $\mathbf{a}$ and $\mathbf{b}$ whose coordinates are given for all $i \in [d]$ by

$$a_i(\mathbf{p}) := r_i^1(\mathbf{p}) \quad \text{and} \quad b_i(\mathbf{p}) := r_i^{\varrho_{\mathbf{p}}}(\mathbf{p}).$$

The formulas for the coordinates a_i and b_i of $\mathbf{a}$ and $\mathbf{b}$ are given in Lemma 17 and Lemma 18 can be used to determine $\varrho_{\mathbf{p}}$ (this characterization of $\varrho_{\mathbf{p}}$ could also be deduced from [35, Corollary 3.1]). Notably, by Lemma 17 we have $a_i a_{i^*}, b_i b_{i^*} \in [0,1)$. We can now reformulate and prove Proposition 15.

Proposition 20. *For any probability vector $\mathbf{p}$ on $[d]$, there exists a unique probability vector $\mathbf{p}'$ on $[d]$ with the same support as $\mathbf{p}$ such that for all $i \in [d]$, we have*

$$a_i(\mathbf{p}) = \left(b_i(\mathbf{p}')\right)^2.$$

It is given by the formula, for all $i \in [d]$,

$$p_i' = \frac{\sqrt{a_i(\mathbf{p})}}{1 - \sqrt{a_i(\mathbf{p})a_{i^*}(\mathbf{p})}} \left(\sum_{j \in [d]} \frac{\sqrt{a_j(\mathbf{p})}}{1 - \sqrt{a_j(\mathbf{p})a_{j^*}(\mathbf{p})}}\right)^{-1}.$$

Proof of Proposition 20. For ease of notation, we set $r_i' = \sqrt{a_i(\mathbf{p})}$. Assume that $\mathbf{p}'$ is a probability vector such that $b_i(\mathbf{p}') = r_i'$ for all $i \in [d]$. By Lemma 17, $(p_i')_{i \in [d]}$ is the probability vector proportional to $(r_i'/(1 - r_i' r_{i^*}'))_{i \in [d]}$ and hence we have uniqueness. We now prove existence. We set $p_i' = 2s r_i'/(1 - r_i' r_{i^*}')$ where s is the normalization constant such that $\mathbf{p}'$ is a probability vector. Now setting

$$(65) \qquad\qquad z = 2s + \sum_{j=1}^d \left(\sqrt{s^2 + p_j' p_{j^*}'} - s\right),$$

we only need to check that $s = 2/\mathcal{R}_{\mathbf{p}'}^z(e, e)$. Indeed if this is the case, Lemma 17 implies that $r_i' = r_i^z(\mathbf{p}')$ and Lemma 18 implies that $z = \varrho_{\mathbf{p}'}$. In view of (65) and of the proof of Lemma

17, we only need to discard the possibility that $s < x_{\min}$ where $x_{\min}$ is the minimizer of $f(x) :=$ $2x + \sum_{j=1}^{d} \left(\sqrt{x^2 + p'_j p'_{j*}} - x \right)$. Our definitions for p'_i and s imply that

(66)
$$r'_i = \frac{\sqrt{s^2 + p'_i p'_{i*}} - s}{p'_{i*}} \text{ if } p'_{i*} > 0 \quad \text{and} \quad r'_i = \frac{p'_i}{2s} \text{ if not.}$$

Since both expressions above are monotone in s, if $s < x_{\min}$, one would have $r'_i > q_i^{\varrho'_{\mathbf{p'}}}(\mathbf{p'})$ whenever $p'_i > 0$ since $r_i^{\varrho'_{\mathbf{p'}}}(\mathbf{p})$ is obtained by substituting s by $x_{\min}$ in (66) (here we use the definition (62) which implies that $\varrho'_{\mathbf{p'}} = f(x_{\min})$, as well as Lemma 17). As $\varrho'_{\mathbf{p'}} \le \varrho_{\mathbf{p'}}$, this also implies that $r'_i > r_i^{\varrho_{\mathbf{p'}}}(\mathbf{p'})$ and thus that

(67)
$$\sum_{i=1}^{d} \frac{(r'_i)^2 (1 - (r'_{i*})^2)}{1 - (r'_i r'_{i*})^2} > 1$$

which yields a contradiction to the definition of r'_i. $\qquad\square$

5.4. **Deducing mixing time from a bounding kernel.** Our aim now is to work with the same stopping time and backbone walk as in Section 4.3 and use the information we have to approximate the transition matrix of the backbone walk $Q_n = P_{n,\mathbf{q}}$ where the probability vector $\mathbf{q}$ was defined below (51), with a power series of $P_{n,\mathbf{p'}}$ the transition matridx of the nearest neighbor random walk associated with $\mathbf{p'}$ of Proposition 20. We further define Q'_n to be the following truncated series (which approximates a multiple of the resolvent of $P_{n,\mathbf{p'}}$ at $z = \varrho_{\mathbf{p'}}$)

$$Q'_n := \frac{1}{\sqrt{k}} \sum_{t=0}^{\lfloor \log k \rfloor^4} \left(\frac{P_{n,\mathbf{p'}}}{\varrho_{\mathbf{p'}}} \right)^t.$$

(The fact that $\varrho_{\mathbf{p'}}$ is positive can be deduced from the expression (61) with a few simple computations. It also follows from the forthcoming Lemma 25).

Proposition 21. *Given $\mathbf{p}$ a probability vector on $[d]$, there exists a real $C = C(\mathbf{p})$ such that for $\mathbf{p'}$ given by Proposition 20, we have, for all $x, y \in V_n$*

$$Q_n(x, y) \le C Q'_n(x, y)$$

We postpone the proof of this proposition to Section 5.6 and deduce Theorem 1 out of it. The proof includes a few technical lemmas whose proofs are postponed to the end of this section.

Proof of Theorem 1. Our first step is to use the comparison above to obtain spectral estimates for Q_n. We cannot control directly the spectral gap but we can estimate the contraction of functions with large variance. More precisely, given a matrix A of size $n \times n$ and $1 \le u \le \sqrt{n}$, we define

(68)
$$\kappa_u(A) := \sqrt{\max_{f : \|f\|_2 \ge \frac{u}{\sqrt{n}} \|f\|_1} \frac{\langle Af, Af \rangle}{\langle f, f \rangle}}.$$

Note that $\kappa_1(A)$ is the operator norm of A and $\kappa_{\sqrt{n}}(A)$ is the square root of the maximal diagonal entry of $A^* A$. For general u, the scalar $\kappa_u(A)$ can be thought of as a kind of pseudo-norm of A

restricted to vectors which are localized in terms of their ℓ^2 over ℓ^1 ratio. The function $u \mapsto \kappa_u(A)$ can be thought of as a spectral analog (for a matrix) of the isoperimetric profile of a graph (if A is the adjacency matrix of a graph, the isoperimetric profile is essentially obtained by restricting the maximum in (68) to functions f which are indicator functions of a subset of vertices).

Lemma 22. *Let A, B be two $n \times n$ matrices such that B is a bistochastic matrix. Assume that for some real $c \geq 0$ and all $x, y \in [n]$, we have $|A(x,y)| \leq cB(x,y)$ then for all $1 \leq u \leq \sqrt{n}$,*

$$(69) \qquad \kappa_u(A) \leq c\sigma(B) + \frac{c}{u},$$

where $\sigma(B) = \|B_{|\mathbf{1}^\perp}\|_{2 \to 2}$ is the singular radius of B.

From Proposition 21, we may apply Lemma 22 when $A = Q_n$ and $B = \alpha Q'_n$, with $\alpha = k^{-1/2} \sum_{t=0}^{\lceil \log k \rceil^4} \varrho_{\mathbf{p}'}^{-t}$ and $c = C\alpha^{-1}$ for the constant C given by Proposition 21. In this case, from the triangle inequality, we have

$$(70) \qquad c\sigma(B) = \left\| \frac{C}{\sqrt{k}} \sum_{t=0}^{\lfloor \log k \rfloor^4} \frac{(P_{n,\mathbf{p}'}^t)_{|\mathbf{1}^\perp}}{\varrho_{\mathbf{p}'}^t} \right\| \leq \frac{C}{\sqrt{k}} \sum_{t=0}^{\lfloor \log k \rfloor^4} \left(\frac{\sigma_{n,\mathbf{p}'}(t)}{\varrho_{\mathbf{p}'}} \right)^t,$$

and, since $\varrho_{\mathbf{p}'} > 0$ (see forthcoming Lemma 25), for some adequate choice of C'

$$c = \frac{C}{\sqrt{k}} \sum_{t=0}^{\lfloor \log k \rfloor^4} \varrho_{\mathbf{p}'}^{-t} \leq e^{C'(\log k)^4}.$$

We now bound (70). For that, we use the next proposition which quantifies the convergence of $\sigma_{\mathbf{p}}(t)$ to $\varrho_{\mathbf{p}}$ in (14).

Proposition 23. *For any probability vector $\mathbf{p}$ and integer $t \geq 1$, we have*

$$\varrho_{\mathbf{p}} \leq \sigma_{\mathbf{p}}(t) \leq (t+1)^{2/t} \varrho_{\mathbf{p}}.$$

From Proposition 23, we deduce that

$$\sum_{t=0}^{\lfloor \log k \rfloor^4} \left(\frac{\sigma_{\mathbf{p}'}(t)}{\varrho_{\mathbf{p}'}} \right)^t \leq ((\log k)^4 + 1)^3,$$

Using Assumption (23) and Lemma 22, for any fixed $k \geq 5$, for all $n \geq n_0(k)$ sufficiently large, we obtain

$$(71) \qquad \kappa_u(Q_n) \leq \frac{(\log k)^{13}}{\sqrt{k}}.$$

Now we want to use this estimate to build an adapted time for the original walk $P_{n,\mathbf{p}}$. The idea is first to iterate Q_n several times in order to contract the ℓ^2 norm below the threshold u and then use the original transition matrix $P_{n,\mathbf{p}}$ to finish the job. For this purpose, for a large integer k (we assume $k > (\log k)^{26}$) which will be conveniently fixed later on, and for $n \geq 3$, we set

$$a_n := \left\lfloor \frac{\log n}{\log k - 26 \log \log k} \right\rfloor \quad \text{and} \quad b_n := \lfloor \log \log n \rfloor.$$

We define $T := b_n + \tau_{a_n}$ where $(\tau_s)_{s \geq 0}$ are as in Subsection 4.3 the successive times of the backbone walk. Our spectral estimates (71) implies that X_T is close to equilibrium:

Lemma 24. *For any fixed integer $k \geq 3$, let a_n, b_n be as above and $T = b_n + \tau_{a_n}$. If Assumption (23) holds, then we have*

$$\lim_{n \to \infty} \max_{x \in V_n} \|\mathbb{P}_x[X_T \in \cdot] - \pi_n\|_{\mathrm{TV}} = 0.$$

To show that

$$\max_{x \in V_n} T_{n,\mathbf{p}}^{\mathrm{mix}}(x, \varepsilon) \leq (1 + \delta)(\log n)/\mathfrak{h}(\mathbf{p})$$

for n sufficiently large, we use Proposition 9 with $t = t_n$ and $s = s_n$ where

$$T = b_n + \tau_{a_n}, \quad t_n := \lfloor (1 + \delta/2)(\log n)/\mathfrak{h}(\mathbf{p}) \rfloor \quad \text{and} \quad s_n := \lfloor (\delta/2)(\log n)/\mathfrak{h}(\mathbf{p}) \rfloor.$$

With this setup, the first term in (33) tends to zero according to Lemma 24. For the third one we need to show that $\sigma_{n,\mathbf{p}}(s_n)$ is bounded away from one. Since (30) holds for $\mathbf{p}'$ we have (cf. (56))

$$(72) \qquad \limsup_{\substack{n \to \infty \\ s \to \infty}} \sigma_{n,\mathbf{p}'}(s) = \varrho_{\mathbf{p}'} < 1.$$

Now since $\mathbf{p}$ and $\mathbf{p}'$ have the same support one can compare $\sigma_{n,\mathbf{p}}(s)$ and $\sigma_{n,\mathbf{p}'}(s)$. More precisely applying [36, Lemma 13.22] to the operators $\mathcal{P}_{\mathbf{p}}^s (\mathcal{P}_{\mathbf{p}}^*)^s$ and $\mathcal{P}_{\mathbf{p}'}^s (\mathcal{P}_{\mathbf{p}'}^*)^s$ yields for every s and n

$$(73) \qquad \frac{1 - \sigma_{n,\mathbf{p}}(s)^{2s}}{1 - \sigma_{n,\mathbf{p}'}(s)^{2s}} \geq \min_{i \in [d]} \left(\frac{p_i}{p_i'} \right)^{2s}.$$

Hence $\limsup_{n \geq 1} \sigma_{n,\mathbf{p}}(s_0) < 1$ for some s_0 and thus from (57) we get èthat

$$(74) \qquad \limsup_{\substack{n \to \infty \\ s \to \infty}} \sigma_{n,\mathbf{p}}(s) < 1.$$

It remains to show that

$$\lim_{n \to \infty} \mathbb{P}[\tau_{a_n} > t - b_n] = 0.$$

From the law of large number and Lemma 14, for any $\delta > 0$, we may choose an integer k sufficiently large such that

$$\lim_{n \to \infty} \mathbb{P}\left[T \leq \left(1 + \frac{\delta}{4} \right) a_n \frac{\log k}{\mathfrak{h}(\mathbf{p})} \right] = 1.$$

This concludes the proof of Theorem 1. $\qquad \square$

Proof of Lemma 22. The statement is an immediate consequence the following functional inequality valid for every f

$$(75) \qquad \sqrt{\langle Af, Af \rangle} \leq c\sigma(B)\|f\|_2 + \frac{c}{\sqrt{n}}\|f\|_1.$$

Since B is bistochastic, the constant functions are left invariant by B and its transpose. It follows that $\sigma(B)$ is the operator norm of B projected on functions with zero sum. Now given f, if $|f|$ is the vector $|f|(x) := |f(x)|$ and $|A|$ is the matrix $|A|(x, y) := |A(x, y)|$, we have

$$\langle Af, Af \rangle \leq \langle |A||f|, |A||f| \rangle \leq c^2 \langle B|f|, B|f| \rangle.$$

34

The orthogonal projection of $|f|$ on zero sums functions is $\underline{f}(x) := |f|(x) - \|f\|_1/n$. We have

(76) $$\langle B|f|, B|f|\rangle = \|f\|_1^2/n + \langle B\underline{f}, B\underline{f}\rangle \leq \|f\|_1^2/n + \sigma(B)^2\|f\|_2^2.$$

We deduce (75) using the triangle inequality, $\sqrt{a^2 + b^2} \leq |a| + |b|$. $\qquad\square$

Proof of Lemma 24. Recall $T = b_n + \tau_{a_n}$. The distribution of X_T can be written as

$$\mathbb{P}_x[X_T \in \cdot] = (P_{n,\mathbf{p}}^{b_n} Q_n^{a_n})(x, \cdot).$$

We first show that for any $x \in V_n$ (recall $u = u_k := e^{(\log k)^5}$)

(77) $$\|Q_n^{a_n}(x, \cdot) - \pi_n\|_2 \leq \frac{2u}{\sqrt{n}}.$$

Since Q_n is a contraction, we note that $\|Q_n^t(x, \cdot) - \pi_n\|_2$ is non-decreasing in t. Moreover,

$$\|Q_n^{t+1}(x, \cdot) - \pi_n\|_2 = \|Q_n(Q_n^t(x, \cdot) - \pi_n)\|_2 \leq \max\left(\kappa_u(Q)\|Q_n^t(x, \cdot) - \pi_n\|_2, \frac{2u}{\sqrt{n}}\right),$$

where we have used that $\|Q_n^t(x, \cdot) - \pi_n\|_1 \leq 2$. Hence, an immediate induction yields for all $t \geq 0$,

$$\|Q_n^t(x, \cdot) - \pi_n\|_2 \leq \max\left(\kappa_u(Q)^t, \frac{2u}{\sqrt{n}}\right).$$

Thus, our bound (71) and our choice for a_n imply (77). To conclude the proof, we use the usual ℓ^2 bound and combine it with (77). This gives

(78)
$$\|P_{n,\mathbf{p}}^{b_n} Q_n^{a_n}(x, \cdot) - \pi_n\|_{\mathrm{TV}} \leq \frac{\sqrt{n}}{2}\|P_{n,\mathbf{p}}^{b_n} Q_n^{a_n}(x, \cdot) - \pi_n\|_2 \leq \frac{\sqrt{n}}{2}\sigma_{\mathbf{p},n}(b_n)^{b_n}\|Q^{a_n}(x, \cdot) - \pi_n\|_2 \leq \sigma_{\mathbf{p},n}(b_n)^{b_n} u.$$

Finally we conclude by using (72) and that b_n tends to infinity. $\qquad\square$

Remark 5.1 (Relaxation of our assumption concerning the spectral radius). *As in Remark 3.1, we denote by $\dim_0(H)$ the flat-dimension of a vector space H of $\mathbb{R}^{V_n}$ and we set $\varrho_{n,\mathbf{p}}^H$ to be the operator norm of $P_{n,\mathbf{p}}$ on the orthogonal of H. We may modify Theorem 1 as follows: if (H_n) is a sequence of invariant vector spaces of $P_{n,\mathbf{p}'}$ such that $\lim_n \varrho_{n,\mathbf{p}'}^{H_n} = \varrho_{\mathbf{p}'}$ and $\dim_0(H_n) \leq n^{o(1)}$ (that is $\lim_n \log \dim_0(H_n)/\log n = 0$) then the conclusion of Theorem 1 holds.*

Indeed, in Lemma 22, if H is an invariant subspace of the bistochastic matrix B and its transpose, then (69) can be improved to $\kappa_u(A) \leq c\varrho_H(B) + c\sqrt{\dim_0(H)}/u$, where $\varrho_H(B)$ is the operator norm of B on the orthogonal of H. Recall that if P_H is the orthogonal projection onto H, then $\|P_H g\|_2 \leq \|g\|_1\sqrt{\dim_0(H)/n}$. Setting $\underline{g} = |f| - P_H|f|$, we may thus replace the bound (76) by $\langle B|f|, B|f|\rangle \leq \|f\|_1^2 \dim_0(H)/n + \langle B\underline{g}, B\underline{g}\rangle \leq \|f\|_1^2 \dim_0(H)/n + \varrho_H(B)\|f\|_2^2$. It gives the claimed improvement of (69). The rest of the argument is essentially unchanged (the sequence b_n has to be chosen so that $\varepsilon_n \log n \ll b_n \ll \log n$).

Remark 5.2 (More quantitative bound on the mixing time). *A more quantitative upper bound on $T_n^{\mathrm{mix}}(\varepsilon)$ can be obtained by choosing k_n tending to infinity, and using a more quantitative version*

of Proposition 6 for anisotropic walks on trees. In the reversible case (14), optimizing all choices of parameters in our proof, we obtain a result of the form

$$T_{n,\mathbf{p}}^{\mathrm{mix}}(\varepsilon) \leq \frac{\log n}{\mathfrak{h}(\mathbf{p})} + C(\log n)^{2/3}$$

provided that $\varrho_{n,\mathbf{p}'}$ converges fast enough to $\varrho_{\mathbf{p}'}$ as n go to infinity (more specifically we require $\varrho_{n,\mathbf{p}'} \leq \varrho_{\mathbf{p}'} + C(\log n)^{-1/3}$). Note that our correction term is larger than $(\log n)^{1/2}$, and thus the proof developed in this section does not allow to obtain the anisotropic counter-part of Equations (38), (39), which allow to describe more accurately the profile of relaxation to equilibrium provided some quantitative information about the convergence (23) is given.

5.5. **Proof of Proposition 23.** We start with a general lemma on the spectral radius of the operator $\mathcal{P}_{\mathbf{p}}$ and the probability of transitions.

Lemma 25. *Let $\mathcal{G}$ be a finitely generated group. For any probability vector $\mathbf{p} \in \ell^2(\mathcal{G})$, any integer $t \geq 1$, any $x \in \mathcal{G}$, we have*

$$\|\mathcal{P}_{\mathbf{p}}^t \delta_x\|_2 \leq \varrho_{\mathbf{p}}^t.$$

Proof. We may assume $x = e$ without loss of generality. Since $\mathcal{P}_{\mathbf{p}}^t$ is the generator of a random walk with spectral radius $\varrho_{\mathbf{p}}^t$ we may also assume that $t = 1$. We have that $\|\mathcal{P}_{\mathbf{p}} \delta_e\|_2 \leq \sigma_{\mathbf{p}}(1) = \|\mathcal{P}_{\mathbf{p}}\|_{2 \to 2}$. If the reversibility condition (14) holds, then $\sigma_{\mathbf{p}}(1) = \varrho_{\mathbf{p}}$ and the lemma follows. In the general case, we use the group structure to obtain the required bound. We first write that for any integer $k \geq 1$,

$$\|\mathcal{P}_{\mathbf{p}} \delta_e\|_2^{2k} = \left(\sum_{x \in \mathcal{G}} \mathcal{P}_{\mathbf{p}}(e, x)^2 \right)^k = \sum_{x_1, \ldots, x_k} \left(\mathcal{P}_{\mathbf{p}}(e, x_1) \cdots \mathcal{P}_{\mathbf{p}}(e, x_k) \right)^2.$$

Using that $\mathcal{P}_{\mathbf{p}}(xg, yg) = \mathcal{P}_{\mathbf{p}}(x, y)$ for all x, y, g in $\mathcal{G}$, we may write

$$\mathcal{P}_{\mathbf{p}}(e, x_1) \cdots \mathcal{P}_{\mathbf{p}}(e, x_k) = \mathcal{P}_{\mathbf{p}}(e, x_1)\mathcal{P}_{\mathbf{p}}(x_1, x_2 x_1) \cdots \mathcal{P}_{\mathbf{p}}(x_{k-1} \cdots x_1, x_k \cdots x_1),$$

and

$$\sum_{x \in \mathcal{G}} \left(\mathcal{P}_{\mathbf{p}}^k(e, x) \right)^2 = \sum_x \left(\sum_{x_1, \ldots, x_{k-1}} \mathcal{P}_{\mathbf{p}}(e, x_1)\mathcal{P}_{\mathbf{p}}(x_1, x_2 x_1) \cdots \mathcal{P}_{\mathbf{p}}(x_{k-1} \cdots x_1, x) \right)^2$$

$$\geq \sum_x \sum_{x_1, \ldots, x_{k-1}} \left(\mathcal{P}_{\mathbf{p}}(e, x_1)\mathcal{P}_{\mathbf{p}}(x_1, x_2 x_1) \cdots \mathcal{P}_{\mathbf{p}}(x_{k-1} \cdots x_1, x) \right)^2$$

$$= \sum_{x_1, \ldots, x_k} \left(\mathcal{P}_{\mathbf{p}}(e, x_1)\mathcal{P}_{\mathbf{p}}(x_1, x_2 x_1) \cdots \mathcal{P}_{\mathbf{p}}(x_{k-1} \cdots x_1, x_k \cdots x_1) \right)^2.$$

We deduce that

$$\|\mathcal{P}_{\mathbf{p}} \delta_e\|_2^{2k} \leq \|\mathcal{P}_{\mathbf{p}}^k \delta_e\|_2^2 \leq \sigma_{\mathbf{p}}(k)^2.$$

We now let k tend to infinity and apply (17). $\qquad\square$

Proposition 23 is now an immediate consequence of the RD property (29) for the free group.

Proof of Proposition 23. Haagerup inequality (that is, RD property for free groups) implies that for any $t \geq 0$,

$$\sigma_{\mathbf{p}}(t)^t = \|\mathcal{P}_{\mathbf{p}}^t\|_{2 \to 2} \leq (t+1)^2 \|\mathcal{P}_{\mathbf{p}}^t \delta_e\|_2,$$

see [27, Lemma 1.4] (the proof is written in the case of the free group, denoted by $\mathcal{G}_{\text{free}}^{d,0}$ with our notation, but applies also to $\mathcal{G}_{\text{free}}^{q_1,q_2}$ with $q_1 + 2q_2 = d$). It remains to use Lemma 25. $\qquad\square$

5.6. **Proof of Proposition 21.** The matrices Q_n and $P_{n,\mathbf{p}'}$ are both defined as the transition kernel corresponding to projections of Markov chains on the group $\mathcal{G}$ on V_n. From (26)-(28), if $\mathbf{q}$ is a finitely supported probability vector on $\mathcal{G}$, for all x, y in V_n,

$$P_{n,\mathbf{q}}(x,y) = \sum_{g \in \mathcal{G}} \mathcal{P}_{\mathbf{q}}(e,g) \mathbf{1}(\varphi_n(g,x) = y),$$

where φ_n is the action of $\mathcal{G}$ on V_n. It is thus sufficient to prove the inequality for the corresponding kernels $\mathcal{Q}$ (as in (51)) and $\mathcal{P}_{\mathbf{p}'}$ on $\mathcal{G}$, that is

$$(79) \qquad \forall x \in \mathcal{G}, \quad \mathcal{Q}(e,x) \leq \frac{C}{\sqrt{k}} \sum_{t=0}^{\lfloor \log k \rfloor^4} \left(\frac{\mathcal{P}_{\mathbf{p}'}}{\varrho_{\mathbf{p}'}} \right)^t (e,x).$$

Since $\mathcal{Q}(e,x) = 0$ for all $x \notin \partial U$, it is sufficient to check (79) for $x \in \partial U$. By Lemma 16, if $z \geq \varrho_{\mathbf{p}}$ and $x = g_i y$ for some $g_i \in \mathcal{A}$, then $\mathcal{R}^z(e,x) \geq c\mathcal{R}^z(e,y)$ for some positive $c = c(\mathbf{p}, z)$. Since $\mathcal{R}^1(e,y) \geq 1/k$ for all $y \in U$, we find for all $x \in \partial U$,

$$\mathcal{Q}(e,x) \leq \frac{1}{k} \leq \frac{C}{\sqrt{k}} \sqrt{\mathcal{R}_{\mathbf{p}}^1(e,x)},$$

with $C = 1/\sqrt{c}$. Thus, from Proposition 15, for some new constant $C = C(\mathbf{p})$, for all $x \in \partial U$,

$$\mathcal{Q}(e,x) \leq \frac{C}{\sqrt{k}} \mathcal{R}_{\mathbf{p}'}^{\varrho_{\mathbf{p}'}}(e,x).$$

To deduce (79) from this last bound, we expand the resolvent as a power series. It requires some care because, when the reversibility condition (14) holds, $z = \varrho_{\mathbf{p}'}$ is precisely the threshold $\varrho'_{\mathbf{p}}$ for the convergence the power series (61).

With the notation of Lemma 16, for any $\mathbf{p}$ and $i \in [d]$, the function $z \mapsto \gamma_i^z$ is decreasing in $z \geq \varrho_{\mathbf{p}}$. Moreover, by Lemma 17, using the strict convexity of the function f there, we have for all $z \geq \varrho_{\mathbf{p}}$, $\gamma_i^{\varrho_{\mathbf{p}}} - \gamma_i^z \leq C\sqrt{z - \varrho_{\mathbf{p}}}$ for some $C = C(\mathbf{p})$ (the inequality is even valid without square-root when $\varrho_{\mathbf{p}} < \varrho'_{\mathbf{p}}$ for an adequate choice of constant). By Lemma 16, it follows that for some new $C = C(\mathbf{p})$ for all $x \in \mathcal{G}$,

$$|\mathcal{R}_{\mathbf{p}}^{\varrho_{\mathbf{p}}}(e,x) - \mathcal{R}_{\mathbf{p}}^z(e,x)| \leq C(|x|+1)\sqrt{z - \varrho_{\mathbf{p}}} \mathcal{R}_{\mathbf{p}}^{\varrho_{\mathbf{p}}}(e,x),$$

where $|x|$ is the distance of x to e in the tree $\mathcal{T}_d$ and where we have used the telescopic sum decomposition (with the convention that a product over an empty set is one)

$$\prod_{i=1}^k a_i - \prod_{i=1}^k b_i = \sum_{j=1}^k \left(\prod_{i=1}^{j-1} a_i \right) (b_j - a_j) \left(\prod_{i=j+1}^k b_i \right).$$

By Lemma 13, the diameter of ∂U being at most $C \log k$, we find that for all $x \in \partial U$, $\mathcal{R}_{\mathbf{p}}^{\varrho_{\mathbf{p}}}(e, x) \leq 2\mathcal{R}_{\mathbf{p}}^{z}(e, x)$ provided that $0 \leq z - \varrho_{\mathbf{p}} \leq c(\log k)^{-2}$ for some positive constant $c = c(\mathbf{p}) > 0$. We now fix $z = \varrho_{\mathbf{p}'} + c(\mathbf{p}')(\log k)^{-2}$. From what precedes, for all $x \in \partial U$,

$$\mathcal{R}_{\mathbf{p}'}^{\varrho_{\mathbf{p}'}}(e, x) \leq 2\mathcal{R}_{\mathbf{p}'}^{z}(e, x) = \frac{2}{z} \sum_{t=0}^{\infty} \left(\frac{\mathcal{P}_{\mathbf{p}'}}{z} \right)^{t} (e, x).$$

By Lemma 25, we have $\mathcal{P}_{\mathbf{p}'}^{t}(e, x) \leq \varrho_{\mathbf{p}'}^{t}$ and, for some new constant $C = C(\mathbf{p}')$, for any $s \geq 0$,

$$\sum_{t=s}^{\infty} \left(\frac{\mathcal{P}_{\mathbf{p}'}}{z} \right)^{t} (e, x) \leq C(\log k)^{2} e^{-\frac{s}{C(\log k)^{2}}}.$$

We now recall that by Proposition 15, for all $x \in \partial U$, $\mathcal{R}_{\mathbf{p}'}^{z}(e, x) \geq \mathcal{R}_{\mathbf{p}'}^{\varrho_{\mathbf{p}'}}(e, x)/2 \geq c/\sqrt{k}$. It follows that if $s = \lfloor C' \log k \rfloor^{3}$ for some large enough constant C', we have

$$\frac{1}{z} \sum_{t=s}^{\infty} \left(\frac{\mathcal{P}_{\mathbf{p}'}}{z} \right)^{t} (e, x) \leq \frac{1}{2} \mathcal{R}_{\mathbf{p}'}^{z}(e, x).$$

Consequently, for this value of s,

$$\mathcal{R}_{\mathbf{p}'}^{z}(e, x) \leq \frac{2}{z} \sum_{t=0}^{s} \left(\frac{\mathcal{P}_{\mathbf{p}'}}{z} \right)^{t} (e, x) \leq \frac{2}{\varrho_{\mathbf{p}'}} \sum_{t=0}^{s} \left(\frac{\mathcal{P}_{\mathbf{p}'}}{\varrho_{\mathbf{p}'}} \right)^{t} (e, x).$$

This concludes the proof of (79) $\qquad \square$

6. Random walks covered by a colored group

6.1. **Minimal mixing time for color covered random walks.** We now present a last extension of our results. As in the setting of Theorem 5, we assume that for a finitely generated non-amenable group $\mathcal{G}$, we have a sequence of finite sets (V_n) with $\#V_n = n$ and (φ_n) a sequence of actions of $\mathcal{G}$ on V_n. Let $r \geq 1$ be an integer. We think of $[r] = \{1, \ldots, r\}$ as a set of colors. An element $\mathbf{p} \in M_r(\mathbb{R})^{\mathcal{G}}$ is written as a matrix-valued vector $\mathbf{p} = (p_g)_{g \in \mathcal{G}}$ with $p_g \in M_r(\mathbb{R})$. The support of $\mathbf{p}$ is then the subset of $\mathcal{G}$ such that p_g is not the null matrix. We consider $\mathbf{p} \in M_r(\mathbb{R})^{\mathcal{G}}$ with finite support such that

$$P_{1,\mathbf{p}} := \sum_{g \in \mathcal{G}} p_g$$

is an irreducible stochastic matrix on $[r]$ with invariant probability measure μ. Then, we denote by $\mathcal{P}_{\mathbf{p}}$ the operator on $\ell^2(\mathcal{G} \times [r])$ defined by

(80)
$$\mathcal{P}_{\mathbf{p}} = \sum_{g \in \mathcal{G}} p_g \otimes \lambda(g),$$

where $\lambda(g)$ is as in (28) and $\otimes$ is the tensor product. In probabilistic terms, $\mathcal{P}_{\mathbf{p}}$ is the transition kernel of a random walk $(\mathcal{X}_t)$ on $\mathcal{G} \times [r]$ where the probability to jump from (g, u) to (h, v) is

$p_{hg^{-1}}(u,v)$. We denote by $\varrho_{\mathbf{p}}$ the spectral radius of $\mathcal{P}_{\mathbf{p}}$ and by $\mathfrak{h}(\mathbf{p})$ the entropy rate of $\mathcal{P}_{\mathbf{p}}$ defined by: for any $u_0 \in [r]$,

$$\mathfrak{h}(\mathbf{p}) = \lim_{t \to \infty} -\frac{1}{t} \sum_{(g,u) \in \mathcal{G} \times [r]} \mathcal{P}_{\mathbf{p}}^t((e,u_0),(g,u)) \log \mathcal{P}_{\mathbf{p}}^t((e,u_0),(g,u)).$$

The fact that $\mathfrak{h}(\mathbf{p})$ does not depend on u_0 is an easy consequence of the assumption that $P_{1,\mathbf{p}}$ is irreducible. Again, if $\mathcal{G}$ is non-amenable and $\varrho_{\mathbf{p}} < 1$ holds, then $\mathfrak{h}(\mathbf{p}) > 0$. Besides, the proof of Shannon-McMillan-Breiman Theorem in [31, Theorem 2.1], actually proves that if $\mathcal{X}_0 = (e,u_0)$, a.s.

$$(81) \qquad\qquad \mathfrak{h}(\mathbf{p}) = \lim_{t \to \infty} -\frac{\log \mathcal{P}_{\mathbf{p}}^t((e,u_0),\mathcal{X}_t)}{t}.$$

With $(S_g)_{g \in \mathcal{G}}$ as in (26), we define the stochastic matrix on $\mathbb{R}^{V_n \times [r]}$

$$(82) \qquad\qquad P_{n,\mathbf{p}} = \sum_{g \in \mathcal{G}} p_g \otimes S_g.$$

This matrix is the transition kernel a Markov chain on $V_n \times [r]$ covered by $(\mathcal{X}_t)$ in the sense that if we define for $(g,u) \in \mathcal{G} \times [r]$ and $x \in V_n$, $\bar{\varphi}_n((g,u),x) := (\varphi_n(g,x),u)$ then $X_t := \bar{\varphi}_n(\mathcal{X}_t,x)$ is a Markov chain with transition matrix $P_{n,\mathbf{p}}$ started at (x,u_0). The measure $\pi_n(x,u) = \mu(u)/n$ is an invariant probability of $P_{n,\mathbf{p}}$. Moreover, since (81) holds, the proof of Proposition 6 actually implies that mixing time of X_t, $T_{n,\mathbf{p}}^{\mathrm{mix}}(x,\varepsilon)$, satisfies for any fixed $\varepsilon \in (0,1)$ and uniformly in $x \in V_n$, the lower bound (22).

This setting allows to consider a random walk on the *n-lift of a base graph*. More precisely, let G_1 be a finite simple connected graph with $d/2$ undirected edges on the vertex set $[r]$. We consider the free group $\mathcal{G}_{\mathrm{free}}$ with $d/2$ generators and their $d/2$ inverses $(g_1, \ldots, g_d)$, that is $g_i^{-1} = g_{i*}$ for some involution on $[d]$ without fixed point. Each generator g_i is associated to a directed edge (u_i,v_i) of G_1 and $g_i^{-1} = (v_i,u_i)$ is the inverse directed edge. We consider the action of $\mathcal{G}_{\mathrm{free}}$ on $[n]$ defined by $\varphi_n(g_i,x) = \sigma_i(x)$ where $(\sigma_1, \ldots, \sigma_d)$ are permutation matrices such that $\sigma_i^{-1} = \sigma_{i*}$. Then, if $E_{k,\ell} \in M_r(\mathbb{R})$ is the canonical matrix defined by $E_{k,\ell}(i,j) = \mathbf{1}_{\{(k,\ell)=(i,j)\}}$, then the graph G_n with vertex set $[n] \times [r]$ and adjacency matrix $\sum_i E_{u_i,v_i} \otimes S_i$ is a simple graph which is called a n-lift (or a n-covering) of G_1: the $[n] \times [r] \to [r]$ map $\psi(x,u) = u$ is n to 1 and, for any (x,u), the image by ψ of the adjacent vertices of (x,u) in G_n coincides with the adjacent vertices of $\psi(x,u)$ in G_1. If d_u is the degree of the vertex u in G_1 and $p_{g_i} = E_{u_i,v_i}/d_{u_i}$ then $P_{1,\mathbf{p}}$ and $P_{n,\mathbf{p}}$ are the transition matrices of the simple random walks on G_1 and G_n respectively.

We are ready to state the analog of Theorem 5.

Theorem 26. *Let $\mathcal{G}$ be a finitely generated non-amenable group with the property RD, (V_n) a sequence of finite sets with $\#V_n = n$ and (φ_n) a sequence of actions of $\mathcal{G}$ on V_n which converges strongly. For any integer $r \geq 1$ and any finitely supported $\mathbf{p} \in M_r(\mathbb{R})^{\mathcal{G}}$ such that $\varrho_{\mathbf{p}} < 1$ and $P_{1,\mathbf{p}}$ is*

an irreducible aperiodic Markov chain, the mixing time of the random walk with transition matrix $P_{n,\mathbf{p}}$ satisfies, for every $\varepsilon \in (0,1)$,

$$\lim_{n \to \infty} \frac{T_{n,\mathbf{p}}^{\mathrm{mix}}(\varepsilon)}{\log n} = \frac{1}{\mathfrak{h}(\mathbf{p})}.$$

Note that in the above statement the RD property and the strong convergence property are defined in terms of scalar valued vectors $\mathbf{p} \in \ell^2(\mathcal{G})$. From [15], an example of application of Theorem 26 is the random walk on a random n-lift of a weighted base graph such that $P_{1,\mathbf{p}}$ is irreducible and aperiodic (see [19] for a recent alternative and independent proof of this case).

6.2. **Proof of Theorem 26.** We let $(\mathcal{X}_t)$ be the random walk with kernel $\mathcal{P}_{\mathbf{p}}$ started from $\mathcal{X}_0 = (e, u_0)$. For $(g, u) \in \mathcal{G} \times [r]$ and $x \in V_n$, we set $\bar{\varphi}_n((g, u), x) = (\varphi_n(g, x), u)$ and let $X_t := \bar{\varphi}_n(\mathcal{X}_t, x)$ be a Markov chain with transition matrix $P_{n,\mathbf{p}}$ started at (x, u_0). We adapt the arguments of Section 4 to our matrix-valued context.

6.2.1. *Relative spectral radius, strong convergence and RD property.* Let $\mathbf{q} = (q_g) \in M_r(\mathbb{R})^{\mathcal{G}}$ with finite support. We define $\ell^2(\mu)$ as the Hilbert space on $\mathbb{R}^r$ endowed with the scalar product $\langle f, g \rangle_\mu = \sum_i \mu(i) \bar{f}(i) g(i)$. Similarly, $\ell_n^2(\mu)$ and $\ell_{\mathcal{G}}^2(\mu)$ are the Hilbert spaces on the vector spaces $\mathbb{R}^{V_n \times [r]}$ and $\mathbb{R}^{\mathcal{G} \times [r]}$ endowed with the scalar products:

$$\langle f, g \rangle_\mu = \sum_{(x,i) \in X \times [r]} \mu(i) \bar{f}(x, i) g(x, i),$$

with $X = V_n$ and $X = \mathcal{G}$ respectively. We note that the subspace of $\mathbb{R}^{V_n \times [r]}$: $H_r = \mathbb{R}^r \otimes \mathbf{1}$ of vectors f of the form for some $g \in \mathbb{R}^r$, $f(x, i) = g(i)$ is an invariant subspace of dimension r for $P_{n,\mathbf{q}}$ and its adjoint in $\ell_n^2(\mu)$. Hence $P_{n,\mathbf{q}}$ admits a direct sum decomposition on $H_r \oplus H_r^\perp$. We note also the restriction of $P_{n,\mathbf{q}}$ to H_r coincides with $P_{1,\mathbf{q}}$. We define the *relative singular radius* as the following operator norm

$$(83) \qquad \bar{\sigma}_{n,\mathbf{q}} := \|(P_{n,\mathbf{q}})_{|H_r^\perp}\|_{\ell_n^2(\mu) \to \ell_n^2(\mu)}.$$

From [41, p256] (see also [44]), if (φ_n) converges strongly then we have

$$(84) \qquad \lim_{n \to \infty} \bar{\sigma}_{n,\mathbf{q}} = \sigma_{\mathbf{q}},$$

where $\sigma_{\mathbf{q}} := \|\mathcal{P}_{\mathbf{q}}\|_{\ell_{\mathcal{G}}^2(\mu) \to \ell_{\mathcal{G}}^2(\mu)}$.

Besides, let $E_{ij} \in M_r(\mathbb{R})$ be the canonical matrix with all entries zero but entry (i, j) equal to 1. The $\ell^2(\mu) \to \ell^2(\mu)$ operator norm of E_{ij} is $\sqrt{\mu(i)/\mu(j)}$. Since $\varrho_{\mathbf{q}}$ coincides with the $\ell_{\mathcal{G}}^2(\mu) \to \ell_{\mathcal{G}}^2(\mu)$ operator norm, from the triangle inequality, we have

$$\sigma_{\mathbf{q}} = \left\| \sum_{i,j \in [r]^2} \sum_{g \in \mathcal{G}} q_g(i, j) E_{ij} \otimes \lambda(g) \right\|_{\ell_{\mathcal{G}}^2(\mu) \to \ell_{\mathcal{G}}^2(\mu)} \leq \sum_{(i,j) \in [r]^2} \sqrt{\frac{\mu(i)}{\mu(j)}} \sigma_{\mathbf{q}(i,j)},$$

where $\mathbf{q}(i,j) = (q_g(i,j)) \in \mathbb{R}^{\mathcal{G}}$ and $\sigma_{\mathbf{q}(i,j)}$ is the singular radius of $\mathcal{P}_{\mathbf{q}(i,j)}$ in $\ell^2(\mathcal{G})$. It follows, that if $\mathcal{G}$ has the RD property and R is the diameter of the support of $\mathbf{q}$ (in the Cayley graph associated to any symmetric generating set $\mathcal{A}$) then, for some constant $C(\mathcal{G}, \mathcal{A}) > 0$,

$$(85) \qquad \sigma_{\mathbf{q}} \leq CR^C \sum_{(i,j) \in [r]^2} \sqrt{\frac{\mu(i)}{\mu(j)}} \sqrt{\sum_{g \in \mathcal{G}} q_g(i,j)^2} \leq Cr^2 R^C \sqrt{\sum_{g \in \mathcal{G}} \|q_g\|^2_{\ell^2(\mu) \to \ell^2(\mu)}},$$

where we have used that $\sqrt{\mu(i)/\mu(j)}|q_g(i,j)| = \|q_g(i,j)E_{ij}\|_{\ell^2(\mu) \to \ell^2(\mu)} \leq \|q_g\|_{\ell^2(\mu) \to \ell^2(\mu)}$.

6.2.2. *Skeleton Walk.* We now adapt the argument of Subsection 4.3. We let $\mathcal{R}_{\mathbf{p}} = (\mathcal{I}_{\mathcal{G} \times [r]} - \mathcal{P}_{\mathbf{p}})^{-1}$ be the Green's operator associated with $\mathcal{P}_{\mathbf{p}}$. For $g, h \in \mathcal{G}$, we denote by $\mathcal{R}_{\mathbf{p}}(g,h) \in M_r(\mathbb{R})$ the matrix whose entry (i,j) is $\mathcal{R}_{\mathbf{p}}((g,i),(h,j))$. For $g \in \mathcal{G}$, we define $u(g) \in M_r(\mathbb{R})$ as the matrix

$$u(g) := \mathcal{R}_{\mathbf{p}}(e,g) = \sum_{t=0}^{\infty} \mathcal{P}_{\mathbf{p}}^t(e,g),$$

where $\mathcal{P}_{\mathbf{p}}^t(g,h) \in M_r(\mathbb{R})$ has entry (i,j) equal to $\mathcal{P}_{\mathbf{p}}^t((g,i),(h,j))$. Given $k \geq 1$, we define the set

$$(86) \qquad U := \{g \in \mathcal{G} : \|u(g)\|_{\ell^2(\mu) \to \ell^2(\mu)} \geq 1/k\}.$$

The backbone walk is the induced walk on the successive exit times from U: $\tau_0 := 0$, $\tau_1 = \tau$ and, for integer $s \geq 1$, $\tau_{s+1} := \inf\{t \geq \tau_k : \mathcal{X}_t \mathcal{X}_{\tau_s}^{-1} \notin U\}$. We define $\mathcal{Q} = \mathcal{P}_{\mathbf{q}}$ as the transition kernel of the random walk $\mathcal{X}_{\tau_m}$.

From (81) and $\varrho_{\mathbf{p}} < 1$, the proofs and statements of Proposition 13 and Lemma 14 continue to hold in our new setting (in (52), we replace $q_g \leq 1/k$ by $\|q_g\|_{\ell^2(\mu) \to \ell^2(\mu)} \leq 1/k$).

6.2.3. *Deducing mixing time from RD property and the strong convergence.* We may now conclude the proof of Theorem 26 by adapting the content of Subsection 4.4.

Proof of Theorem 26. We fix $\varepsilon \in (0,1)$, $\delta > 0$ and $(x, u_0) \in V_n \times [r]$ and prove that for n sufficiently large

$$T_{n,\mathbf{p}}^{\mathrm{mix}}((x, u_0), \varepsilon) \leq (1 + \delta) \log n / \mathfrak{h}(\mathbf{p}).$$

Let (τ_m) and U be as above for some large k to be chosen. We set $m := \lfloor (1 + \delta/4)(\log n) / \log k \rfloor$.

For integer $s \geq 1$, the relative s-th singular radius is

$$\bar{\sigma}_{n,\mathbf{p}}(s) := \|(P_{n,\mathbf{p}}^s)_{|H_r^\perp}\|_{\ell_n^2(\mu) \to \ell_n^2(\mu)}^{1/s} \quad \text{and} \quad \sigma_{\mathbf{p}}(s) := \|\mathcal{P}_{\mathbf{p}}^s\|_{\ell_{\mathcal{G}}^2(\mu) \to \ell_{\mathcal{G}}^2(\mu)}^{1/s}.$$

From (84), for all $s \geq 1$, $\lim_n \bar{\sigma}_{n,\mathbf{p}}(s) = \sigma_{\mathbf{p}}(s) < 1$. Since $\varrho_{\mathbf{p}} < 1$ and $\lim_{s \to \infty} \sigma_{\mathbf{p}}(s) = \varrho_{\mathbf{p}}$, we deduce that for all $s \geq s_0$ large enough and all $n \leq n_0$ large enough, $\sigma_{n,\mathbf{p}}(s) \leq 1 - \delta_0$ for some $\delta_0 > 0$ (we argue as below (57)). Moreover, since $P_{1,\mathbf{p}}$ is irreducible and aperiodic, we have that $\sigma_{1,\mathbf{p}} < 1$. We deduce that for $\delta = \min(\delta_0, 1 - \sigma_{1,\mathbf{p}}) > 0$, for all $s \geq s_0$ and all $n \geq n_0$

$$(87) \qquad \sigma_{n,\mathbf{p}}(s) = \|(P_{n,\mathbf{p}}^s)_{|\mathbf{1}^\perp}\|_{\ell_n^2(\mu) \to \ell_n^2(\mu)}^{1/s} = \max(\bar{\sigma}_{n,\mathbf{p}}(s), \sigma_{1,\mathbf{p}}(s)) \leq 1 - \delta.$$

We use Proposition 9 for the walk $X_t = \bar{\varphi}_n(\mathcal{X}_t, x)$ with

$$T = \tau_m + s, \quad t = \lfloor (1 + \delta) \log n / \mathfrak{h} - 2 \log \log n \rfloor \quad \text{and} \quad s = \lfloor \log \log n \rfloor.$$

41

For our choice of s, it follows from (87) that the third term in (33) is smaller than $\varepsilon/3$. It remains to prove that for n sufficiently large

$$(88) \qquad \mathbb{P}[\tau_m > t] \leq \varepsilon/3 \quad \text{and} \quad \|P_{n,\mathbf{p}}^s Q_n^m((x,u_0),\cdot) - \pi_n\|_{\mathrm{TV}} \leq \varepsilon/3.$$

where $Q_n = P_{n,\mathbf{q}}$ is the Markov chain of the induced walk $X_{\tau_m} = \bar{\varphi}_n(\mathcal{X}_{\tau_m}, x)$ on $V_n \times [r]$. For the first inequality of (88), we choose $k(\delta)$ sufficiently large and it is a consequence of Lemma 14 and the law of large numbers.

The second inequality of (88) is obtained using spectral estimates for $Q_n = P_{n,\mathbf{q}}$. The Cauchy-Schwarz inequality gives

$$(89) \qquad \|P_{n,\mathbf{p}}^s Q_n^m((x,u_0),\cdot) - \pi_n\|_{\mathrm{TV}} \leq C_0 \sqrt{n} \|Q_n^m P_{n,\mathbf{p}}^s f\|_{\ell_n^2(\mu)},$$

with $C_0 = \sqrt{r/\min_i \mu(i)}$ and $f(y,v) = \delta_{(x,u_0)}(y,v)/\mu(v) - 1/n$. Let Π_H be the orthogonal projection in $\ell_n^2(\mu)$ onto a vector space H. We find

$$(90) \qquad \|Q_n^m P_{n,\mathbf{p}}^s f\|_{\ell_n^2(\mu)} \leq \|Q_n^m P_{n,\mathbf{p}}^s \Pi_{H_r} f\|_{\ell_n^2(\mu)} + \|Q_n^m P_{n,\mathbf{p}}^s \Pi_{H_r^\perp} f\|_{\ell_n^2(\mu)}.$$

We now compute a spectral bound of the two terms on the right-hand side of (90). We first observe that $\|f\|_{\ell_n^2(\mu)} \leq 1$ and $\|\Pi_{H_r} f\|_{\ell_n^2(\mu)} \leq C/\sqrt{n}$ with $C = 1/\sqrt{\min_i \mu(i)}$. Since $\langle f, \mathbf{1} \rangle_\mu = 0$, we find from (87) and the fact that Q_n is a contraction in $\ell_n^2(\mu)$,

$$(91) \qquad \|Q_n^m P_{n,\mathbf{p}}^s \Pi_{H_r} f\|_{\ell_n^2(\mu)} \leq \|P_{n,\mathbf{p}}^s \Pi_{H_r} f\|_{\ell_n^2(\mu)} \leq \frac{C}{\sqrt{n}} \sigma_{n,\mathbf{p}}(s)^s \leq \frac{C}{\sqrt{n}}(1-\delta)^s.$$

We now give a bound of the second term on the right-hand side of (90). From (85) and Proposition 13, we have for some constant C

$$\sigma_{\mathbf{q}} \leq C k^{-1/2} (\log k)^C.$$

From (84) we deduce that for all n large enough,

$$\bar{\sigma}_{n,\mathbf{q}} \leq 2C k^{-1/2} (\log k)^C.$$

Since $\|f\|_{\ell_n^2(\mu)} \leq 1$, $P_{n,\mathbf{p}}^s \Pi_{H_r^\perp} = \Pi_{H_r^\perp} P_{n,\mathbf{p}}^s$ and $P_{n,\mathbf{p}}$ is a contraction in $\ell_n^2(\mu)$, we deduce that

$$(92) \qquad \|Q_n^m P_{n,\mathbf{p}}^s \Pi_{H_r^\perp} f\|_{\ell_n^2(\mu)} \leq \bar{\sigma}_{n,\mathbf{q}}^m \|f\|_{\ell_n^2(\mu)} \leq 2C k^{-1/2}(\log k)^C$$

Equation (89) together with (90)-(91)-(92) guaranties that X_{τ_m+s} is close to equilibrium in total variation. It concludes the proof of (88). $\qquad \square$

7. Proof of Lemma 19

Let us consider $\mathcal{A}$ the set of finite words in the alphabet $[d]$, $\mathcal{B}$ the set of words without repetition in $[d]$ and for a fixed involution $*$ on $[d]$, $\mathcal{C}$ the set of finite words in which the patterns ii^* and i^*i do not appear.

Given $\alpha = (\alpha_i)_{i=1}^d$ a set of non-negative numbers in $[0,1)^d$, we define the function $F_\alpha(i_1.\ldots.i_t) = \alpha_{i_1} \cdots \alpha_{i_t}$. We have immediately

$$\text{(93)} \qquad \sum_{\mathcal{A}} F_\alpha(\mathbf{i}) < \infty \Leftrightarrow \sum_{[d]} \alpha_i < 1.$$

Now a word $\mathbf{i} \in \mathcal{A}$ can be encoded by a word $\mathbf{j} \in \mathcal{B}$ and and a sequence $(n_t)_{t=1}^{|\mathbf{j}|}$ which counts how many time each letter is repeated. For this reason we have, given $\beta = (\beta_i)_{i=1}^d$, $\sum_{\mathcal{A}} F_\beta(\mathbf{i}) = \sum_{\mathcal{B}} F_{\beta'}(\mathbf{i})$, where $\beta_i' = \sum_{n \geq 1} \beta_i^n = \beta_i/(1-\beta_i)$. Hence taking $\beta_i = \alpha_i/(1+\alpha_i)$ we obtain

$$\text{(94)} \qquad \sum_{\mathcal{B}} F_\alpha(\mathbf{i}) < \infty \Leftrightarrow \sum_{[d]} \frac{\alpha_i}{1+\alpha_i} < 1.$$

Finally to encode a word in $\mathbf{i} \in \mathcal{C}$, we first consider a finite word $\mathbf{j}$ without repetition in $[d']$ where d' is the number of conjugation classes for $*$. Let $\mathcal{B}'$ be the set of such words $\mathbf{j}$. Then, to encode $\mathbf{i}$, we have to replace each of the letter of $\mathbf{j}$ by a pattern. If the conjugation class $j \in [d']$ is a single element $\{i\}$ in $[d]$, there is only one possible pattern which is i. We thus define the weight of j as $\gamma_j := \alpha_i$. Otherwise, the conjugation class $j \in [d']$ is a pair $\{i, i^*\}$. Then the possible patterns are $(i^*)^n$ and i^n, with $n \geq 1$. This gives a total weight $\gamma_j := \alpha_i/(1-\alpha_i) + \alpha_{i^*}/(1-\alpha_{i^*})$. We thus have $\sum_{\mathbf{i} \in \mathcal{C}} F_\alpha(\mathbf{i}) = \sum_{\mathbf{j} \in \mathcal{B}'} F_\gamma(\mathbf{j})$. In particular the sum is finite if and only if

$$\text{(95)} \qquad \sum_{j \in [d']} \frac{\gamma_j}{1+\gamma_j} = \sum_{i \in [d]} \left(\frac{\alpha_i}{1+\alpha_i} \mathbf{1}_{\{i=i^*\}} + \frac{1}{2} \frac{\alpha_i(1-\alpha_i)^{-1} + \alpha_{i^*}(1-\alpha_{i^*})^{-1}}{1+\alpha_i(1-\alpha_i)^{-1} + \alpha_{i^*}(1-\alpha_{i^*})^{-1}} \mathbf{1}_{\{i \neq i^*\}} \right) < 1.$$

This is the required statement. $\qquad \square$

REFERENCES

[1] C. A. Akemann and P. A. Ostrand. Computing norms in group C^*-algebras. *Amer. J. Math.*, 98(4):1015–1047, 1976.

[2] 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.

[3] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, pages 333–348, 1986.

[4] N. Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. Theory of computing (Singer Island, Fla., 1984).

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

[6] A. Amit and N. Linial. Random graph coverings. I. General theory and graph connectivity. *Combinatorica*, 22(1):1–18, 2002.

[7] A. Amit and N. Linial. Random lifts of graphs: edge expansion. *Combin. Probab. Comput.*, 15(3):317–332, 2006.

[8] A. Avez. Entropie des groupes de type fini. *C. R. Acad. Sci. Paris Sér. A-B*, 275:A1363–A1366, 1972.

[9] R. Basu, J. Hermon, and Y. Peres. Characterization of cutoff for reversible Markov chains. *Ann. Probab.*, 45(3):1448–1487, 2017.

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

[11] N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly. Random walks on the random graph. *Ann. Probab.*, 46(1):456–490, 2018.

[12] S. Blachère, P. Haïssinsky, and P. Mathieu. Asymptotic entropy and Green speed for random walks on countable groups. *Ann. Probab.*, 36(3):1134–1152, 2008.

[13] C. Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. arXiv:1502.04482, 2015.

[14] C. Bordenave, P. Caputo, and J. Salez. Random walk on sparse random digraphs. *Probab. Theory Related Fields*, 170(3-4):933–960, 2018.

[15] C. Bordenave and B. Collins. Eigenvalues of random lifts and polynomials of random permutation matrices. *Ann. of Math. (2)*, 190(3):811–875, 2019.

[16] P. Cartier. *Harmonic analysis on trees.* Amer. Math. Soc., Providence, R.I., 1973.

[17] T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli. Weighted expanders and the anisotropic Alon–Boppana theorem. *European Journal of Combinatorics*, 25(5):735 – 744, 2004.

[18] I. Chatterji. Introduction to the rapid decay property. In *Around Langlands correspondences*, volume 691 of *Contemp. Math.*, pages 53–72. Amer. Math. Soc., Providence, RI, 2017.

[19] G. Conchon-Kerjan. Cutoff for random lifts of weighted graphs. arXiv:1908.02898.

[20] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. U.S.A.*, 93(4):1659–1664, 1996.

[21] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 57(2):159–179, 1981.

[22] A. Figà-Talamanca and T. Steger. Harmonic analysis for anisotropic random walks on homogeneous trees. *Mem. Amer. Math. Soc.*, 110(531):xii+68, 1994.

[23] J. Friedman. Relative expanders or weakly relatively Ramanujan graphs. *Duke Math. J.*, 118(1):19–35, 2003.

[24] P. Gerl and W. Woess. Local limits and harmonic functions for nonisotropic random walks on free groups. *Probab. Theory Relat. Fields*, 71(3):341–355, 1986.

[25] J. Geronimus. On a set of polynomials. *Ann. of Math. (2)*, 31(4):681–686, 1930.

[26] R. I. Grigorchuk and A. Żuk. On the asymptotic spectrum of random walks on infinite families of graphs. In *Random walks and discrete potential theory (Cortona, 1997)*, Sympos. Math., XXXIX, pages 188–204. Cambridge Univ. Press, Cambridge, 1999.

[27] U. Haagerup. An example of a nonnuclear C^*-algebra, which has the metric approximation property. *Invent. Math.*, 50(3):279–293, 1978/79.

[28] U. Haagerup and F. Larsen. Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras. *J. Funct. Anal.*, 176(2):331–367, 2000.

[29] J. Hermon. Cutoff for Ramanujan graphs via degree inflation. *Electron. Commun. Probab.*, 22:Paper No. 45, 10, 2017.

[30] J. Hermon. A technical report on hitting times, mixing and cutoff. *ALEA Lat. Am. J. Probab. Math. Stat.*, 15(1):101–120, 2018.

[31] V. A. Kaĭmanovich and A. M. Vershik. Random walks on discrete groups: boundary and entropy. *Ann. Probab.*, 11(3):457–490, 1983.

[32] H. Kesten. Full Banach mean values on countable groups. *Math. Scand.*, 7:146–156, 1959.

[33] H. Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92:336–354, 1959.

[34] F. Ledrappier. Some asymptotic properties of random walks on free groups. In *Topics in probability and Lie groups: boundary theory*, volume 28 of *CRM Proc. Lecture Notes*, pages 117–152. Amer. Math. Soc., Providence, RI, 2001.

[35] F. Lehner. On the computation of spectra in free probability. *J. Funct. Anal.*, 183(2):451–471, 2001.

[36] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[37] L. Lovász and M. D. Plummer. *Matching theory.* AMS Chelsea Publishing, Providence, RI, 2009. Corrected reprint of the 1986 original [MR0859549].

[38] E. Lubetzky. private communication.

[39] E. Lubetzky and Y. Peres. Cutoff on all Ramanujan graphs. *Geom. Funct. Anal.*, 26(4):1190–1216, 2016.

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

[41] J. A. Mingo and R. Speicher. *Free probability and random matrices*, volume 35 of *Fields Institute Monographs.* Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017.

[42] B. Mohar. A strengthening and a multipartite generalization of the Alon-Boppana-Serre theorem. *Proc. Amer. Math. Soc.*, 138(11):3899–3909, 2010.

[43] J. Petersen. Die Theorie der regulären graphs. *Acta Math.*, 15(1):193–220, 1891.

[44] G. Pisier. A simple proof of a theorem of Kirchberg and related results on C^*-norms. *J. Operator Theory*, 35(2):317–335, 1996.

[45] S. Sodin. Random matrices, nonbacktracking walks, and orthogonal polynomials. *J. Math. Phys.*, 48(12):123503, 21, 2007.