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 $T_d$. 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 $T_d$), and we obtain a lower bound of the form $\frac{1}{h} \log n$, where $h$ is the entropy rate associated with $\mathcal{G}$. We call this the entropic lower bound.

It was recently proved that in the case $\mathcal{G} = T_d$, 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 the random walks on $G_n$ under which the random walk exhibit 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.

2010 Mathematics Subject Classification: 05C81, 60J10.

Keywords: Expander Graphs, Mixing time, Spectral Gap, Ramanujan Graph, Entropic time.

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 $3 \leq d \leq n - 1$ 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 $(X_t)_{t \geq 0}$ be the simple random walk on $G_n$, which is the Markov process with transition matrix, for all $x, y \in V_n$

$$P_n(x, y) = \frac{1_{\{x, y\} \in E_n}}{d}.$$ 

The uniform measure on $V_n$ denoted by $\pi_n$ is reversible for the process. Furthermore if $G_n$ is connected, then $\pi_n$ is the unique invariant measure of $P_n$. $G_n$ is not bipartite, then $P_n^t(x, \cdot)$ converges to $\pi_n$ 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 $\pi_n$ in terms of the total variation distance. More formally, the total variation mixing time associated
with threshold $\epsilon \in (0, 1)$ and initial condition $x \in V_n$, is defined by

$$T_n^{\text{mix}}(x, \epsilon) := \inf \{ t \in \mathbb{N} : d_n(x, t) < \epsilon \},$$

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

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

The worst-case mixing time is classically defined as

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

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 [29, Lemma 12.16]) allows to control the distance in function of the spectral radius of $P_n$: for all $x \in V_n$,

$$d_n(x, t) \leq \frac{\sqrt{n-1}}{2} \rho_n,$$

where

$$\rho_n = \left\| (P_n)_{1-} \right\|$$

is the operator norm of $P_n$ restricted to functions with zero sum (this is also the second largest eigenvalue of $P_n$ in absolute value counting multiplicities). This yields in particular that

$$T_n^{\text{mix}}(\epsilon) \leq \frac{1}{\left\| \log \rho_n \right\|} \left( \frac{1}{2} \log n - \log(2\epsilon) \right).$$

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

On the other hand, a naive lower bound of the same order of $T_n^{\text{mix}}(\epsilon)$ can be obtaind by using the elementary fact that the graph distance $\text{Dist}(x, X_t)$ between $X_t$ and the initial condition $x$ is stochastically dominated by a random walk on non-negative integers, started at 0, with jump probabilities $1/d$ to the left and $(d - 1)/d$ to the right (on positive integers). Thus, $\text{Dist}(x, X_t)$ remains in a ball of radius $r$ 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 $x \in V_n$

$$T_n^{\text{mix}}(x, 1 - \epsilon) \geq \frac{d}{(d-2) \log (d-1)} \left( \log n - C \epsilon \sqrt{\log n} \right).$$

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 [32].

However, an important observation is that the factor in front of $\log n$ in (3) and (4) cannot match: from Alon-Boppana lower bound [1, 35], we have for any sequence of graph $\liminf_n \rho_n \geq \rho := \frac{2\sqrt{d-1}}{d}$,
or, more precisely, there exists a constant $C = C(d)$ such that for every $n$ and every $d$-regular graph on $n$ vertices

$$\rho_n \geq \rho - \frac{C}{(\log n)^2}. \tag{5}$$

The number $\rho = 2\sqrt{d-1}/d$ is the spectral radius of the simple random walk on the infinite $d$-regular tree $T_d$ (and incidentally also for the biased random walk on integers used in the lower-bound strategy). A (non-bipartite) graph such that $\rho_n \leq \rho$ is called a Ramanujan graph. Hence a natural question is the following:

if a sequence of graphs on $n$ vertices has an asymptotic minimal spectral radius in the sense that $\rho_n = (1 + o(1))\rho$, does it also have a minimal mixing time in the sense that $T_n^{\text{mix}}(\varepsilon) = (1 + o(1))\frac{(d-2)}{d\log(d-1)}\log n$ for any fixed $\varepsilon \in (0, 1)$?

An affirmative answer was given to this question in [31] (see also [25]):

\textbf{Theorem A \([31]\).} Let $d \geq 3$ be an integer and let $(G_n)$ be a sequence of $d$-regular graphs on $n$ vertices such that their spectral radius satisfy $\lim_{n \to \infty} \rho_n = \rho = 2\sqrt{d-1}/d$. Then for any $\varepsilon \in (0, 1)$, we have

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

\textbf{Remark 1.1.} The result above remains of course valid if our sequence $G_n$ is not defined for every $n \in \mathbb{N}$ but only on an infinite subsequence of $\mathbb{N}$ provided that $\rho_n$ converges along this subsequence. 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 \textit{cutoff phenomenon}: a sequence of finite Markov chains $P_n$ exhibits cutoff if at first order $T_n^{\text{mix}}(\varepsilon)$ does not depend on $\varepsilon \in (0, 1)$, that is, for any $\varepsilon \in (0, 1)$, $\lim_{n \to \infty} T_n^{\text{mix}}(\varepsilon)/T_n^{\text{mix}}(1-\varepsilon) = 1$. Since its original discovery by Diaconis, Shashahani and Aldous in the context of card shuffling [19, 2, 3], this phenomenon has attracted much attention. We refer to [18, 29] 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, 13].

As a warmup, we provide a novel proof of Theorem A which is simpler than that presented in [31] and [25] (independently observed by Lubetzky [30]). A more precise version of Theorem A will be proved in Proposition 9 below (it notably allows to obtain the second order term in the asymptotic development of $T_n^{\text{mix}}(\varepsilon)$). With our approach we can also relax a little bit the assumption and allow the presence of $n^\alpha$ eigenvalues outside the interval $[-\rho, \rho]$, with $\alpha \in (0, 1)$ small enough, at the cost of discarding a small proportion of possibly bad starting points. More precisely, given $(G_n)$ a sequence of $d$-regular graphs on $n$ vertices, we define the upper semi-continuous function $I : [0, 1] \to [0, 1]$, which can be interpreted as an asymptotic density of eigenvalues on log-log scale

$$I(u) = \inf_{\varepsilon \downarrow 0} \frac{\log \left( \sum_{\lambda \in \text{Sp}(P_n)} |\lambda| < \varepsilon \right) \text{dim}(E_\lambda)}{\log n}, \tag{7}$$

$$\lim_{n \to \infty} T_n^{\text{mix}}(\varepsilon) = (1 + o(1))\frac{(d-2)}{d\log(d-1)}\log n$$

for any fixed $\varepsilon \in (0, 1)$?
where \( \dim(E^\lambda_n) \) denotes the dimension of the eigenspace corresponding to \( \lambda \).

**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 such that their spectral radius satisfy for all \( n \), \( \rho_n \leq 1 - \delta \) and for all \( u > \rho \),

\[
I(u) \leq 1 - 2 \frac{\log(u/\rho + \sqrt{(u/\rho)^2 - 1})}{\log(d - 1)}.
\]

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

\[
\lim_{n \to \infty} \# \left\{ x \in V_n : \frac{T_n^{\text{mix}}(x, \varepsilon)}{\log n} \geq (1 + \eta) \frac{d}{(d - 2) \log(d - 1)} \right\} / n^{1-\varepsilon}\delta = 0.
\]

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 \( T_n^{\text{mix}}(x, \varepsilon) \) does not depend on \( x \), and (9) implies that \( \lim_{n \to \infty} T_n^{\text{mix}}(\varepsilon)/\log n = d/(d - 2) \log(d - 1) \). See Remark 3.1 for a variant of Theorem B which allows to control \( T_n^{\text{mix}}(\varepsilon) \) at the cost of modifying the definition 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 with biased directions. For \( d \in \mathbb{N} \), we set \([d] = \{1, \ldots, d\}\). One way to define a biased 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 in \( d \) sets of edges \((E_{n,i})_{i \in [d]}\) where each vertex of \( V_n \) being 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 \( \ast : i \mapsto i^\ast \) of \([d] = \{1, \ldots, d\}\). We are going to make make the weaker assumption that \( G_n \) is a Schreier graph: the graph may have loops or multiple edges and we assume that its adjacency matrix may be written as

\[
\sum_{i=1}^{d} S_i,
\]

where, for each \( i \), \( S_i \) is a permutation matrix of a permutation \( \sigma_i \) on \( V_n \) and \( \sigma_i^\ast = \sigma_i^{-1} \). In graphical terms, this is equivalent to assume that the set of edges \( E_n \) admits a partition \( E_n = \cup_{j=1}^{q} E_{n,j} \) such that for each \( j \), \( E_{n,j} \) is a disjoint union of cycles of length at least 2 (where a cycle of length 2 is a single edge) and each vertex \( x \in V_n \) is adjacent to a cycle of \( E_{n,j} \). Indeed, \( q \) is the number of equivalence classes of the involution and the sets \( E_{n,j} \) are the cycle decompositions of the permutations (which coincide if \( \sigma_i = \sigma_i^{-1} \)). The case described above where \( E_n \) is partitionned in \( d \) types simply corresponds to the case were \( \ast \) is the identity. In Definition 2 below, we will give an extended definition of a Schreier graph.

If \( d \) is even, it is a standard exercise to check that any \( d \)-regular graph is a Schreier graph for some collection of \( d/2 \) permutations and their inverses. A more explicit example of such graph is any Cayley graph of a finitely generated group and a symmetric set of generators of size \( d \): the
permutations $S_i$ corresponds to the (left or right) multiplication by an element of the symmetric set of generators. We note that decomposing $d \geq 3$ as $d = 2k + l$, with $k, l$ non-negative integers, the infinite $d$-regular tree $T_d$ is the Cayley graph generated by $k$ free copies of $\mathbb{Z}$ and $l$ free copies of $\mathbb{Z}/2\mathbb{Z}$ with their natural generators.

Now given a Schreier graph $G_n$ with $\#V_n = n$ and given a probability vector $p = (p_1, \ldots, p_d)$ (with positive coordinates summing to one) such that

$$p_i > 0 \quad \text{and} \quad p_i^* = p_i, \quad \text{for all } i \in [d],$$

we consider the matrix

$$P_{n,p} = \sum_{i=1}^{d} p_i S_i.$$

Note that by construction $P_{n,p}$ is a symmetric Markovian matrix. This is the transition kernel of a random walk on $G_n$ which is called the anisotropic random walk. Again, $\pi_n$, the uniform measure on $V_n$, is reversible for this process. The spectral radius of $P_{n,p}$ is defined as the operator norm of $P_{n,p}$ projected onto the orthogonal of constant functions

$$\rho_{n,p} = \| (P_{n,p})_{1} \|,$$

(it coincides with the second largest eigenvalue of $P_{n,p}$ in absolute value counting multiplicities).

From what precedes, we may also define the anisotropic random walk on $T_d$ with probability vector $p = (p_1, \ldots, p_d)$. We denote by $P_{p}$ its transition kernel. We refer to see [28, 22, 20] for properties of the anisotropic random walks on $T_d$. From [24, 16], the Alon-Bopanna lower bound for the spectral radius of $P_{n,p}$ is

$$\rho_{n,p} \geq (1 + o(1))\rho_p,$$

where $\rho_p$ is the spectral radius of $P_{p}$, given by the classical Akemann-Ostrand formula [1],

$$\rho_p = \min_{s > 0} \left\{ 2s + \sum_{i=1}^{d} \left( \sqrt{s^2 + p_i^2} - s \right) \right\}.$$

The mixing time of the random walk admits a minimal asymptotic value. To make this more precise, we have to present the lower bound which is obtained using a generalization of the strategy yielding (4). Consider $(X_t)_{t \geq 0}$ an anisotropic random walk on $T_d$ with transition kernel $P_{p}$ and starting from the root of $T_d$ denoted by $e$. The entropy rate $h(p)$ of $P_p$ is defined as

$$h(p) := \lim_{t \to \infty} -\frac{1}{t} \sum_{x \in T_d} P_{p}^t(e, x) \log P_{p}^t(e, x),$$

We have $h(p) > 0$ as soon as $\sum_i 1_{\{p_i > 0\}} \geq 3$. From Shannon-McMillan-Breiman Theorem [20, Theorem 2.1], we have a.s.

$$h(p) = \lim_{t \to \infty} -\frac{\log P_{p}^t(e, X_t)}{t}.$$
In words, after a time \( t \), the random walk \( X_t \) typically remains within a set of size \( \exp(h(p)(t+o(t))) \) and the random walk is roughly uniformly distributed on this set. For any time \( t > 0 \) and \( x \in V_n \), the entropy of \( P_{t,n}^p(x,\cdot) \) is at most the entropy of \( P_t^p(e,\cdot) \) (see Proposition 6 below). As a consequence, we have that for any fixed \( \varepsilon \in (0,1) \), uniformly in \( x \in V_n \),

\[
T_{n,p}^{\text{mix}}(x,1-\varepsilon) \geq (1 + o(1)) \frac{\log n}{h(p)}.
\]

18

In the spirit of Theorem A, for a given probability vector \( p \), 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 \( \rho_{n,p} = (1 + o(1))\rho_p \), does it also have an asymptotic minimal mixing time in the sense that \( T_n^{\text{mix}}(\varepsilon) = (1 + o(1))(\log n)/h(p) \) for any fixed \( \varepsilon \in (0,1) \)?

The answer turns out to be somewhat more involved.

**Theorem 1.** Let \( d \geq 3 \) be an integer, \( * \) an involution on \([d]\) and let \( p \) be a probability vector on \([d] \) which satisfies the condition (11). Then, there exists another probability vector \( p' \) which satisfies the condition (11) such that the following holds. If a sequence of Schreier graphs \( G_n \) on \( n \) vertices as in (10) satisfies

\[
T_n^{\text{mix}}(\varepsilon) = (1 + o(1))(\log n)/h(p)
\]

for any fixed \( \varepsilon \in (0,1) \), then for every \( \varepsilon \in (0,1) \)

\[
\lim_{n \to \infty} T_n^{\text{mix}}(\varepsilon) = \frac{1}{h(p)}.
\]

The condition (19) can be thought as a Ramanujan property for the anisotropic random walk with probability \( p' \). In some cases, this condition (19) can be relaxed to allow \( n^{o(1)} \) eigenvalues outside the interval \([-\rho_{p'},\rho_{p'}] \); see Remark 5.1 below. An explicit expression for the vector \( p' \) is provided in the proof. In particular we have that \( p' = p \) if and only if \( p \) is the uniform vector. This result is thus a generalization of Theorem A. We believe that our result is sharp in the following sense: we conjecture that for any \( q \neq p' \), there exists a sequence of Schreier graphs \( G_n \) on \( n \) vertices such that \( \lim_{n \to \infty} \rho_{n,q} = \rho_q \) and \( \lim_{n \to \infty} \rho_{n,p'} \neq \rho_{p'} \). Among those graphs, we conjecture that there are graphs such that \( \liminf_{n \to \infty} T_n^{\text{mix}}(\varepsilon)/\log n > 1/h(p) \).

For \( p \neq \) different from the uniform vector, a source of example for Theorem 1 is in [14]. Up to the involution, we consider independent permutations \( \sigma_i \) on \([n] \) vertices: if \( i \neq i^* \), \( \sigma_i \) is a uniform permutation on \( n \) elements and, if \( i^* = i \), we take \( n \) even and \( \sigma_i \) is a uniform matching on \( n \) elements (where a matching is an involution without fixed point). Then, in probability, the condition (19) is true for any probability vector \( p' \) which satisfies the condition (11).

**Minimal mixing time for covered random walks.** We now present another extension. We start with a refinement of the notion of Schreier graph.

**Definition 2 (Group action, covering and Schreier graph).** Let \( \mathcal{G} \) be a finitely generated group with unit \( e \), \( V \) a finite set and a map \( \varphi : \mathcal{G} \times V \to V \) is an action of \( \mathcal{G} \) on \( V \), if we have for all \( x \in V \),
$g, h \in G, \varphi(e, x) = x$ and $\varphi(gh, x) = \varphi(g, \varphi(h, x))$. For any $g \in G$, 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 $G$ then the Schreier graph of $(G, \mathcal{A}, \varphi)$ 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 $(G, \mathcal{A}, \varphi)$, we say that $(G, \mathcal{A})$ is a covering of $G$.

Let us check that this definition extends the definition of Schreier graphs defined above. If the adjacency matrix of a finite graph $G = (V, E)$ is of the form (10), then $G$ is the Schreier graph of $(S_V, \mathcal{A}, \varphi)$ where $S_V$ is the symmetric group on $V$, $\mathcal{A} = \{\sigma_1, \ldots, \sigma_d\}$ and the covering map is $\varphi(\sigma, x) = \sigma(x)$. Conversely, if $G$ is the Schreier graph of $(G, \mathcal{A}, \varphi)$ as in Definition 2 with $\mathcal{A} = \{a_1, \ldots, a_d\}$ then its adjacency matrix is of the form (10) where the involution $\ast: i \mapsto i^\ast$ is defined as $i^\ast = j$ if and only if $a_j = a_i^{-1}$. Note that if the involution $\ast: i \mapsto i^\ast$ on $[d]$ has $l$ fixed points and $k + l$ equivalence classes then $G$ is $d$-regular with $d = 2k + l$. As already pointed, the infinite tree $T_d$ is the Cayley graph of the group $G_{\text{free}}$ generated by $k$ free copies of $\mathbb{Z}$ and $l$ free copies of $\mathbb{Z}/2\mathbb{Z}$ with their natural generators denoted $A_{\text{free}}$. By considering the group homomorphism from $G_{\text{free}}$ to $G$ which maps $A_{\text{free}}$ to $\mathcal{A}$, we deduce that all Schreier graphs are covered by $(G_{\text{free}}, A_{\text{free}})$ with the corresponding involution.

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

We are now interested in the more specific situation where, for some finitely generated non-amenable group $G$, we have a sequence of finite sets $(V_n)$ with $\#V_n = n$ and $(\varphi_n)$ a sequence of actions of $G$ on $V_n$. Let $p \in l^2(G)$ be a given finitely supported probability vector such that

\begin{equation}
\text{the support of } p \text{ generates } G \quad \text{and} \quad p_g = p_g^{-1} \quad \text{for all } g \in G.
\end{equation}

Then, we are interested in the random walk on $V_n$ with transition matrix

\begin{equation}
P_{n, p} = \sum_{g \in G} p_g S_g,
\end{equation}

where for $g \in G$, $S_g$ is as in Definition 2. Note that if the support of $p$ is contained in a finite symmetric set $\mathcal{A}$, then $P_{n, p}$ is an anisotropic random walk on the Schreier graph of $(G, \mathcal{A}, \varphi_n)$. 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.

Similarly, we denote by $\mathcal{P}_p$ the kernel of the corresponding random walk on $G$ defined as

\begin{equation}
\mathcal{P}_p = \sum_{g \in G} p_g \lambda(g),
\end{equation}

where, for $g \in G$, $\lambda(g)$ is the left multiplication operator (or the left regular representation of $g$) defined by: for all $h \in G$, $\lambda(g)(h) = gh$. Let $\rho_p$ be the spectral radius of $\mathcal{P}$ and let $\rho_{n, p}$ be the
spectral radius of $P_{n,p}$ defined in (13). From [24], the Alon-Boppana lower bound (14) is still valid. Moreover, from Kesten’s Theorem [27], if $G$ is non-amenable and if (21) holds, we have $\rho_p < 1$.

Similarly, we denote by $h(p)$ the entropy rate of $P_p$ defined by (15). From Avez’ Theorem [8], if $G$ is non-amenable and (21) holds, then $h(p) > 0$. From Proposition 6 below, the mixing time of the anisotropic random walk on $G_n$ is lower bounded in terms of $h(p)$: we have that for any fixed $\varepsilon \in (0, 1)$, uniformly in $x \in V_n$, (16) holds.

For a given probability vector $p$ supported by a generator, a natural question is thus the following: Are there spectral conditions for a sequence of actions $(\phi_n)$ of $G$ on $V_n$ with $\#V_n = n$ to guarantee that the anisotropic random walk on $V_n$ has an asymptotic minimal mixing time in the sense that $T_n^{\text{mix}}(\varepsilon) = (1 + o(1))(\log n)/h(p)$ for any fixed $\varepsilon \in (0, 1)$?

Our answer to this question is based on two notions of group algebra which we now define.

**Definition 3 (RD property).** A finitely generated group $G$ has the RD property if for every finitely supported $p \in \ell^2(G)$, the spectral radius $\rho_p$ of $P_p$ is well controlled by the $\ell^2$-norm of $p$ in the following sense: for any finite symmetric generating set $A$ of $G$, there exists a constant $C = C(G, A)$ such that

$$\rho_p \leq CR^C \|p\|_2,$$

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

We refer to [17] for an introduction to the RD property. We note that free groups and hyperbolic groups satisfy the RD property. Observe also that it is sufficient to check (24) for a single finite symmetric generating set $A$ of $G$.

Our second notion is the strong convergence of operators algebras which we define here in our specific Markovian setting. It can thought as an analog of the Ramanujan property for a sequence of group actions on finite sets.

**Definition 4 (Strong convergence).** Let $G$ be a finitely generated group, $(V_n)$ a sequence of finite sets and $(\phi_n)$ a sequence of covering maps of $G$ on $V_n$. We say that the covering maps $(\phi_n)$ converge strongly (on Markovian operators) if for every finitely supported probability vector $p \in \ell^2(G)$ such that the right-hand side of (21) holds, we have

$$\lim_{n \to \infty} \rho_{n,p} = \rho_p.$$

where $\rho_p$ is the spectral radius of $P_p$ defined in (23) and $\rho_{n,p}$ is the spectral radius of $P_{n,p}$ defined in (22) and (13).

We are now ready to state our final result.

**Theorem 5.** Let $G$ be a finitely generated non-amenable group with the property RD, $(V_n)$ a sequence of finite sets with $\#V_n = n$ and $(\phi_n)$ a sequence of actions of $G$ on $V_n$ which converges...
strongly. Then for every finitely supported probability vector $p \in \ell^2(\mathcal{G})$ such that (21) holds the mixing time of the random walk with transition matrix $P_{n,p}$ satisfies, for every $\varepsilon \in (0,1)$,

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

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 $n^{o(1)}$; see Remark 4.2 below.

A source of example for Theorem 5 is in [14]: random actions of the free group are strongly convergent. More precisely, let $*: i \mapsto i^*$ be an involution on $[d]$ with $l$ fixed points, let $\mathcal{G}_{\text{free}}$ be the group generated by $k = (d - l)/2$ free copies of $\mathbb{Z}$ and $l$ free copies of $\mathbb{Z}/2\mathbb{Z}$, and let $\mathcal{A}_{\text{free}} = \{a_1, \ldots, a_d\}$ be the natural set of generators. Up to the involution, we consider independent permutations $\sigma_{n,i}$ on $[n]$ vertices: if $i \neq i^*$, $\sigma_{n,i}$ is a uniform permutation on $n$ elements and, if $i^* = i$, we take $n$ even and $\sigma_{n,i}$ is a uniform matching on $n$ elements. We consider the action of $\mathcal{G}_{\text{free}}$ on $V_n = \{1, \cdots, n\}$ defined by $\varphi_n(a_i, x) = \sigma_{n,i}(x)$. Then, in probability, this sequence of actions is strongly convergent.

**Minimal mixing time for color covered random walks.** In this paper, we will also consider yet another extension which notably allows 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 [21], 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 (18) only stated in this introduction, and provide a general result (Proposition 7) 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 $k$-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 II concerning anisotropic walk, by combining the ideas of Section 4 with an analysis of the resolvent of the anisotropic random walk on $T_d$.

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

**Notation.** For $k$ a positive integer, we set $[k] = \{1, \ldots, k\}$. If $V$ is a countable set and $M$ is a bounded operator in $\ell^2(V)$, we use the notation: for all $x, y \in V$, $M(x, y) = \langle e_x, Me_y \rangle$ where $e_x$ is the coordinate vector at $x$. The integer part of real number $t$ is denoted by $\lfloor t \rfloor$. 
Acknowledgment. CB warmly thanks Nalini Anantharaman and Justin Salez for enlightening
discussions on anisotropic random walks. HL acknowledges support from a Productivity Grant
from CNPq and by a JCNE grant from FAPERJ. CB is supported by French ANR grant ANR-16-
CE40-0024-01.

2. Preliminaries

2.1. The entropic time lower bound. For the sake of completeness we provide a complete proof
of the entropic lower bound which is only sketched in the introduction. The result is stated in the
more general setup of Theorem 5. Recall that $G$ is a finitely generated non-amen able 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 $G$ on $V_n$, that $P_{n,p}$, $P_p$
denote the transitions matrix on $V_n$ and $G$ respectively defined by (22) and (23), and that $h(p)$ is
the entropy rate associated with $P_p$.

Proposition 6. 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 $P_{n,p}$ satisfies for any $\varepsilon \in (0, 1)$
$$\lim_{n \to \infty} \inf_x \frac{T_{n,p}^{\text{mix}}(x, 1 - \varepsilon)}{\log n} \geq \frac{1}{h(p)}.$$  

Proof. Let $(X_t)$ denote the random walk on $G$ starting from the unit $e$ and with transition $P_p$. Its
distribution is denoted by $\mathbb{P}$. The result is an almost direct consequence of the Shannon-McMillan-
Breiman Theorem [26, Theorem 2.1], which states that $\log P_p(e, X_t)$ concentrates around its mean;
see [17]. In particular given $\varepsilon, \delta > 0$ we have for all $t$ sufficiently large
$$\mathbb{P}\left[ \log P_p^t(e, X_t) < -(1 + \delta)h(p)t \right] \leq \varepsilon/2.$$  

Thus if one sets
$$V_\delta(t) := \left\{ g \in G : P_p^t(e, g) \geq e^{-(1+\delta)h(p)t} \right\}$$
we have
$$|V_\delta(t)| \leq e^{(1+\delta)h(p)t} \quad \text{and} \quad \mathbb{P} [X_t \notin V_\delta(t)] \leq \varepsilon/2.$$  

Now given $x \in V_n$ arbitrary, we consider $X_t := \varphi_n(X_t, x)$, which is a random walk with transition
matrix $P_{n,p}$ started at $x$, and let $V_\delta(t) := \left\{ \varphi_n(g, x) : g \in V_\delta(t) \right\}$, the image of $V_\delta(t)$ by the action.
We have, for all $t$ sufficiently large
$$\pi_n(V_\delta(t)) = \frac{|V_\delta(t)|}{n} \leq \frac{|V_\delta(t)|}{n} \leq \frac{e^{(1+\delta)h(p)t}}{n},$$
and $P_{n,p}^t(x, V_\delta(t)) = \mathbb{P}(X_t \in V_\delta(t)) \geq 1 - \varepsilon/2$. Thus we have
$$\|P_{n,p}^t(x, \cdot) - \pi_n\|_{TV} \geq 1 - \varepsilon/2 - e^{(1+\delta)h(p)t}/n.$$  

Considering $t = \lfloor \log(n\varepsilon/2)/(1+\delta)h(p) \rfloor$, we conclude that $T_{n,p}^{\text{mix}}(x, 1 - \varepsilon) \geq 1 - \varepsilon$. We obtain the
desired result. \qed
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 \(x \in V\), \(P_x\) denotes the distribution of this process with initial condition \(X_0 = x\).

A classical way to obtain mixing time upper-bound is via the use of strong stationnary time (see [29, Chapter 6]). A strong stationnary 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 [29, Lemma 6.11] says that if \(T\) is a strong stationnary time for (1) then (the bound is in fact proved for the separation distance which is larger)

\[ \|P^t(x,\cdot) - \pi\|_{TV} \leq \|\nu - \pi\|_{TV} + P_x[T > t]. \]

A careful inspection of the proof in [29] 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 [9] to compare mixing time 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.

**Proposition 7.** Let \((X_t)\) be a finite reversible Markov chain with transition matrix \(P\) with spectral radius \(\rho\) and equilibrium measure \(\pi\). If \(T\) is a stopping time for \(X\) and \(P_x(X_T \in \cdot) = \nu\), then we have for any non-negative integers \(t\) and \(s\):

\[ \|P^{t+s}(x,\cdot) - \pi\|_{TV} \leq \|\nu - \pi\|_{TV} + P_x[T > t] + 3\rho^{2s/3}. \]

**Proof.** The main ingredient of our proof is [9, Corollary 2.4], which we reformulate as follows. Let \(\Omega\) be the state space of the Markov chain. Given a set \(A \subset \Omega\), \(s \geq 0\) and \(\varepsilon > 0\) we set

\[ U(A,s,\varepsilon) := \{ y \in \Omega : |P^t(y, A) - \pi(A)| > \varepsilon, \forall t \geq s \}. \]

Then we have

\[ \pi(U(A,s,\varepsilon)) \leq 2\varepsilon^{-2} \rho^{2s} \]

From the definition of total variation distance we have

\[ \nu(U(A,s,\varepsilon)) \leq 2\varepsilon^{-2} \rho^{2s} + \|\nu - \pi\|_{TV}. \]

For every \(x\), \(t\) and \(s\), using the triangle inequality and the fact that \(X_T \sim \nu\) we obtain that

\[ P^{t+s}(x, A) - \pi(A) \leq \sum_{i=1}^{t} \sum_{y \notin U} P_x(T = i ; X_T = y)(P^{s+t-i}(y, A) - \pi(A)) + P_x[T > t] + \nu(U). \]

From the definition of \(U(A,t,\varepsilon)\) the double sum above is smaller than \(\varepsilon\). Thus, from (26), we obtain (maximizing over \(A\))

\[ \|P^{t+s}(x,\cdot) - \pi\|_{TV} \leq P_x[T > t] + \|\nu - \pi\|_{TV} + 2\varepsilon^{-2} \rho^{2s} + \varepsilon, \]

From the definition of total variation distance we have

\[ \nu(U(A,s,\varepsilon)) \leq 2\varepsilon^{-2} \rho^{2s} + \|\nu - \pi\|_{TV}. \]

For every \(x\), \(t\) and \(s\), using the triangle inequality and the fact that \(X_T \sim \nu\) we obtain that

\[ P^{t+s}(x, A) - \pi(A) \leq \sum_{i=1}^{t} \sum_{y \notin U} P_x(T = i ; X_T = y)(P^{s+t-i}(y, A) - \pi(A)) + P_x[T > t] + \nu(U). \]

From the definition of \(U(A,t,\varepsilon)\) the double sum above is smaller than \(\varepsilon\). Thus, from (26), we obtain (maximizing over \(A\))

\[ \|P^{t+s}(x,\cdot) - \pi\|_{TV} \leq P_x[T > t] + \|\nu - \pi\|_{TV} + 2\varepsilon^{-2} \rho^{2s} + \varepsilon, \]
and one can conclude by choosing \( \varepsilon = \varrho^{2s/3} \).

3. Simple random walks on Ramanujan graphs revisited

3.1. Sketch of proof of Theorem \( \textbf{A} \) and Theorem \( \textbf{B} \). In order to prove Theorem \( \textbf{A} \) and Theorem \( \textbf{B} \) we apply Proposition 7 for a stopping time defined using a coupling between the random walk on \( G_n \) and that on \( T_d \), the infinite \( d \)-regular tree. This coupling is defined thanks to a covering map from \( T_d \) to \( G_n \).

We denote by \( e \) the root vertex of \( T_d \). Let \( X \) be the simple random walk on \( T_d \) starting from \( e \). Given \( x \in V_n \), we fix a local graph homeomorphism \( \varphi : T_d \rightarrow G_n \) (each vertex \( v \) in \( T_d \) 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 \( X_t := \varphi(X_t) \).

For a well chosen integer \( k \geq 1 \), we define the stopping time \( \tau \) as

\[
\tau = \inf\{t \geq 0 : \text{Dist}(X_t, e) = k\},
\]

where \( \text{Dist}(v, o) \) is the distance of the vertex \( v \) in \( T_d \) to the root \( e \). With \( k = \frac{\log n}{\log(d - 1)}(1 + o(1)) \), we show that at the time \( \tau \), \( X_\tau = \varphi(X_\tau) \) is close to equilibrium. More precisely, we use that the distribution of \( X_\tau \) can be expressed as an explicit polynomial of the transition matrix \( P_n \) (cf. Lemma 8), and thus its \( \ell_2 \)-norm can be controlled in terms of the spectral radius of \( P_n \) (cf. Lemma 10). This spectral bound turns out to be optimal.

Then the proof is concluded easily by using Proposition 7 and the fact that the detailed behavior of \( \tau \), 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 9 below (which can also be obtained using methods from [31]). We note also that the variables \( X_\tau \) and \( \tau \) are independent and Proposition 7 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)

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

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

\[
Q_k(x, y) := \frac{\#\text{NB}_k(x, y)}{\#\text{NB}_k(x)} = \frac{\#\text{NB}_k(x, y)}{d(d - 1)^{k-1}}.
\]
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, 37] and [15] for an early reference) will help us to control the spectral radius of $Q_k$ in terms of that of $P$.

**Lemma 8.** For every integer $k$, there exists a polynomial $p_k$ such that $Q_k = p_k(P)$ for every graph $G$. More precisely we have

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

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

$U_{-1} = 0, \quad U_0 = 1, \quad$ and $\quad U_{k+1} = 2XU_k - U_{k-1}$.\]

Proof. It is sufficient to check that the identity $Q_k = p_k(P)$ is valid on the $d$-regular infinite tree $T_d$. We set $\bar{Q}_k := d(d-1)^{k-1}Q_k$ and $\bar{P} := dP$. Using that $\bar{Q}_k(x,y) = 1_{\text{Dist}(x,y) = k}$ it is simple to check that

$$\bar{Q}_{k+1} = dP\bar{Q}_k - \bar{Q}_{k-1}.$$\]

The result follows then by induction on $k$. We find $p_1 = x, \quad p_2 = x^2d/(d-1) - 1/(d-1)$ and, from (28),

$$p_{k+1} = \frac{d}{d-1}xp_k - \frac{1}{d-1}p_{k-1}.$$\]

The conclusion follows. \]

The polynomials $(p_k)$ are called the Geronimus polynomials (in reference to [23]) or the non-backtracking polynomials.

**3.3. Proof of Theorem A.** Recall that $\rho_n$ denotes the spectral radius for $P_n$ restricted to non-constant function and let

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

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

**Proposition 9.** 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 $\delta_n$ tending to zero such that for all $\varepsilon \in (0,1)$ for all $n$ sufficiently large (depending on $\varepsilon$)

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

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

$$s \mapsto P\left[Z \geq \frac{(d-2)^{3/2}}{2\sqrt{d(d-1)}}s\right].$$
In particular if \( \lim_{n \to \infty} \eta_n \log n = 0 \), then

\[
T_{n}^{\text{mix}}(\varepsilon) = \frac{d}{(d - 2) \log(d - 1)} \log n + \Phi(\varepsilon) \sqrt{\log n} + o(\sqrt{\log n}).
\]

Note that the upper-bound in (30) is an immediate consequence of (29), while the lower bound (displayed in [31, Fact 2.1]) which is valid for any \( d \)-regular graph and follows from the argument sketched in the introduction. We note also that it follows from [12] 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 [33] from Proposition 9.

The remainder of this subsection is devoted to the proof of Proposition 9. The proof includes a technical lemma whose proof is postponed to the end of the Section.

**Proof of Proposition 9.** We apply the content of the previous subsection to our problem. Let \( x \) be in \( V_n \) and \( \varphi : T_d \to G_n \) be as in Section 3.1 be a local graph homeomorphism such that \( \varphi(e) = x \), where \( e \) is the root of \( T_d \). Let \( \mathcal{X}_t \) be the simple random walk on \( T_d \) started at the root vertex \( e \). Then \( X_t := \varphi(X_t) \) is a simple random walk on \( G_n \) starting from \( x \). For an integer \( k \) to be chosen later on, let \( \tau \) be defined as in (27). Since non-backtracking paths in a tree are geodesic paths, it is immediate to see that the distribution of \( X_{\tau} \) 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 \( \ell_2 \) upper-bound on total variation distance (2) applied for \( t = 1 \) yields

\[
\|Q_{k,n}(x, \cdot) - \pi_n\|_{TV} \leq r_{k,n} \sqrt{n}.
\]

where, using Lemma 8

\[
r_{k,n} := \max_{\lambda \in \text{Sp}(Q_{k,n}) \setminus \{ 1 \}} |\lambda| = \max_{\lambda \in \text{Sp}(P_n) \setminus \{ 1 \}} p_k(\lambda).
\]

Hence if one sets

\[
k = k_n := \min \left\{ k : r_{k,n} \leq \frac{1}{\sqrt{n \log n}} \right\},
\]

we deduce from (31) that \( \|Q_{k,n,n}(x, \cdot) - \pi_n\|_{TV} \leq (\log n)^{-1} \).

We now apply Proposition 7 for \( T = \tau \). We obtain that, provided that \( \rho_n \leq 1 - \delta \) (which is true for all \( n \) large enough if \( \eta_n \to 0 \) for e.g. \( \delta = 1/20 \)), for all \( t \geq 0 \),

\[
d_n(x, t + s) \leq \|Q_{k,n,n}(x, \cdot) - \pi_n\|_{TV} + \mathbb{P}[\tau \geq t] + 3(1 - \delta)^{2s/3}.
\]

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

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

we obtain

\[
T_n^{\text{mix}}(\varepsilon) \leq t_n(\varepsilon) + s_n.
\]
Now, the central limit theorem for the biased random walk on the line implies that
\[
\frac{\text{Dist}(X_t, o) - t((d - 2)/d)}{2\sqrt{d - 1}/d}
\]
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 10.** For any integer \( d \geq 3 \), there exists a constant \( C \) such that for all \( n \) sufficiently large we have
\[
k_n \leq \begin{cases} 
\frac{\log n}{\log(d-1)} + C (\log \log n), & \text{if } \eta \leq (\log n)^{-2} (\log \log n)^2, \\
\frac{\log n}{\log(d-1)} (1 + C \sqrt{\eta}), & \text{if } \eta \geq (\log n)^{-2} (\log \log n)^2.
\end{cases}
\]

The above estimates combined with the use of the central limit theorem (details are left to the reader) imply that
\[
t_n(\varepsilon) \leq \left( \frac{d}{(d - 2) \log(d - 1)} + C \sqrt{\eta n} \right) \log n + (\Phi(\varepsilon) + \delta_n) \sqrt{\log n}.
\]
This concludes the proof of Proposition 9 (provided that Lemma 10 has been established). \( \square \)

**Proof of Lemma 10.** We use the following classic identities, valid for all \( \theta \in \mathbb{R} \setminus \{0\} \) and \( k \in \mathbb{N} \),
\[
U_k(\cosh \theta) = \frac{\sinh((k + 1)\theta)}{\sinh \theta} \quad \text{and} \quad U_k(\cos \theta) = \frac{\sin((k + 1)\theta)}{\sin \theta}.
\]

We note that \( U_k \) is either an even or an odd function (depending on the parity of \( k \)). We thus have for any \( \lambda \),
\[
|p_k(\lambda)| \leq \frac{1}{d(d - 1)^{k/2}[d - 1]|U_k(\lambda/\rho)| + |U_{k - 2}(\lambda/\rho)|]
\]
Using the fact (it can be checked using (33) that \( |U_k(x)| \leq k + 1 \) on \([0, 1]\) and \( U_k(x) \) is increasing on \([1, \infty)\) we obtain that
\[
\max_{\lambda \in \text{Sp}(P_n) \setminus 1} |p_k(\lambda)| \leq \frac{1}{d(d - 1)^{k/2}[d - 1]U_k(1 + \eta_n) + U_{k - 2}(1 + \eta_n)].
\]
and hence \( r_{k,n} \leq (d - 1)^{-k/2}U_k(1 + \eta_n) \). Using the identity (33) we obtain that there exists a constant \( C \) such that
\[
r_{k,n} \leq \frac{C}{(d - 1)^{k/2}} \min(\eta^{-1/2}, k) e^{\eta^{-1/2}}.
\]
This is sufficient to obtain the desired estimate on \( k_n \). \( \square \)

### 3.4. Proof of Theorem B
Let \( \eta > 0 \). To prove Theorem B use (32) with \( k_n \) replaced by \( k_n' = \frac{\log n}{\log(d-1)}(1 + \eta/2) \). By the law of large number \( \tau = \tau(k) \) is asymptotically equivalent to \( k d / (d - 2) \). Hence, to prove Theorem B it is sufficient to show that, there exists \( c > 0 \) such that for all \( n \) large enough, we have \( \|Q_{k_n',n}(x, \cdot) - \pi_n\|_{\text{TV}} \leq n^{-c\eta} \) for a least \( n^{1 - 2c\eta} \) vertices \( x \) in \( V_n \). It is thus sufficient to show that for all \( n \) sufficiently large,
\[
\sum_{x \in V_n} \|Q_{k_n,n}(x, \cdot) - \pi_n\|_{\text{TV}} \leq n^{1 - 3c\eta}.
\]
To take into account the information we have about the multiplicity of eigenvalues, we must be more precise than (31) in our decomposition. If for \( \lambda \in \text{Sp}(P_n) \setminus \{1\} \) we let \( \alpha_\lambda(x) \) be the square norm of the projection of the coordinate vector at \( x \) on \( E_\lambda^n \), the eigenspace of \( P_n \) corresponding to \( \lambda \), that is

\[
\alpha_\lambda(x) := \max_{f \in E_\lambda^n} \frac{|f(x)|^2}{\sum_{y \in V_n} |f(y)|^2}.
\]

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

\[
(2\|Q_{k,n}(x, \cdot) - \pi_n\|_{TV})^2 \leq \sum_{y \in V_n} \left( Q_{k,n}(x, y) - \frac{1}{n} \right)^2 = \sum_{\lambda \in \text{Sp}(P_n) \setminus \{1\}} n p_k(\lambda)^2 \alpha_\lambda(x).
\]

Hence, averaging over \( x \), we have

\[
\frac{1}{n} \sum_{x \in V_n} \|Q_{k,n}(x, \cdot) - \pi_n\|_{TV}^2 \leq \sum_{\lambda \in \text{Sp}(P_n) \setminus \{1\}} p_k(\lambda)^2 \dim(E_\lambda^n).
\]

Using the fact (recall (34)) that \( p_k(\lambda) \leq (d - 1)^{-k/2}(k + 1) \) when \( \lambda \leq \rho \), we obtain that

\[
\sum_{\lambda \in \text{Sp}(P_n) \cap [-\rho, \rho]} p_k(\lambda)^2 \dim(E_\lambda^n) \leq (d - 1)^{-k}(k + 1)^2 n.
\]

For \( \lambda \not\in [-\rho, \rho] \), as a consequence of (33) we have

\[
\limsup_{k \to \infty} \frac{1}{k} \log |p_k(\lambda)| \leq \lim_{k \to \infty} \frac{1}{k} \log U_k(|\lambda|/\rho) - \frac{1}{2} \log(d - 1)
\]

\[
= \log \left( |\lambda|/\rho - \sqrt{(|\lambda|/\rho)^2 - 1} \right) - \frac{1}{2} \log(d - 1).
\]

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

\[
\limsup_{n \to \infty} \frac{1}{\log n} \log \sum_{\lambda \in \text{Sp}(P_n) \setminus [-\rho, \rho]} p_k(\lambda)^2 \dim(E_\lambda^n)
\]

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

(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 (37) is at most \( c_0 \eta \) with

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

Together with (36), it concludes the proof of (35) with \( c = c_0/4 \).
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 \( \dim_0(H) = n \max_{x \in V_n} \| P_H e_x \|_2^2 \) where \( P_H \) is the orthogonal projection onto \( H \). This definition implies \( \dim_0(H) \geq \dim(H) \), \( \dim_0(\text{span}(\pi_n)) = 1 \) and \( \dim_0(\text{span}(e_x)) = n \). 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) but by replacing in (7) \( \dim(E_{\lambda}^n) \) by \( \dim_0(E_{\lambda}^n) \). The proof of Theorem B actually proves that (6) holds if \( \rho_n < 1 - \delta \) and for all \( u > \rho \), (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 \( p \in \ell^2(\mathcal{G}) \) and we denote by \((X_t)_{t \geq 0}\) the random walk on \( \mathcal{G} \) with transition kernel \( P_p \) started at \( X_0 = e \), the unit of \( \mathcal{G} \). The underlying probability distribution of the process \((X_t)_{t \geq 0}\) on \( \mathcal{G}^N \) will be denoted by \( \mathbb{P} \). Finally, if \( \varphi_n \) is the action of \( \mathcal{G} \) on \( V_n \) as in Theorem 5 given a fixed \( x \in V_n \), we set \( X_t = \varphi_n(X_t, x) \). Then \((X_t)_{t \geq 0}\) is a trajectory of the Markov chain on \( V_n \) with transition kernel \( P_{n,p} \) and initial condition \( x \).

4.2. Proof strategy. Our strategy shares some similarity with that adopted in the Ramanujan case: we try to build a skeleton walk using stopping times which are defined in terms of the walks performed on the covering.

The two important properties that our skeleton walk must satisfies are the following

(i) At each step, one jumps more or less uniformly to one of \( k \) neighbors for a large \( k \).

(ii) The Poincaré constant associated with the skeleton 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 skeleton walk that jumps close to uniformly on large sets, we perform an explicit construction based on the Green’s operator associated to \( P_p \).

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

4.3. Construction of the skeleton walk from the Green’s operator. Given \( k \) a large integer, our task is to find a stopping time \( \tau \) for the process \((X_t)\) starting from \( X_0 = e \) such that \( X_\tau \) is close to be uniformly distributed on a set of \( k \) vertices. We denote by \( A = \{a_1, \ldots, a_d\} \) be the symmetric support of \( P_p \).

We define \( \Gamma \) as the Cayley graph of \( \mathcal{G} \) generated by \( A \). By construction \((X_t)\) is a random walk on \( \Gamma \). We are going to choose our stopping time of the form

\[
\tau := \inf \{ t \geq 0 : X_t \notin U \},
\]
where \( U \) is a finite and contains \( e \). Notably, \( \tau \) is almost surely finite and \( X_\tau \) is supported on the set \( \partial U \) 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 \( R_p = (I - P_p)^{-1} \) be the Green’s operator associated with \( P_p \). We define \( u \) to be the image of the coordinate vector at \( e \) by \( R_p \). This corresponds to the expected number of visit at \( g \) starting from \( e \):
\[
u(g) := R_p(e,g) = \sum_{t=0}^{\infty} P^t_p(e,g).
\]
And given \( k \geq 1 \) we define the set
\[
U := \left\{ g \in G : \nu(g) > \frac{1}{k} \right\}.
\]

Our skeleton 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 \( s \geq 1, \tau_{s+1} := \inf\{ t \geq \tau_k : X_tX_{\tau_{s+1}}^{-1} \notin U \} \). We finally set \( Y_s := X_{\tau_s} \).

We denote by \( Q \) the transition kernel associated with the Markov chain \( (Y_s) \): for any \( g, h \in G \),
\[
Q(g,h) = P(\nu_\tau = hg^{-1}).
\]
By construction, we have
\[
Q = P_q \quad \text{where, for all } g \in G, \quad q_g := Q(e,g) = P[Y_1 = g] = P[X_\tau = g].
\]
Note also that from (21), \( U^{-1} = U \) and the vector \( q = (q_g)_{g \in G} \) satisfies the right-hand side of (21).

We let \( (Y_s) \) be the induced walk by \( (Y_s) \) on \( V_n, Y_s := \phi_n(Y_s, x) \), and let \( Q_n \) denote the associated transition kernel. The following result establishes that \( U \) has the desired property.

**Proposition 11.** Assume that \( G \) is non-amenable and that \( p \) satisfies (21). Then there exists a constant \( C \) such that for every integer \( k \geq 2 \), the set defined by Equation (38) satisfies \( e \in U \), \( \#U \leq Ck \log k \), \( \text{diam}(U) \leq C \log k \) (where \( \text{diam} \) denotes the diameter for the graph distance in \( \Gamma \)) and such that for \( q \) defined by (39),
\[
\forall g \in \partial U, \quad q_g \leq \frac{1}{k}.
\]

**Proof.** By definition of the function \( u \), we have \( e \in U \) and for any \( g \in G \), \( P[X_\tau = g] \leq u(g) \leq 1/k \), as requested. We now check that the cardinality of \( U \) is controlled by \( k \log k \). This is a simple consequence of the fact that \( G \) is non-amenable, which implies that \( \rho_p < 1 \). Notably
\[
P^t_p(g,h) \leq \|P^t_p\| = \rho^t_p,
\]
and thus, if $|g|$ is graph the distance of $g$ and $e$ in $\Gamma$,

$$u(g) = \sum_{t \geq |g|} P^t_p(e, g) \leq (1 - \rho_p)^{-1} \rho_p^{|g|}.$$  

This implies that $U$ is included in the ball $B_r$ of radius $r = C_1 \log k$ around the unit $e$ (we may take for instance $C_1 = 2/|\log \rho_p|$). For any $b \in \mathbb{N}$, we find

$$\frac{\#U}{k} \leq \sum_{g \in U} u(g) \leq \sum_{g \in B_r} \sum_{t=1}^{\infty} P^t_p(e, g) \leq \sum_{t=1}^{br} \sum_{g \in B_r} P^t_p(e, g) + \sum_{t=br+1}^{\infty} \sum_{g \in B_r} P^t_p(e, g) \leq br + \sum_{t=br+1}^{\infty} (\#B_r) \rho_p^t.$$

We choose $b > 0$ such that $(d - 1) \rho_p^b < 1$. 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 $C_2 = 2bC_1$).

### 4.4. Deducing mixing time from RD property and the strong convergence.

To compare the original walk with skeleton walk, the first requirement is to control how much time each skeleton step requires on average. This can be deduced from the definition of the entropy of $G$.

**Lemma 12.** For any $\varepsilon > 0$, there exists $k(\varepsilon) > 1$ such that for all integers $k \geq k(\varepsilon)$,

$$\mathbb{E}[\tau] \leq (1 + \varepsilon) \log k \log(\rho_p).$$

**Proof.** Given $t_1 < t_2 < \infty$ we decompose the expectation in three contribution ($\tau \leq t_1$, $\tau \in (t_1, 2]$, $\tau > t_2$) and obtain

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

We set

$$t_1 := (1 + \varepsilon/2) \log k \rho_p$$

and

$$t_2 := C \log k$$

for some adequate constant $C$, and prove that the second and third term in (41) are smaller than $(\varepsilon/4)(\log k / \rho_p)$. We start by bounding the tail probability of $\tau$. Recalling that $\rho_p$ is the spectral radius of $P_p$,

$$\mathbb{P}(\tau > t) \leq \mathbb{P}(|X_t| \in U) \leq \sum_{g \in U} P^t_p(e, g) \leq \#U \rho_p^t.$$

Hence, for any $s > 0$,

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

By Proposition [I], 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
\[
\mathbb{E} \left[ \tau \mathbf{1}_{(\tau > t_1)} \right] \leq \frac{1}{\log k} \leq \frac{\varepsilon \log k}{4h(p)}.
\]
Now to control the second term, we need to show that \( \mathbb{P}(\tau > t_1) \leq \varepsilon/(4Ch(p)) \). Set
\[
H = \left\{ g \in \mathcal{G} : \mathcal{P}^t_{g}(e,g) \leq e^{-(1-\varepsilon/3)t_1b(p)} \right\},
\]
and arguing as above,
\[
\mathbb{P}(\tau > t_1) \leq \mathbb{P}(X_{t_1} \in U) \leq \sum_{g \in \mathcal{G} \cap H} \mathcal{P}^t_{g}(e,g) + \mathbb{P}(X_{t_1} \notin H).
\]
Now, from (17), if \( k \) is large enough, \( \mathbb{P}(X_{t_1} \notin H) \leq \varepsilon/(8Ch(p)) \) and, by Proposition 11
\[
\sum_{g \in \mathcal{G} \cap H} \mathcal{P}^t_{g}(e,g) \leq \#Ue^{-(1-\varepsilon/3)t_1b(p)} \leq (Ck \log k)k^{-(1+\varepsilon/10)} \leq \varepsilon/(8Ch(p))
\]
The latter is at most \( \varepsilon \) if \( k \) is large enough. \( \square \)

**Remark 4.1.** The above proof actually shows that conclusion of Lemma 12 is true for any exit time from a set of cardinality \( k^{1+o(1)} \). On the other hand (17) and the lower bound \( v(X_\tau) \geq \mathcal{P}^t_{g}(e,X_\tau) \) imply easily that \( \mathbb{E}[\tau] \geq (1 - \varepsilon)(\log k)/b(p) \) for all \( k \) large enough. Hence, our set \( U \) maximizes asymptotically the mean exit time (among all sets of cardinality \( k^{1+o(1)} \)).

All ingredients are now gathered to conclude.

**Proof of Theorem 2.** We fix \( \varepsilon \in (0,1) \), \( \delta > 0 \) and \( x \in V_n \) arbitrary and prove that for \( n \) sufficiently large
\[
T_{\text{mix}}^{\text{mix}}(x,\varepsilon) \leq (1 + \delta) \log n/b(p).
\]
Considering \( \tau \) constructed with \( U \) from Proposition 11 for some large \( k \) which we are going to choose depending on \( \delta \) but not on \( n \), and setting \( m := \lfloor (1 + \delta/4)(\log n)/\log k \rfloor \), we use Proposition 7 for the walk \( X_\tau := \varphi_n(X_\tau, x) \)
\[
T = \tau_m, \quad t = \lfloor (1 + \delta) \log n/b(p) - \log \log n \rfloor \quad \text{and} \quad s = \lfloor \log \log n \rfloor.
\]

As \( \lim_n \rho_{n,p} = \rho_{p} < 1 \), the third term in (25) is smaller than \( \varepsilon/3 \), we are left with controlling the two first terms, that is proving that for \( n \) sufficiently large (recall the definition of \( Q_n \) below Equation (39)).
\[
\mathbb{P}[\tau_m > t] \leq \varepsilon/3 \quad \text{and} \quad \|Q_n^m(x, \cdot) - \pi_n\|_{TV} \leq \varepsilon/3.
\]
For the first inequality, choosing \( k(\delta) \) sufficiently large Lemma 12 guaranties that \( t_n \geq (1 + \delta/2)m\mathbb{E}[\tau] \) and hence the first inequality in (42) is a consequence of the law of large numbers.

The second inequality is obtained using spectral estimates for \( Q_n \). Since \( G \) has the RD property (24), we deduce from Proposition 11 that for some constants \( C, C' \) (depending on \( p \))
\[
\rho_q \leq C(\log k)^C \left( k^{-2}\#\partial U \right)^{1/2} \leq C'k^{-1/2}(\log k)^{C+1/2}.
\]
Now, the assumption that \((\varphi_n)\) converges strongly applied to \(q\) implies that for all \(n\) large enough (depending on \(k\)), the spectral radius \(\rho_{n,q}\) of \(Q_n\) satisfies
\[
\rho_{n,q} \leq 2C^t k^{-1/2} (\log k)^{C+1/2}
\]
Then, we use the usual \(\ell^2\)-distance bound and we obtain that for any \(m \geq 1\),
\[
\|Q^m_n(x, \cdot) - \pi_n\|_{TV} \leq \sqrt{n}\rho_{n,q}^m,
\]
and we can conclude replacing \(m\) by its value. \(\square\)

**Remark 4.2** (Relaxation of the definition of the spectral radius.). We may relax a little bit the assumption of strong convergence. If \(H\) a vector space of \(\mathbb{R}^\nu\) which is invariant by \(P_{n,p}\), we set \(\rho^H_{n,p}\) to be the operator norm of \(P_{n,p}\) on the orthogonal of \(H\). Recall the definition of the flat-dimension in Remark 3.2. Now, we say that the sequence of actions \((\varphi_n)\) converges relatively strongly if for any finitely supported probability vector \(p \in \ell^2(G)\) such that (21) holds, we have \(\limsup \rho_{n,p} < 1\) and \(\lim_n \rho^H_{n,p} = \rho_p\) 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 3 also holds under this weaker assumption. Indeed, we simply replace the bound (13) by the bound valid for any invariant vector space \(H\) of \(Q_n\) which contains \(\pi_n\):
\[
\|Q^m_n(x, \cdot) - \pi_n\|_{TV} \leq \sqrt{n}\|Q^m_n(x, \cdot) - \pi_n\|_2 \leq \sqrt{n}\rho_{n,q}^m (\sqrt{\dim_0(H)} + 1)/\sqrt{n} + \sqrt{n} (\rho^H_{n,q})^m,
\]
which follows directly from the spectral theorem and the observation that, if \(P_H\) is the orthogonal projection onto a vector space \(H\), then,
\[
\|P_H f\|_2 \leq \sum_x |f(x)||P_H e_x|_2 \leq \|f\|_1 \sqrt{\dim_0(H)/n}.
\]
Finally, we notice that if \(\dim_0(H) = n^{o(1)}\) and \(\limsup \rho_{n,q} < 1\) then the first term on the right-hand side of (44) 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 \(G\) as the group obtained by \(k\) free copies of \(Z\) and \(l\) free copies of \(Z/2Z\) where \(k + l\) is the number of equivalence classes of the involution, as detailed below Definition 2. We denote by \(A = \{g_1, \ldots, g_d\}\) its natural set of generators. The probability vector \(p = (p_1, \ldots, p_d)\) as in Theorem 1 is identified with a vector in \(\ell^2(G)\) 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 \((X_t)_{t \geq 0}\) the random walk on \(G\) with transition kernel \(P_p\) started at \(X_0 = e\), the unit of \(G\). The underlying probability distribution of the process \((X_t)_{t \geq 0}\) on \(G^\mathbb{N}\) will be denoted by \(\mathbb{P}(\cdot)\). Finally, we define \(\varphi_n\) as the action of \(G\) on \(V_n\) such that for all \(i \in [d]\), \(S_{g_i} = S_i\) where \(S_i\) is as in (12) and \(S_g\) is the permutation matrix associated to \(\varphi_n(g, \cdot)\). Finally, given \(x \in V_n\) we set \(X_t = \varphi_n(X_t, x)\), that is \((X_t)_{t \geq 0}\) is a trajectory of the Markov chain on \(V_n\) with transition kernel \(P_{n,p}\) with initial condition \(x\).
5.2. **Organization of the proof.** Our idea is to use the same stopping time strategy that for the previous section. The important difference is that instead of using the RD property to conclude, we want to show that the generator of the skeleton random walk can be reasonably approximated by a polynomial in \( P_{n, p'} \), the generator of the random walk with anisotropy given by \( p' \). Our first job is thus to identify the value of \( p' \) which is possible. This is performed by studying the resolvent of the process which is done in Section 5.3. An explicit expression for \( p' \) is given Proposition 17.

Then in Section 4.4 we use our result concerning resolvents to obtain a relevant bound on the kernel of the skeleton walk (Proposition 18). Combining this with a few \( \ell_2 \) computation (Lemma 19), this allows to prove Theorem 1 adapting the approach used for Theorem 5.

5.3. **The relation between \( p \) and \( p' \) via resolvent.** The resolvent of \( \mathcal{P}_p \) is defined for \( z \notin \sigma(\mathcal{P}_p) \)

\[
R_p^z = (zI - \mathcal{P}_p)^{-1}.
\]

We also consider \( R_{\mathcal{P}}^p \) (that is \( z = \rho_p \)) that can be defined by continuity (it is an unbounded operator). In the above expression, \( I \) is the identity operator in \( \mathcal{G} \).

As our group is non-amenable, the vector \( (R_p^1(e, x))_{x \in \mathcal{G}} \) is very close to be integrable (it belongs to \( \ell^{1+}(\mathcal{G}) := \cap_{\epsilon>0} \ell^{1+\epsilon}(\mathcal{G}) \) but not to \( \ell^1(\mathcal{G}) \)), while \( (R_{\mathcal{P}}^p(e, x))_{x \in \mathcal{G}} \) 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 13.** *Given \( p \) a probability vector on \([d]\), there exists a unique probability vector \( p' \) and a real \( C = C(p) \) such that for all \( x, y \in \mathcal{G} \),

\[
R_p^1(x, y) = C \left( R_{p'}^{q_i}(x, y) \right)^2.
\]

As a consequence of tree structure the Cayley graph associated with \((\mathcal{G}, A)\), that can be identified with the regular tree \( \mathcal{T}_d \), the resolvent function admits a simple “multiplicative” expression (this is a well established result that can be found e.g. in [22] or [20]): \( R_p^z(e, x) \) can be obtained by multiplying \( R_p^z(e, e) \) by a quantity \( q_i^z(p) \) for each edge of type \( i \) which is crossed on the minimal path linking \( e \) to \( x \). Hence to prove Proposition 13, we need to find a probability vector \( p' \) such that for all \( i \in [d] \), \( q_i^z(p) = \left( q_i^z(p') \right)^2 \).

We need some extra-notation to give an expression for the coefficients \( q_i \). Let us denote by \( R_{\mathcal{P}}^z, i = (zI - \mathcal{P}_{p,i})^{-1} \) the resolvent of the operator \( \mathcal{P}_{p,i} = \mathcal{P}_p - p_i(\delta_e \otimes \delta_{g_i} + p\delta_{g_i} \otimes \delta_e) \) obtained from \( \mathcal{P}_p \) by removing the transitions between \( e \) and \( g_i \). Finally we let \( \gamma_i^z \) be the diagonal coefficient of \( R_{\mathcal{P}}^z, i \):

\[
\gamma_i^z = \gamma_i^z(p) := R_{\mathcal{P}}^z(e, e).
\]

**Lemma 14** ([see Lemma 3 in [22] / Proposition 3.4 in [20]]). *For any reduced word \( x = g_{i_1} \ldots g_{i_n} \in \mathcal{G} \) written in reduced form (that is \( g_{i_{k+1}} \neq g_{i_k}^z \) for all \( k \)) and \( z \notin \sigma(\mathcal{P}_p) \), we have

\[
R_p^z(e, x) = R_p^z(e, e) \prod_{t=1}^n p_t \gamma_{i_t}^z.
\]
Moreover,

\[
R_{p}^{z}(e,e) = \left( z - \sum_{j \in [d]} p_{j}^{2} \gamma_{j}^{*} \right)^{-1} \quad \text{and} \quad \gamma_{i}^{z} = \left( z - \sum_{j \neq i} p_{j}^{2} \gamma_{j}^{*} \right)^{-1}.
\]

As we shall check, \( R_{p}^{z}(e,e) \) and \( \gamma_{i}^{z} \) have a proper limit at \( z = \rho_{p} \). The above lemma allows to compute explicitly the resolvent operator.

Lemma 15. Let \( z \in [\rho_{p}, \infty) \). If \( s = 1/(2R_{p}^{z}(e,e)) \) and \( q_{i} = p_{i} \gamma_{i}^{z} \), we have

\[
q_{i} = \frac{\sqrt{s^{2} + p_{i}^{2}} - s}{p_{i}} \quad \text{and} \quad p_{i} = \frac{2sq_{i}}{1 - q_{i}}.
\]

Moreover, \( s = 1/(2R_{p}^{z}(e,e)) \) is the largest solution of the equation for \( s \in (0, \infty) \):

\[
z = 2s + \sum_{j=1}^{d} \left( \sqrt{s^{2} + p_{j}^{2}} - s \right).
\]

Proof. From (15), we have \( 2 = s - \sum_{j} p_{j} q_{j} \) and \( p_{i} q_{i}^{2} + 2s q_{i} - p_{i} = 0 \). The formulas follows immediately. It remains to prove that \( s \) is the largest solution of \( f(x) = z \) with \( f(x) = 2x + \sum_{j} \left( \sqrt{x^{2} + p_{j}^{2}} - x \right) \). Since \( f \) strictly convex, the equation \( f(s) = z \) has zero, one or two solutions. We have \( f(0) = 1, f'(0) = 2 - d < 0 \) and \( f \) diverges at infinity. Hence, \( f(s) = z \) has one solution for \( z \geq 1 \), and, from (15), two for \( z \in (\rho_{p}, 1) \), say \( s_{-}(z) < s_{+}(z) \) and one at \( z = \rho_{p} \). At the branching point, \( z = 1, s_{-}(1) = 0 \) and \( s_{+}(1) \geq s(\rho_{p}) > 0 \). Since \( z \mapsto R_{p}^{z}(e,e) \) is continuous in \( z \geq \rho_{p} \), we deduce the claimed statement.

The above lemma has some implication for the value of \( q_{i} = p_{i} \gamma_{i}^{z} \) for \( z \in \{\rho_{p}, 1\} \).

Lemma 16. Let \( z \in [\rho_{p}, \infty) \) and \( q_{i} = p_{i} \gamma_{i}^{z} \). We have

\[
\sum_{i=1}^{d} \frac{q_{i}}{1 + q_{i}} = 1 \iff z = 1 \quad \text{and} \quad \sum_{i=1}^{d} \frac{q_{i}^{2}}{1 + q_{i}^{2}} = 1 \iff z = \rho_{p}.
\]

Proof. Let \( s = 1/(2R_{p}^{z}(e,e)) \). By Lemma (15) since \( 2s = z - \sum_{j} p_{j} q_{j} \) and \( p_{i}/(2s) = q_{i}/(1 - q_{i}^{2}) \), we have

\[
\sum_{i} \frac{q_{i}}{1 + q_{i}} = \sum_{i} \frac{p_{i}(1 - q_{i})}{2s} = \frac{1 - (z - 2s)}{2s}.
\]

We get the first claim.

For the second claim, by Lemma (15) we have

\[
\frac{1}{q_{i}} = \frac{p_{i}}{\sqrt{s^{2} + p_{i}^{2}} - s} = \frac{\sqrt{s^{2} + p_{i}^{2}} + s}{p_{i}}.
\]

It follows that

\[
\frac{q_{i}^{2}}{1 + q_{i}^{2}} = \frac{1}{q_{i}^{2} + 1} = \frac{p_{i}^{2}}{p_{i}^{2} + \left( \sqrt{s^{2} + p_{i}^{2}} + s \right)^{2}} = \frac{p_{i}^{2}}{2\sqrt{s^{2} + p_{i}^{2}} \left( \sqrt{s^{2} + p_{i}^{2}} + s \right)}.
\]
We apply again (46) and we sum over \(i\). We find
\[
\sum_i \frac{q_i^2}{1 + q_i^2} = \frac{1}{2} \sum_i \left(1 - \frac{s}{\sqrt{s^2 + p_i^2}}\right).
\]
The right-hand side is equal to 1 if and only if
\[
2 + \sum_i \left(\frac{s}{\sqrt{s^2 + p_i^2}} - 1\right) = 0.
\]
or equivalently if \(s > 0\) is a local extremum of the function \(f(x) = 2x + \sum_i \left(\sqrt{x^2 + p_i^2} - x\right)\). This function \(f\) is strictly convex and diverges at infinity, so it has a unique local minimum, say \(s^*\). Recall that, by Lemma 15, we have \(z = f(s)\). Finally, by (15), \(s = s^*\) is equivalent to \(z = \rho_p\). It gives the second claim.

Now we are ready to identify the value of \(p'\) which is such that (13) holds. We set, for \(i \in [d]\), \(q_i^p(p) = p_i\gamma_i^p(p)\) and we introduce the vectors \(a\) and \(b\) whose coordinates are given for all \(i \in [d]\) by
\[
a_i(p) := q_i^1(p) \quad \text{and} \quad b_i(p) := q_i^\rho_p(p).
\]

The formulas for the coordinates \(a_i\) and \(b_i\) of \(a\) and \(b\) are given in Lemma 15, notably we have \(a_i, b_i \in [0, 1]\). We can now reformulate and prove Proposition 13.

**Proposition 17.** For any probability vector \(p\) on \([d]\), there exists a unique probability vector \(p'\) on \([d]\) such that for all \(i \in [d]\), we have
\[
a_i(p) = \left(b_i(p')\right)^2.
\]
It is given by the formula, for all \(i \in [d]\),
\[
p_i' = \frac{\sqrt{a_i(p)}}{1 - a_i(p)} \left(\sum_{j \in [d]} \frac{\sqrt{a_j(p)}}{1 - a_j(p)}\right)^{-1}.
\]

**Proof of Proposition 17.** For ease of notation, we set \(q_i' = \sqrt{a_i(p)}\). Assume that \(p'\) is a probability vector such that \(b_i(p') = q_i'\) for all \(i \in [d]\). By Lemma 15, we necessarily have \(p_i' = 2sq_i/(1 - q_i^2)\) where \(2s = 1/R^\rho_p(e,e)\). It implies the claimed uniqueness. We now prove existence. We set \(p_i' = 2sq_i/(1 - q_i^2)\) where \(s > 0\) is the normalization constant such that \(p'\) is a probability vector. We have \(q_i' = (\sqrt{s^2 + p_i^2} - s)/p_i\). We set \(z = f(s)\) with \(f(x) = 2x + \sum_j \left(\sqrt{x^2 + p_j^2} - x\right)\). Since \(q_i = \sqrt{a_i(p)}\), by the first statement of Lemma 16
\[
\sum_i \frac{q_i^2}{1 + q_i^2} = 1.
\]
We note that (17) is valid for any \(s, p_i's, q_i's\) such that for all \(i\), \(q_i = (\sqrt{s^2 + p_i^2} - s)/p_i\). Thus, the argument below (17) implies that \(s\) is the unique global minimum of \(f\). Hence, by (15), we have...
Finally, we note that the equation \( \rho_{p'} = f(s) \) has a unique solution in \( s \), It follows by Lemma 15 that \( 2s = 1/R_{p'}^\rho (e,e) \) and thus \( q'_i = \rho_{p'}^\rho (p') \). It concludes the proof. \( \square \)

5.4. Deducing mixing time from a bounding kernel. Our aim now is to work with the same stopping time and skeleton walk of Section 4.3 and use the information we have to approximate the transition matrix of the skeleton walk \( Q_n = P_{n,q} \) (defined below (39)), with a power series of \( P_{n,p'} \) the transition matrix of the nearest neighbor random walk associated with \( p' \) of Proposition 17. We further define \( Q'_n \) to be the following truncated series (which approximates a multiple of the resolvent of \( P_{n,p'} \) at \( z = \rho_{p'} \))

\[
Q'_n := \frac{1}{\sqrt{k}} \sum_{t=0}^{[\log k]^4} \left( \frac{P_{n,p'} \rho_{p'}}{\rho_{p'}} \right)^t.
\]

**Proposition 18.** Given \( p \) a probability vector on \( [d] \), there exists a real \( C = C(p) \) such that for \( p' \) given by Proposition 17, we have, for all \( x,y \in V_n \)

\[
Q_n(x,y) \leq CQ'_n(x,y)
\]

We postpone the proof of this proposition to Section 5.5 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 be we can obtain estimate for contraction of function with large variance. Given a matrix \( A \) of size \( n \times n \) and \( 1 \leq u \leq \sqrt{n} \), we define

\[
\kappa_u(A) := \sqrt{\max_{f \neq 0} \frac{\langle Af, Af \rangle}{\langle f, f \rangle} \frac{\|f\|_2^2}{\|f\|_1}}
\]

Note that \( \kappa_1(A) \) is the operator norm of \( A \) and \( \kappa_{\sqrt{n}}(A) \) is the maximal diagonal entry of \( M \) in absolute value. For general \( u \), the scalar \( \kappa_u(A) \) can be thought as a kind of pseudo-norm of \( A \) restricted to vectors which are localized in terms of their \( \ell^2 \) over \( \ell^1 \) ratio.

**Lemma 19.** Let \( A, B \) be two \( n \times n \) matrices such that \( B \) is a bistochastic matrix and, for some real \( c \geq 0 \), for all \( x,y \in [n] \), \( |A(x,y)| \leq cB(x,y) \) then for all \( 1 \leq u \leq \sqrt{n} \),

\[
\kappa_u(A) \leq c\rho(B) + \frac{c}{u},
\]

where \( \rho(B) \) is the second largest singular value of \( B \).

From Proposition 18 we may apply Lemma 19 when \( A = Q_n \) and \( B = \alpha Q'_n \), with \( \alpha = k^{-1/2} \sum_{t=0}^{[\log k]^4} \rho_{p'}^{-t} \) and \( c = C\alpha^{-1} \) for the constant \( C \) given by Proposition 18. In this case we have

\[
c\rho(B) = C \frac{[\log k]^4}{\sqrt{k}} \sum_{t=0}^{25} \left( \frac{\rho_{n,p'}}{\rho_{p'}} \right)^t
\]
and for some adequate choice of $C'$

\[
(51) \quad c = \frac{C}{\sqrt{k}} \sum_{t=0}^{|\log k|^4} \rho_p^{-t} \leq e^{C'(|\log k|^4)}.
\]

To ensure that the second term in (19) is small we set $u = u_k := e^{(|\log k|^5)}$. Then, from Assumption (19) and Lemma 19 for any fixed $k$, for all $n \geq n_0(k)$ sufficiently large, we obtain

\[
(52) \quad \kappa_u(Q_n) \leq \frac{(|\log k|^5)}{\sqrt{k}}.
\]

Now we want to use this estimate to build a stopping time for the original walk $P_n,p$. The idea is first to iterate $Q_n$ several times in order to contract the $\ell^2$ norm below the threshold $u$ and then use the original transition matrix $P_n,p$ to finish the job. For this purpose, for a fixed integer $k \geq 3$ which will be defined later on, and $n \geq 3$, we set

\[
a_n := \left\lfloor \frac{\log n}{\log k - 4 \log \log k} \right\rfloor \quad \text{and} \quad b_n := \lfloor \log \log n \rfloor.
\]

We define $T := b_n + \tau_{a_n}$. Our spectral estimates (52) implies that $X_T$ is close to equilibrium:

**Lemma 20.** For any fixed integer $k \geq 3$, let $a_n, b_n$. If Assumption (19) holds, then we have

\[
\lim_{n \to \infty} \max_{x \in V_n} \|P_x[X_{b_n + \tau_{a_n}} \in \cdot] - \pi_n\|_{TV} = 0.
\]

To show that

\[
\max_{x \in V_n} T_{n,p}^{\text{mix}}(x, \varepsilon) \leq (1 + \delta)(\log n)/\mathbb{h}(p)
\]

for $n$ sufficiently large, we use Proposition 7 with $T = b_n + \tau_{a_n}$, $t_n := [(1 + \delta/2)(\log n)/\mathbb{h}(p)]$ and $s_n := [(\delta/2)(\log n)/\mathbb{h}(p)]$.

With this setup, the first term in (25) tend to zero according to Lemma 20 and the third one $(\rho_{p,n}^{2s_n/3})$ tends to zero because $\limsup_{n \to \infty} \rho_{n,p} < 1$ any $p$ satisfying (11) (this is a consequence of (19) and Cheegars inequality see [29, Theorem 13.4]). We can choose 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 12 for any $\delta > 0$, we may choose an integer $k$ sufficiently large such that

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

This concludes the proof of Theorem 1.\[ \square \]

**Proof of Lemma 19.** The statement is an immediate consequence the following functional inequality valid for every $f$

\[
(53) \quad \sqrt{\langle Af, Af \rangle} \leq c \rho(B)\|f\|_2 + \frac{c}{\sqrt{n}}\|f\|_1.
\]
Since $B$ is bistochastic, the constant functions are left invariant by $B$ and its transpose. It follows that $\rho(B)$ 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

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

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

$$\langle B|f|, B|f| \rangle = \langle |f|_1^2/n + B\bar{f}, B\bar{f} \rangle \leq \|f\|^2_1/n + \rho(B)^2 \|f\|^2_2. \tag{54}$$

We deduce (53) using the triangle inequality, $\sqrt{a^2 + b^2} \leq |a| + |b|$. □

**Proof of Lemma 20.** The distribution of $X_{b_n + \tau_n}$ can be written as a product of $P_{n, p}$ and $Q_n$ (defined as the transition Kernel of :

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

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

$$\|Q_n^a(x, \cdot) - \pi_n\|_2 \leq \frac{2u}{\sqrt{n}}. \tag{55}$$

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 (52) and our choice for $a_n$ imply (55). To conclude the proof, we use the usual $\ell^2$ bound and combine it with (55), this gives

$$\|P_{n, p}^b Q_{n, p}^a(x, \cdot) - \pi_n\|_{TV} \leq \frac{\sqrt{n}}{2} \|P_{n, p}^b Q_{n, p}^a(x, \cdot) - \pi_n\|_2 \leq \frac{\sqrt{n}}{2} \rho_{n, p}^b \|Q_{n, p}^a(x, \cdot) - \pi_n\|_2 \leq \rho_{n, p}^b u. \tag{56}$$

Finally, we conclude using that $b_n$ tends to infinity and that $\limsup_n \rho_{n, p} < 1$ (which follows from assumption (19) by [29, Theorem 13.4]). □

**Remark 5.1** (Relaxation of our assumption concerning the spectral radius). As in in Remark 3.1, we denote by $\text{dim}_0(H)$ the flat-dimension of a vector space $H$ of $\mathbb{R}^V_n$ and we set $\rho_{n, p}^H$ to be the operator norm of $P_{n, p}$ on the orthogonal of $H$. We may modify Theorem 4 as follows: if $(H_n)$ is a sequence of invariant vector spaces of $P_{n, p'}$ such that $\lim_n \rho_{n, p'}^{H_n} = \rho_{p'}$ and $\text{dim}_0(H_n) \leq n^o(1)$ (that is $\lim \log \text{dim}_0(H_n)/\log n = 0$) then the conclusion of Theorem 4 holds.

Indeed, in Lemma 19, if $H$ is an invariant subspace of the bistochastic matrix $B$ and its transpose, then (49) can be improved in $\kappa_u(A) \leq c \rho_H(B) + c \sqrt{\text{dim}_0(H)/n}$, where $\rho_H(B)$ is the operator norm of $B$ on the orthogonal of $H$. Recall that if $P_H$ is the orthogonal projection onto $H$, then $\|P_H g\|_2 \leq \|g\|_1 \sqrt{\text{dim}_0(H)/n}$. Setting $g = |f| - P_H |f|$, we may thus replace the bound (54) by $\langle B|f|, B|f| \rangle \leq \|f\|^2_1 \text{dim}_0(H)/n + \langle B \bar{f}, B \bar{f} \rangle \leq \|f\|^2_1 \text{dim}_0(H)/n + \rho_H(B) \|f\|^2_2$. It gives the claimed
improvement of \([49]\). 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 \(T_n^{\text{mix}}(\varepsilon)\) can be obtained by choosing \(k_n\) tending to infinity, and using a more quantitative version of Proposition \([6]\) for anisotropic walks on trees. Optimizing all choices of parameters in our proof, we obtain a result of the form

\[
T_n^{\text{mix}}(\varepsilon) \leq \frac{\log n}{b(p)} + C(\log n)^{2/3}.
\]

We have thus not obtained in this section the anisotropic counter-part of Equations \([29], [30]\), which allow to describe more accurately the profile of relaxation to equilibrium providing some quantitative information about the convergence \([19]\) is given.

### 5.5. Proof of Proposition \([18]\)

The matrices \(Q_n\) and \(P_{n,p}\) both being defined as the kernel corresponding to projections of Markov chains on the free group \(G\) on \(V_n\), it is sufficient to prove the inequality for the corresponding kernels \(Q\) (as in \([30]\)) and \(P_{p'}\) on \(G\), that is

\[
\forall x \in G, \quad Q(e, x) \leq \frac{C}{\sqrt{k}} \sum_{t=0}^{\left\lfloor \log k\right\rfloor} \left(\frac{p_{p'}}{\rho_{p'}}\right)^t (e, x).
\]

Since \(Q(e, x) = 0\) for all \(x \notin \partial U\), it is sufficient to check \((57)\) for \(x \in \partial U\). By Lemma \([14]\) if \(z \geq \rho_p\) and \(x = g_iy\) for some \(g_i \in A\), then \(R^z(e, x) \geq cR^z(e, y)\) for some positive \(c = c(p, z)\). Since \(R^1(e, y) \geq 1/k\) for all \(y \in U\), we find for all \(x \in \partial U\),

\[
Q(e, x) \leq \frac{1}{k} \leq \frac{C}{\sqrt{k}} \sqrt{R^1_p(e, x)},
\]

with \(C = 1/\sqrt{c}\). Thus, from Proposition \([13]\) for some new constant \(C = C(p)\), for all \(x \in \partial U\),

\[
Q(e, x) \leq \frac{C}{\sqrt{k}} R^{\rho_{p'}}(e, x).
\]

To deduce \((57)\) from this last bound, it remains to develop in series the resolvent. It requires some care because \(z = 1/\rho_{p'}\) is precisely the radius convergence of the power series \(\sum_{t \geq 0} (zP_{p'})^t\).

With the notation of Lemma \([14]\) for any \(p\) and \(i \in [d]\), the function \(z \mapsto \gamma_i^z\) is decreasing in \(z \geq \rho_p\). Moreover, by Lemma \([15]\) we have \(|\gamma_i^{\rho_p} - \gamma_i^z| \leq C\sqrt{z - \rho_p}\) for some new \(C = C(p)\). By Lemma \([14]\) it follows that for some new \(C = C(p)\) for all \(x \in G\),

\[
|\mathcal{R}_{p}^{\rho_p}(e, x) - \mathcal{R}_{p}^{z}(e, x)| \leq C(|x| + 1) \sqrt{z - \rho_p}\mathcal{R}_{p}^{\rho_p}(e, x),
\]

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

\[
\prod_{i=1}^{k} a_i - \prod_{i=1}^{k} b_i = \sum_{j=1}^{k-1} \left(\prod_{i=1}^{j-1} a_i\right) (b_j - a_j) \left(\prod_{i=j+1}^{k} b_i\right).
\]
By Lemma 11 the diameter of \( \partial U \) being at most \( C \log k \), we find that for all \( x \in \partial U \), \( R_p^z(e, x) \leq 2R_p^z(e, x) \) provided that \( 0 \leq z - \rho_p \leq c(\log k)^{-2} \) for some positive constant \( c = c(p) > 0 \). We now fix \( z = \rho_{p'} + c(p')(\log k)^{-2} \). From what precedes, for all \( x \in \partial U \),
\[
R_{p'}^z(e, x) \leq 2R_p^z(e, x) = \frac{2}{z} \sum_{t=0}^{\infty} \left( \frac{P_{p'}}{z} \right)^t (e, x).
\]
Since \( \|P_{p'}\| = \rho_{p'} \), we have \( P_{p'}(e, x) \leq \rho_{p'}^t \) and, for some new constant \( C = C(p') \), for any \( s \geq 0 \),
\[
\sum_{t=s}^{\infty} \left( \frac{P_{p'}}{z} \right)^t (e, x) \leq C(\log k)^2 e^{-s/(C(\log k)^2)}.
\]

We now recall that by Proposition 13 for all \( x \in \partial U \), \( R_p^z(e, x) \geq R_{p'}^z(e, x)/2 \geq c/\sqrt{k} \). It follows that if \( s = [C' \log k]^3 \) for some large enough constant \( C' \), we have
\[
\frac{1}{z} \sum_{t=s}^{\infty} \left( \frac{P_{p'}}{z} \right)^t (e, x) \leq \frac{1}{2} R_{p'}^z(e, x).
\]
Consequently, for this value of \( s \),
\[
R_{p'}^z(e, x) \leq \frac{2}{z} \sum_{t=0}^{s} \left( \frac{P_{p'}}{z} \right)^t (e, x) \leq \frac{2}{\rho_{p'}} \sum_{t=0}^{s} \left( \frac{P_{p'}}{\rho_{p'}} \right)^t (e, x).
\]
This concludes the proof of (57).

\[\Box\]

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 \( G \), we have a sequence of finite sets \( (V_n) \) with \( \#V_n = n \) and \( (\varphi_n) \) a sequence of actions of \( G \) on \( V_n \). Let \( r \geq 1 \) be an integer. We think \( |r| = \{1, \ldots, r\} \) as a set of colors. An element \( p \in M_r(\mathbb{R})^G \) is written as a matrix-valued vector \( p = (p_g)_{g \in G} \) with \( p_g \in M_r(\mathbb{R}) \). The support of \( p \) is then the subset of \( G \) such that \( p_g \) is not the null matrix. We consider \( p \in M_r(\mathbb{R})^G \) with finite support such that
\[
P_{1,p} := \sum_{g \in G} p_g
\]
is an irreducible Markovian matrix on \( [r] \) with invariant probability measure \( \pi \) and which satisfies the matrix-valued analog of (21):
\[
(58) \quad \text{support of } p \text{ generates } G \quad \text{and for all } (u, v) \in [r]^2, g \in G: \pi(u)p_g(u, v) = \pi(v)p_{g^{-1}}(v, u).
\]
In particular, \( P_{1,p} \) is reversible with respect to \( \pi \). Then, we denote by \( P_p \) the operator on \( \ell^2(G \times [r]) \) defined by
\[
(59) \quad P_p = \sum_{g \in G} p_g \otimes \lambda(g),
\]
where $\lambda(g)$ is as in [23] and $\otimes$ is the tensor product. In probabilistic terms, $P_P$ is the transition kernel of a random walk $(X_t)$ on $G \times [r]$ where the probability to jump from $(g, u)$ to $(h, v)$ is $p_{hg^{-1}}(u, v)$. We denote by $h(p)$ the entropy rate of $P_P$ defined by: for any $u_0 \in [r]$,

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

The fact that $h(p)$ does not depend on $u_0$ is an easy consequence of the assumption that $P_{1, p}$ is irreducible. Again, if $G$ is non-amenable and (58) holds, then $h(p) > 0$. Besides, the proof of Shannon-McMillan-Breiman Theorem in [26, Theorem 2.1], actually proves that if $X_0 = (e, u_0)$, a.s.

$$h(p) = \lim_{t \to \infty} -\frac{\log P^t_P((e, u_0), X_t)}{t}.$$  

With $(S_g)_{g \in G}$ as in [22], we define the Markovian matrix on $[\mathbb{R}^{V_n \times [r]}$

$$P_{n, p} = \sum_{g \in G} p_g \otimes S_g.$$

This matrix is the transition kernel a Markov chain on $V_n \times [r]$ covered by $(X_t)$ in the sense that if we define for $(g, u) \in G \times [r]$ and $x \in V_n$, $\varphi_n((g, u), x) := (\varphi_n(g, x), u)$ then $X_t := \varphi_n(X_t, x)$ is a Markov chain with transition matrix $P_{n, p}$ started at $(x, u_0)$. The measure $\pi_n(x, u) = \pi(u)/n$ is an invariant probability of $P_{n, p}$ and from (58). $P_{n, p}$ is reversible with respect to $\pi_n$. Moreover, since (60) holds, the proof of Proposition 5 actually implies that mixing time of $X_t, T_{n, p}^{\text{mix}}(x, \varepsilon)$, satisfies for any fixed $\varepsilon \in (0, 1)$ and uniformly in $x \in V_n$, the lower bound (18).

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 $G_{\text{free}}$ with $d/2$ generators and their $d/2$ inverses $(g_1, \ldots, g_d)$, that is $g_i^{-1} = g_i^*$ for some involution. 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 $G_{\text{free}}$ 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 $E_{k, l} \in M_r(\mathbb{R})$ is the matrix defined by $E_{k, l}(i, j) = 1_{(k, l) = (i, j)}$, 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 $\psi$ 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 $p_{g_i} = E_{u_i, v_i}/d_u$, then $P_{1, p}$ and $P_{n, p}$ are the transition matrices of the simple random walks on $G_1$ and $G_n$ respectively. In this case the condition (58) is fulfilled with $\pi(u) = d_u/d$.

We are ready to state the analog of Theorem 5.

**Theorem 21.** Let $G$ be a finitely generated non-amenable group with the property RD, $(V_n)$ a sequence of finite sets with $\#V_n = n$ and $(\varphi_n)$ a sequence of actions of $G$ on $V_n$ which converges strongly. For any integer $r \geq 1$ and any finitely supported $p \in M_r(\mathbb{R})^G$ such that $P_{1, p}$ is an
irreducible and aperiodic Markov chain and such that (58) holds, the mixing time of the random walk with transition matrix \( P_{n,p} \) satisfies, for every \( \varepsilon \in (0,1) \),

\[
\lim_{n \to \infty} \frac{T_{n,p}^{\text{mix}}(\varepsilon)}{\log n} = \frac{1}{h(p)}.
\]

Note that in the above statement the RD property and the strong convergence property are defined in terms of scalar valued vectors \( p \in \ell^2(G) \). From [14], an example of application of Theorem 21 is the simple random walk on a random \( n \)-lift of a connected non-bipartite base graph.

6.2. Proof of Theorem 21. We let \( (X_t) \) be the random walk with kernel \( P_p \) started from \( X_0 = (e,u_0) \). For \((g, u) \in G \times [r] \) and \( x \in V_n \), we set \( \bar{\varphi}_n((g, u), x) = (\varphi_n(g, x), u) \) and let \( X_t := \bar{\varphi}_n(X_t, x) \) be a Markov chain with transition matrix \( P_{n,p} \) 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 \( q = (g) \in M_r(\mathbb{R})^G \) with finite support such that the right-hand side of (58) holds. We define \( \ell^2(\pi) \) as the Hilbert space on \( \mathbb{R}^r \) endowed with the scalar product \( (f, g)_{\pi} = \sum_i \pi(i) f(i) g(i) \). Similarly, \( \ell^2_0(\pi) \) and \( \ell^2_\emptyset(\pi) \) are the Hilbert spaces on the vector spaces \( \mathbb{R}^{V_n \times [r]} \) and \( \mathbb{R}^{G \times [r]} \) endowed with the scalar products:

\[
(f, g)_{\pi} = \sum_{(x,i) \in X \times [r]} \pi(i) f(x, i) g(x, i),
\]

with \( X = V_n \) and \( X = G \) respectively. Then, the right-hand side of (58) asserts that \( P_{n,q} \) and \( P_q \) are self-adjoint operators on \( \ell^2_0(\pi) \) and \( \ell^2_\emptyset(\pi) \) respectively. We note that the subspace of \( \mathbb{R}^{V_n \times [r]} \) \( H_r = \mathbb{R}^r \otimes 1 \) of vectors \( f \) of the form for some \( g \in \mathbb{R}^r \), \( f(x, i) = g(i) \) is an invariant subspace of \( P_{n,q} \) of dimension \( r \). Hence \( P_{n,q} \) admits a direct sum decomposition on \( H_r \oplus H_r^\perp \). We note also the restriction of \( P_{n,q} \) to \( H_r \) coincides with \( P_1,q \). It follows that all eigenvalues of \( P_1,q \) are also in the spectrum of \( P_{n,q} \). We define the relative spectral radius as the following operator norm

\[
\bar{\rho}_{n,q} := \| (P_{n,q})_{|H_r^\perp} \|_{\ell^2_0(\pi) \to \ell^2_\emptyset(\pi)} = \max_{\lambda \in \text{Sp}(P_{n,q}_{|H_r^\perp})} |\lambda|.
\]

From [34] p256 (see also [36]), if \( (\varphi_n) \) converges strongly then we have

\[
\lim_n \bar{\rho}_{n,q} = \rho_q,
\]

where \( \rho_q \) is the spectral radius of \( P_q \). Besides, since \( \rho_q \) coincides with the \( \ell^2_0(\pi) \to \ell^2_\emptyset(\pi) \) operator norm, from the triangle inequality, we have

\[
\rho_q \leq \sum_{(i,j) \in [r]^2} \sqrt{\frac{\pi(i)}{\pi(j)}} \rho_{q(i,j)},
\]

where \( q(i, j) = (q_g(i, j)) \in \mathbb{R}^G \) and \( \rho_{q(i,j)} \) is the spectral radius of \( P_{q(i,j)} \) in \( \ell^2(G) \). It follows, that if \( G \) has the RD property and \( R \) is the diameter of the support of \( q \) (in the Cayley graph associated
We use Proposition 7 for the walk

\[ (66) \]

\[ \rho_q \leq CR^C \sum_{(i,j) \in [r]^2} \sqrt{\frac{\pi(i)}{\pi(j)}} \left( \sum_{g \in G} q_g(i,j) \right)^2 \leq CR^C \sum_{g \in G} \|q_g\|^2_{\ell^2(\pi) \to \ell^2(\pi)}. \]

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

\[ u(g) := R_p(e,g) = \sum_{t=0}^{\infty} P^t_p(e,g). \]

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

\[ (65) \]

\[ U := \{ g \in G : \|u(g)\|_{\ell^2(\pi) \to \ell^2(\pi)} \geq 1/k \}. \]

The skeleton 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 \( Q = P_\pi \) as the transition kernel of the random walk \( \mathcal{X}_{\tau_m} \).

From (60) and \( \rho_p < 1 \) (since \( G \) is non-amenable), the proofs and statements of Proposition 11 and Lemma 12 continue to hold in our new setting (in (61), we replace \( q_g \leq 1/k \) by \( \|q_g\|_{\ell^2(\pi) \to \ell^2(\pi)} \leq 1/k \)).

We observe that \( Q = P_\pi \) satisfies the right-hand side of (68). Indeed since for all \( g \in G \), the adjoint of \( p_g \) in \( \ell^2(\pi) \) is \( p_g^{-1} \), we have \( \|u(g)\|_{\ell^2(\pi) \to \ell^2(\pi)} = \|u(g^{-1})\|_{\ell^2(\pi) \to \ell^2(\pi)} \) and \( U = U^{-1} \). Moreover, for \( g \in G \), \( q_g(i,j) \) can be written as a sum of weighted paths from \((e,i)\) to \((g,j)\) of the form \( p_{g_1}(i_0,i_1) \cdots p_{g_k}(i_{k-1},i_k) \) with \( i_0 = i \) and \( i_k = j \) and \( g_k \cdots g_1 = g \).

6.2.3. Deducing mixing time from RD property and the strong convergence. We may now conclude the proof of Theorem 21 by adapting the content of Subsection 4.4.

Proof of Theorem 21. We fix \( \epsilon \in (0,1), \delta > 0 \) and \((x,u_0) \in V_n \times [r]\) and prove that for \( n \) sufficiently large

\[ T_{\text{mix}}^n((x,u_0),\epsilon) \leq (1 + \delta) \log n/h(p). \]

Let \( (\tau_m) \) and \( U \) be as above for some large \( k \) to be chosen. We set \( m := \lceil (1 + \delta/4)(\log n)/\log k \rceil \). We use Proposition 7 for the walk \( X_t = \mathcal{X}_{\tau_m}(\mathcal{X}_t,x) \) with

\[ T = \tau_m + s, \quad t = \lceil (1 + \delta) \log n/\log - 2 \log \log n \rceil \quad \text{and} \quad s = \lceil \log \log n \rceil. \]

As \( \lim_n \rho_n.p = \rho_p < 1 \) (from (63)) and \( \rho_{1.p} < 1 \) (by assumption \( P_{1.p} \) is irreducible and aperiodic), we have for some \( \delta > 0 \) and all \( n \) large enough that

\[ (66) \]

\[ \rho_{n,p} = \|P_{n,p}\|_{\ell^2(\pi) \to \ell^2(\pi)} = \max(\tilde{\rho}_{n,p}, \rho_{1,p}) < 1 - \delta. \]
It follows that the third term in (25) is smaller than $\varepsilon/3$. It remains to prove that for $n$ sufficiently large

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

where $Q_n = P_{n,q}$ is the Markov chain of the induced walk $X_{\tau_m} = \bar{\varphi}_n(X_{\tau_m}, x)$ on $V_n \times [r]$. For the first inequality of (67), we choose $k(\delta)$ sufficiently large and it is a consequence of Lemma 12 and the law of large numbers.

The second inequality of (67) is obtained using spectral estimates for $Q_n = P_{n,q}$. The Cauchy-Schwarz inequality gives

$$\|P_{n,p}^* Q_n^m((x,u_0),\cdot) - \pi_n\|_{TV} \leq C_0 \sqrt{n} \|Q_n^m P_{n,p}^* f\|_{\ell_2^2(\pi)},$$  

with $C_0 = \sqrt{r/\min_i \pi(i)}$ and $f(y,v) = \delta_{(x,u_0)}(y,v)/\pi(v) - 1/n$. Let $\Pi_H$ be the orthogonal projection in $\ell_2^2(\pi)$ onto a vector space $H$. We find

$$\|Q_n^m P_{n,p}^* f\|_{\ell_2^2(\pi)} \leq \|Q_n^m P_{n,p}^* \Pi_H, f\|_{\ell_2^2(\pi)} + \|Q_n^m P_{n,p}^* \Pi_{H^\perp}, f\|_{\ell_2^2(\pi)}.$$

We now compute a spectral bound of the two terms on the right-hand side of (69). We first observe that $\|f\|_{\ell_2^2(\pi)} \leq 1$ and $\|\Pi_H, f\|_{\ell_2^2(\pi)} \leq C/\sqrt{n}$ with $C = 1/\sqrt{\min_i \pi(i)}$. Since $\langle f, 1 \rangle = 0$, we find from (66) and the fact that $Q_n$ is a contraction in $\ell_2^2(\pi)$,

$$\|Q_n^m P_{n,p}^* \Pi_H, f\|_{\ell_2^2(\pi)} \leq \|Q_n^m P_{n,p}^* \Pi_H, f\|_{\ell_2^2(\pi)} \leq \frac{C}{\sqrt{n}} \rho_{n,p} \leq \frac{C}{\sqrt{n}} (1 - \delta)^s.$$

We now give a bound of the second term on the right-hand side of (69). From (64) and Proposition 11 we have for some constant $C$

$$\rho_{n,q} \leq C k^{-1/2}(\log k)^C.$$

From (63) we deduce that for all $n$ large enough,

$$\tilde{\rho}_{n,q} \leq 2Ck^{-1/2}(\log k)^C.$$

Since $\|f\|_{\ell_2^2(\pi)} \leq 1$, $P_{n,p}^* \Pi_{H^\perp} = \Pi_{H^\perp} P_{n,p}^*$ and $P_{n,p}$ is a contraction in $\ell_2^2(\pi)$, we deduce that

$$\|Q_n^m P_{n,p}^* \Pi_{H^\perp}, f\|_{\ell_2^2(\pi)} \leq \tilde{\rho}_{n,q}^m \|f\|_{\ell_2^2(\pi)} \leq 2Ck^{-1/2}(\log k)^C.$$

Equation (68) together with (69)-(70)-(71) guaranties that $X_{\tau_n+s}$ is close to equilibrium in total variation. It concludes the proof of (67). \qed

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

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

[14] C. Bordenave and B. Collins. Eigenvalues of random lifts and polynomial of random permutations matrices. https://arxiv.org/abs/1801.00876, 2018.

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

[16] 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.

[17] 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.

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

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

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

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

[22] 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.

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

[24] R. I. Grigorchuk and A. Žuk. 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.

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

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

[27] H. Kesten. Symmetric random walks on groups. Trans. Amer. Math. Soc., 92:336–354, 1959.
[28] 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.

[29] 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.

[30] E. Lubetzky. Private communication.

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

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

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

[34] 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.

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

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

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