ON SENSITIVITY OF UNIFORM MIXING TIMES

JONATHAN HERMON

Abstract. We show that the order of the $L_\infty$-mixing time of simple random walks on a sequence of uniformly bounded degree graphs of size $n$ may increase by an optimal factor of $\Theta(\log \log n)$ as a result of a bounded perturbation of the edge weights. This answers a question and a conjecture of Kozma.

Keywords: Sensitivity; mixing-time; sensitivity of mixing times; hitting times.
MSC class: 60J10.

1. Introduction

An important question is whether mixing times are robust under small changes to the geometry of the Markov chain. For instance, can bounded perturbations of the edge weights change the mixing time by more than a constant factor? Similarly, how far apart can the mixing times of lazy simple random walks on two roughly-isometric graphs of bounded degree be? A related question is whether mixing times can be characterized up to universal constants (perhaps only under reversibility) using geometric quantities or extremal characterizations which are robust.

Different variants of this question were asked by various authors such as Pittet and Saloff-Coste [15], Kozma [10, p. 4], Diaconis and Saloff-Coste [5, p. 720] and Aldous and Fill [2, Open Problem 8.23] (the last two references ask for an extremal characterization of the $L_\infty$-mixing time in terms of the Dirichlet form).

Denote the $L_p$ mixing time of lazy simple random walk on a finite connected simple graph $G$ by $\tau_p(G)$ (see (1.3)). Kozma [10] made the following conjecture:

Conjecture 1.1 ([10]). Let $G$ and $H$ be two finite $K$-roughly isometric graphs (see Definition 1.6) of maximal degree $\leq d$. Then for some $C(K, d)$ depending only on $(K, d)$,

$$\tau_\infty(G) \leq C(K, d) \tau_\infty(H).$$

Our main result, Theorem 1, asserts that this conjecture is false. We shall only consider the following particularly simple type of rough isometries. Let $G_1 := (V_1, E_1)$ be some graph. Let $G_2 = (V_2, E_2)$ be a graph obtained from $G_1$ by “stretching” some of the edges of $G_1$ by a factor of at most $K$ (we say that $G_2$ is a $K$-stretch of $G_1$). That is, for some $E \subset E_1$ we replace each edge $\{u, v\} \in E$ by a path of length at most $K$ (whose end-points are still denoted by $u$ and $v$). Note that $V_1 \subset V_2$. The identity map is a $K$-rough isometry of $G_1$ and $G_2$.

Theorem 1. There exist two families of uniformly bounded degree simple connected graphs $(G_n = (V_n, E_n))_{n \geq 0}$ and $(G'_n)_{n \geq 0}$ and some $c > 0$ such that $|V_n| \to \infty$ and for each $n$, $G'_n$ is a 2-stretch of $G_n$ and

$$\tau_1(G'_n) \geq c \tau_\infty(G_n) \log \log |V_n|.$$  

(1.1)

Remark 1.2. The log $\log |V_n|$ term in (1.1) is optimal. This follows from the fact that the Log-Sobolev constant, and as shown in [10], also the spectral-profile, provide bounds on $\tau_\infty(G)$ for a graph $G = (V, E)$ of maximal degree $d$ which are sharp up to a $C_d \log \log |V|$ factor. These bounds

University of Cambridge, Cambridge, UK. E-mail: jonathan.hermon@statslab.cam.ac.uk. Financial support by the EPSRC grant EP/L018896/1.

3That is, using quantities which can change by at most some bounded factor under a bounded perturbation of the edge weights, or for lazy simple random walk on a bounded degree graph, under rough-isometries.
of geometric bounds on mixing-times for general (weighted) trees under bounded perturbation of the edge weights by at most some multiplicative constant factor can change these geometric bounds only by a constant factor. Theorem 1 serves as a cautionary note on the possibility of developing sharp geometric bounds on mixing-times.}

In contrast with Theorem 1, many well-known families of graphs are robust. Robustness of the $L_\infty$ and $L_1$ mixing times for general (weighted) trees under bounded perturbation of the edge weights was established in [8] (by Peres and the author) and [14] (by Peres and Sousi), respectively. Robustness of $L_1$ mixing times for general trees under rough-isometries was recently established in [11] by Addario-Berry and Roberts. Some other examples, which we borrow from [6] (where robustness of the $L_1$ mixing time is considered, however apart from example (2) below, the analysis of the $L_\infty$ mixing time is identical) are collected in the next proposition.

**Proposition 1.5.** The following families of graphs are robust: (1) Tori $\mathbb{Z}^d_n$ for every fixed $d$; (2) The giant component of a supercritical Erdős-Rényi random graph $\mathcal{G}(n, c/n)$ for a
fixed $c > 1$; (3) Maximal connected component of a critical Erdős-Rényi random graph $G(n, c/n)$; (4) The hypercube $\{0, 1\}^n$.

1.1. Definitions. Given a (weighted) network $(V, E, (c_{e})_{e \in E})$, where each edge $\{u, v\} \in E$ is endowed with a conductance (weight) $c_{u,v} = c_{v,u} > 0$ (with the convention that $c_{u,v} = 0$ if $\{u, v\} \notin E$), a lazy random walk on $G = (V, E)$, $(X_{t})$, repeatedly does the following: when the current state is $v \in V$, the random walk will stay at $v$ with probability $1/2$ and move to vertex $u$ (such that $\{u, v\} \in E$) with probability $c_{u,v} / (2c_{v})$, where $c_{v} := \sum_{w \in V} c_{v,w}$. The default choice for $c_{u,v}$ is 1 (in which case, we say that the random walk is unweighted), which corresponds to lazy simple random walk on $G$ (in which at each step the walk with equal probability either stays put or moves to a new vertex, chosen from the uniform distribution over the neighbors of its current position). Its stationary distribution is given by $\pi(x) := c_{x}/c_{V}$, where $c_{V} := \sum_{v \in V} c_{v} = 2 \sum_{e \in E} c_{e}$. This is a reversible Markov chain, i.e. $\pi(u)P(u, v) = \pi(v)P(v, u)$ for all $u, v \in V$, where throughout $P$ denotes the transition matrix of the walk.

We denote by $P_{\pi}^{t_{x}}$ (resp. $P_{x}$) the distribution of $X_{t}$ (resp. $(X_{t})_{t \geq 0}$), given that the initial state is $x$. Let $\mu, \nu$ be two distributions on the state space $\Omega$. Denote $a_{\mu, \nu, \pi} := \frac{\|\mu \circ \pi(x)\|_{\pi(x)}}{\|\pi(x)\|_{\pi(x)}}$. The family of $L_p$ distances is defined as follows.

$$
\|\mu - \nu\|_{p, \pi, \Omega} := \begin{cases} 
(\sum_{x} \pi(x)a_{\mu, \nu, \pi}^{p})^{1/p}, & 1 \leq p < \infty, \\
\max_{x \in \Omega} a_{\mu, \nu, \pi}, & p = \infty,
\end{cases} \tag{1.2}
$$

($p = 1$ gives twice the total-variation distance, i.e. $\|\mu - \nu\|_{1, \pi, \Omega} = 2\|\mu - \nu\|_{TV}$. The $\varepsilon$-$L_p$-mixing-time of the chain is defined as

$$
\tau_{p}(\varepsilon) := \min\{t : \max_{\pi} \|P_{\pi}^{t} - \pi\|_{p, \pi, \Omega} \leq \varepsilon\}. \tag{1.3}
$$

When $\varepsilon = 1/2$ we omit it from the notation and terminology (i.e. the $L_p$ mixing time is defined as the $1/2$-$L_p$ mixing time).

**Definition 1.6.** Let $G_{i} := (V_{i}, E_{i})$ ($i = 1, 2$) be two finite graphs. For $u, v \in V_{i}$, let $d_{i}(u, v)$ be graph distance (w.r.t. $G_{i}$) between $u$ and $v$ (i.e. the number of edges along the shortest path in $G_{i}$ between $u$ and $v$). We say that $f : V_{1} \rightarrow V_{2}$ is a $K$-rough isometry of $G_{1}$ and $G_{2}$ if

1. $\forall u, v \in V_{1}$, \hspace{1cm} \(d_{1}(u, v) - 1) / K \leq d_{2}(f(u), f(v)) \leq K(d_{1}(u, v) + 1).$

2. For every $w \in V_{2}$, there exists some $v \in V_{1}$ such that $d_{2}(f(v), w) \leq K$.

We say that $G_{1}$ and $G_{2}$ are $K$-roughly isometric if there exists such $f$ as above.

We use the convention that $C, C', C_{1}, \ldots$ (resp. $c, c', c_{1}, \ldots$) denote positive absolute constants which are sufficiently large (resp. small). Different appearances of the same constant at different places may refer to different numeric values.

1.2. The transitive case. The following question is re-iterated from [6] (over there it is asked for the $L_{1}$ mixing time) and essentially also from [15].

**Question 1.7.** Are transitive graphs robust?

Let $G = (V, E)$ be a finite connected transitive graph. Denote the eigenvalues of $I - P$ (where $P$ is the transition matrix of lazy simple random walk on $G$) by $0 = \lambda_{1} < \lambda_{2} \leq \cdots \leq \lambda_{|V|}$. Since for transitive graphs the mixing time is the same from all initial states, averaging over the starting position yields that (cf. [2] p. 284]) for all $x$ and $t \geq 0$

$$
\|P_{x}^{2t} - \pi\|_{\infty, \pi} = \|P_{x}^{2t} - \pi\|_{2, \pi}^{2} = \sum_{y} \pi(y)\|P_{y}^{t} - \pi\|_{2, \pi}^{2} = \sum_{i=2}^{|V|} \lambda_{i}^{2t}
$$

(transitivity is used only in the middle equality). Since the eigenvalues of $I - P$ are robust (e.g. [2] Corollary 8.4), it follows that transitive graphs are robust under bounded perturbations which
preserve transitivity. A positive answer to Question 1.7 will be obtained by a positive answer to the following question. Is it the case that for transitive graphs, also after a bounded perturbation, the ratio of the $L_2$ mixing time starting from the worst initial point with that starting from the best initial point is bounded?

1.3. Related work. It is classical that under reversibility the $L_1$ mixing time can be characterized using hitting times of sets which are “worst” in some sense (e.g. [11, Ch. 24]). Thus in order to show that it is not robust, it suffices to construct an example in which hitting times are not robust. As we explain below, this is somewhat easier.

Recently, the author and Peres obtained a characterization of the $L_2$ mixing time in terms of hitting time distributions. Namely, Theorem 1.1 in [12] asserts that under reversibility the $L_2$ mixing time is within some universal constant from the minimal time $t$ such that for every set $A$ of stationary probability at most $1/2$, the probability that $A$ is not escaped from by time $t$ is at most $\pi(A) + \frac{1}{2}\sqrt{\pi(A)(1 - \pi(A))}$. While we do not use this characterization as part of our analysis of the example from the proof of Theorem [11] it guided us in its construction and in the choices of certain parameters.

Ding and Peres [6] constructed a sequence of bounded degree graphs satisfying that the order of the total-variation mixing times strictly increases as a result of a certain sequence of bounded perturbations of the edge weights. Their construction was refined by the author and Peres in [9] (Theorem 3), which contains various additional results concerning sensitivity of mixing times and the cutoff phenomenon under small changes to the geometry of the chain.

Our construction from the proof of Theorem 1 uses a key observation from [6]. Namely, that the harmonic measure of the walk on a tree may change drastically as a result of a bounded perturbation, and that this can be used to create examples in which hitting times are not robust in the following sense. The order of the expected hitting time of some large set $A$ which is “worst” in some sense (both before and after the perturbation), starting from the worst initial state, may change as a result of the perturbation. The idea of exploiting the non-robustness of the harmonic measure was originally used by Benjamini [4] to study instability of the Liouville property.

Both in Ding and Peres’ construction and in our construction the chain mixes rapidly (there in total-variation and here in $L_2$) once it reaches a certain “huge” expander, $H$ (and because $H$ carries most of the stationary probability of the walk, the walk cannot mix before reaching it). We use the fact that the harmonic measure is sensitive (in the sense mentioned in the previous paragraph) in order to create “shortcuts” to $H$ for the walk on the original graph, $G_n$, which are essentially “invisible” for the walk on the 2-stretched graph, $G_n'$.

In order to change the mixing time in total-variation it suffices to change the order of the expected hitting time of $H$, starting from the worst initial state. However, this does not suffice in order to change the order of the $L_\infty$ mixing time. Thus we will work much harder and show that for every initial vertex $x$ the following holds. The probability that the walk on the original graph does not reach $H$ in some $t \ll \tau_1(G_n')$ steps (through one of the aforementioned shortcuts) is much smaller than the $L_\infty$ distance from stationarity of the distribution of the walk at time $t$, conditioned to not reach $H$ by that time. In order to achieve this, we will plant more shortcuts to $H$ in regions in which the chain mixes slower.

In [11] Kozma constructed a sequence of finite reversible Markov chains satisfying that $\rho^{(n)} \geq c\tau_\infty^{(n)} \log \log(\min_x \pi_n(x))$, for all $n$, where $\rho^{(n)}$ is the spectral profile bound (see (2.6)) on the $L_\infty$ mixing time of the $n$-th chain in the sequence, $\tau_\infty^{(n)}$. In Kozma’s construction there are $n - \lceil \log n \rceil + 1$ “islands” $H_{\lceil \log n \rceil}, \ldots, H_n$, each of size $2^{2n}$. The weights are chosen so that the stationary distribution is the uniform distribution. In his example the only way for the chain to “escape” from some $H_t$ is by moving to a random state picked according to the uniform distribution on the state space, which occurs in each step w.p. $n2^{-n}$.
As in Kozma’s construction there is a weighted edge between all pairs of vertices, it is not clear that such a construction is possible in the bounded degree setup. Nevertheless, our example uses several ideas from Kozma’s construction.

- In Kozma’s construction, each $H_i$ has vertex set $A_i \times B_i$. The network on $H_i$ can be described as a “Cartesian product” of the complete graphs on $A_i$ and $B_i$, in which the walk updates its $B_i$ co-ordinate at a rate $\alpha_i(n) = o(1)$.

  In our construction we use a bounded degree analog of the aforementioned network, which we denote by $U_i$. Namely, we replace the complete graphs by expanders. In order to delay the rate of transitions along the expander on the co-ordinate corresponding to $B_i$ in Kozma’s construction, while keeping the graph unweighted, we stretch each of its edges by a factor $\ell_i(n) \to \infty$.

- In Kozma’s example, for all $i$, $|A_i| = 2^{2^i-2}$ and so $\pi(A_{i+1}) \approx \pi(A_i)^2$. Moreover, each $A_i$ has roughly the same contribution to the spectral profile bound (see (2.0)). This is achieved by tuning the rates $\alpha_i(n)$ in an appropriate manner. Namely, by setting $\alpha_{i+1}(n) = 2\alpha_i(n)$, for all $i$. As noted in Remark 1.3 in some sense, such behavior of the spectral profile is necessary in order for it to overshoot $\tau_\infty$ by an optimal factor (of order $\log \log |V|$). In our construction we will take $\ell_i(n) = 2\ell_{i+1}(n)$ in order to obtain the same effect.

- Recall that in Kozma’s construction each pair of vertices are connected by a weighted edge, which has the effect of bringing the chain to stationary “at once” at a fixed rate, and this is the only way the chain can escape from the “island”, $H_i$, it started at. In our construction we need to somehow imitate this behavior (in a bounded degree, unweighted fashion). At the same time, as noted in Remark 1.3 we need that after stretching some of the edges by a factor of two, the walk must sequentially move through all of the islands in order to mix in total-variation.

  In order to achieve this behavior, we “stitch” the “islands” together so that the walk can escape from each of them either to a huge expander $H$ (once it is reached, the walk mixes rapidly) or to an island with adjacent index. The islands are glued to each other and to the expander $H$ in a way that allows us to manipulate (by stretching some edges) the probability of escaping an “island” by reaching the expander $H$.

2. Preliminaries

Recall that the spectral gap of a reversible Markov chain with transition matrix $P$ on a finite state space is defined as the smallest non-zero eigenvalue of $I - P$. We say that a graph $G$ is a $\lambda$-expander if the spectral gap of lazy simple random walk on $G$, denoted by $\lambda(G)$, is at least $\lambda$. We say that a sequence of graphs $(G_n)_{n \in I}$ is an expander family if $\inf_{n \in I} \lambda(G_n) > 0$. As mentioned in §1.3 we shall use expanders as building blocks in our construction. By abuse of terminology, below we often refer to a single graph as an “expander”. What we actually mean by that is that all of the graphs we refer to as expanders in the construction of the family of graphs we construct form together an expander family.

**Definition 2.1.** Consider a Markov chain chain on a finite state space $\Omega$ with transition matrix $P$ and stationary distribution $\pi$. We define the Cheeger constant of the chain as

$$
\Phi := \min_{A : 0 < \pi(A) \leq 1/2} Q(A) / \pi(A), \quad \text{where } Q(A) := \sum_{x \in A, y \notin A} \pi(x) P(x,y).
$$

The following is the well-known discrete analog of Cheeger inequality.

**Theorem 2.2** (e.g. [11], Theorem 13.14). If $P$ is reversible then

$$
\Phi^2 / 2 \leq \lambda \leq 2\Phi.
$$

By (2.1) a sequence of graphs $(G_n)_{n \in I}$ is an expander family iff $\inf_{n \in I} \Phi(G_n) > 0$.

The following proposition will be useful in what comes.
Proposition 2.3. There exists a constant $c_d > 0$ (depending only on $d$) such that if $H$ is a simple graph of maximal degree $d$ and $G$ is a $K$-stretch of $H$, then

$$
\Phi(G) \geq c_d \Phi(H)/K \quad \text{and so} \quad \lambda(G) \geq c_d^2 \Phi^2(H)/(2K^2).
$$

(2.2)

Proof. The argument in [12, Claim 2.2] covers the case in which $|V(G)| \leq \frac{d}{2}|V(H)|$. We shall reduce the general case to this case. It is easy to see that the minimum in the definition of $\Phi$ is always attained by a connected set. For $e = \{u, v\} \in E(H)$ let $\gamma_e$ be the collection of internal vertices along the segment from $u$ to $v$ in $G$ which replaced the edge $e$ (if $e$ was not stretched, then $\gamma_e$ is empty). Let $\pi_G$ be the stationary distribution of the walk on $G$. Similar reasoning as in [12, Claim 2.2] shows that for some connected set $B$

(i) $Q_G(B)/\pi_G(B) \leq c_d \Phi(G)$ (where $Q_G(B)$ denotes $Q$ w.r.t. the graph $G$).

(ii) There is at most one $e \in E(H)$ so that $\gamma_e \setminus B$ an $\gamma_e \cap B$ are both non-empty. However, even if such $e$ exists, $\gamma_e \cap B$ is connected.

The operation of contracting a set of vertices $D$ is defined as follows. Replace all of $D$ by a single vertex $x$ and for every edge $\{u, v\}$ with $u$ in $D$ and $v \not\in D$, replace it with an edge $\{x, v\}$ and if also $v \in D$ replace it by a loop at $x$ of weight 2. Let $G_2$ be the network obtained by contracting each $\gamma_e$, for all $e' \in E(H)$, apart from $\gamma_e$ for the $e$ so that $\gamma_e \setminus B$ an $\gamma_e \cap B$ are both non-empty (if such $e$ exists). If such $e$ exists, we contract $\gamma_e \cap B$ and also $(\gamma_e \setminus B) \cup \{v\}$, where $v$ is the endpoint of $e$ incident to $(\gamma_e \setminus B)$.

Let $G_3$ be the graph obtained by deleting all loops from $G_2$. It is straightforward to check that $\Phi(G_2) \geq \Phi(G_3)/(2K) \geq c_d^2 \Phi(H)/K$ (where the second inequality follows from [12, Claim 2.2]). Conversely, by (i)-(ii), $\Phi(G_2) \leq Q_G(B)/\pi_G(B) \leq c_d \Phi(G)$ (where the first inequality is obtained by considering the set $B'$ in $G_2$ obtained from $B$ by replacing each $\gamma_e' \subset B$ by the corresponding vertex in $G_2$) and so indeed $\Phi(G) \geq c_d \Phi(H)/K$. \hfill \Box

The Poincaré (spectral gap) inequality asserts that when time is scaled according to the inverse of the spectral gap, the $L_2$ distance from stationarity of every distribution decays exponentially in the number of (scaled) time units.

Lemma 2.4. Let $(\Omega, P, \pi)$ be a finite lazy irreducible reversible Markov chain with spectral gap $\lambda$. Let $\mu$ be a distribution on $\Omega$. Then

$$
\|P^t_{\mu} - \pi\|_{2, \pi} \leq e^{-\lambda t} \|\mu - \pi\|_{2, \pi}, \quad \text{for all } t \geq 0.
$$

(2.3)

Definition 2.5. Let $(\Omega, P, \pi)$ be reversible. Let $A \subsetneq \Omega$. We define $\lambda(A)$ (“the spectral gap of the set $A$”) to be the smallest eigenvalue of the sub stochastic matrix obtained by restricting $I - P$ to $A$. Similarly, define $\Phi(A) = \min_{B \subset A} Q(B)/\pi(B)$.

The following extension of (2.1) is due to Goel et al. [7, (1.4) and Lemma 2.4]. For every irreducible reversible chain, and every set $A$ with $\pi(A) \leq 1/2$ we have that

$$
\Phi^2(A)/4 \leq \lambda(A) \leq \Phi(A).
$$

(2.4)

The hitting time of a set $D$ is defined as $T_D := \inf \{t : X_t \in D\}$. Using the spectral decomposition of $P_A$, the restriction of $P$ to the set $A$ (e.g. [3, Lemma 3.8]), it is easy to show that for every set $A$ and $a, a' \in A$ we have that $P_A^t(a, a') \leq \frac{\pi(a')}{\pi(a)} e^{-\lambda(A)t}$ and so

$$
\forall a \in A, \quad P_a[T_{\Omega \setminus A} > t] = \sum_{b \in A} P_A^t(a, b) \leq |A| \max_{b \in A} \sqrt{\frac{\pi(b)}{\pi(a)}} e^{-\lambda(A)t}.
$$

(2.5)

Finally, we recall that the spectral profile upper bound on $\tau_{\infty}$ is [7, (4)]

$$
\rho := 8\lambda^{-1} \log 2 + \int_{\min_{x} \pi(x)}^{1/2} \frac{4dv}{v \Lambda(v)}, \quad \text{where } \Lambda(v) := \inf_{A \subset \Omega \mid \pi(A) \leq v} \lambda(A).
$$

(2.6)
3. Proof of Theorem \[\textbf{1}\]

3.1. The construction. For notational convenience we often omit ceiling signs. As described in §1.3 we shall construct graphs \(U_n, \ldots, U_1\). For all \(i\) we will have

\[
2^{2n_i + 5n_i} \leq |V(U_i)| \leq C2^{2n_i + 6n_i}.
\]

We will then “stitch” them together to obtain the ultimate graph \(G_n = (V_n, E_n)\). Hence

\[
n2^{2n_i + 5n_i} \leq |V_n| \leq Cn2^{2n_i + 6n_i}. \tag{3.1}
\]

For all \(i\), the graph \(U_i\) will be a Cartesian product of an expander \(H_i = (V(H_i), E(H_i))\) with a graph \(W_i\) (obtained by making a small modification to a certain tree \(T_i\)), of sizes

\[
2^{2n_i - 2n_i} \leq |V(H_i)| \leq 2^{2n_i - 2n_i + n_i - 1}, \quad 2^{2n_i + 1 + 5n_i} \leq |V(W_i)| \leq C2^{2n_i + 5n_i}, \tag{3.2}
\]

Step 1.1 Let \(1 \leq i \leq n\). We now construct the tree \(T_i\) which shall have roughly \(2^{2n_i + 1}\) good leafs \(GL_i\) and roughly \(2^{2n_i - 1}\) bad leafs \(BL_i\) (apart from \(i = 1\) which only has good leafs). It will be obtained by stretching the edges of a tree \(T_{bs,i}\) (bs stands for “before stretching”) which is a “binary tree”, rooted at \(a_i\), whose good leafs, \(GL_i\), are of depth \(2^{n_i}\), while its bad leafs, \(BL_i\), are of some other depth, \(j_i\). The sets of good leafs of the two trees \(T_{bs,i}\) and \(T_i\) are the same, and likewise for the sets of bad leafs. Note that usually a finite binary tree is defined so that all of its leafs are of the same depth, while here, crucially, the leaf set is partitioned into two sets of different depths. By abuse of terminology, we still refer to such a tree as a finite binary tree.

- We first describe the construction of \(T_i\) as it is simpler. Take a binary tree, \(T_{bs,1}\), of depth \(2^{n_1}\) rooted at \(a_1\). Denote its leafs by \(GL_1\). Then stretch each of its edges by a factor of \(2^{n_1}\).

- We now construct \(T_i\) for \(1 < i \leq n\) in several steps. Before describing \(T_{bs,i}\) we consider a binary tree of depth \(2^{n_i/2}\), rooted at \(a_i\), denoted by \(T_{fh,i}\) (fh stands for “first half”, as \(T_{bs,i}\) is the “first half” of \(T_{bs,i}\)).

Step 1.2 For every vertex \(u\) which is not a leaf of \(T_{fh,i}\) we distinguish its two children by left and right child. For every vertex \(u \in T_{fh,i}\) let \(Left(u)\) (resp. \(Right(u)\)) be the number of left (resp. right) children along the path from \(a_i\) to \(u\). Let

\[
g(u) = \text{Left}(u) - \text{Right}(u). \tag{3.3}
\]

Denote the \(k\)-th level of a rooted tree \(T\) by \(L_k(T)\). We partition the leaf set of \(T_{fh,i}\), \(L_{2^{n_i/2}}(T_{fh,i})\), into two parts: \(\text{GMP}_i\), the set of good middle points and \(\text{BMP}_i\), the set of bad middle points (they are “middle points” w.r.t. \(T_{bs,i}\)) defined as follows:

\[
\text{GMP}_i := \{u \in L_{2^{n_i/2}}(T_{fh,i}) : g(u) \leq 2^{n_i - 6}\} \quad \text{and} \quad \text{BMP}_i := L_{2^{n_i/2}}(T_{fh,i}) \setminus \text{GMP}_i.
\]

We now extend \(T_{fh,i}\) so that the good leafs, \(GL_i\) (resp. bad leafs \(BL_i\)), of the resulting tree, \(T_{bs,i}\), will be the leafs which are decedents of \(\text{GMP}_i\) (resp. \(\text{BMP}_i\)).

- Attach to each vertex in \(\text{GMP}_i\) a binary tree of depth \(2^{n_i} - 2^{n_i/2}\). This makes the total number of leafs that have a vertex in \(\text{GMP}_i\) as an ancestor \(2^{2^{n_i/2}}(1 - o(1))\).

- Attach to each vertex in \(\text{BMP}_i\) a binary tree of depth \(\lceil \log_2(2^{2^{n_i/2}}/|\text{BMP}_i|) \rceil\) so that the set of leafs that have a vertex in \(\text{BMP}_i\) as an ancestor, \(BL_i\), is of size

\[
2^{2^{n_i/2} - 1} \leq |BL_i| \leq 2^{2^{n_i/2} - 1} + 1. \tag{3.4}
\]

Call the resulting tree \(T_{bs,i}\).

Step 1.3 Denote the union of \(\text{GMP}_i\) with the collection of vertices of \(T_{bs,i}\) which have an ancestor in \(\text{GMP}_i\) by \(\text{GS}_i\) (a shorthand for “good side”).
• For each $1 < i \leq n$, stretch each edge of $\mathcal{T}_{bs,i}$ that both of its end-points lie in $\text{GS}_i$ by a factor of $2^{5n}$. Stretch each of the rest of the edges of $\mathcal{T}_{bs,i}$ by a factor of

$$q_i := \lceil 2^{5n - \frac{7}{2}} \rceil.$$

Call the resulting tree $\mathcal{T}_i$.

Step 2 We now modify the tree $\mathcal{T}_i$ in the region close to its leafs in order to make it an expander.

We note that for $i = 1$ the below modification is done only to the region close to $\text{GL}_1$ as $\mathcal{T}_1$ only has good leafs. Let $G_{E_i} = (V(G_{E_i}), E(G_{E_i}))$ and $B_{E_i} = (V(B_{E_i}), E(B_{E_i}))$ (E stands for expander) be 3-regular expanders of size $|\text{GL}_i|/2$ and $|\text{BL}_i|/2$, resp.. Let $\mathcal{P}_{GL_i}$ (resp. $\mathcal{P}_{BL_i}$) (a shorthand for “parents of good (resp. bad) leafs” w.r.t. $\mathcal{T}_{bs,i}$), be the collection of the $|\text{GL}_i|/2$ (resp. $|\text{BL}_i|/2$) vertices of distance $2^{5n}$ (resp. $q_i$) w.r.t. $\mathcal{T}_i$ from $\text{GL}_i$ (resp. $\text{BL}_i$). We naturally identify $\mathcal{P}_{GL_i}$ (resp. $\mathcal{P}_{BL_i}$) with the collection of parents w.r.t. $\mathcal{T}_{bs,i}$ of the vertices in $\text{GL}_i$ (resp. $\text{BL}_i$). Identify each vertex of $\mathcal{P}_{GL_i}$ (resp. $\mathcal{P}_{BL_i}$) with a vertex of $G_{E_i}$ (resp. $B_{E_i}$) in a bijective manner. Let $\mathcal{T}_v = (V(\mathcal{T}_v), E(\mathcal{T}_v))$ be the induced tree at $v$ w.r.t. $\mathcal{T}_i$ (i.e. the induced tree on the set of vertices that the path from them to $o_i$ goes through $v$). For each $u, v \in \mathcal{P}_{GL_i}$ (resp. $\mathcal{P}_{BL_i}$) such that $\{u, v\} \in E(G_{E_i})$ (resp. $E(B_{E_i})$), we connect each vertex $w \in V(\mathcal{T}_u)$ to $\phi_{u,v}(w) \in V(\mathcal{T}_v)$ by an edge, where $\phi_{u,v}$ is the trivial isomorphism of $\mathcal{T}_u$ and $\mathcal{T}_v$. Call the resulting graph $W_i = (V(W_i), E(W_i))$.

Step 3 Let $1 \leq i \leq n - 1$. Denote $s_i := \prod_{j=i+1}^{n} |\text{BL}_i|$. Note that by (3.4)

$$2^{3n-22n+i} \leq s_i \leq 2^{3n-22n+i+n-i}.$$

Let $H_i = (V(H_i), E(H_i))$ be a 3-regular expander of size $s_i$. Let $U_i = W_i \times H_i$ (a Cartesian product of $W_i$ and $H_i$). That is, $V(U_i) = V(W_i) \times V(H_i)$ and $\{(u, h), (u', h')\} \in E(U_i)$ if either $u = u'$ and $\{h, h'\} \in E(H_i)$ or $h = h'$ and $\{u, u'\} \in E(W_i)$.

We now “stitch” together the graphs $U_n, \ldots, U_1$. We refer to $R_i := \{o_i\} \times V(H_i)$ (for $i = n$, $R_n := \{o_n\}$) as the roots of $U_i$. We refer to the set

$$\text{Bad}_i := B_{L_i} \times V(H_i) \quad \text{(for $i = n$, $\text{Bad}_n := \text{BL}_n$)}$$

as the bad leafs of $U_i$ (even though it is not a tree). Similarly, we refer to

$$\text{Good}_i := G_{L_i} \times V(H_i) \quad \text{(for $i = n$, $\text{Good}_n := \text{GL}_n$)}$$

as the good leafs of $U_i$.

Note that $|\text{Bad}_{i+1}| = |R_i| = s_i$, for all $i$. We shall connect $U_{i+1}$ to $U_i$ for all $1 \leq i < n$, by identifying $\text{Bad}_{i+1}$ with $R_i$ (step 4). We shall also connect all of the $U_i$’s “at once” by connecting $\text{Good} := \bigcup_{i=1}^{n} \text{Good}_i$ using one “huge” expander $H$ of size $|\text{Good}|$ (step 5).

Step 4 For all $i < n$, we take an arbitrary bijection $\phi : \text{Bad}_{i+1} \rightarrow R_i$. We then replace each $u \in \text{Bad}_{i+1}$ and $\phi(u)$ with a new vertex, $u'$, that the set of edges which are incident to it is the union of the edges which are incident to $u$ (in $U_{i+1}$) and to $\phi(u)$ (in $U_i$). By abuse of notation we shall not distinguish between the set $\{u' : u \in \text{Bad}_{i+1}\}$ and the sets $\text{Bad}_{i+1}$ and $R_i$.

Step 5 We identify the set Good with a 3-regular expander $H = (V(H), E(H))$ of size $|\text{Good}|$ as follows. We label Good by the set $V(H)$ and connect $u, v \in \text{Good}$ if $\{u, v\} \in E(H)$. Call the obtained graph $G_n = (V_n, E_n)$.

Step 6 We now describe $G'_n = (V'_n, E'_n)$. The only difference between it and $G_n$ is that for all $1 < i \leq n$ we perform the following step in between step 1.1 and step 1.2:
Consider the set of vertices $V(\mathcal{T}_{\text{bd},i})$ in the tree $\mathcal{T}_{\text{bd},i}$. For every non-leaf vertex $u \in V(\mathcal{T}_{\text{bd},i})$ and its right child $v$, we stretch the edge $\{u, v\}$ by a factor of 2. Call the obtained graph $\mathcal{T}_{\text{bd},i}'$. The remaining steps are analogous to the ones in the construction of $G_n$. We spell them out for the sake of concreteness.

The leaf set of $\mathcal{T}_{\text{bd},i}'$ can be identified with that of $\mathcal{T}_{\text{bd},i}$. Hence we may partition the leaf set of $\mathcal{T}_{\text{bd},i}'$ into the same two sets GMP and BMP, defined in step 1.2 using the function $g$, taken again w.r.t. the tree $\mathcal{T}_{\text{bd},i}$.

After this is done, we can proceed with step 1.2 and obtain the tree $\mathcal{T}_{\text{bs},i}'$.

We denote the tree of index $i$ obtained at the end of step 1.3 (when before step 1.2 one performs the aforementioned intermediate step) by $\mathcal{T}_i'$. The only difference between $\mathcal{T}_i'$ and $\mathcal{T}_i$ is that the edges that were stretched by a factor 2 in $\mathcal{T}_{\text{bd},i}'$ will be stretched ultimately by a total factor of $2q_i$ rather than just $q_i$ (as in $\mathcal{T}_i$).

The construction of $G_n'$ is concluded by completing the remaining steps of the construction of $G_n$ with $\mathcal{T}_i'$ now playing the role of $\mathcal{T}_i$ in construction of $G_n$.

### 3.2. Analysis of the construction

We start with the analysis of $G_n'$. We write $P'$ and $E'$ to denote probabilities and expectations w.r.t. the walk on either $G_n'$, or $\mathcal{T}_i'$, for some $i$ (where the identity of the graph will be clear from context, and otherwise specified). Denote the stationary distribution of the walk on $G_n'$ by $\pi'$. We will show that (for all sufficiently large $n$)

$$\tau_1(G_n') \geq (n - 1)2^{18n - 2} =: \tau$$

(3.6)

Denote Nice$_i = \text{GMP} \times \mathcal{V}(H_i)$ for $2 \leq i \leq n$ and Nice$_1 = \mathcal{R}_i$. Denote the connected component of o$_n$ w.r.t. the cut Nice := $\bigcup_{i=1}^n$Nice$_i$ of the graph $G_n'$ by Small. Recall that for every distribution $\mu$ on the state space we have that $\frac{1}{2}\|\mu - \pi'\|_1 = \|\mu - \pi'\|_{TV} = \max_A \mu(A) - \pi'(A)$, where $\|\mu - \pi'\|_{TV}$ is the total variation distance. It follows that

$$\|P_{o_n}(X_{\tau} \in \cdot) - \pi'(\cdot)\|_{TV} \geq P_{o_n}'[X_{\tau} \in \text{Small}] - \pi'(\text{Small}) \geq P_{o_n}'[T_{\text{Nice}} > \tau] - C_02^{-2n/2},$$

where we have used the following estimate in the last inequality

$$\pi'(\text{Small}) \leq \max_{u_n} \frac{\text{deg}(u)}{\text{deg}(v)} |\text{Small}|/|V(G_n')| \leq 15|\text{Small}|/|V(G_n')| \leq C_02^{-2n/2}.$$

Hence, in order to prove (3.6) it suffices to show that

$$P_{o_n}'[T_{\text{Nice}} \geq \tau] = 1 - o(1).$$

(3.7)

In order to establish (3.7) we use the following lemma.

**Lemma 3.1.** Uniformly in $2 \leq i \leq n$ and $r \in \mathcal{R}_i$, we have that

$$P_r'[T_{\text{Bad}} \leq 2^{18n - 2} \mid T_{\text{Bad},i} < T_{\text{Nice}}] \leq P_{o_n}'[T_{\text{BL},i} \leq 2^{18n - 2} \mid T_{\text{BL},i} < T_{\text{GMP},i}] = o(1/n),$$

(3.8)

$$P_r'[T_{\text{Nice}} < T_{\text{Bad},i}] \leq P_{o_n}'[T_{\text{GMP},i} < T_{\text{BL},i}] + o(1/n) = o(1/n),$$

(3.9)

where in the l.h.s of both (3.8) and (3.9) the probability is taken w.r.t. the walk on $G_n'$ and in the middle terms w.r.t. the walk on $\mathcal{T}_i'$.

As $R_i = \text{Bad}_{i+1}$ for all $1 \leq i < n$, it follows from Lemma 3.1 that w.p. $1 - o(1)$, started from $o_n$, before the walk on $G_n'$ reaches Nice it has to make its way from $R_{i+1}$ to $R_i$ for all $1 \leq i < n$, and each of these $n - 1$ “stages” will take it at least $2^{18n - 2}$ steps (in which case, it must be the case that $\tau_1(G_n') \geq (n - 1)2^{18n - 2}$, as desired). We believe that the assertion of Lemma 3.1 is intuitive and that from a high level perspective the proof is not complicated. For the sake of completeness we choose to present a relatively detailed proof of Lemma 3.1.

**Proof of Lemma 3.1** The first inequalities in both (3.8) and (3.9) are obtained via a straightforward coupling argument and symmetry (of the sets Nice$_i$, Bad$_i$ and $R_i$ w.r.t. the $H_i$ co-ordinate of the walk). The additive $o(1/n)$ term in the middle term of (3.9) is there to cover the following two scenarios w.r.t. the walk on $G_n'$, started from $r \in \mathcal{R}_i$:
• $T_{\text{jb}^n_{j < i}, \text{Nice}_j} < T_{\text{Nice}_j \cup \text{Bad}_j}$.
• $T_{\text{PBL}_j} < T_{\text{Nice}} < T_{\text{Bad}}$.

Working out the details of the aforementioned couplings is left as an exercise.

We now prove the equality in (3.8). Recall the notation from step 6 of the construction. Let $(X_t^k)_{s \geq 0}$ be lazy simple random walk on $T'_k$. We may view the walk $(X_t^k)_{s \geq 0}$ only when it visits distinct vertices of $T_{bs,i}$ (recall that $V(T_{bs,i}) \subset V(T'_k \setminus V(T'_k))$). That is, consider the walk $Y_j = X_t^k$, for $j \geq 0$, where $S_0 = T_{V(T_{bs,i})}$ and for $j \geq 1$

$$S_j = \inf\{s > S_{j-1} : X_s^k \in V(T_{bs,i}) \setminus \{X_{S_{j-1}}^k\}\}.$$

For a set $D \subset V(T_{bs,i})$ denote its hitting time w.r.t. $(Y_s)$ by

$$\tau_D := \inf\{s : Y_s \in D\}.$$

Recall that $V(T_{bs,i}) \setminus \text{GS}_i$ (where $\text{GS}_i$ is defined at step 1.3 of the construction) is the set of vertices of $T_{bs,i}$ which are connected in $T'_k$ to their neighbors w.r.t. $T_{bs,i}$ by paths of length either $q_i$ or $2q_i$. It is easy to see that

(a) If $Y_{j-1} \in V(T_{bs,i}) \setminus \text{GS}_i$, then the law of $S_j - S_{j-1}$ stochastically dominates the law of the time it takes lazy simple random walk on $Z$ to reach distance $q_i$ from its starting position (call this law $\xi$). Moreover, started from $o_i$, conditioned on $\tau_{\text{GMP}} \geq j$, we have that $S_j$ stochastically dominates

$$S'_j := \sum_{k=1}^j \xi_k,$$

where $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with law $\xi$.

(b) Started from $o_i$, conditioned on $\tau_{\text{BL}_i} < \tau_{\text{GMP}}$, we have that the law of $\tau_{\text{BL}_i}$ stochastically dominates the law of the hitting time of the $2^{2n+i-2}$-th level of a binary tree, by simple random walk started from its root.

(c) The law of the last hitting time is highly concentrated around $3 \times 2^{2n+i-2}$. Hence

$$\mathbb{P}_{o_i}[\tau_{\text{BL}_i} \leq b_i \mid \tau_{\text{BL}_i} < \tau_{\text{GMP}}] \leq 2^{-c_02^{2n+i}} = o(1/n), \quad \text{where } b_i := 3 \times 2^{2n+i-3}.$$ 

Note that (using (3.5))

$$2^{18n-2} \leq \frac{2}{3}b_i q_i^2 = \frac{2}{3} \mathbb{E}[S'_{b_i}]$$

Using (a)-(c) above, it is not hard to verify that, for all $2 < i \leq n$,

$$\begin{align*}
P'_{o_i}[\tau_{\text{BL}_i} \leq 2^{18n-2} \mid \tau_{\text{BL}_i} < \tau_{\text{GMP}}] &\leq P'_{o_i}[\tau_{\text{BL}_i} \leq b_i \mid \tau_{\text{BL}_i} < \tau_{\text{GMP}}] + P'_{o_i}[S_{b_i} \leq 2^{18n-2} \mid \tau_{\text{BL}_i} < \tau_{\text{GMP}}] \\
&\leq 2^{-c_02^{2n}} + \mathbb{P}[S'_{b_i} \leq \frac{2}{3} \mathbb{E}[S'_{b_i}]] = o(1/n).
\end{align*}$$

where the first probability is taken w.r.t. the walk on $T'_k$ and $P'_{o_i}[S_{b_i} \leq 2^{18n-2} \mid \tau_{\text{BL}_i} < \tau_{\text{GMP}}]$ w.r.t. $(X_t^k)_{s \geq 0}$ (in the sense that $(S_j)$ and the walk $(Y_j)$ are both determined by $(X_s^k)_{s \geq 0}$). This concludes the proof of (3.8). We now prove the equality in (3.9).

Fix $1 < i \leq n$. Consider an infinite tree $T'$ obtained by starting with $T$, an infinite binary tree, rooted at $o$, and stretching every edge between each vertex and its left (resp. right) child by a factor of $q_i$ (resp. $2q_i$). Let $v_k$ (resp. $u_k$) be the last (resp. first) vertex in $L_k(T)$ to be visited by a simple random walk on $T'$ started from $o$.

As before, for every $v \in V(T)$ let $g(v) = \text{Left}(v) - \text{Right}(v)$ (this function is defined w.r.t. $T$). Using a network reduction and the well-known connection between random walks and electrical networks (cf. Example 9.9 for a similar but simpler network reduction), we argue that

**Lemma 3.2.** $(g(v_{k+1}) - g(v_k))_{k \geq 0}$ is a sequence of i.i.d. random variables, each equals to 1 w.p. $\frac{\sqrt{2}}{1+\sqrt{2}}$ and to $-1$ w.p. $\frac{1}{1+\sqrt{2}}$. 

We do not care so much about the exact value of the constant $\frac{\sqrt{2}}{1+\sqrt{2}}$. The important point is that it is larger than $1/2$. Before proving the lemma, we explain how it implies (3.9). It follows from Lemma 3.2 that $g(v_{2n+1-i})$ is highly concentrated around $2^{2n+i-2} \left( \frac{\sqrt{2}}{1+\sqrt{2}} \right)$. It is not hard to verify that this implies that $g(v_{2n+1-i}) \geq 2^{2n+i-5}$ w.p. $1 - o(1/n)$ (uniformly in $i$).

Using a coupling argument (in which the walks on $T'$ and on $T_i'$, started from $o$ and $o_i$, resp., are coupled so that they follow the same trajectory until they reach the $2^{2n+i-2}$th level of the corresponding non-stretched trees) we obtain (3.9) (we leave the details as an exercise). This concludes the proof of $\tau_1(G_n) \geq (n-1)2^{18n-2}$.

**Proof of Lemma 3.3.** The fact that $(g(v_{k+1}) - g(v_k))_{k \geq 0}$ is a sequence of i.i.d. random variables follows by symmetry and the definition of the sequence $(v_i)_{i \geq 0}$. Let $w$ be the effective conductance (e.g. [11] Ch. 9) from the root of $T'$ to “infinity” (i.e. the limit of the effective conductance between the root and $L_\ell(T')$ as $\ell \to \infty$). Similarly, let $w_R$ and $w_L$ be the effective conductances between the root and infinity in the right and left subtrees of $T'$ (where the right subtree is obtained by deleting from $T'$ the left child of the root and all of its descendants, and the left subtree is similarly defined), resp. Then $w = w_R + w_L$ and by symmetry $\frac{1}{w_R} = 2q_1 + \frac{1}{w}$ and $\frac{1}{w_L} = \frac{1}{w}$. The proof is concluded by solving these equations (which yields $w = \frac{1}{\sqrt{2}q_1}$ and $w_L = \frac{1}{(1+\sqrt{2})w}$) noting that $\mathbb{P}[g(v_1) - g(v_0) = 1] = \frac{\sqrt{2}}{\sqrt{q_1}}$, where the probability is taken w.r.t. SRW on $T'$, started at its root. □

We now show that $\tau_\infty(G_n) \leq C_12^{18n}$. By standard results it suffices to show that there exist $C > 0$ and $\ell > 1$ such that for all $n$,

$$\tau_\ell(G_n) \leq [C2^{18n}] =: t$$

($C$ shall and $\ell$ shall be determined later, and throughout we assume that $C$ is sufficiently large and $\ell - 1$ is sufficiently small so that all of the equations involving them below are satisfied). This follows from the following standard fact (e.g. [10] Lemma 2.4.6).

**Fact 3.3.** For every reversible Markov chain we have that for all $a > 0$

$$2\tau_2(\sqrt{a}) = \tau_\infty(a),$$

$$\forall 1 < \ell < 2, \quad \tau_2(a^{m_\ell}) \leq m_\ell \tau_\ell(a), \quad \text{where} \quad m_\ell := 1 + \lfloor (2-\ell)/(2\ell - 2) \rfloor.$$

We argue that the spectral gap of $G_n$, denoted by $\lambda(G_n)$, satisfies

$$\lambda(G_n) \geq c_1/(\max q_i)^2 = c_1 q_2^{-2} \geq c_2 2^{-16n}.$$  

This follows from Proposition 2.3, as $G_n$ is a $q_2$-stretch of a bounded degree expander.

We start with an elementary observation which shall be used below repeatedly. Let $x \in V_n$ and $t > 0$. Let $A$ be some event (which is determined by $(X_0, \ldots, X_t)$). Denote its complement by $A^c$. Then for all $1 < \ell \leq 2$

$$||P^\ell_x - \pi||_{\ell,x} = P_x[A]||P^\ell_x|\cdot | A| - \pi(\cdot)||_{\ell,x} + P_x[A^c]||P^\ell_x|\cdot | A^c| - \pi(\cdot)||_{\ell,x}.$$  

In particular, (using the fact that for every distribution $\mu$ on $V_n$ and every $\ell > 1$ we have that $||\mu - \pi||_{\ell,x} \leq C'\mathbb{V}_n(\ell^{-1}/\ell)$ if $P_x[A^c] \ll \mathbb{V}_n[\ell^{-1}/\ell]$, we may neglect the second term in the r.h.s. above, and concentrate on bounding $P_x[A]||P^\ell_x|\cdot | A| - \pi(\cdot)||_{\ell,x}$ from above.

We now argue that starting from a vertex in the expander $H$ the walk mixes in $L_2$ in at most $2^{17n}$ steps (this is a wasteful estimate, but it suffices for our purposes). Hence, by the Markov property and the triangle inequality, it suffices to show that for $t = [C2^{18n}]$,

$$\forall x, \quad P_x[T_{V(H)} > t]||P^t_x|\cdot | T_{V(H)} > t| - \pi(\cdot)||_{2,x} \leq 1/4.$$  

**Lemma 3.4.**

$$\lim_{n \to \infty} \max_{x \in V(H)} ||P^t_x - \pi||_{2,x} = 0.$$
Proof: Let
\[ J := \text{PGL}_n \cup (\cup_{i=1}^{n-1} \text{PGL}_i \times V(H_i)) \]
be the collection of vertices of distance (w.r.t. \( G_n \)) \( 2^n \) from \( V(H) \). Each vertex in \( J \) is connected to two vertices in \( V(H) \) by a path of length \( 2^n \). Note that (started from \( V(H) \)) by time \( 2^{14n} \) the walk reaches \( J \) w.p. at least \( 1 - C_2 |V_n|^{-2} \). Hence (by the above discussion regarding (3.15)) we can neglect the case this fails. Let \( i \) be the index such that \( X_{T_j} \in V(U_i) \). Trivially, by the time the walk can cross a path of length \( 2^n \) it must make at least \( 2^n \) steps. But this means that, w.p. at least \( 1 - C_2 |V_n|^{-2} \), the walk will make at least \( 2^{n-6} \) steps along each of the expanders \( H_i \) and \( G_{E_i} \) between the last visit to \( V(H) \) prior to \( T_j \) and \( T_{j-1} \). Using (3.11)-(3.12) and (2.8) it is not hard to show that this means that at time \( 2^{14n} \) the \( L_2 \) distance of the walk (started from \( H \)) from \( \pi \) is \( O(n2^n) \) (as we later perform a similar more subtle calculation, we leave this as an exercise). Finally, the claim is obtained using (3.11) and the Poincaré (spectral gap) inequality (2.9). \( \square \)

To conclude the proof we now verify (3.13). The case that \( x \in V(U_1) \) can be treated separately in a similar manner to the analysis below. The following lemma asserts that w.l.o.g. we may assume that the initial position of the walk is in \( R_i \) for some \( i \).

Lemma 3.5. Provided that \( \ell - 1 \) is sufficiently small and \( C \) is sufficiently large, for all \( 1 < i \leq n \) and \( x \in V(U_i) \) (uniformly)
\[
\mathbb{P}_x[T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2 || P^d_x[ \cdot | T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2] - \pi(\cdot) ||_{\ell, \pi} = o(1) \tag{3.14} \]

Proof: We first argue that for all \( 1 < i \leq n \) and \( x \in V(U_i) \) (provided that \( C \) is taken to be sufficiently large)
\[
\mathbb{P}_x[T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2] \leq C_3 2^{-2^{n+i}}. \tag{3.15} \]
Let \( Z_i \) (resp. \( \tilde{Z}_i \)) be the number of steps the walk made by time \( t/2 \) along the \( W_i \) (resp. \( H_i \)) co-ordinate of \( U_i \). As \( \mathbb{P}_x[Z_i < t/30] < C_2 |V(G_n)|^{-2} \), in order to prove (3.15) it suffices to show that (provided that \( C \) is taken to be sufficiently large)
\[
\max_{u \in V(W_i)} \mathbb{P}_u[T_{\text{GL}_i \cup \text{BL}_i \cup \{o_i\}} > t/30] \leq 2^{-2^{n+i}} \tag{3.16} \]
where the probability is taken w.r.t. the walk on \( W_i \). Let \( \tilde{P}_i \) and \( \tilde{\pi}_i \) be the transition matrix and stationary distribution (resp.) of lazy simple random walk on \( W_i \). Let \( \tilde{\lambda}_i \) be the smallest eigenvalue of the substochastic matrix obtained by restricting \( I - \tilde{P}_i \) to \( V(W_i) \) \( \setminus (\text{GL}_i \cup \text{BL}_i \cup \{o_i\}) \). Using (2.9) and an obvious extension of Proposition (2.8) it is not hard to verify that
\[
\tilde{\lambda}_i \geq C_2^{-1}(16n) \tag{3.17} \]

The proof of (3.16) (and so also of (3.15)) can now be concluded using (2.9).

Our next goal is to show that
\[
||P_x[X_{t/2} \in \cdot | T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2] - \pi(\cdot)||_{\ell, \pi} \leq C_7 2^{(\ell - 1)2^{n+i+2}} \tag{3.18} \]
Observe that this, in conjunction with (3.15), implies the assertion of the lemma.

For every \( (u, h) \in V(W_i) \times V(H_i) \) denote
\[
p_{u, h} := P_x[X_{t/2} \in (u, h) | T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2] \]
As before, since \( \mathbb{P}_x[\tilde{Z}_i < t/20] < C_2 |V(G_n)|^{-2} \) we can neglect the case that \( \tilde{Z}_i < t/20 \). Hence we can make the following estimate: for every \( h \in V(H_i) \) we have that for all \( 1 < i \leq n \) and \( x \in V(U_i) \)
\[
\sum_{u \in V(W_i)} p_{u, h} \leq C_4 \sum_{u \in V(W_i)} P_x[X_{t/2} \in (u, h) | T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2, \tilde{Z}_i \geq t/20]. \]
Note that the conditioning on \( T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2 \), does not affect the projection of the walk onto its \( H_i \) co-ordinate, given the number of steps made on that co-coordinate. Also observe that conditioned on \( T_{V(H)} > t/2, T_{R_i \cup R_{i-1}} > t/2 \), the aforementioned projection (up to time \( t/2 \), viewed in times in which the \( H_i \) co-ordinate changes) is itself a random walk on the expander \( H_i \).
This, in conjunction with the fact that the $L_\infty$ mixing time of SRW on $H_i$ is at most $t/20$ (to see this, use (2.3), (3.10) and the fact that $|V(H_i)| = |R_i| \leq 2^{2n-2^{2n+1}+n-i}$, implies that

$$
\sum_{u \in V(W_i)} P_x[X_{t/2} \in (u,h) \mid T_{V(H)} > t/2, T_{R_i\cup R_{i-1}} > t/2, \tilde{Z}_i \geq t/20] \leq C_4/|V(H_i)|.
$$

Combining the last two inequalities, we get that for every $h \in V(H_i)$, for all $1 < i \leq n$ and $x \in V(U_i)$

$$
\sum_{u \in V(W_i)} p_{u,h} \leq C_4/|V(H_i)|. \tag{3.18}
$$

Since $\sum_{i=1}^{k} x_i$ subject to the constraints $x_i \in \mathbb{R}_+$, for all $i$, and $\sum_{i=1}^{k} x_i = a > 0$ is at most $a^\ell$, provided that $\ell \geq 1$, using (3.18) we get that for all $1 < i \leq n$ and $x \in V(U_i)$

$$
\|P_x[X_{t/2} \in \cdot \mid T_{V(H)} > t/2, T_{R_i\cup R_{i-1}} > t/2] - \pi(\cdot)\|_{\ell,\pi} \leq C_5 \sum_{h \in V(H_i)} 1/|V(G_n)| \sum_{u \in V(W_i)} p_{u,h}/(1/|V(G_n)|)^\ell
$$

$$
\leq C_6 |V(G_n)|^{-1} \sum_{h \in V(H_i)} |V(H_i)|^{-\ell} = C_6 \left(\frac{|V(G_n)|}{|V(H_i)|}\right)^{\ell-1} \leq C_7 2^{(\ell-1)2^{2n+i+2}},
$$

as desired. \hfill \square

Using (3.14), in order to prove (3.13) it suffices to show that for all $i$ and all $x \in R_i$ we have (provided that $\ell - 1$ is sufficiently small and $C$ is sufficiently large) that

$$
P_x[T_{V(H)} > t/2]||P_x^{t/2}[\cdot \mid T_{V(H)} > t/2] - \pi(\cdot)\|_{\ell,\pi} \leq 1/8. \tag{3.20}
$$

The case $i = 1$ can be treated separately in a similar manner to the analysis below. Hence below we assume that $i > 1$ and $x \in R_i$.

We argue that (provided that $C$ is sufficiently large) for all $i > 1$ and $x \in R_i$,

$$
c_3 P_x[T_{V(H)} > \frac{t}{2}] \leq P_x[T_{R_{i-1}} \leq T_{V(H)}] \leq \frac{|BMP_i|}{|BMP_i \cup BMP_1|} = \frac{|BMP_i|}{2^{2^{2n+i+1}}} \leq 2^{-c_2 2^{2n+i}}. \tag{3.21}
$$

The proof of the first inequality in (3.21) is similar to the proof of (3.15) and hence omitted. Finally, using a coupling argument, the second inequality in (3.21) can be proven by noting that started from $a_i$, the hitting distribution of $GMP_i \cup BMP_1 = \mathcal{L}_{2^{2n+i-2} \times 0}(T_i) = \mathcal{L}_{2^{2n+i-2} \cup \mathcal{B}_{0,i}}$ w.r.t. simple random walk on $T_i$ is the uniform distribution on this set. We omit the details.

It is easy to verify that starting at $x \in R_i$, the walk will make by time $t/2$ at least $2^{5n}$ consecutive steps on either $U_i$ or $U_{i+1}$ w.p. at least $1 - C_6 |V(G_n)|^{-2}$. On this event, as before, the walk will make by time $t/2$ at least $2^{5n-6}$ steps on either $H_i$ or on both of $H_{i+1}$ and $BE_{i+1}$, (in one of its visits to either $U_i$ or $U_{i+1}$, resp., in which it stays in it for at least $2^{5n}$ steps) w.p. at least $1 - C_6 |V(G_n)|^{-2}$. Using the same reasoning as in (3.18)-(3.19), we get that for all $i \geq 1$ and $x \in R_i$,

$$
||P_x^{t/2}[\cdot \mid T_{V(H)} > t/2] - \pi(\cdot)\|_{\ell,\pi} \leq C_7 2^{(\ell-1)2^{2n+i+2}}. \tag{3.22}
$$

We leave the verification of (3.22) as an exercise. Using (3.21), (3.22), we get that for all $i > 1$ and $x \in R_i$ (provided that $\ell - 1$ is sufficiently small and $C$ is sufficiently large),

$$
P_x[T_{V(H)} > t/2]||P_x^{t/2}[\cdot \mid T_{V(H)} > t/2] - \pi(\cdot)\|_{\ell,\pi} \leq 1/8. \tag{3.23}
$$

This concludes the proof of $\tau_2(G_n) \leq \lceil C 2^{18n} \rceil$. \hfill \square
4. Proof of Proposition 1.5

(1) Torus $\mathbb{Z}_n^d$: (the argument is almost identical to that in [6]) The lower bound follows from the fact that the inverse of the spectral gap (which is robust and up to a $\log 2$ factor, bounds $\tau_1$ from below) is at least $c_d n^2$, while the upper bound can be deduced from the Morris and Peres’ Evolving sets bound [13], which is also robust. (2) The giant component of an Erdős-Rényi supercritical random graph $G(n, c/n)$: the existence of paths of length $\Theta(\log n)$ implies that the inverse of the Log-Sobolev constant, is at least $c \log n$ (this can be derived using Lemma 4.2 in [7]). This provides a robust lower bound on $\tau_\infty$. The upper bound follows from the fact that in this case the spectral gap satisfies $\lambda := \Theta(1/\log^2 n)$ and so $\tau_\infty \leq C \lambda^{-1} \log n$ (since $c$ is fixed, the average degree is uniformly bounded). Finally, for example (3) and (4) the argument is identical to that in [6] and hence omitted.

Acknowledgements

The author is grateful to Gady Kozma and Yuval Peres for useful discussions. The author would like to thank them and the anonymous referee for making suggestions which improved the presentation of the note.

References

[1] Addario-Berry L., and Roberts M. I. (2017). Mixing time bounds via bottleneck sequences. *Journal of Statistical Physics.* 1–27. DOI:10.1007/s10955-017-1917-5
[2] Aldous D., and Fill. J. (2002). Reversible Markov chains and random walks on graphs. Unfinished manuscript. Available at [http://www.stat.berkeley.edu/~aldous/RWG/book.html](http://www.stat.berkeley.edu/~aldous/RWG/book.html).
[3] Basu R., Hermon J., and Peres Y. (2017). Characterization of cutoff for reversible Markov chains. *Ann. Probab.* 45, no. 3, 1448–1487. DOI:10.1214/16-AOP1090 MR3650406
[4] Benjamini I. (1991). Instability of the Liouville property for quasi-isometric graphs and manifolds of polynomial volume growth. *Journal of Theoretical Probability* 4 no. 3 631–637. DOI:10.1007/BF01210328 MR1115166
[5] Diaconis P., and Saloff-Coste L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability* 6 no. 3 695–750. DOI:10.1214/aoap/1034968224 MR1410112
[6] Ding J., and Peres Y. (2013). Sensitivity of mixing times. *Electronic Communications in Probability* 18, 61–6. DOI:10.1214/ECP.v18-2765 MR3141797
[7] Goel S., Montenegro R., and Tetali P. (2006). Mixing time bounds via the spectral profile. *Electron. J. Probability* 11 no. 1 1–26. DOI:10.1214/EJP.v11-300 MR2199053
[8] Hermon J., and Peres Y. (2017). A characterization of $L_2$ mixing and hypercontractivity via hitting times and maximal inequalities. *Probability Theory and Related Fields* 1–32. DOI:10.1007/s00440-017-0769-x
[9] Hermon J., and Peres Y. (2016). On sensitivity of mixing times and cutoff. ArXiv preprint [arXiv:1610.04357](https://arxiv.org/abs/1610.04357)
[10] Kozma G. (2007). On the precision of the spectral profile. *Latin American Journal of Probability and Mathematical Statistics* 3 321–329. MR2372888
[11] Levin D., and Peres Y. (2017). *Markov chains and mixing times.* Vol. 107. American Mathematical Society, Providence, RI. With contributions by Elizabeth L. Wilmer and a chapter by James G. Propp and David B. Wilson.
[12] Lubetzky E., and Sly A. (2011). Explicit expanders with cutoff phenomena. *Electronic Journal Probability* 16 no. 15 419–435. DOI:10.1214/EJP.v16-869 MR2774096
[13] Morris B., and Peres Y. (2005) Evolving sets, mixing and heat kernel bounds. *Probability Theory and Related Fields* 133 no. 2 245–266. DOI:10.1007/s00440-005-0434-7 MR2198701
[14] Peres Y., and Sousi P. (2015) Mixing times are hitting times of large sets. *J. Theoret. Probab.* 28 no. 2 488–519. DOI:10.1007/s10959-013-0497-9 MR3370663.
[15] Pittet C., and Saloff-Coste L. (2000) On the stability of the behavior of random walks on groups. *Journal of Geometric Analysis* 10 no. 4 713–737. DOI:10.1007/BF02921994 MR1817783
[16] Saloff-Coste L. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*. Lecture notes in Maths., Vol. **1665** Springer Berlin, 301–413. DOI:10.1007/BFb0092621 MR1490046