Sensitivity of mixing times of Cayley graphs

Jonathan Hermon and Gady Kozma

Abstract. We show that the total variation mixing time is not quasi-isometry invariant, even for
Cayley graphs. Namely, we construct a sequence of pairs of Cayley graphs with maps between them
that twist the metric in a bounded way, while the ratio of the two mixing times goes to infinity.
The Cayley graphs serving as an example have unbounded degrees. For non-transitive graphs, we
construct *bounded degree* graphs for which the mixing time from the worst starting point for one
graph is asymptotically smaller than the mixing time from the best starting point of the random
walk on a network obtained by increasing some of the edge weights from 1 to $1 + o(1)$.

1 Introduction

There are numerous works aiming at sharp geometric bounds on the mixing time of
a finite Markov chain. Examples include Morris and Peres' evolving sets bound [39],
expressed in terms of the expansion profile, and the related bound by Fountoulakis and
Reed [19]. The sharpest geometric bounds on the uniform (a.k.a. L_∞) mixing time are
given in terms of the log-Sobolev constant (see [16] for a survey on the topic) and the
spectral profile bound, due to Goel et al. [20]. Both determine the uniform mixing
time up to a multiplicative factor of order $\log \log[1/\min \pi(x)]$, where throughout
π denotes the stationary distribution (see [16, 32]). The reader is not familiar with
mixing time definitions can find them in Section 2.2. Other notions and definitions
used below can be found in Sections 1.5 and 1.7.

This type of geometric bounds on mixing times are robust under bounded per-
turbations of the edge weights, and in the bounded degree setup, also under quasi-
isometries. That is, changing some of the edge weights by at most some multiplicative
constant factor can change these geometric bounds only by some corresponding con-
stant factor. A natural question, with obvious implications to the potential sharpness
of such geometric bounds, is whether mixing times are themselves 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 simple random walks (SRWs)
on two quasi-isometric graphs of bounded degree be? Different variants of this
question were asked by various authors such as Pittet and Saloff-Coste [42, Section 6],
Diaconis and Saloff-Coste [16, p. 720], and Aldous and Fill [3, Open Problem 8.23].

Ding and Peres [17] 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.[1] In [24], a similar example is constructed in which the uniform mixing time is sensitive under bounded perturbations of the edge weights, as well as under a quasi-isometry. All these examples are based on the "perturbed tree" example of T. Lyons [37] (simplified by Benjamini [7]). In particular, they are highly non-transitive, and a priori it appears as if what makes such examples work could not be imitated by a transitive example. It remained an open problem to determine whether the total variation mixing time of random walk on vertex-transitive graphs is robust under small perturbations. This was asked by Ding and Peres [17, Question 1.4] (see also [32, p. 3] and [42, Section 6]). In this paper, we give a negative answer to this question, even when the small perturbation preserves transitivity.

We denote the group of permutations of n elements by $\mathfrak{S}_n$. Recall that a transposition is an element of $\mathfrak{S}_n$ which exchanges two values and keeps all the rest fixed.

Theorem 1.1   *There exist a pair of sequences of sets of transpositions S_n and S'_n such that the Cayley graphs $\mathrm{Cay}(\mathfrak{S}_n, S_n)$ and $\mathrm{Cay}(\mathfrak{S}_n, S'_n)$ are $(3,0)$-quasi-isometric and*

$$t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S'_n)) \gtrsim t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n)) \log\log\log|\mathfrak{S}_n|.$$

Further, $S_n \subset S'_n \subset S_n^3 := \{xyz : x, y, z \in S_n\}$.

Of course, $\log\log\log|\mathfrak{S}_n| \asymp \log\log n$. We formulated the theorem in this way because the size of the group is the more natural object in this context. Let us remark that probably the ratio of mixing time in our example is indeed $\asymp \log\log\log|\mathfrak{S}_n|$, but for brevity, we prove only the lower bound.

The mixing times in Theorem 1.1 are the total variation ones. In what comes, whenever we write mixing time without mentioning the metric, it is always the total variation mixing time. The behavior described in Theorem 1.1 cannot occur for the uniform mixing times which in the transitive setup is quasi-isometry invariant (see Theorem 2.5).

1.1   Variations on a theme

A related question, asked by Itai Benjamini (private communication) is whether there exists some absolute constant $C > 0$ such that for every finite group G for all two symmetric sets of generators S and S' such that $S \subset S'$ we have that the mixing time of SRW on the Cayley graph of G with respect to S' is at most $C\frac{|S'|}{|S|}$ times the mixing time of SRW on the Cayley graph of G with respect to S (a set S is called *symmetric* if $S = S^{-1} := \{s^{-1} : s \in S\}$). Our example also disproves this. In fact, $S \subset S' \subseteq S^3$ and $|S'| - |S| \le \sqrt{|S|}$, where $S^i := \{s_1 \cdots s_i : s_1, \ldots, s_i \in S\}$ for $i \in \mathbb{N}$. The definition of an (a, b)-quasi-isometry (see Section 1.5) gives that if $S \subseteq S' \subseteq S^i$ then $\mathrm{Cay}(G, S)$ and $\mathrm{Cay}(G, S')$ are $(i, 0)$-quasi-isometric.

---

[1]Their construction was refined by Hermon and Peres in [26, Theorem 3], so that the mixing time changes by an order of $\log|V|$, which is optimal. The same paper contains various additional results concerning sensitivity of mixing times and of the cutoff phenomenon under small changes to the geometry of the chain.

The reason that $|S'| - |S| \le \sqrt{|S|}$ is explained in the proof sketch section below—both share a complete graph on some set K with $|K| \asymp n$. Hence, this complete graph has an order of $\asymp n^2$ edges, and there are only $o(n)$ additional edges. We could have increased S_n by including in it all $|K|!$ permutation of the elements in K, while keeping $S'_n \setminus S_n$ the same set (of size $o(n)$), thus making $\frac{|S'_n| - |S_n|}{|S_n|}$ tremendously smaller.

We will also be interested in weighted versions of the problem, as these allow us to define "weak" perturbations in a natural way. Let Γ be a group, and let $W \coloneqq (w(s) : s \in S)$ be symmetric weights (i.e., $w(s) = w(s^{-1})$) such that the support of W generates Γ. The discrete-time lazy random walk on Γ with respect to W is the process with transition probabilities $P(g, g) = 1/2$ and $P(g, gs) = \frac{w(s)}{2 \sum_{r \in S} w(r)}$ for all $g, s \in \Gamma$. We denote its TV (total variation) mixing time by $t_{\mathrm{mix}}(\mathrm{Cay}(\Gamma, W))$. In continuous time, let $R \coloneqq (r(s) : s \in S)$ be symmetric rates. The continuous-time random walk on Γ with respect to R is the process that has infinitesimal transitions rates $r(s)$ between g and gs for all $g, s \in \Gamma$. Denote its mixing time by $t_{\mathrm{mix}}(\mathrm{Cay}(\Gamma, R))$. As in the unweighted case, due to the group symmetry the invariant distribution is uniform and the TV distance between it and the distribution of the walk at some given time is independent of the initial state.

Recall that $\mathfrak{S}_n$ is the symmetric group (the group of permutations of n elements). The following is the promised weighted version of our main result.

Theorem 1.2 *For every $f : \mathbb{N} \to [1, \infty)$ satisfying that $1 \ll f(n) \le \log\log\log n$, there exist a sequence $(S_n)_{n=3}^{\infty}$ of sets of transpositions $S_n \subset \mathfrak{S}_n$ and a sequence of weights $(W_n)_{n=3}^{\infty}$, such that $W_n = (w_n(s))$ is supported on S_n and satisfies that $1 \le w_n(s) \le 1 + (f(n!)/\log\log n)^{1/4}$ for all $s \in S_n$, and such that*

$$(1.1) \qquad t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, W_n)) \gtrsim t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n)) f(n!).$$

Similarly, in continuous time, if we set $R_n = W_n$ (for the above W_n), we get that

$$(1.2) \qquad t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, R_n)) \gtrsim t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n)) f(n!).$$

We remark that the power $1/4$ is not optimal (it was not a priority for us to optimize it). As before, $|S_n| \asymp n^2$.

1.2 A non-transitive example

Our third result shows that if one is willing to consider non-transitive instances, then indeed one can have a bounded degree example whose (usual worst-case) mixing time is of strictly smaller order than the mixing time starting from the best initial state (i.e., the one from which the walk mixes fastest) after a small perturbation. In all previous constructions of graphs with a sensitive mixing time, there was a large set that starting from it, the walk mixes rapidly both before and after the perturbation, and the mixing time is governed by the hitting time of this set (which is sensitive by construction). In particular, the mixing time started from the best initial state is not sensitive.

Let G be a connected graph. Let $W = (w(e) : e \in E(G))$ be positive edge weights. Consider the *lazy random walk* $(X_k)_{k=0}^{\infty}$ on G, i.e., the process with transition probabilities $P(x, y) = \frac{w(xy)}{2 \sum_z w(xz)}$ and $P(x, x) = \frac{1}{2}$ for all neighboring $x, y \in G$.

For $x \in G$, we define the mixing time starting from x by

$$t_{\mathrm{mix}}(G, W, x) := \min\{k : \|\mathbb{P}_x(X_k = \cdot) - \pi\|_{\mathrm{TV}} \le 1/4\}.$$

With this definition, the usual mixing time $t_{\mathrm{mix}}(G, W)$ (see Section 2.2) is equal to $\max_x t_{\mathrm{mix}}(G, W, x)$.

Theorem 1.3 *There exist a sequence of finite graphs $L_n = (V_n, E_n)$ of diverging sizes and uniformly bounded degree (i.e., $\sup_n \max_{v \in V_n} \deg v < \infty$) and a sequence of some symmetric edge weights $W_n = (w_n(e) : e \in E_n)$ such that $1 \le w_n(e) \le 1 + \delta_n$ for all $e \in E_n$ and such that*

$$(1.3) \qquad\qquad \max_{x \in V_n} t_{\mathrm{mix}}(L_n, 1, x) \le \delta_n \min_{x \in V_n} t_{\mathrm{mix}}(L_n, W_n, x)$$

for some $\delta_n \to 0$.

It follows from Theorem 1.3 that the average TV mixing time, by which, we mean $\inf\{t : \sum_x \pi(x)\|\mathbb{P}_x(X_t = \cdot) - \pi\|_{\mathrm{TV}} \le 1/4\}$, can be sensitive to perturbations. This is in contrast with the average L_2 mixing time (see Section 2.2). This gives a negative answer to a question of Addario-Berry (private communication).

As in Theorem 1.1, the change in the order of the mixing time in Theorem 1.3 (the inverse of the δ_n in (1.3)) is $o(\log\log\log |V_n|)$. If we replace the condition $w_n \le 1 + \delta_n$ with $w_n \le 1 + c$, then the change in the order of the mixing time can be as large as $\log\log\log |V_n|$.

Let us quickly sketch the construction of Theorem 1.3 (full details are in Section 4). Let n be some number, and let S_n be the set of transpositions from Theorem 1.2. Let H be a large, fast mixing graph, and let A be some subset of the vertices of H with $|A| = |S_n|$ and with the vertices of A far apart from one another. The graph L of Theorem 3 has as its vertex set $\mathfrak{S}_n \times H$ (we are using here the same notation for the graph and its set of vertices). We choose the edges of L such that random walk on L has the following behavior. Its H projection is just SRW on the graph H. Its $\mathfrak{S}_n$ projection is also SRW on $\mathrm{Cay}(\mathfrak{S}_n, S_n)$, but slowed down significantly. Any given transposition $s \in S_n$ can be applied only when a corresponding vertex of A is reached in the second coordinate. The perturbation goes by perturbing only the $\mathfrak{S}_n$ projection. We defer all other details to Section 4.

1.3 A proof sketch

We will now sketch the proof of our main result, Theorem 1.1 (the proof of Theorem 1.2 is very similar). Readers who intend to read the full proof can safely skip this section.

Random walk on $\mathrm{Cay}(\mathfrak{S}_n, S_n)$ with S_n composed of transpositions is identical to the *interchange process* on the graph G which has n vertices and $\{x, y\}$ is an edge of G if and only if the transposition $(x, y) \in S_n$. Hence, we need to construct two graphs G and G' on n vertices, estimate the mixing time of the two interchange processes and show that the corresponding Cayley graphs are quasi-isometric.

Our two graphs have the form of "gadget plus complete graph." Namely, there is a relatively small part of the graph D which we nickname "the gadget" and all vertices in $G \backslash D$ are connected between them. While D and the corresponding D' in G' will

be small (we will have $|D| = |D'|$), they dominate the mixing time of the interchange process.

To describe the gadget, let $u \in \mathbb{N}$ and $\varepsilon \in (0, \frac{1}{2})$ be some parameters. The gadget will have u "stages" $H_1, \ldots, H_u$ (the gadget is almost $\cup_{i=1}^u H_i$ but not quite). We obtain each H_i by "stretching" the edges of some graph H_i' which is a union of binary trees of depth $s_i := 4^{i-1}u$ (note that H_i' has the depth exponential in i and hence has volume doubly exponential in i). To get H_i, replace each edge of H_i' with a path of length $\ell_i := 2^{u+1-i}$. Namely, for each edge $\{x, y\}$ of H_i', we add $\ell_i - 1$ new vertices (denote them by $v_1, \ldots, v_{\ell_i-1}$, and denote also $v_0 = x$ and $v_{\ell_i} = y$) and connect v_j to v_{j+1} for all $j \in \{0, \ldots, \ell_i - 1\}$; and remove the edge $\{x, y\}$.

We still need to explain how many trees are in each H_i and how they are connected to one another and to the rest of the graph. For this, we need the parameter ε, which at this point can be thought of as a sufficiently small constant. For each of the vertices in each of the trees (before stretching), we label the children arbitrarily "left" and "right." For each leaf $x \in H_i$, we define $g(x)$ to be the number of left turns in the (unique) path from the root to x. We now let

$$(1.4) \qquad B_i := \left\{ x \text{ leaf of } H_i : g(x) > \left(\frac{1}{2} + \varepsilon \right) s_i \right\}.$$

The sets B_i are used twice. First, we use them to decide how many trees will be in each H_i. For $i = 1$, we let H_1 be one tree. For every $i > 1$, we let H_i have $|B_{i-1}|$ trees, and identify each point of B_{i-1} with one of the roots of one of the trees in H_i. Second, we use the B_i to connect the H_i to the complete graph. Every leaf of H_i which is not in B_i is identified with a vertex of the complete graph (the complete graph K will be of size $n - o(n)$, much larger than $\cup_{i=1}^u H_i$ which will be of size $O(n^{1/4})$, and so most of the vertices of K are not identified with a vertex of the gadget). This terminates the construction of G (see Figure 1). Experts will clearly notice that this is a variation on the perturbed tree idea. In other words, while the perturbed tree itself (as noted above) is highly non-transitive, one can use it as a basis for transitive example by examining the interchange process on it.

The graph G' is almost identical, the only difference is that in each path corresponding to a left turn we add short bridges. Namely, examine one such path and denote its vertices $v_0, \ldots, v_{\ell_i}$ as above. Then in G', we add edges between v_{2j} and v_{2j+2} for all $j \in \{0, \ldots, \ell_i/2 - 1\}$.

Why this choice of parameters? It is motivated by a heuristic that for such graphs, namely, a gadget connected to a large complete graph, the mixing time of the interchange process is the time *all* particles have left the gadget (they do not have to all be outside the gadget at the same time, it is enough that each particle left the gadget at least once by this time). See Section 1.4 for some context for this heuristic. Thus, we are constructing our H_i such that the time that it takes all particles to leave H_i is approximately independent of i. Indeed, the time a particle takes to traverse a single stretched edge is approximately $\ell_i^2 \asymp 4^{u-i}$ while each tree of H_i has depth $4^{i-1}u$ (in the sense that this is the depth of the tree before its edges have been stretched) and the particle has to traverse all levels of H_i, so it exits H_i after time approximately $4^{u-1}u$, which is independent of i. And this holds for all particles simultaneously because the probability that a particle takes $\lambda \cdot 4^{u-1}u$ time to traverse the tree (for some
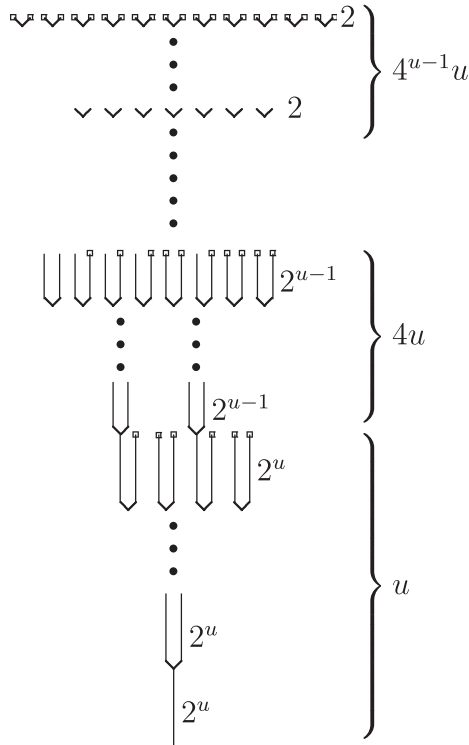
Figure 1: The gadget. Vertices marked with small squares actually belong to the complete graph rather than to the gadget.

$\lambda > 1$) is exponentially small in the number of layers $4^{i-1}u$, and hence that would not happen to any of the particles in the tree, which has approximately $2^{4^{i-1}u}$ particles, if λ is sufficiently large. In the roughest possible terms, the growing height of the trees is dictated by the growing number of vertices (which must grow because H_i has many more roots than H_{i-1}, since each $x \in B_{i-1}$ is a root of H_i) while the decreasing stretching balances the growing height to get approximately uniform expected exit time. The only exception is H_1, whose height is not dictated by the number of roots (clearly, as there is only one), but by the stretching.

With the definitions of G and G' done, estimating the mixing times is relatively routine, so we make only two remarks in this quick sketch. How do we translate the fact that all particles visited the complete graph into an upper bound on the mixing time? We use a *coupling* argument. We couple two instances σ and σ' of the interchange process (in continuous time) using the same clocks and letting them walk identically unless $\sigma(x) = \sigma'(y)$ for an edge $\{x, y\}$ that is about to ring, in which case we apply the transposition to exactly one of σ or σ', reducing the number of disagreements (this coupling involves a standard trick of doubling the rates, and censoring each step with probability 1/2). The fact that the complete graph is much larger and has many more edges simplifies our analysis (the reader can find the details of the coupling in Section 3.2).

The lower bound for the mixing time on G' uses the standard observation that adding those edges between v_{2j} and v_{2j+2} makes the left turn more likely to be taken than the right turns, transforming B_i from an atypical set (with respect to the hitting distribution of the leaf set of H_i) to a typical one, and hence, the particle that started at the root of H_1 has high probability to traverse *all* H_i before entering the complete graph for the first time. This, of course, takes it $4^{u-1}u^2$ time units (compare to the mixing time bound of $4^{u-1}u$ for the interchange process on G). Of course, the mixing time of the interchange process on G' is also bounded by the time that all particles leave the gadget, but we found no way to use this. We simply bound the time a single particle leaves the gadget and get our estimate.

1.4 The mixing time of the interchange process

Since our proof revolves around estimating the mixing time of the interchange process on some graph, let us spend some time on a general discussion of this topic. We first mention some conjectures relating the mixing time of the interchange process on a finite graph G to that of $|G|$ independent random walks on G.

Given a finite graph $G = (V, E)$ and edge rates R, the corresponding n-fold product chain is the continuous-time Markov chain on V^n satisfying that each coordinate evolves independently as a random walk on G with edge rates R. This is a continuous-time walk on the n-fold Cartesian product of G with itself, whose symmetric edge rates R_n are given by

$$R_n((v_1, \ldots, v_n), (v_1, \ldots, v_{k-1}, v'_k, v_{k+1}, \ldots, v_n)) \coloneqq R(v_k, v'_k)$$

for all $v_1, \ldots, v_n, v'_k \in V$ and $k \in [n]$. We shall refer to this Markov chain as n independent random walks on G with edge rates R and denote its (TV) mixing time by $t_{\text{mix}}(n \text{ independent RWs on } G, R)$. As usual, the mixing time is defined with respect to the worst starting tuple of n points, which turns out to be when they all start from the worst point for a single walk on G with edge rates R.

Oliveira [40] conjectured that there exists an absolute constant $C > 0$ such that the TV mixing time of the interchange process on an n-vertex graph G with rates R, i.e., $t_{\text{mix}}(\text{Cay}(\mathfrak{S}_n, R))$, is at most $Ct_{\text{mix}}(n \text{ independent RWs on } G, R)$. See [30, Conjecture 2] and [29, Question 1.12] for two different strengthened versions of this conjecture. See [30] for a positive answer for high dimensional products.

For the related *exclusion* process, some progress on Oliveira's conjecture is made in [29]. Returning to the interchange process, in the same paper, the following more refined question is asked [29, Question 1.12]: Is $t_{\text{mix}}(\text{Cay}(\mathfrak{S}_n, R))$ equal up to some universal constants to the mixing time of n independent random walks on (G, R) starting from n *distinct* locations? (see [29] for precise definitions). We see that our result is related to finding some graphs G such that the mixing time of $|G|$ independent SRW with edge rates 1 on G, starting from distinct initial locations, is sensitive under small perturbations. In fact, the graphs we construct in this paper satisfy this property too, but in the interest of brevity, we will not prove this claim (the proof is very similar to the one for the interchange process we do provide). This conjectured relation between the exclusion process and independent random walks is behind the heuristic we employed (and mentioned in Section 1.3) to construct our example.

As we now explain, if we did not require the initial locations to be distinct (as is the case in Oliveira's conjecture) such sensitivity could not occur. It is easy to show (e.g., [29]) that when $|G| = n$,

$$\frac{1}{4} t_{\text{rel}}(G, R) \log n \leq t_{\text{mix}}(n \text{ independent RWs on } G, R) \leq 4 t_{\text{rel}}(G, R) \log n,$$

where $t_{\text{rel}}(G, R)$ is the *relaxation time* of (G, R), defined as the inverse of the second smallest eigenvalue of $-\mathcal{L}$, where $\mathcal{L}$ is the infinitesimal Markov generator of the walk (G, R). The relaxation time is robust under small perturbations (see Section 2.1), and hence so is $t_{\text{mix}}(n \text{ independent RWs on } G, R)$. Our result that the mixing time is sensitive does not contradict Oliveira's conjecture, as he conjectured only an upper bound (which, in our case, is sharp for neither S_n nor S'_n).

Loosely speaking, in order to make the mixing time of n independent random walks starting at distinct locations of smaller order than (the robust quantity) $t_{\text{rel}}(G, R) \log n$ it is necessary that the eigenvector corresponding to the minimal eigenvalue of $-\mathcal{L}$ be localized on a set of cardinality $n^{o(1)}$. This is a crucial observation in tuning the parameters in our construction, which explains why for smaller areas of the graph (namely, H_i with small index i) we "stretch" edges by a larger factor. This is the opposite of what is done in [24].

Lastly, we comment that in contrast with a single random walk, in order to change the mixing time of n independent random walks, starting from n distinct initial locations, it does not suffice for the perturbation only to change the typical behavior of the walk, but rather it is necessary that it significantly changes the probabilities of some events in some sufficiently strong quantitative manner. See [24] for a related discussion, about why it is much harder to construct an example where the uniform mixing time is sensitive than it is to construct one where the TV mixing time is sensitive.

1.5 Quasi-isometries and robustness

Since we hope this note will be of interest to both group theory and Markov chain experts, let us take this opportunity to compare two similar notions related to comparison of the geometry of two graphs or of two reversible Markov chains. The first is the notion of quasi-isometry which is more geometric in nature. The second is the notion of robustness which is more analytic. In particular, we are interested in properties which are preserved by these notions.

This discussion is an important part of the background, but let us advise the readers that it is not necessary to appreciate our results, as they apply in both cases. For example, Theorem 1.1 shows that the mixing time is neither quasi-isometry invariant nor robust.

A quasi-isometry (defined first in [23]) between two metric spaces X and Y is a map $\phi : X \to Y$ such that for some numbers (a, b) we have

$$\forall u, v \in X \qquad \frac{d(u, v) - b}{a} \leq d(\phi(u), \phi(v)) \leq a d(u, v) + b,$$

where d denotes the distance (in X or in Y, as appropriate). Further, we require that, for every $y \in Y$, there is some $x \in X$ such that $d(\phi(x), y) \leq a + b$. We say that X and Y are

(a, b) quasi-isometric if such a ϕ exists. (Our choice of definition is unfortunately only partially symmetric. If $\phi : X \to Y$ is an (a, b) quasi-isometry then one may construct a quasi-isometry $\psi : Y \to X$ with the same a but perhaps with a larger b.)

For a property of random walk that is defined naturally on infinite graphs, we say that it is *quasi-isometrically invariant* if whenever G and H are two quasi-isometric infinite graphs, the property holds for G if and only if it holds for H (the graphs are made into metric spaces with the graph distance). Examples include a heat kernel on-diagonal upper bound of polynomial type [12], an off-diagonal upper bound [21], and a corresponding lower bound [10, 22]. A particularly famous example is the Harnack inequality [5]. For a quantitative property of random walk naturally defined on finite graphs, such as the mixing time, one says that it is invariant to quasi-isometries if, whenever G and H are (a, b)-quasi-isometric, the property may change by a constant that depends only on a and b and not on other parameters. Similar notions may be defined for Brownian motion on Riemannian manifolds, and one may even ask questions like "if a manifold M is quasi-isometric to a graph G and Brownian motion on M satisfies some property, does random walk on G satisfy an equivalent property?" and a number of examples of this behavior are known.

The notion of *robustness* does not have a standard definition, and in particular, the definitions in [17] and [24] differ (and also differ from the definition we will use in this paper). Nevertheless, they all have a common thread: a definition for Markov chains that implies that the property in question is preserved under quasi-isometry of graphs of bounded degree, but that makes sense also without any a priori bound on the transition probabilities. Here, we will use the following definition. Let $\mathcal{M}$ be the set of finite state Markov chains. We say that a $q : \mathcal{M} \to [0, \infty]$ is *robust* if, for any $A \in (0, 1]$, there exists some $K \in (0, 1]$ such that the following holds. Assume M and M' are two irreducible reversible Markov chains on the same finite state space V with stationary distributions π and π' and transition matrices P and P' satisfying

$$
(1.5) \qquad \begin{aligned}
\forall\, x \in V, &\qquad A\pi'(x) \leq \pi(x) \leq \tfrac{1}{A}\pi'(x), \\
\forall\, f \in \mathbb{R}^V, &\quad A\mathcal{E}'(f, f) \leq \mathcal{E}(f, f) \leq \tfrac{1}{A}\mathcal{E}'(f, f),
\end{aligned}
$$

where $\mathcal{E}(f, f)$ and $\mathcal{E}'(f, f)$ are the corresponding Dirichlet forms, namely,

$$
\mathcal{E}(f, g) := \frac{1}{2} \sum_{u, v \in V} \pi(u) P(u, v)(f(u) - f(v))(g(u) - g(v))
$$

and similarly for $\mathcal{E}'$. Then $q(M) \geq Kq(M')$.

We also define robustness for Markov chains in continuous time, and in this case, we replace $P(u, v)$ above with $\mathcal{L}(u, v)$ which is the infinitesimal rate of transition from u to v, but otherwise the definition remains the same.

If P and P' are SRWs on (a, b) quasi-isometric graphs with the same vertex set (with the quasi-isometry being the identity), whose maximal degrees are at most D, then (1.5) holds with some A depending only on (a, b, D) [14]. Thus, a robust quantity is also quasi-isometry invariant between graphs of bounded degree on the same vertex set.

Each notion has its advantages and disadvantages relative to the other notion. Quasi-isometry has the flexibility that the spaces compared need not be identical or

even of the same type, indeed the fact that a Lie group (a continuous metric space, indeed a manifold) is quasi-isometric to any cocompact lattice of it (a discrete metric space) plays an important role in group theory. Robustness has the advantage that unbounded degrees are handled seamlessly.

Returning to our results, since the examples of our Theorem 1.1 are not of bounded degree, it is natural to ask if they satisfy a comparison of Dirichlet form of the form (1.5). In fact, this is true because in said examples our pair of sets of generators S_n and S'_n (from the statement of Theorem 1.1) satisfy for all n that $S_n \subset S'_n$ and that any $s' \in S'_n \backslash S_n$ can be written as $s_1(s')s_2(s')s_3(s') \in S_n^3 = \{xyz : x, y, z \in S_n\}$ in a manner satisfying that

$$(1.6) \qquad \max_{s \in S_n} \sum_{s' \in S'_n \backslash S_n} \sum_{i=1}^{3} \mathbb{1}\{s_i(s') = s\} \le 2.$$

It is standard and not difficult to see that (1.6) implies the comparison of Dirichlet forms condition (1.5) (see, e.g., [8, Theorem 4.4]). Thus, the examples of Theorem 1.1 also satisfy (1.5) with A being a universal constant. We remark that, in general, $S \subset S' \subseteq S^3$ is sufficient for deriving (1.5) only with an A that may depend on $|S'|$.

1.6 Remarks and open problems

We start with a remark on the Liouville property problem, a problem which for us was a significant motivation for this work. An infinite graph with finite degrees is called Liouville if every bounded harmonic function is constant (a function f on the vertices of a graph is called harmonic if $f(x)$ is equal to the average of f on the neighbors of x for all x).

An open problem in geometric group theory is whether the Liouville property is quasi-isometry invariant in the setup of Cayley graphs (and, in the spirit of the aforementioned question of Benjamini, whether it is preserved under deletion of some generators, possibly by passing to a subgroup, if the smaller set of generators does not generate the group). The problem of stability of the Liouville property is related to that of mixing times. Indeed, the example of Lyons [37] mentioned above which is a base for all previous examples for sensitivity was in fact an example for the instability of the Liouville property (for non-transitive graphs).

A result of Kaimanovich and Vershik (see [31] or [35, Chapter 14]) states that for Cayley graphs, the Liouville property is equivalent to the property of the walk having zero speed. Of course, our graphs being finite means there is no unique number to be designated as "speed," as in the Kaimanovich–Vershik setting. But still it seems natural to study the behavior of $\mathrm{dist}(X_t, 1)$ as a function of t, where X_t is the random walk, 1 is the identity permutation (and the starting point of the walker), and dist is the graph distance with respect to the relevant Cayley graph (with respect to S_n or S'_n, as the case may be). Interestingly, perhaps, the functions increase linearly for the better part of the process for both our S_n and S'_n, so we cannot reasonably claim we show some version of instability for the speed for finite graphs. (We will not prove this claim, but it is not difficult.)

Due to the relation to the Liouville problem, there is interest in reducing the degrees in Theorem 1.1. We note that since our S_n is a set of transpositions, we must

have $|S_n| \le \binom{n}{2} \asymp \left(\frac{\log|\mathfrak{S}_n|}{\log\log|S_n|}\right)^2$. As explained in the proof sketch section above, in our construction, there is a set $K \subset [n] := \{1, \ldots, n\}$ such that $|K| = n(1 - o(1))$ and all of the transpositions of the form (a, b) with $a, b \in K$ belong to S_n. Hence $|S_n| \asymp n^2$.

Let us mention two possible approaches to reduce the size of S_n. The first is to replace the complete graph over K in the construction by an expander. In this case, we will have $|S_n| \asymp n$. It seems reasonable that this approach works, but we have not pursued it. Let us remark at this point that the mixing time of the interchange process on an expander is not known, with the best upper bound being $\log^2 n$ [4] (see also [29]).

The second, and more radical, is to replace the $\binom{|K|}{2}$ transpositions corresponding to pairs from K by some number (say m, but importantly independent of n) of random permutations of the set K, obtained by picking m independent random perfect matchings of the set A, and for each perfect matching taking, the permutation that transposes each matched pair. (If $|K|$ is odd, we keep one random element unmatched.) Note that the Cayley graph is no longer an interchange process, and that approximately n^2 elements have been replaced by a constant number. The degree would still be unbounded because of the other part of the graph. Again, we did not pursue this approach. One might wonder if it is possible to replace the entire graph, not just K, by matchings, but this changes the mixing time significantly.

Question 1.1 *Can one take the set of generators S_n to be of constant size? (Certainly, not with transpositions but with general subsets of $\mathfrak{S}_n$, or with other groups). If not, can one take $|S_n|$ to diverge arbitrarily slowly as a function of $|G_n|$? Is there a relation between the degree of the graph and the maximal amount of distortion of the mixing time which is possible?*

A related question is the following.

Question 1.2 *Does the aforementioned question of Benjamini have an affirmative answer for bounded degree Cayley graphs?*

Here are two questions about the sharpness of our $\log\log\log$ term.

Question 1.3 *Does there exist a sequence of finite groups G_n of diverging sizes, and sequences of generators $S_n \subset S'_n \subseteq S_n^i$ for some $i \in \mathbb{N}$ (independent of n) for all n, such that $|S'_n| \lesssim |S_n|$ and*

$$(1.7) \qquad t_{\mathrm{mix}}(G_n, S'_n) \gtrsim t_{\mathrm{mix}}(G_n, S_n) \log|G_n|?$$

Question 1.4 *Can one have in the setup of Theorem* 1.3

$$(1.8) \qquad \min_{x \in V_n} t_{\mathrm{mix}}(G_n, W_n, x) \gtrsim t_{\mathrm{mix}}(G_n) \log|V_n|?$$

The opposite inequalities to (1.7) and (1.8) hold since the spectral gap is a quasi isometry invariant (see Section 2.1) and on the other hand determines the mixing time of a random walk on an n-vertex graph up to a factor $2 \log n$ (see, e.g., [33, Section 12.2]).

Our last question pertains to Theorem 2.5. It is inspired by a question of Itai Benjamini on the Liouville property in the infinite setting.

Question 1.5 *Let $G = (V, E)$ be a finite connected vertex-transitive graph. Is the uniform (or L^2) mixing time robust under bounded perturbations of the edge weights? (certainly, this is open only when the perturbation does not respect the transitivity). Likewise, does there exist some $C(a, b, d) > 0$ (independent of G) such that if the degree of G is d and G is (a, b)-quasi-isometric to G' (which, again, need not be vertex-transitive), then the uniform mixing times of the SRWs on the two graphs can vary by at most a $C(a, b, d)$ factor?*

We end the introduction with a few cases for which the mixing time is known to be robust. Robustness of the TV and L_∞ mixing times for all reversible Markov chains under changes to the holding probabilities (i.e., under changing the weight of each loop by at most a constant factor) was established in [41] by Peres and Sousi and in [28] by Hermon and Peres. Boczkowski, Peres, and Sousi [9] constructed an example demonstrating that this may fail without reversibility. Robustness of the TV and L_∞ mixing times for general (weighted) trees under bounded perturbations of the edge weights was established in [41] by Peres and Sousi and in [28] by Hermon and Peres. Robustness of TV mixing times for general trees under quasi-isometries (where one of the graphs need not be a tree, but is "tree-like" in that it is quasi-isometric to a tree) was established in [1] by Addario-Berry and Roberts.

In many cases, known robust quantities provide upper and lower bounds on the mixing time which are matching up to a constant factor. For example, in the torus $\{1, \ldots, \ell\}^d$ with nearest neighbor lattice edges, the mixing time is bounded above by the isoperimetric profile bound on the mixing time [39] and below by the inverse of the spectral gap. For a fixed d, both bounds are $\Theta(\ell^2)$. As both quantities are robust, we get that any graph quasi-isometric to the torus would have mixing time $\Theta(\ell^2)$, as in the torus. In fact, the same holds for bounded degree Cayley graphs of moderate growth (see, e.g., [29, Section 7]). Moderate growth is a technical condition, due to Diaconis and Saloff-Coste [15], who determined the order of the mixing time and the spectral gap for such Cayley graphs. Breuillard and Tointon [11] showed that for Cayley graphs of bounded degree this condition is equivalent in some precise quantitative sense to the condition that the diameter is at least polynomial in the size of the group.

Lastly, in a recent work [36], R. Lyons and White showed that for finite Coxeter systems increasing the rates of one or more generators does not increase the L_p distance between the distribution of the walk at a given time t and the uniform distribution for any $p \in [1, \infty]$. Since multiplying all rates by exactly a factor C changes the mixing time by exactly a factor $1/C$, this implies that the mixing time is robust under bounded permutations of the rates of the generators.

1.7 Notation

We denote $[n] = \{1, \ldots, n\}$. We denote by $\mathbb{P}_v$ probabilities of random walk starting from v, which should be a vertex of the relevant graph. We denote by c and C arbitrary positive universal constants which may change from place to place. We will use c for constants which are small enough and C for constants which are large enough. We will occasionally number them for clarity. We denote $X \lesssim Y$ for $X \leq CY$ and $X \asymp Y$ for $X \lesssim Y$ and $Y \lesssim X$. We denote $X \ll Y$ for $X = o(Y)$. Throughout, we do not distinguish

between a graph G and its set of vertices, denoting the latter by G as well. The set of edges of G will be denoted by $E(G)$.

2 Preliminaries

Definition 2.1 Let Γ be a finitely generated group, and let S be a finite set of generators satisfying $s \in S \iff s^{-1} \in S$. We define the Cayley graph of Γ with respect to S, denoted by $\mathrm{Cay}(\Gamma, S)$, as the graph whose vertex set is G and whose edges are

$$\{(g, gs) : g \in \Gamma, s \in S\}.$$

Definition 2.2 Let G be a weighted graph, and let $(r(e)_{e \in E(G)})$ be the weights. The interchange process on G is a continuous-time process in which particles are put on all vertices, all different. Each edge e of G is associated with a Poisson clock which rings at rate $r(e)$. When the clock rings, the two particles at the two vertices of e are exchanged.

The interchange process is always well defined for finite graphs (which is what we are interested in here). For infinite graphs, there are some mild conditions on the degrees and on r for it to be well defined. The interchange process on a graph G of size n is equivalent to a random walk in continuous time X_t on $\mathfrak{S}_n$ with the generators S of $\mathfrak{S}_n$ being all transpositions (xy) (in cycle notation) for all (xy) which are edges of G. The rate of the transposition (xy) is $r(xy)$. The position of the ith particle at time t is then $X_t^{-1}(i)$, where the inverse is as permutations.

2.1 Comparison of Dirichlet forms

Recall the condition (1.5) for comparison of Dirichlet forms. When it holds it implies a comparison of the eigenvalues: If $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and $0 = \lambda'_1 \le \lambda'_2 \le \cdots \le \lambda'_n$ are the eigenvalues of $I - P$ and $I - P'$, respectively, then under (1.5) (see, e.g., [3, Corollary 8.4] or [8, Corollary 4.1]),

$$(2.1) \qquad\qquad A\lambda_i \le \lambda'_i \le \lambda_i/A \quad \text{for all } i.$$

The same inequality holds for the eigenvalues of the Markov generators $-\mathcal{L}$ and $-\mathcal{L}'$ in continuous time (that is, $\mathcal{L}(x, y) = r(xy)$ for $x \neq y$ and $\mathcal{L}(x, x) = -\sum_{y: y \neq x} r(xy)$, where $r(xy)$ is the rate of the edge (xy) and with the convention that $r(xy) = 0$ if $xy \notin E$). The proof is the same as in the discrete case (see, again, [8, Corollary 4.1]). The quantity λ_2 is called the *spectral gap*. It follows that it is robust.

2.2 Mixing times

We now define the relevant notions of mixing: total variation, L_2 and uniform. We start with the total variation mixing time which is the topic of this paper, and which we will simply call *the* mixing time.

Definition 2.3 Let X_t be a Markov chain on a finite state space (in continuous or discrete time) with stationary measure π, and denote the probability that $X_t = y$ conditioned on $X_0 = x$ by $P_t(x, y)$. Then the mixing time is defined by

$$t_{\text{mix}} = \max_x \inf\{t \geq 0 : \|P_t(x, \cdot) - \pi\|_{\text{TV}} \leq \tfrac{1}{4}\}.$$

In discrete time, we often assume that X_t is lazy, i.e., that at each step, $\mathbb{P}(X_{t+1} = X_t) \geq \tfrac{1}{2}$, and we will not state this explicitly. In particular, the mixing time in Theorem 1.1 is for the lazy chain. (Without laziness issues of bipartiteness and near bipartiteness pop up, which have little theoretical or practical interest; see, e.g., [6, Remark 1.9] and [27, 41].)

The other notions we are interested in are the L_2 and uniform mixing time and the average L_2 mixing time. Here are the relevant definitions.

Definition 2.4 Let X_t, π and $P_t(x, y)$ be as above. Then the L_2 mixing time, the L^∞ (or uniform) mixing time and the average L_2 mixing time are, respectively,

$$t_{\text{mix}}^{(2)} := \min\{t : \max_x \|\mathbb{P}_x(X_t = \cdot) - \pi\|_{2,\pi} \leq 1/2\},$$

$$t_{\text{mix}}^{\text{unif}} := \min\left\{t : \max_{x,y}\left|\frac{\mathbb{P}_x(X_t = y)}{\pi(y)} - 1\right| \leq 1/4\right\},$$

$$t_{\text{mix}}^{\text{ave}} := \min\left\{t : \sum_x \pi(x)\|\mathbb{P}_x(X_t = \cdot) - \pi\|_{2,\pi}^2 \leq 1/4\right\}.$$

Here and below,

$$\|\mu - \pi\|_{2,\pi}^2 := \sum_x \pi(x)\left(\frac{\mu(x)}{\pi(x)} - 1\right)^2 = -1 + \sum_x \frac{\mu(x)^2}{\pi(x)}.$$

The constants $\tfrac{1}{2}$ and $\tfrac{1}{4}$ do not play an important role and were chosen for convenience. We remark that in the reversible setting the L^2 and the L^∞ mixing times satisfy $t_{\text{mix}}^{\text{unif}} = 2t_{\text{mix}}^{(2)}$, while even without reversibility $t_{\text{mix}}^{\text{unif}} \leq 2t_{\text{mix}}^{(2)}$. See [20, Equation (2.2)] and [38, Equation (8.5)] for a proof in continuous time. The proof in discrete time is similar.

In the remainder of this section, we show the following.

Theorem 2.5 *The average L_2 mixing time is robust for reversible Markov chains in continuous time.*

An immediate corollary is that the (usual, not averaged) L_2 mixing time is robust in the transitive setup, under perturbations that preserve transitivity (in the discrete time case assuming the holding probabilities are bounded away from 0). By the remark above, the same holds for the uniform mixing time. Theorem 2.5 is not needed for the proofs of our main results. We added it for the sake of completeness. The proof is similar to the one in [42].

Proof Let $\mathcal{L}$ be the Markov generator, and let $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$ be the eigenvalues of $-\mathcal{L}$. Denote $P_t = e^{t\mathcal{L}}$. Then,

$$\|\mathbb{P}_x(X_t = \cdot) - \pi\|_{2,\pi}^2 + 1 = \sum_y \frac{\mathbb{P}_x(X_t = y)^2}{\pi(y)}$$

$$(2.2) \qquad \overset{(*)}{=} \sum_y \frac{\mathbb{P}_x(X_t = y)\mathbb{P}_y(X_t = x)}{\pi(x)} = \frac{\mathbb{P}_x(X_{2t} = x)}{\pi(x)},$$

where in $(*)$ we used reversibility.

Hence,

$$\sum_x \pi(x) \|\mathbb{P}_x(X_t = \cdot) - \pi\|_{2,\pi}^2 = -1 + \sum_x \mathbb{P}_x(X_{2t} = x)$$

$$(2.3) \qquad = \mathrm{Tr}(P_{2t}) - 1 = \sum_{i=2}^n \exp(-2\lambda_i t).$$

Recalling the definition of the average L_2 mixing time, we get

$$t_{\mathrm{mix}}^{\mathrm{ave}} = \inf\left\{ t : \sum_x \mathbb{P}_x[X_{2t} = x] \le 5/4 \right\}$$

$$(2.4) \qquad = \inf\left\{ t : \sum_{i=2}^n \exp(-2\lambda_i t) \le 1/4 \right\}.$$

Using (2.1) concludes the proof. ∎

Remark 2.6 The same calculations can be done in discrete time, leading to analogs of (2.3) and (2.4): $\sum_x \pi(x)\|\mathbb{P}_x(X_t = \cdot) - \pi\|_{2,\pi}^2 = \sum_{i=2}^n \beta_i^{2t}$ and so

$$t_{\mathrm{mix}}^{\mathrm{ave}} = \inf\left\{ t \in \mathbb{N} : \sum_{i=2}^n \beta_i^{2t} \le 1/4 \right\},$$

where $1 = \beta_1 > \beta_2 \ge \cdots \ge \beta_n > -1$ are the eigenvalues of the transition matrix P (assuming P is irreducible and aperiodic). This would allow to conclude a similar result in discrete time if it weren't for values of β_i close to either 0 or -1. Both problems can be resolved by adding laziness, but in the interest of brevity, we skip the details.

2.3 Geometric notions

Recall from Section 1.5, the definition of (a, b)-quasi-isometry for metric spaces, and that when we say that graphs are (a, b)-quasi-isometric, we are treating them as metric spaces with the graph distance as the metric.

Definition 2.7 Consider a reversible Markov chain on a finite state space Ω with transition matrix P (in continuous time, with generator $\mathcal{L}$) and stationary distribution π. We define the *Cheeger constant* of the chain as

$$\Phi := \min_{A:\, 0 < \pi(A) \le 1/2} Q(A, A^c)/\pi(A), \quad \text{where}$$

$$Q(A, A^c) := \begin{cases} \sum_{x \in A, y \notin A} \pi(x)P(x, y), & \text{in discrete time,} \\ \sum_{x \in A, y \notin A} \pi(x)\mathcal{L}(x, y), & \text{in continuous time.} \end{cases}$$

We will also need a version for a subset of the graph (this is the discrete analog of Dirichlet boundary conditions).

Definition 2.8 Let Ω, P, $\mathcal{L}$ and π be as above. Let $A \subsetneq \Omega$. We define $\Phi(A) := \min_{B \subset A} Q(B, B^c)/\pi(B)$.

Further, we define $\lambda(A)$ to be the smallest eigenvalue of the substochastic matrix obtained by restricting $I - P$ (resp. $-\mathcal{L}$) to A.

The following discrete version of Cheeger's inequality under Dirichlet boundary conditions is well known (see, e.g., [20, (1.4) and Lemma 2.4]. For every irreducible discrete- or continuous-time reversible chain, and every set A with $\pi(A) \leq 1/2$, we have that

$$(2.5) \qquad \Phi^2(A)/4 \leq \lambda(A) \leq \Phi(A), \qquad \frac{\Phi^2(A)}{4 \max_{a \in A} |\mathcal{L}(a, a)|} \leq \lambda(A) \leq \Phi(A)$$

in discrete and continuous time, respectively.

Lemma 2.9 Let G be a finite graph, v a vertex of G and $A_1, \ldots, A_k$ the components of $G \backslash \{v\}$, i.e., of G after removal of the vertex v and all adjoining edges. Let $w_i \in A_i$ be vertices. Then the probability that random walk starting from v hits $\{w_1, \ldots, w_k\}$ at w_i is proportional to the effective conductance from v to w_i.

For a gentle introduction to electrical networks, see [18].

Proof Denote by T_{w_i} the hitting time of w_i and by T_W the hitting time of the set $\{w_1, \ldots, w_k\}$. If the walker returns to v before T_W, the process begins afresh, so it is enough to consider only the last excursion from v. In other words, the probabilities are proportional to the conditioned probabilities $\mathbb{P}_v(T_{w_i} = T_W \mid T_W < T_v)$ (we define T_v to be the return time to v). Since each w_i is in a different component of $G \backslash \{v\}$, these conditional probabilities are proportional to $\mathbb{P}_v(T_{w_i} < T_v)$. These are proportional to the effective conductance (see [35, Exercise 2.47]). ∎

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 $uv \in E$ by a path of length at most K (whose endpoints are still denoted by u and v). Note that $V_1 \subset V_2$. The identity map is a $(K, 0)$-quasi-isometry of G_1 and G_2.

Lemma 2.10 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

$$(2.6) \qquad\qquad\qquad \Phi(G) \geq c_d \Phi(H)/K,$$

where $\Phi(G)$ and $\Phi(H)$ are the Cheeger constants of G and H, respectively.

This is well known and easy to see (see, e.g., [24, Proposition 2.3] for a proof). We finish this section with a simple lemma on stretched trees.

Lemma 2.11 Let T be a finite binary tree of height ℓ, let $f : \{1, \ldots, \ell\} \to \mathbb{N}$ be non-increasing, and let G be the graph one gets by stretching each edge between levels $h - 1$ and h of T to a path of length $f(h)$. Then, for every v in level h of T, we have

$$\mathbb{P}_G(\text{random walk starting from } v \text{ hits the root before hitting level } \ell) \leq 2^{-h}.$$

Proof The symmetry of the problem allows us to identify all the vertices in each level of T (before stretching). Consider the probability that random walk starting from v hits level ℓ before hitting the root. After the identification, we have the following.

- Level ℓ is just one vertex (which we also denote by ℓ).
- Removing the vertex corresponding to level h (which we also denote by v) disconnects the root from ℓ.

Hence, Lemma 2.9 may be used. Suppressing the dependence on ℓ, denote the resistances from the root and from v to ℓ by R_1 and R_2, respectively. Then the probability to hit the root before hitting ℓ is R_2/R_1. These resistances can be computed directly using parallel-series laws. Indeed, the resistance of $f(i)$ edges in a series is $f(i)$ and the resistance of 2^i parallel connections of this kind between i and $i+1$ is $2^{-i}f(i)$. All in all, we get

$$R_1 = \sum_{i=1}^{\ell} \frac{f(i)}{2^i} \geq \sum_{i=1}^{\ell} \frac{f(i+h)}{2^i} \geq \sum_{i=1}^{\ell-h} \frac{f(i+h)}{2^i} = 2^h R_2,$$

where the first inequality follows because f is non-increasing. The assertion of lemma follows. ∎

2.4 A tail estimate for hitting times

Recall that the *hitting time* of a set D is defined as $T_D := \inf\{t \geq 0 : X_t \in D\}$. Denote π conditioned on A by π_A, i.e., $\pi_A(a) = \mathbb{1}\{a \in A\}\frac{\pi(a)}{\pi(A)}$. Using the spectral decomposition of P_A (the restriction of the transition matrix P to the set A) with respect to the inner product $\langle f, g\rangle_{\pi_A} := \sum_a \pi_A(a)f(a)g(a)$, we get (see, e.g., [3, Chapter 3] or [6, Lemma 3.8]),

$$(2.7) \qquad \min_{x,y \in A} \frac{\pi(x)}{|A|\pi(y)} \sum_{a \in A} \mathbb{P}_a[T_{A^c} > t] \leq \mathbb{P}_{\pi_A}[T_{A^c} > t] \leq e^{-\lambda(A)t}$$

in discrete or continuous time.

3 Proof of Theorems 1.1 and 1.2

Throughout this section, we consider the interchange process on a graph G in continuous time in which all edges ring at rate 1 (Theorems 1.1 and 1.2 are formulated in discrete time, but translating the mixing time from continuous time to discrete time is simple and we explain this for Theorem 1.1 at the end of its proof, the explanation there holds for Theorem 1.2 equally). Since the claims of both theorems are asymptotic, we may and will assume that n is sufficiently large.

Let us start the proof by recalling elements of the construction already discussed in the proof sketch in Section 1.3 and in other places in the introduction. We need to find a set of transpositions $S_n \subset \mathfrak{S}_n$ such that $t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n))$ is small compared to either $t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S'_n))$ for a second set of transpositions S'_n such that $S_n \subseteq S'_n \subseteq S_n^3$ (in Theorem 1.1) or to $t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n, W_n))$ for some weights W_n (in Theorem 1.2). We describe our set of transpositions using a graph G on n vertices, whose edges are the transpositions. The construction has two parameters, $u \in \mathbb{N}$ and $\varepsilon \in (0, \frac{1}{2})$ (both

will be chosen later). We designate u parts of G and call them $H_1, \ldots, H_u$ (we will use H_i to denote both a subset of $[n]$ and the induced subgraph, and we will now describe them as graphs, thus describing also a part of G). The H_i are constructed inductively as follows. The induction base, H_1 is a binary tree of depth u whose edges have been replaced by paths of length 2^u. To define H_{i+1} given H_i we label, in each vertex of each of the trees used to construct H_i one child as "left" and the other as "right." We denote, for each leaf v of H_i, the number of left children on the path from the root to v by $g(v)$. Recall the definition of the bad leaves B_i,

$$(3.1) \qquad B_i := \left\{ x \text{ leaf of } H_i : g(x) > \left(\frac{1}{2} + \varepsilon \right) s_i \right\}$$

from (1.4). We define H_{i+1} as a forest of $|B_i|$ binary trees of depth s_{i+1}, with each edge replaced by a path of length ℓ_{i+1}, with

$$s_i := 4^{i-1} u \qquad \ell_i := 2^{u+1-i}$$

and each tree rooted at a point of B_i (so $H_{i+1} \cap H_i = B_i$ as sets). This terminates the description of the H_i. All this, we remind, was already discussed in Section 1.3 with some additional explanations and motivation (and a figure depicting the gadget $\bigcup H_i$ on page 30).

We now claim that, uniformly in ε,

$$2^{4^{u-1}u} \leq \left| \bigcup_{i=1}^{u} H_i \right| \leq 2^{4^u u}.$$

Indeed, the first inequality is clear because H_u has at least one root (since $\varepsilon < \frac{1}{2}$) and the second inequality comes from

$$|H_i| \leq \ell_i 2^{s_i+1} |B_{i-1}| \leq 2^{u+2-i+4^{i-1}u} \prod_{j=1}^{i-1} 2^{4^{j-1}u} \leq 2^{u+2-i+4^{i-1}u(1+1/3)} \leq 2^{3 \cdot 4^{i-1}u},$$

which can be summed readily to give

$$(3.2) \qquad \sum_{j=1}^{i} |H_j| \leq 2^{4^i u},$$

and in particular, the case $i = u$ is what we need. Hence, we may choose some $u = u_n$ such that $u 4^u \asymp \log n$ (in particular, $u \asymp \log \log n$) such that

$$(3.3) \qquad \sum_{i=1}^{u} |H_i| \leq n^{1/4},$$

regardless of ε (we need here $n \geq 65{,}536$ to have $2^{4^u u} \leq n^{1/4}$ for $u = 1$). Fix such a u for the rest of the proof.

The subgraph $\bigcup H_i$ is the "gadget," and the rest of the graph G is a complete graph on a set of vertices K. The gadget connects to the complete graph via the good leaves of the H_i (and all the leaves of the last one, H_u) so we define K to also include those

vertices. Thus, we define

$$K := \left([n] \setminus \bigcup_{i=1}^{u} H_i \right) \cup \left(\bigcup_{i=1}^{u-1} \left(\{\text{leaves of } H_i\} \setminus B_i \right) \right) \cup \{\text{leaves of } H_u\}.$$

Let the edges of G be all the edges of all the H_i union with a complete graph on K, i.e.,

$$E(G) := \left(\bigcup_{i=1}^{u} E(H_i) \right) \cup \{\{x, y\} : x, y \in K\}.$$

This finishes the construction of G (except for the choice of ε), and hence of S_n. We delay the definitions of S'_n and W_n to Section 3.3.

Below, when we want to emphasize the dependence on u and ε, we will write $G_n(u, \varepsilon)$ for G and $S_n(u, \varepsilon)$ for S_n. We will also denote $G_n(\varepsilon) := G_n(u_n, \varepsilon)$ and $S_n(\varepsilon) := S_n(u_n, \varepsilon)$ (recall that u_n is the value we fixed above such that $u_n 4^{u_n} \asymp \log n$).

3.1 An upper bound for the time to exit the gadget

Throughout the proofs, we will pick the parameter ε so that $\varepsilon > u^{-1/3}$.

Lemma 3.1 *The expected exit time from H_i, starting from a worst initial state in H_i (i.e., the one maximizing this expectation), denoted by L_i, satisfies (uniformly in i)*

$$(3.4) \qquad\qquad L_i \asymp \ell_i^2 s_i = u 4^u =: L.$$

As this lemma is standard, we only sketch its proof.

Proof sketch Examine the random walk X on H_i, and let $\sigma_0, \sigma_1, \ldots$ be the times when it reaches a vertex of degree 3 (we require also $X_{\sigma_{i+1}} \neq X_{\sigma_i}$). Between σ_i and σ_{i+1}, the walk is in a part of the graph which is simply three paths of length ℓ_i. By symmetry, it reaches each of the 3 ends of these lines with equal probability. Hence, X_{σ_i} is identical to a random walk on a binary tree of depth s_i. The distance of random walk on a binary tree from the root has the same distribution as a random walk on $\mathbb{N}$ with a drift toward infinity, and hence, a simple calculation shows that the expected exit time is Cs_i. To get back to random walk on H_i, we note that, even if we condition on $X_{\sigma_0}, X_{\sigma_1}, \ldots$, then the local symmetry says that the times $\sigma_{i+1} - \sigma_i$ are independent of X_{σ_i} and of one another. For each i, we have $\mathbb{E}(\sigma_{i+1} - \sigma_i) \asymp \ell_i^2$, because this is the same as the exit time from the interval $\{0, \ldots, \ell_i\}$, where the walk exits 0 at rate 3 (and the other vertices at rate 2), again by the symmetry. ∎

Below, we employ the notation $L = u 4^u$ from the above lemma. Let $T_K := \inf\{t : X_t \in K\}$ be the hitting time of the complete graph K. Recall that we have fixed a choice of $u = u_n$ satisfying that $u 4^u \asymp \log n$.

Proposition 3.2 *There exist some constants C and c' such that, for every ε and n, we have that the graph $G_n(\varepsilon)$ satisfies for all $i \in [u]$ that*

$$(3.5) \qquad\qquad \sum_{v \in H_i} \mathbb{P}_v \left[T_K > \frac{CL}{\varepsilon^4} \right] \leq C \frac{1}{|H_i|^{c'\varepsilon^2}}.$$

Consequently, if E is the event that, for all $i \in [u]$, all particles whose initial location is in H_i hit the complete graph K before time CL/ε^4, then $\lim_{n\to\infty} \mathbb{P}(E) = 1$, uniformly in ε.

We recall our standing assumptions that n is sufficiently large and that $\varepsilon > u_n^{-1/3}$ (in particular, "uniformly in ε" above means "uniformly in $\varepsilon \in (u^{1/3}, \frac{1}{2})$"). Let us remark that the ε^4 term is not optimal, but this is not a priority for us.

Proof of Proposition 3.2   The assertion of the last sentence of the proposition follows from (3.5) by a union bound over the particles (recall that in the interchange process, each particle is performing a random walk). We also need here our assumption that $\varepsilon > u^{-1/3}$, as it gives $\sum_{i=1}^{u} 1/|H_i|^{c\varepsilon^2} = o(1)$, since $|H_i| \geq 2^{s_i} = 2^{4^{i-1}u}$.

Thus, we need to verify (3.5). Let $m = m(\varepsilon) \geq 1$ be some integer parameter to be fixed later. Let

$$\rho_i(t) := \sum_{v\in H_i} \mathbb{P}_v[\min\{T_{[n]\setminus W_i}, T_K\} > t], \qquad W_i = W_i(m) := \bigcup_{j\in[i-m,i+m]\cap[u]} H_j,$$

where $T_{[n]\setminus W_i}$ is the hitting time of $[n]\setminus W_i$ (or the exit time of W_i, if you prefer). The proof of Proposition 3.2 is concluded by combining the following two lemmas. Indeed, let m be the minimal value which satisfies the requirement of Lemma 3.4, so $4^m \asymp \varepsilon^{-2}$. We use the same value of m in Lemma 3.3 and get that for $t > CL/\varepsilon^4$ we have (3.6). Combining this with (3.7) gives the proposition. ∎

Lemma 3.3   *For all $i \in [u]$ and all $t \geq C_1 16^m L$ for some C_1 sufficiently large,*

$$(3.6) \qquad\qquad \rho_i(t) \leq \frac{1}{|H_i|^2}.$$

(As usual, C_1 is an absolute constant. In particular, it depends on neither i nor t.)

Lemma 3.4   *There exist absolute constants $C, c > 0$ such that for all $\varepsilon \in (0, 1/2)$, if $4^m \varepsilon^2 \geq C$, then, for all $n \geq N_0$, the graph $G(n, u, \varepsilon)$ satisfies for every $i \in [u]$ that*

$$(3.7) \qquad\qquad \sum_{v\in H_i} \mathbb{P}_v[T_{[n]\setminus W_i} < T_K] \leq \frac{C}{|H_i|^{c\varepsilon^2}}.$$

Proof of Lemma 3.3   Let $\mathcal{M}$ be the restriction of the Markov generator $\mathcal{L}$ to $W_i\setminus K$ (i.e., this is the generator of the chain killed upon exiting $W_i\setminus K$). Let λ be the smallest eigenvalue of $-\mathcal{M}$. It will be convenient to extend the definition $\ell_i = 2^{u+1-i}$ also to negative i. We now claim that

$$(3.8) \qquad\qquad \lambda \gtrsim \ell_{i-m}^{-2}.$$

To see this, let W be an arbitrary connected component of W_i. We first apply Lemma 2.10 to W. Since it is a piece of an infinite binary tree with edges stretched to various extents, but not more than ℓ_{i-m}, and since the infinite tree has positive Cheeger constant, we get that the Cheeger constant of W is at least ℓ_{i-m}^{-1}. Applying Cheeger's inequality (2.5) to $W\setminus K$ embedded in an infinite, stretched tree shows (3.8).

Using this, we get that

$$
\begin{aligned}
\rho_i(t) &\overset{(2.7)}{\leq} |W_i| \exp(-\lambda t) \overset{(3.8,3.2)}{\leq} 2^{4^{i+m}u} \exp(-c\ell_{i-m}^{-2} t) \\
&\overset{(*)}{\leq} 2^{4^{i+m}u} \exp(-c4^{i-u-m-1} \cdot C_1 16^m \cdot u4^u) \overset{(3.2)}{\leq} \frac{1}{|H_i|^2},
\end{aligned}
$$

(3.9)

where the inequality marked $(*)$ follows from the definitions of ℓ_i and L (recall that $L = u4^u$) and from the bound on t in the statement of the lemma. In the last inequality, we also use that C_1 is sufficiently large. ∎

Proof of Lemma 3.4 We divide the event $T_{[n]\setminus W_i} < T_K$ into two cases: that the random walk hits H_{i+m+1} before hitting K, and that it hits H_{i-m-1} before hitting K. Denote these two events by $\mathcal{U}$ and $\mathcal{D}$, respectively (notice that if $i \geq u - m$ then $\mathcal{U}$ is empty and if $i \leq m + 1$ then $\mathcal{D}$ is empty). The letters $\mathcal{U}$ and $\mathcal{D}$ stand for "up" and "down," with the orientation being as in Figure 1 (page 30).

We first handle $\mathcal{U}$. For $\mathcal{U}$ to happen there must be some time $\sigma < T_K$ such that $X_\sigma \in B_{i+m-1} \subset H_{i+m}$ and, further, the walker is contained in H_{i+m} between time σ and the first hitting time to the set of leaves of H_{i+m} which (on the event $\mathcal{U}$) occurs at B_{i+m}. Assume such a σ exists and examine the walker between σ and $T_{B_{i+m}}$. The walker is not simple (because being after σ conditions it to not return to the roots of H_{i+m}) but this is not important for us. The symmetry of the tree implies that at the first time after σ that the walker visits a leaf of H_{i+m}, the difference between the number of left and right turns along the path the walker takes is distributed like a sum of i.i.d. ± 1 variables (giving equal probability to each value). In particular, the probability that the target leaf is in B_{i+m} is

$$
\mathbb{P}\left(\mathrm{Bin}(s_{i+m}, \tfrac{1}{2}) > \left(\tfrac{1}{2} + \varepsilon \right) s_{i+m} \right) \leq \exp(-c\varepsilon^2 s_{i+m}).
$$

Our assumption that a time σ exists only reduces the probability further so we get, for every $v \in H_i$, that $\mathbb{P}_v(\mathcal{U}) \leq \exp(-c\varepsilon^2 s_{i+m})$. Summing over v gives

$$
\sum_{v \in H_i} \mathbb{P}_v(\mathcal{U}) \leq |H_i| \exp(-c\varepsilon^2 s_{i+m}) \overset{(3.2)}{\leq} 2^{4^i u} \exp(-c\varepsilon^2 4^{i+m}u),
$$

and we see that if m satisfies $4^m \geq 2/c\varepsilon^2$ then this sum is smaller than, say, $1/|H_i|^2$. Require m to satisfy that, but do not fix its value yet (there will be a similar requirement below).

We move to the estimate of $\mathcal{D}$. We use Lemma 2.11 and get that for any v in level h of H_i (we are counting levels before stretching here) or in the path between level h and $h + 1$, we have

$$
\mathbb{P}_v(\mathcal{D}) \leq 2^{-h - \sum_{j=i-m}^{i-1} s_j}.
$$

(Note that Lemma 2.11 measures a larger event. Indeed, $\mathcal{D}$ is the event to hit the root of our tree before hitting level $i + m$ or K, so it is smaller than the event to hit the root before $i + m$, which is what is measured by Lemma 2.11.) The number of vertices at

level h, or in a path between level h and $h+1$, is $|B_{i-1}| \cdot 2^h \cdot \ell_i$, so we get

$$(3.10) \qquad \sum_{v \in H_i} \mathbb{P}_v(\mathcal{D}) \le \sum_{h=0}^{s_i-1} |B_{i-1}| 2^h \ell_i \cdot 2^{-h - \sum_{j=i-m}^{i-1} s_j} = s_i \ell_i \frac{|B_{i-1}|}{2^{\sum_{j=i-m}^{i-1} s_j}}.$$

Denote $p_j := \mathbb{P}\big(\mathrm{Bin}(s_j, \tfrac{1}{2}) > (\tfrac{1}{2} + \varepsilon)s_j\big)$. Then $|B_j| \le |B_{j-1}|2^{s_j} p_j$ and further $p_j \le \exp\big(-c\varepsilon^2 s_j\big)$. Iterating this gives

$$|B_{i-1}| \le |B_{i-m-1}|2^{\sum_{j=i-m}^{i-1} s_j} \prod_{j=1}^{i-1} p_j \le |B_{i-m-1}|2^{\sum_{j=i-m}^{i-1} s_j} \exp\big(-c\varepsilon^2 s_{i-1}\big).$$

Substituting this in (3.10) gives

$$\sum_{v \in H_i} \mathbb{P}_v(\mathcal{D}) \le s_i \ell_i |B_{i-m-1}| \exp(-c\varepsilon^2 s_{i-1}) \overset{(3.2)}{\le} s_i \ell_i 2^{4^{i-m-1}u} \exp(-c\varepsilon^2 s_{i-1}).$$

We see that taking m so that $4^m \varepsilon^2$ is sufficiently large makes the term $2^{4^{i-m-1}u} = 2^{s_{i-1}4^{-m+1}}$ negligible compared to the exponential (recall that $s_i = 4^{i-1}u$). This is the last requirement from m and we may fix its value. Further, our standing assumption that $\varepsilon > u^{-1/3}$ means that the $s_i \ell_i$ terms are also negligible with respect to $\exp(-c\varepsilon^2 s_{i-1})$. Hence

$$\sum_{v \in H_i} \mathbb{P}_v(\mathcal{D}) \lesssim \exp(-c'\varepsilon^2 s_{i-1}),$$

as needed. The lemma is thus proved, and so is Proposition 3.2. ∎

Having established in Proposition 3.2 that the walker hits K, we now show that it remains there for a considerable amount of time.

Lemma 3.5 *Let $t := C_1 L/\varepsilon^4$ for some C_1 sufficiently large. For every $x \in [n]$, let $N(x)$ be the amount of time a walker starting from x spends in K up to time t. Then*

$$\mathbb{P}\big[\exists x \in [n] \text{ s.t. } N(x) < \tfrac{2}{3}t\big] \to 0$$

as $n \to \infty$, uniformly in $\varepsilon > u^{-1/3}$ (but not necessarily in C_1).

Proof Let $q = C_2 L/\varepsilon^4$, where C_2 is the constant from Proposition 3.2, denoted there by C. Apply Proposition 3.2 after some arbitrary time s. We get that during the interval $[s, s+q]$, the probability that all particles in $[n]\setminus K$ hit K is at least

$$1 - C \sum_{i=1}^{u} \frac{1}{|H_i|^{c\varepsilon^2}} > 1 - C2^{-u/4}.$$

Hence, for any fixed value of C_1, we can apply this for $s = 0, q, 2q, \dots, q(\lfloor t/q \rfloor - 1)$ (just a constant number of times, in fact $\lfloor C_1/C_2 \rfloor$) and get that with probability going to 1, all events happen simultaneously. In other words, no particle spent more than $2q$ consecutive time units in any visit of $[n]\setminus K$.

Let us now bound the number of possible visits. We will show that a.a.s. as $n \to \infty$ no particle makes more than one visit to $[n]\setminus K$ by time t after reaching K for the first time. For this purpose, denote ∂K to be all points of K with a neighbor in $[n]\setminus K$ (namely, leaves of H_i which are not in B_i for $i < u$ and all leaves of H_u).

Suppose a particle is at time 0 at some $x \in K$. Let us first bound the number of jumps it does up to time t. Since the degrees of our graph are all bounded by $|K|$, this number is stochastically dominated by an appropriate Poisson variable, and in particular, the probability that the particle performed more than $2t|K|$ jumps is $o(1/n)$. Adding the restriction that the jump would be to a vertex of K only reduces the number further, so we get the same bound for the number of jumps to vertices of K.

Among the first $2t|K|$ jumps to vertices of K, the number of jumps to ∂K is stochastically dominated by $\mathrm{Bin}(2t|K|, \frac{|\partial K|}{|K|-1})$. Hence, the probability that more than $4t|\partial K|$ of them are to ∂K is $o(1/n)$ (where we used that $|\partial K| \gtrsim n^c$, which follows from our choice of u).

Examine now the first $4t|\partial K| + 2$ visits to ∂K (not necessarily up to time t, all of them). The probability that at least two of the following jumps were away from K is at most $(4t|\partial K| + 2)^2/|K|^2 = o(1/n)$, where we used the fact that by (3.3) $|\partial K| \leq n^{1/4}$ and $|K| \geq n - n^{1/4}$, as well as $t \lesssim \log^2 n$ (recall that $L \asymp \log n$ and $\varepsilon > u^{-1/3} \asymp (\log\log n)^{-1/3}$). In the case that indeed no more than 1 of these jumps went to $[n]\backslash K$, we get that the first $4t|\partial K| + 2$ visits to ∂K include all the visits to ∂K up to time t: no more than $4t|\partial K|$ visits from K and no more than 2 visits from $[n]\backslash K$ (the first hitting of K and the first return to K).

Combining everything together, we see that after first reaching K (which a.a.s. all particles do by time q) a.a.s. all particles leave K at most once by time t and during such excursion they each spend at most $2q$ time units away from K. Taking C_1 to be large enough in terms of C_2 concludes the proof. ∎

3.2 The coupling

Denote the transposition (x, y) by τ_{xy}. Consider two initial configurations σ and σ' of the interchange process. We now define a coupling $((\sigma_t)_{t\geq 0}, (\sigma'_t)_{t\geq 0})$ of the interchange processes starting from these initial states. We make the edges ring at rate 2, but when an edge rings, it is ignored with probability $1/2$. We use the same clocks for both systems. If, at time t, an edge $e = xy$ rings and $\sigma_{t-}(x) = \sigma'_{t-}(y)$ (where $\sigma_{t-}(x) \coloneqq \lim_{\delta\to 0^+} \sigma_{t-\delta}(x)$, as usual) or $\sigma_{t-}(y) = \sigma'_{t-}(x)$, then with probability $1/2$, we set $\sigma_t = \sigma_{t-} \circ \tau_e$ and $\sigma'_t = \sigma'_{t-}$; and with probability $1/2$, we set $\sigma_t = \sigma_{t-}$ and $\sigma'_t = \sigma'_{t-} \circ \tau_e$ (either way, the number of disagreements decreases). If $\sigma_{t-}(x) \neq \sigma'_{t-}(y)$ and $\sigma_{t-}(y) \neq \sigma'_{t-}(x)$, then with probability $1/2$, we set $\sigma_t = \sigma_{t-} \circ \tau_e$ and $\sigma'_t = \sigma'_{t-} \circ \tau_e$; and with probability $1/2$, we set $\sigma_t = \sigma_{t-}$ and $\sigma'_t = \sigma'_{t-}$.

We see that for all i once the particle labeled i is coupled in the two systems, it remains coupled. That is, if $\sigma_t^{-1}(i) = (\sigma'_t)^{-1}(i)$, then, for $t' > t$, we also have $\sigma_{t'}^{-1}(i) = (\sigma'_{t'})^{-1}(i)$. Whenever the position of particle i is adjacent in one system is adjacent to the current position of particle i in the other system (i.e., $\sigma_t^{-1}(i)(\sigma'_t)^{-1}(i) \in E_n$), the infinitesimal rate in which they are coupled is 2.

Lemma 3.6 *There exists a C such that*

$$t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n(\varepsilon, u_n))) \leq \frac{Cu_n 4^{u_n}}{\varepsilon^4},$$

under our usual assumption that $\varepsilon > u_n^{-1/3}$.

Proof Lemma 3.5 shows that a.a.s. indeed all particles in one system spend at least $\frac{2}{3}$ of the time in K by time $C_1 L/\varepsilon^4$ for any C_1 sufficiently large. By a union bound, this applies to both systems in the above coupling. On this event (occurring for both systems), for each i, the particle labeled i has to spend at least $1/3$ of the time by time $C_1 L/\varepsilon^4$ in K simultaneously in both systems. Since the particle gets coupled with rate 2 during these times, a standard argument shows that the conditional probability of particle i not getting coupled is at most

$$(3.11) \qquad \exp(-cC_1 L/\varepsilon^4) \le \exp(-cC_1 L) \overset{(*)}{\le} n^{-cC_1},$$

where the inequality marked by $(*)$ follows since $L = u4^u \asymp \log n$ (see just above (3.3)). If C_1 is sufficiently large, this will be $\ll 1/n$ and we may apply a union bound and get that a.a.s. all particles are coupled by time $C_1 L/\varepsilon^4$. As the initial states σ and σ' are arbitrary, this implies that the mixing time is at most $C_1 L/\varepsilon^4$ (see, e.g., [33, Theorem 5.4]). ∎

3.3 The perturbation

In this section, we analyze the perturbed versions of S_n, lower bound their mixing time, and thus conclude the proofs of Theorems 1.1 and 1.2. The following convention will be useful here and in other places in the paper. Thus, we make special note of it

Definition 3.7 We call an edge of H_i that belongs to a path that is a stretching of a left edge (of H_i') a "left edge." Similarly for right edges.

Do not be confused with the definition of g. It is still the case that g counts left edges before stretching, not all left edges of H_i.

Proof of Theorem 1.2 Recall that we are given a function $1 \ll f(n) \le \log\log\log n$ and we need to construct generators S_n and weights $W_n = (w_n(s))_{s \in S}$ satisfying $1 \le w_n(s) \le 1 + (f(n!)/\log\log n)^{1/4}$ such that

$$t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n, W_n)) \gtrsim t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n))f(n!).$$

Define $\varepsilon := c_1(f(n!)/\log\log n)^{1/4}$, where c_1 is a universal positive constant that will be fixed soon (but let us already require $c_1 < \frac{1}{4}$). The requirement $\varepsilon > u^{-1/3}$ will be satisfied for n sufficiently large. We use the set $S_n(u, \varepsilon)$ defined above with $u = u_n$ and this ε (we remind that $u_n \asymp \log\log n$).

Denote, for any $\delta > 0$, $W(\delta, n) = (w(s))_{s \in S_n}$ with

$$w(s) := \begin{cases} 1 + \delta, & s \text{ is a left edge}, \\ 1, & \text{otherwise} \end{cases}$$

("otherwise" referring to both right edges and to edges of K). We will take $\delta = \varepsilon/c_1$ in what follows. The notation $W(\delta, n)$ will be reused below in the proof of Theorem 1.1, but there we will take $\delta = 3$, so let us proceed under the assumption $\delta \le 3$, which holds under the definitions of δ and ε above too.

Recall the notation $g(v)$ for the number of left children in a path from the root to v (before stretching). Examine first an infinite binary tree where each left child has weight $1 + \delta$ for some $\delta > 0$, and each right child has weight 1 (denote this object by

$\mathcal{T}_\delta$). Let Y_k be the *last* vertex in the kth level visited by random walk on it. By [26, Fact 4.1(2a)] (proved in the appendix of [26]), $g(Y_k)$ has the same distribution as the sum of k independent $\{0,1\}$-variables taking the value 1 with probability

$$\frac{\sqrt{1+\delta}}{1+\sqrt{1+\delta}} = \frac{1}{2} + \delta/8 + O(\delta^2).$$

This fact holds also for random walk on $\mathcal{T}_\delta$ started from either child of the root and conditioned not to return to the root. The proof in [26] applies to this case verbatim. Now, if c_1 is sufficiently small, then

$$\eta := \frac{\sqrt{1+\varepsilon/c_1}}{1+\sqrt{1+\varepsilon/c_1}} > \frac{1}{2} + 3\varepsilon$$

(recall that $\delta = \varepsilon/c_1$ is bounded above by 3). Fix c_1 to satisfy this property.

Still on the infinite tree $\mathcal{T}_\delta$, denote by Y_k^*, the vertex where the walker is at on the *first* time, it hits level k, in other words, the hitting point. It is straightforward to see that $\mathbb{P}(|g(Y_k^*) - g(Y_k)| > \lambda) \le 2e^{-c\lambda}$ for every λ, where the (nonnegative) constant c is independent of ε.

Information on Y_k^* can already be translated to our graphs H_i, because random walk on H_i, when considered only at times when it reaches a vertex before stretching, is identical to random walk on a piece of $\mathcal{T}_{\varepsilon/c_1}$ (say, by Lemma 2.9). We get that a random walk starting from a root of H_i and conditioned not to go to H_{i-1} before leaving H_i (for $i = 1$, an unconditioned walker) has, when it exits H_i that g is distributed like $\mathrm{Bin}(s_i, \eta)$ plus a quantity with a uniform exponential tail (uniform in both i and the value attained by the $\mathrm{Bin}(s_i, \eta)$ random variable).

A similar argument shows that, now on our graphs H_i, if Z_i is the first vertex, the walker is in among the roots of H_{i+1} (which of course is also a leaf of H_i) and Z_i^* is the last vertex, the walker is in H_i (say, before hitting the leaves of H_{i+1} or K for the first time), then $|g(Z_i) - g(Z_i^*)|$ is bounded with an exponential tail (uniformly in i).

We may now finish the proof of the theorem. Indeed, let X be the particle that was at time 0 at the root of H_1. Let T be the time X hits the leaves of H_1. We see that, if $\lambda > 0$ is some sufficiently small constant, then

$$\mathbb{P}(T \ge \lambda u 4^u) > 1 - Ce^{-cs_1} = 1 - Ce^{-cu},$$
$$\mathbb{P}(g(X_T) > (\tfrac{1}{2} + 2\varepsilon)s_1) > 1 - Ce^{-cs_1\varepsilon^2} > 1 - Ce^{-cu^{1/2}},$$

where the last inequality is due to $\varepsilon \gtrsim (\log\log n)^{-1/4}$. In particular, with the same probability X_T is in B_1. The same X still has that $g > (\tfrac{1}{2} + \varepsilon)s_1$ when leaving H_1 (again with probability $> 1 - Ce^{-cs_1\varepsilon^2}$), and then hits the leaves of H_2 after another at least $\lambda u 4^u$ time units, and hits B_2, and so on. We get that, at time $\lambda u^2 4^u$, this particle is still inside the gadget, with probability at least $1 - Cue^{-cu^{1/2}}$. This of course means the walk on $\mathfrak{S}_n$ is not yet mixed. Hence

$$t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n, W_n)) \gtrsim u^2 4^u.$$

With Lemma 3.6, we get that

$$t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n, W_n)) \gtrsim u^2 4^u \gtrsim t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n)) \cdot u \varepsilon^4 \asymp t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n)) f(n!)$$

as claimed (in the last "$\asymp$" we used $u \asymp \log \log n$). This concludes the proof. ∎

Proof of Theorem 1.1 Recall from the proof sketch Section 1.3 that S_n' is created by adding to each path of S_n that came from stretching a left edge, edges between even vertices (initially at distance two from one another). The parallel–serial laws show that the resistance of a path of length $2N$ to which such edges have been added is $\frac{2}{3}N$. Examining a walker only at times where it is in vertices that were not added in the stretching process, we see that its walk is exactly identical to a walk on $\mathrm{Cay}(\mathfrak{S}_n, S_n, W(3, n))$, where $W(3, n)$ is from the previous proof. Hence choosing $\varepsilon = 3c_1$, we get, as in the previous proof, $t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n')) \gtrsim u^2 4^u$.

This almost finishes the proof of Theorem 1.1. The only remaining issue to address is that Theorem 1.1 is formulated in discrete time, while we worked all along in continuous time. This is not a problem. Indeed, if P is a transition matrix and I is the identity matrix, then the total variation mixing time $t_{\mathrm{mix}}^{\delta\,\mathrm{lazy}}$ of the δ-lazy chain with transition matrix $\delta I + (1 - \delta)P$ and that of the continuous-time chain with generator $\mathcal{L} = P - I$, denoted by $t_{\mathrm{mix}}^{\mathrm{ct}}$, satisfy

$$(3.12) \qquad \frac{\delta}{C(1-\delta)}\left(t_{\mathrm{mix}}^{\delta\,\mathrm{lazy}} - c_\delta\right) \leq t_{\mathrm{mix}}^{\mathrm{ct}} \leq \frac{C}{(1-\delta)}\left(t_{\mathrm{mix}}^{\delta\,\mathrm{lazy}} + c_\delta\right)$$

for an absolute constant $C > 0$ and a constant $c_\delta > 0$, independent of the Markov chain. The case $\delta = 1/2$ follows directly from [33, Theorem 20.3] and the argument extends to all $\delta \in (0, 1)$. For much finer relations between the two mixing times in the reversible setup, see [6, 13, 27].

In our case, we estimated the continuous time mixing time with the rates equal to 1, while the generator $\mathcal{L}$ has rates $1/|S_n|$ or $1/|S_n'|$, as the case may be. Multiplying all the rates by a constant changes the mixing time by the same constant, so we get

$$t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n)) \overset{(3.12)}{\lesssim} |S_n| t_{\mathrm{mix}}^{\mathrm{ct}}(\mathrm{Cay}(\mathfrak{S}_n, S_n)) \lesssim |S_n| u 4^u \asymp n^2 \log n$$

$$t_{\mathrm{mix}}(\mathrm{Cay}(\mathfrak{S}_n, S_n')) \overset{(3.12)}{\gtrsim} |S_n'| t_{\mathrm{mix}}^{\mathrm{ct}}(\mathrm{Cay}(\mathfrak{S}_n, S_n, W(3, n))) \gtrsim |S_n'| u^2 4^u \asymp n^2 \log n \log \log n.$$

The theorem is thus proved. ∎

4 Proof of Theorem 1.3

Recall that we wish to construct a sequence of graphs G_n with bounded degrees and weights with $1 \leq w_n(e) \leq 1 + o(1)$ such that the mixing time of G_n is significantly smaller than the mixing time of the weighted version.

As a building block in our construction, we will need the auxiliary graph described in the following lemma, whose proof is deferred to Section 4.4.

Lemma 4.1 *There exists an absolute constant $\mu > 0$ such that, for every m, there exists a graph H of maximal degree 6 with $|H| \asymp 2^{10m}$ containing two disjoint sets of vertices B*

and W of sizes $|B| = 2^m$ and $|W| = 2^{10m}$ such that lazy SRW on H satisfies that

$$(4.1) \qquad \mathbb{P}_b\big[T_{B\setminus\{b\}} < T_W\big] \lesssim 2^{-4m} \qquad \forall b \in B,$$

$$(4.2) \qquad \mathbb{E}_h\big[T_{B\setminus\{h\}}\big] \asymp |H|/2^m \qquad \forall h \in H,$$

$$(4.3) \qquad \mathbb{P}_h\big[T_{B\setminus\{h\}} > |H|/2^m\big] \gtrsim 1 \qquad \forall h \in H,$$

$$(4.4) \qquad \mathbb{P}_b\big[T_W < T_b \mid T_W < T_{B\setminus\{b\}}\big] \geq \mu \qquad \forall b \in B.$$

Moreover, the last probability is the same for all $b \in B$. Lastly, for all $w \in W$ starting from w the hitting distribution of B uniform.

As usual, T_W, T_B, etc. are the hitting times of B, W, etc.

4.1 The construction

Let $G = G_n((\log\log n)^{-1/8})$, i.e., the graph from the construction of Theorem 1.2 with the parameter ε from the construction taken to be $(\log\log n)^{-1/8}$ (G is the graph on which the interchange process is performed, so $E(G)$ is a set of transpositions of $\mathfrak{S}_n$). Let m satisfy that $2^{m-1} < |E(G)| \leq 2^m$. Let H be the graph from Lemma 4.1 with this m (so $|H| \asymp 2^{10m} \asymp n^{20}$). Let $A \subseteq B$ (B from the statement of Lemma 4.1) be some arbitrary set of size $|E(G)|$, and let $\tau : A \to E(G)$ be some arbitrary bijection.

We now construct our graph, which we denote by L. We take the vertex set to be $H \times \mathfrak{S}_n$. We define the edges implicitly by describing the transition probabilities of the random walk. Let $\{a, b\} \in E(H)$. If $a \notin A$, we set $P((a, \sigma), (b, \sigma)) = \frac{1}{\deg a}$ for all $\sigma \in \mathfrak{S}_n$, where $\deg a$ is the degree of a in H. If $a \in A$ and $b \in H \setminus A$, we set $P((a, \sigma), (b, \sigma)) = \frac{1}{2\deg a}$, while $P((a, \sigma), (a, \sigma \circ \tau_a)) = \frac{1}{2}$ (recall that τ_a is the transposition corresponding to a). No other transitions have positive probability. Below, we consider the mixing time of the continuous time version of P or the discrete time mixing time of $\frac{1}{2}(I + P)$. We shall denote either mixing time by $t_{\text{mix}}(L)$.

This chain $(X_t, \sigma_t)_{t \geq 0}$ can be described as follows: We have a random walk $(X_t)_{t \geq 0}$ on H and an "interchange process" σ_t on G which evolves in slow motion. Whenever the walk X_t on H is at some vertex $a \in A$, it either stays put or makes a random walk step on H. If it stays put in $a \in A$, then it also makes one step of the interchange process, updating its state to $\sigma_t \circ \tau_a$.

4.2 Analysis of the example

We first define a sequence of random times. Recall the set W from the construction of H in Lemma 4.1. Let $S_1 := \inf\{t \geq 0 : X_t \in W\}$ and $T_1 := \inf\{t > S_1 : X_t \in B\}$. Inductively, set

$$S_{i+1} := \inf\{t > T_i : X_t \in W\} \quad \text{and} \quad T_{i+1} := \inf\{t > S_i : X_t \in B\}.$$

Let $\mathcal{J}$ be the event that, for all $i \leq n^4$, the walk does not visit $B \setminus \{X_{T_i}\}$ between time T_i and S_{i+1}. By (4.1), $\mathbb{P}_y(\mathcal{J}) > 1 - O(n^{-4})$ for all $y \in L$.

Let $Z_i := X_{T_i}$ and $\widehat{\sigma}_i := \sigma_{T_i}$ (σ_t being the second coordinate of the chain (X_t, σ_t)). By Lemma 4.1, we have that $Z_1, \ldots, Z_{n^4}$ are i.i.d. uniform on B. Under $\mathcal{J}$, the behavior

of the permutation σ in the time interval $[T_i, T_{i+1})$ is quite simple: if $Z_i \in B \backslash A$, then it does not change at all in this interval, and if $Z_i \in A$, then it is composed with τ_{Z_i} with probability $\frac{1}{2}$ for each time $t \in [T_i, T_{i+1})$ when $X_t = Z_i$. This, together with (4.4) imply that, still under $\mathcal{J}$, $(\widehat{\sigma}_i)_{i=1}^{n^4}$ evolves precisely like a lazy version of the discrete-time interchange process on G. The laziness has two sources: the probability to hit $B \backslash A$ (which gives laziness $|B \backslash A|/|B|$, which is bounded above by $\frac{1}{2}$), and an additional laziness coming from the event of applying the transposition τ_{Z_i} an even number of times between T_i and S_{i+1}. We use here the fact that the probability in (4.4) is the same for all $a \in A$. In other words, we have a coupling of $\widehat{\sigma}_i$ and lazy interchange on G which succeeds (i.e., the two processes are the same) with probability $1 - O(n^{-4})$.

Let r be the $\frac{1}{4}$ total variation mixing time of this lazy discrete-time interchange process on G. To estimate r, note that by Lemma 3.6, the mixing time of the interchange process is at most $Cu4^u \varepsilon^{-4} \asymp \log n (\log \log n)^{1/2}$ (recall that $u4^u \asymp \log n$). Using (3.12), we may translate this to the mixing time of the lazy discrete-time interchange process and get that $r \lesssim n^2 \log n (\log \log n)^{1/2}$ (recall that $|E(G)| \asymp n^2$). In particular, $r < n^4$ for all sufficiently large n.

Thus, under $\mathcal{J}$, we have that $X_{T_{r+1}}$ has its first coordinate uniform on B and its second approximately uniform on $\mathfrak{S}_n$ and independent of the first coordinate. Removing the requirement that we are on $\mathcal{J}$, the distribution of $X_{T_{r+1}}$ is still approximately uniform (in the TV distance) on the same set, simply because $\mathbb{P}(\mathcal{J}) > 1 - Cn^{-4}$.

In the language of [34], T_{r+1} is an approximate forget time. As we recall below, by combining results from [2] and [34], this implies that

$$(4.5) \qquad t_{\mathrm{mix}}(L) \lesssim \max_y \mathbb{E}_y(T_{r+1}) \overset{(4.2)}{\asymp} (r+2)\frac{|H|}{2^m} \lesssim n^{20} \log n \sqrt{\log \log n}$$

(we have $r + 2$ rather than $r + 1$ in the third expression, to account for the time until the walk hits B for the first time. This is also why we formulated (4.2) for every $h \in H$ and not just for $b \in B$).

Thus, we need only describe briefly the results of [2] and [34]. In [34], the authors define the mixing time differently from us (see the definition of $\mathcal{H}$ in [34, Section 2.3]). We will adopt their notation and call this quantity $\mathcal{H}$. (We will not define $\mathcal{H}$ here as this would take too much space. The reader can find the definition, together with many illuminating examples, in [34].) As for the approximate forget time, it is denoted in [34] by $\mathcal{F}_\varepsilon$ (also in Section 2.3 there). Finally, the result that $\mathcal{F}_\varepsilon \asymp \mathcal{H}$ is a combination of Theorems 3.1 and 3.2 in [34].

As for [2], it defines τ_1 which is the continuous time mixing time, and τ_2 which is the same as $\mathcal{H}$, and [2, Theorem 5] states that $\tau_1 \asymp \tau_2$ (see also [41] where $\mathcal{H}$ is denoted by t_{stop}). Thus, we get

$$t_{\mathrm{mix}}(L) = \tau_1 \asymp \tau_2 = \mathcal{H} \asymp \mathcal{F}_\varepsilon \lesssim \max_y \mathbb{E}_y(T_{r+1}),$$

which justifies the first inequality of (4.5) and finishes the estimate of $t_{\mathrm{mix}}(L)$.

Remark 4.2 An alternative proof that replaces the results of [34] with a coupling argument is a follows. Using the specific construction of the graph H, the expectation of the time required in order to couple the H coordinate is at most $\max_{b \in B} \mathbb{E}_b[T_W]$ (cf. the coupling for lazy SRW on a finite d-ary tree in [33, Section 5.3.4]). The above

analysis allows one to then couple the $\mathfrak{S}_n$ coordinate with the additional amount of time required having expectation at most $Cn^{20} \log n \sqrt{\log \log n}$.

4.3 The perturbation

Recall from the previous section, the stopping times T_i and S_i, the notation $Z_i = X_{T_i}$, $\widehat{\sigma}_i := \sigma_{T_i}$ and the event $\mathcal{J}$. For every $a \in A$ such that the edge that corresponds to τ_a is a left edge (recall Definition 3.7), we increase the weight of the edges $((a, \sigma), (a, \sigma \circ \tau_a))$ to $1 + \theta\varepsilon$ for some θ sufficiently large, to be fixed later. Here, ε is as in Section 4.1, namely, $(\log \log n)^{-1/8}$.

To analyze the effect of this perturbation fix $i < n^4$, assume $Z_i \in A$, and denote $(a, \sigma) := (Z_i, \widehat{\sigma}_i)$. We need to examine the number of times the walker traversed the edge $((a, \sigma), (a, \sigma \circ \tau_a))$ between T_i and S_{i+1}. Denote this number by N. Clearly, if N is even, then $\widehat{\sigma}_{i+1} = \sigma$ and, otherwise, it is $\sigma \circ \tau_a$. Let p_{even} be the probability that N is even. Let $q := \mathbb{P}[X_{T_{i+1}} = a \mid X_{T_i} = a]$. Let β be the probability that after jumping away from a the walk returns to a before hitting W. By a first-step analysis,

$$p_{\text{even}} = q(1 - p_{\text{even}}) + (1 - q)(1 - \beta(1 - p_{\text{even}})).$$

Solving yields that

$$p_{\text{even}} = 1 - \frac{q}{1 + q - (1 - q)\beta}.$$

Conveniently, the perturbation does not affect β, it only affects q, increasing it from $\frac{1}{2}$ to $q_0 := (1 + \theta\varepsilon)/(2 + \theta\varepsilon)$. Hence

$$\frac{\mathrm{d}}{\mathrm{d}q} p_{\text{even}} = -\frac{1 - \beta}{(1 + q - (1 - q)\beta)^2}.$$

The last derivative is negative and is bounded away from 0.

It follows from this that the perturbation decreases the probability p_{even} by an additive term which is $\Theta(\theta\varepsilon)$. Thus, we see that the effect of this perturbation on the induced random walk on $\text{Cay}(\mathfrak{S}_n, S_n)$ is to increase the probability that left edges are taken by $\Theta(\theta\varepsilon)$. Recall from the proof of Theorem 1.2, the notation $W(\delta, n)$ for the weights on S_n which give added weight δ to the left edges. Denote

$$\delta = \frac{q_0}{1 + q_0 - (1 - q)\beta} \bigg/ \frac{1/2}{3/2 - \beta/2} - 1 = O(\theta\varepsilon).$$

Again, this gives a coupling between $\widehat{\sigma}_i$ to a random walk on $\text{Cay}(\mathfrak{S}_n, S_n, W(\delta, n))$ which succeeds with probability $\mathbb{P}(\mathcal{J}) = 1 - O(n^{-4})$ (the probability of $\mathcal{J}$ is not affected by the perturbation).

The only condition to apply the analysis of Theorem 1.2 is $\varepsilon/c_1 \le \delta \le 3$, where c_1 is from the proof of Theorem 1.2. Taking θ sufficiently large will ensure the condition $\delta \ge \varepsilon/c_1$ while the condition $\delta \le 3$ holds for n sufficiently large. Fix θ to satisfy this requirement. Thus, the analysis of the proof of Theorem 1.2 shows that the particle that was at the root of H_1 at time 0 (H_1 from the construction of G, and unrelated to the H from Lemma 4.1) is still in the gadget after $r' := cn^2u^24^u \asymp n^2 \log n \log \log n$ steps of the induced random walk, for c sufficiently small. Since the coupling between
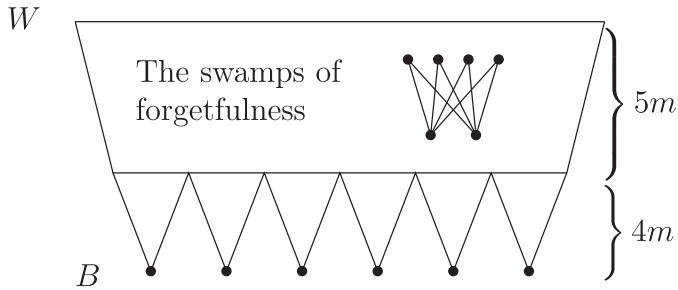
Figure 2: The clock graph. The triangles emanating from vertices of B are the trees, the area above them is the swamp.

$\widehat{\sigma}_i$ and the interchange process succeeds with high probability, this shows the same behavior for $\widehat{\sigma}_i$. This of course means that the random walk on L is not mixed. Using (4.3), we see that with high probability, by time $cr'|H|/2^m$, the induced walk still did not do r' steps, so we get

$$t_{\mathrm{mix}}(L, \text{perturbed weights}) \gtrsim cr'|H|/n^2 \asymp n^{20}\log n \log\log n$$

proving Theorem 1.3.   □

4.4 Proof of Lemma 4.1

Let $s \in \mathbb{N}$. For $0 \le \ell \le s$, we denote $\mathcal{A}_\ell := \{u^k_{i_1,\ldots,i_\ell} : i_1,\ldots,i_\ell \in [4], k \in [2^{s-\ell}]\}$ (for $\ell = 0$, this simply means $\mathcal{A}_0 = [2^s]$). For all $\ell \le s-1$, $i_1,\ldots,i_\ell \in [4]$ and $k \in [2^{s-\ell-1}]$, we connect both $u^k_{i_1,\ldots,i_\ell}$ and $u^{k+2^{s-\ell-1}}_{i_1,\ldots,i_\ell}$ to $u^k_{i_1,\ldots,i_\ell,1}$, $u^k_{i_1,\ldots,i_\ell,2}$, $u^k_{i_1,\ldots,i_\ell,3}$, and $u^k_{i_1,\ldots,i_\ell,4}$.

We start the construction of G with 2^m binary trees of depth $4m$. The set B is taken to be the collection of the 2^m roots. We label the union of the leaves of these trees by $[2^{5m}]$ so that each tree occupies an interval of values and identify it with $\mathcal{A}_0$ (with $s = 5m$, of course). Denote $W := \mathcal{A}_{5m}$. This terminates the construction of the graph from the statement of Lemma 4.1, denoted by H. The construction is depicted in Figure 2 with the trees depicted as triangles. The area above them, nicknamed "the swamps of forgetfulness," is composed of elements as in Figure 2, namely, two vertices below, four vertices above and all edges between them. These elements have the property that the particle forgets one bit whenever it traverses such an element, be it in the up or down direction. When the particle has traversed the swamp fully, it has completely forgotten its starting point. This construction is borrowed from [25, Section 6.2].

All of (4.1)–(4.4) follow because the distance from the roots behaves like random walk on $\mathbb{N}$ with a drift. Equation (4.1) follows because this requires to get to distance $4m$ from the roots and then back up. Equations (4.2) and (4.3) follow because with positive probability the walker hits W and then needs to back up $9m$ levels. Equation (4.4) is the easiest of the four, given (4.1).

Lastly, the claim that from every $v \in W$, the hitting distribution of B is uniform follows from the symmetries of the graph. Indeed, let $\varepsilon_1,\ldots,\varepsilon_{5m} \in \{0,1\}$, and let φ_k

be the map of adding the ε_i to the binary digits, namely,

$$\varphi_k\Big(\sum_{i=0}^{\ell-1} b_i 2^i\Big) = \sum_{i=0}^{\ell-1}(b_i + \varepsilon_i \bmod 2)2^i.$$

Then it is easy to check that the map ψ that takes $u^k_{i_1,\ldots,i_\ell}$ to $u^{\varphi_{5m-\ell}(k)}_{i_1,\ldots,i_\ell}$ is an automorphism of $\bigcup \mathcal{A}_\ell$ (as a graph). If, in addition, $\varepsilon_1 = \cdots = \varepsilon_{4m} = 0$ then this map, restricted to $\mathcal{A}_0$, has the property that if i and j are leaves of the same binary tree, then so are $\psi(i)$ and $\psi(j)$, and then ψ may be extended to an automorphism of the graph H. By appropriately choosing $\varepsilon_{4m+1}, \ldots, \varepsilon_{5m}$, one may get an automorphism ψ that takes b to b' for any two points of B. This shows the uniformity claim.

References

[1] L. Addario-Berry and M. I. Roberts, *Mixing time bounds via bottleneck sequences*. J. Stat. Phys. **173**(2018), nos. 3–4, 845–871.

[2] D. Aldous, *Some inequalities for reversible Markov chains*. J. London Math. Soc. **25**(1982), no. 3, 564–576.

[3] D. Aldous and J. Fill, *Reversible Markov chains and random walks on graphs*. Unfinished manuscript (2002). https://www.stat.berkeley.edu/~aldous/RWG/book.pdf

[4] G. Alon and G. Kozma, *Comparing with octopi*. Ann. Inst. Henri Poincaré Probab. Stat. **56**(2020), no. 4, 2672–2685.

[5] M. T. Barlow and M. Murugan, *Stability of the elliptic Harnack inequality*. Ann. Math. **187**(2018), no. 3, 777–823.

[6] R. Basu, J. Hermon, and Y. Peres, *Characterization of cutoff for reversible Markov chains*. Ann. Probab. **45**(2017), no. 3, 1448–1487.

[7] I. Benjamini, *Instability of the Liouville property for quasi-isometric graphs and manifolds of polynomial volume growth*. J. Theoret. Probab. **4**(1991), no. 3, 631–637.

[8] N. Berestycki, *Mixing times of Markov chains: Techniques and examples*. Unpublished lecture notes (2022). https://homepage.univie.ac.at/nathanael.berestycki/wp-content/uploads/2022/05/mixing3.pdf

[9] L. Boczkowski, Y. Peres, and P. Sousi, *Sensitivity of mixing times in Eulerian digraphs*. SIAM J. Disc. Math. **32**(2018), no. 1, 624–655.

[10] S. Boutayeb, *Heat kernel lower Gaussian estimates in the doubling setting without Poincaré inequality*. Publ. Mat. **53**(2009), no. 2, 457–479.

[11] E. Breuillard and M. C. Tointon, *Nilprogressions and groups with moderate growth*. Adv. Math. **289**(2016), 1008–1055.

[12] E. A. Carlen, S. Kusuoka, and D. W. Stroock, *Upper bounds for symmetric Markov transition functions*. Ann. Inst. H. Poincaré Probab. Statist. **23**(1987), no. 2, 245–287.

[13] G.-Y. Chen and L. Saloff-Coste, *Comparison of cutoffs between lazy walks and Markovian semigroups*. J. Appl. Probab. **50**(2013), no. 4, 943–959.

[14] P. Diaconis and L. Saloff-Coste, *Comparison theorems for reversible Markov chains*. Ann. Appl. Probab. **3**(1993), no. 3, 696–730.

[15] P. Diaconis and L. Saloff-Coste, *Moderate growth and random walk on finite groups*. GAFA **4**(1994), no. 1, 1–36.

[16] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*. Ann. Appl. Probab. **6**(1996), no. 3, 695–750.

[17] J. Ding and Y. Peres, *Sensitivity of mixing times*. Electron. Commun. Probab. **18**(2013), 1–6.

[18] P. G. Doyle and J. L. Snell, *Random walks and electric networks*, Carus Mathematical Monographs, 22, Mathematical Association of America, Washington, DC, 1984.

[19] N. Fountoulakis and B. A. Reed, *Faster mixing and small bottlenecks*. Probab. Theory Related Fields **137**(2007), nos. 3–4, 475–486.

[20] S. Goel, R. Montenegro, and P. Tetali, *Mixing time bounds via the spectral profile*. Electron. J. Probab. **11**(2006), 1–26.

[21] A. A. Grigor'yan, *The heat equation on noncompact Riemannian manifolds*. [Russian] Mat. Sb. **182**(1991), no. 1, 55–87. English translation in *Math. USSR — Sb.* **72**(1992), no. 1, 47–77.

[22] A. A. Grigor'yan, J. Hu, and K.-S. Lau, *Heat kernels on metric spaces with doubling measure*. In: C. Bandt, M. Zähle, and P. Mörters (eds.), Fractal geometry and stochastics IV, Progress in Probability, 61, Birkhäuser Verlag, Basel, 2009, pp. 3–44.

[23] M. Gromov, *Hyperbolic manifolds, groups and actions*. In: Riemann: Surfaces and related topics: Proceedings of the 1978 Stony Brook Conference (State University of New York, Stony Brook, NY, 1978), Annals of Mathematics Studies, 97, Princeton University Press, Princeton, NJ, 1981, pp. 183–213.

[24] J. Hermon, *On sensitivity of uniform mixing times*. Ann. Inst. Henri Poincaré Probab. Stat. **54**(2018), no. 1, 234–248.

[25] J. Hermon, H. Lacoin, and Y. Peres, *Total variation and separation cutoffs are not equivalent and neither one implies the other*. Electron. J. Probab. **21**(2016), no. 44, 1–36.

[26] J. Hermon and Y. Peres, *On sensitivity of mixing times and cutoff*. Electron. J. Probab. **23**(2016), no. 25, 1–34.

[27] J. Hermon and Y. Peres, *The power of averaging at two consecutive time steps: Proof of a mixing conjecture by Aldous and Fill*. Ann. Inst. Henri Poincaré Probab. Stat. **53**(2017), no. 4, 2030–2042.

[28] J. Hermon and Y. A. Peres, *Characterization of L_2 mixing and hypercontractivity via hitting times and maximal inequalities*. Probab. Theory Related Fields **170**(2018), nos. 3–4, 769–800.

[29] J. Hermon and R. Pymar, *The exclusion process mixes (almost) faster than independent particles*. Ann. Probab. **48**(2020), no. 6, 3077–3123.

[30] J. Hermon and J. Salez, *The interchange process on high-dimensional products*. Ann. Appl. Probab. **31**(2021), no. 1, 84–98.

[31] V. A. Kaimanovich and A. M. Vershik, *Random walks on discrete groups: Boundary and entropy*. Ann. Probab. **11**(1983), no. 3, 457–490.

[32] G. Kozma, *On the precision of the spectral profile*. ALEA Lat. Am. J. Probab. Math. Stat. **3**(2007), 321–329.

[33] D. Levin and Y. Peres, Markov chains and mixing times. 2nd ed., with contributions by E. L. Wilmer and a chapter by J. G. Propp and D. B. Wilson, American Mathematical Society, Providence, RI, 2017.

[34] L. Lovász and P. Winkler, *Mixing times*. In: D. Aldous and J. Propp (eds.), Microsurveys in discrete probability, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 41, American Mathematical Society, Providence, RI, 1998, pp. 85–133.

[35] R. Lyons and Y. Peres, Probability on trees and networks, Cambridge Series in Statistical and Probabilistic Mathematics, 42, Cambridge University Press, New York, 2016.

[36] R. Lyons and G. White, *Monotonicity for continuous-time random walks*. Preprint, 2021. arXiv:2109.10386

[37] T. Lyons, *Instability of the Liouville property for quasi-isometric Riemannian manifolds and reversible Markov chains*. J. Differ. Geom. **26**(1987), no. 1, 33–66.

[38] R. Montenegro and P. Tetali, *Mathematical aspects of mixing times in Markov chains*. Found. Trends Theor. Comput. Sci. **1**(2006), no. 3, 237–354.

[39] B. Morris and Y. Peres, *Evolving sets, mixing and heat kernel bounds*. Probab. Theory Related Fields **133**(2005), no. 2, 245–266.

[40] R. Oliveira, *Mixing of the symmetric exclusion processes in terms of the corresponding single-particle random walk*. Ann. Probab. **41**(2013), no. 2, 871–913.

[41] Y. Peres and P. Sousi, *Mixing times are hitting times of large sets*. J. Theoret. Probab. **28**(2015), no. 2, 488–519.

[42] C. Pittet and L. Saloff-Coste, *On the stability of the behavior of random walks on groups*. J. Geom. Anal. **10**(2000), no. 4, 713–737.

Department of Mathematics, University of British Columbia, Vancouver, BC, Canada
e-mail: jhermon@math.ubc.ca

Department of Mathematics, Weizmann Institute of Science, Rehovot, Israel
e-mail: Gady.Kozma@weizmann.ac.il