

On sensitivity of mixing times and cutoff*

Jonathan Hermon[†] Yuval Peres[‡]

Abstract

A sequence of chains exhibits (total variation) cutoff (resp., pre-cutoff) if for all $0 < \epsilon < 1/2$, the ratio $t_{\text{mix}}^{(n)}(\epsilon)/t_{\text{mix}}^{(n)}(1-\epsilon)$ tends to 1 as $n \rightarrow \infty$ (resp., the limsup of this ratio is bounded uniformly in ϵ), where $t_{\text{mix}}^{(n)}(\epsilon)$ is the ϵ -total variation mixing time of the n th chain in the sequence. We construct a sequence of bounded degree graphs G_n , such that the lazy simple random walks (LSRW) on G_n satisfy the “product condition” $\text{gap}(G_n)t_{\text{mix}}^{(n)}(\epsilon) \rightarrow \infty$ as $n \rightarrow \infty$, where $\text{gap}(G_n)$ is the spectral gap of the LSRW on G_n (a known necessary condition for pre-cutoff that is often sufficient for cutoff), yet this sequence does not exhibit pre-cutoff.

Recently, Chen and Saloff-Coste showed that total variation cutoff is equivalent for the sequences of continuous-time and lazy versions of some given sequence of chains. Surprisingly, we show that this is false when considering separation cutoff.

We also construct a sequence of bounded degree graphs $G_n = (V_n, E_n)$ that does not exhibit cutoff, for which a certain bounded perturbation of the edge weights leads to cutoff and increases the order of the mixing time by an optimal factor of $\Theta(\log |V_n|)$. Similarly, we also show that “lumping” states together may increase the order of the mixing time by an optimal factor of $\Theta(\log |V_n|)$. This gives a negative answer to a question asked by Aldous and Fill.

Keywords: reversible Markov chains; simple random walk; mixing time; cutoff, pre-cutoff; perturbations; sensitivity of cutoff; separation cutoff; counter-example.

AMS MSC 2010: 60J10.

Submitted to EJP on November 30, 2016, final version accepted on February 27, 2018.

Supersedes arXiv:1610.04357.

1 Introduction

Consider a reversible irreducible lazy discrete-time Markov chain $X = (X_t)_{t \geq 0}$, defined on a finite state space Ω (we call a chain finite if Ω is finite). Let P and π denote the transition matrix and the unique reversible probability measure associated

*The work on this paper started while the first author was an intern at Microsoft Research. The first author would like to thank Microsoft Research, as well as EPSRC grant EP/L018896/1 for financial support.

[†]University of Cambridge, Cambridge, UK. E-mail: jonathan.hermon@statslab.cam.ac.uk

[‡]Microsoft Research, Redmond, Washington, USA.

E-mail: peres@microsoft.com

to X , respectively (we denote such a chain by (Ω, P, π)). In particular, the *laziness* and *reversibility* assumptions are (resp.) that $P(x, x) \geq 1/2$ and $\pi(x)P(x, y) = \pi(y)P(y, x)$ for all $x, y \in \Omega$. To avoid periodicity and near-periodicity issues, one often considers the lazy version of a discrete time Markov chain, $(X_t^L)_{t=0}^\infty$, obtained by replacing P with $P_L := \frac{1}{2}(I + P)$. Periodicity issues can be avoided also by considering the continuous-time version of the chain, $(X_t^c)_{t \geq 0}$. This is a continuous-time Markov chain whose heat kernel is defined by $H_t(x, y) := \sum_{k=0}^\infty \frac{e^{-tk}}{k!} P^k(x, y)$. It is a classic result of probability theory that for any initial condition the distribution of both X_t^L and X_t^c converge to π when t tends to infinity. The object of the theory of Mixing times of Markov chains is to study the characteristic of this convergence (see [16] for a self-contained introduction to the subject).

For any two distributions μ, ν on Ω , their *total variation distance* is defined as $\|\mu - \nu\|_{TV} := \frac{1}{2} \sum_x |\mu(x) - \nu(x)|$. The worst-case total variation distance at time t is defined as $d(t) := \max_{x \in \Omega} \|P_x^t - \pi\|_{TV}$, where we denote by P_x^t (resp. P_x) the distribution of X_t (resp. $(X_t)_{t \geq 0}$), given that $X_0 = x$. The ϵ total variation *mixing time* is defined as

$$t_{\text{mix}}(\epsilon) := \inf \{t : d(t) \leq \epsilon\},$$

and for the continuous-time chain as $t_c(\epsilon) := \inf \{t : \max_x \|H_t(x, \cdot) - \pi(\cdot)\|_{TV} \leq \epsilon\}$.

Similarly, the (worst-case) *separation distance* from stationarity at time t (resp. for the continuous-time chain) is defined as

$$d_{\text{sep}}(t) := 1 - \min_{x, y \in \Omega} P^t(x, y)/\pi(y), \quad d_{\text{sep}}^c(t) := 1 - \min_{x, y \in \Omega} H_t(x, y)/\pi(y)$$

and the ϵ *separation time* (the “ ϵ separation mixing time”) is defined as

$$t_{\text{sep}}(\epsilon) := \inf \{t : d_{\text{sep}}(t) \leq \epsilon\}, \quad (\text{resp. } t_{\text{sep},c}(\epsilon) := \inf \{t : d_{\text{sep}}^c(t) \leq \epsilon\}.$$

When $\epsilon = 1/4$ we omit it from the above notation.

In this work we consider lazy simple random walks (**LSRW**) on a sequence of finite (uniformly) bounded degree connected graphs, $G_n := (V_n, E_n)$, whose sizes tend to infinity and also lazy random walks on a sequence of networks obtained from them via a bounded perturbation of their edge weights (we defer the formal definition to § 1.3).

Following [10], where Ding and the second author showed that a bounded perturbation can increase the order of the total variation mixing time (we state their result in more details in § 2), we study (by constructing relevant examples) the possible effects of such bounded perturbations on the convergence to the stationary distribution (of the corresponding lazy random walks on the perturbed networks compared to the original LSRWs). In particular, our Theorem 3 asserts that such bounded perturbations can increase the order of the total variation mixing-times by an optimal (as explained in (1.6) below) factor of $\Theta(\log |V_n|)$.

While the aforementioned result is merely an improvement and a simplification of [10, Theorem 1.1], various aspects related to sensitivity of mixing times and the *cutoff phenomenon* (of LSRW on a sequence of uniformly bounded degree graphs $G_n = (V_n, E_n)$) are considered in this work for the first time:

- (1) We consider $o(1)$ -perturbations (in which the weight of each edge may increase only by a $1 + o(1)$ multiplicative factor) and show that they may increase the order of the mixing time by a factor whose order is arbitrarily close to $\log |V_n|$ (part (b) of Theorem 3).
- (2) We consider sensitivity of mixing-times under *lumping* (Definition 1.5). We show that (even in the bounded degree unweighted setup) lumping together pairs of vertices may increase the order of the mixing time by an optimal factor of $\Theta(\log |V_n|)$

(part (c) of Theorem 3). This provides a negative answer to a question by Aldous and Fill [2, Problem 4.45] (Problem 1.6 here).

- (3) We show that (in the above setup) the mixing time of the lazy non-backtracking random walk may be larger than that of the LSRW by a factor of $\Theta(\log |V_n|)$ (Remark 4.6). A similar example (in which the ratio of the mixing times is $o(\log |V_n|)$) was recently constructed by Hubert Lacoïn et al. during an AIM workshop on mixing times of Markov chains.
- (4) We show that the occurrence/non-occurrence of cutoff/pre-cutoff (see (1.1)-(1.2)) is sensitive under $o(1)$ perturbations of the edge weights (Theorems 2 and 3). We also show that even in the above setup, the *product condition* (1.3) need not imply pre-cutoff (Theorem 2).
- (5) Perhaps our most surprising result (Theorem 1) is that the occurrence of *separation cutoff* (1.1) for the sequence of discrete-time lazy chains, does not imply the same for the associated sequence of continuous-time chains (this can be interpreted as a “sensitivity” result w.r.t. the choice of discrete/continuous time¹). This is in contrast with the case of total variation cutoff [7] due to Chen and Saloff-Coste.

In [13] the first author constructed a sequence of pairs of 2-roughly isometric graphs $G_n = (V_n, E_n), G'_n = (V'_n, E'_n)$ of uniformly bounded degree with $|V_n| \rightarrow \infty$, whose ℓ_∞ -mixing times differ by an optimal factor of $\Theta(\log \log |V_n|)$. In this paper we study the convergence to the stationary distribution only w.r.t. the total variation distance and the separation distance.

1.1 The general moral of our results

We now discuss the moral of our results. An important question is whether mixing times are robust. A related question is whether they can be characterized (perhaps only up to some universal constants and only under reversibility) using a geometric quantity which is robust. Different variants of this question were asked by various authors such as Pittet and Saloff-Coste [23], Kozma [15, p. 4], Diaconis and Saloff-Coste [8, p. 720] and Aldous and Fill [2, Open Problem 8.23] (Kozma conjectured that the ℓ_∞ -mixing time is robust under rough-isometries for LSRWs on bounded degree graphs, and the last two references ask for an extremal characterization of the ℓ_∞ -mixing time in terms of the Dirichlet form).

There are numerous works aiming at sharp geometric bounds on mixing-times such as the Fountoulakis-Reed bound [11] (recently refined by Addario-Berry and Roberts [1]) on the total variation mixing time and Morris and Peres’ evolving sets bound [20] on the ℓ_∞ -mixing time, both expressed in terms of the expansion profile of the graph. The sharpest geometric bounds on the ℓ_∞ -mixing time are given by the spectral profile, due to Goel et al. [12] and by the Log-Sobolev bound [8, Corollary 3.11] due to Diaconis and Saloff-Coste. Because these bounds involve geometric quantities, they are robust under small changes to the geometry, like bounded perturbations of the edge weights or (in the bounded degree unweighted setup) under rough-isometries.

Our results strengthen the cautionary note of Ding and Peres [10] (and also of [13]) on the possibility of developing a sharp geometric bound on mixing-times. Indeed, any sharp bound would have to distinguish in some cases between the LSRW on a graph and the walk obtained by some $o(1)$ -perturbation.

Although receiving much attention, the investigation of the cutoff phenomenon has progressed mostly through the study of examples (or of certain classes of chains) rather than by developing general theory. Our results concerning sensitivity of the cutoff

¹A similar sensitivity holds w.r.t. the choice of holding probability (see Remark 1.2).

phenomenon (namely, Theorem 1 and parts of Theorems 2 and 3) demonstrate some difficulties in developing such general theory.

Theorem 1 demonstrates that despite the fact that the separation and the total variation distances are intimately related to one another (e.g. [14, Eq. (1.5), (1.7) and (1.8)]), the former may exhibit surprising behaviors which the latter cannot exhibit. For more on this point see [14, Remark 1.4 and § 2.4].

1.2 Cutoff and pre-cutoff

Before stating our results concerning the cutoff phenomenon we must first give a few definitions. Next, consider a sequence of chains, $((\Omega_n, P_n, \pi_n) : n \in \mathbb{N})$, each with its corresponding worst-distances from stationarity $d^{(n)}(t)$, $d_{\text{sep}}^{(n)}(t)$, its mixing and separation times $t_{\text{mix}}^{(n)}$, $t_{\text{sep}}^{(n)}$, etc.. Loosely speaking, the total variation (resp. separation) **cutoff phenomenon** is said to occur when over a negligible period of time, known as the **cutoff window**, the worst-case total variation distance (resp. separation distance) drops abruptly from a value close to 1 to near 0. In other words, one should run the n -th chain until time $(1 - o(1))t_{\text{mix}}^{(n)}$ (resp. $(1 - o(1))t_{\text{sep}}^{(n)}$) for it to even slightly mix in total variation (resp. separation), whereas running it any further after time $(1 + o(1))t_{\text{mix}}^{(n)}$ (resp. $(1 + o(1))t_{\text{sep}}^{(n)}$) is essentially redundant. Formally, we say that the sequence exhibits a **total variation cutoff** (resp. **separation cutoff**) if the following sharp transition in its convergence to stationarity occurs:

$$\forall \varepsilon \in (0, 1/2], \quad \lim_{n \rightarrow \infty} t_{\text{mix}}^{(n)}(\varepsilon)/t_{\text{mix}}^{(n)}(1 - \varepsilon) = 1 \quad \left(\text{resp.} \quad \lim_{n \rightarrow \infty} t_{\text{sep}}^{(n)}(\varepsilon)/t_{\text{sep}}^{(n)}(1 - \varepsilon) = 1 \right). \quad (1.1)$$

We say that the sequence exhibits a **(total variation) pre-cutoff** if

$$\sup_{0 < \varepsilon < 1/2} \limsup_{n \rightarrow \infty} t_{\text{mix}}^{(n)}(\varepsilon)/t_{\text{mix}}^{(n)}(1 - \varepsilon) < \infty. \quad (1.2)$$

The notions of total variation and separation cutoff for the corresponding sequence of continuous-time chains are defined in an analogous manner ($\lim_{n \rightarrow \infty} t_c^{(n)}(\varepsilon)/t_c^{(n)}(1 - \varepsilon) = 1$ and $\lim_{n \rightarrow \infty} t_{\text{sep},c}^{(n)}(\varepsilon)/t_{\text{sep},c}^{(n)}(1 - \varepsilon) = 1$, resp., for all $\varepsilon \in (0, 1)$). One can also consider the mixing time and separation-time for the sequence of associated lazy chains and define the two notions of cutoff for it. Recently, Chen and Saloff-Coste [7] showed that if $t_c^{(n)} \rightarrow \infty$ then the sequence of the associated continuous-time chains exhibits total variation cutoff iff the same holds for the sequence of the associated lazy chains. A natural question (L. Saloff-Coste, private communication) is whether the same is true for cutoff in separation. Surprisingly, this turns out to be false.

Theorem 1. There exists a sequence of reversible chains so that the lazy chains exhibit separation cutoff but the associated continuous-time chains do not exhibit cutoff.

Remark 1.1. The example we give for Theorem 1 is intimately related to Example 3 in [14] and could be transformed into an example involving simple random walks on a sequence of bounded degree graphs using a similar construction as Example 5 in [14].

Remark 1.2. The δ -lazy version of a chain with transition matrix P is obtained by replacing P with $\delta I + (1 - \delta)P$ (where I is the identity matrix). Chen and Saloff-Coste [7] showed that given a sequence of chains, the corresponding sequence of $1/2$ -lazy chains exhibits total variation cutoff iff the same holds for the corresponding sequence of p -lazy chains, for all $p \in (0, 1)$. One can use the idea behind the construction from the proof of Theorem 1 in order to construct a family of reversible examples demonstrating that for all $p \neq q \in (0, 1)$ it is possible that the sequence of q -lazy versions of a certain sequence of chains exhibits separation cutoff, while the sequence of p -lazy versions does

not exhibit separation cutoff or vice-versa². The necessary adaptations are described in § 3.3.

Question 1.3. Is it the case that separation cutoff for the sequence of continuous-time chains implies the same for the sequence of lazy chains?

Recall that if (Ω, P, π) is a finite reversible irreducible lazy chain, then P is self-adjoint w.r.t. the standard inner-product induced by π (see Definition 6.1) and hence has $|\Omega|$ real eigenvalues, $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} \geq 0$ (where $\lambda_2 < 1$ and $\lambda_{|\Omega|} \geq 0$ by irreducibility and laziness). The *spectral-gap* and *relaxation-time* are defined as $\text{gap} := 1 - \lambda_2$ and $t_{\text{rel}} := (1 - \lambda_2)^{-1}$, resp.. We denote the relaxation-time of LSRW on $G = (V, E)$ by $t_{\text{rel}}(G)$.

In 2004 [21], during an AIM workshop on the cutoff phenomenon, the second author introduced the so called **product condition**:

$$t_{\text{rel}}^{(n)} = o(t_{\text{mix}}^{(n)}), \tag{1.3}$$

a necessary condition for pre-cutoff (by (1.5); e.g. [16, Proposition 18.4]), and suggested that it is also a sufficient condition for cutoff for many “nice” families of reversible chains. In general, the product condition does not always imply cutoff. Aldous and Pak (private communication via P. Diaconis) have constructed relevant (reversible) examples (see [16, Chapter 18], Pak’s example is described in § 2). This left open the problem of identifying general classes of chains for which the product condition is indeed sufficient for cutoff. This was verified e.g. for lazy birth and death chains [9] and recently for lazy weighted walks on trees [5].

In Aldous’ example the graph supporting the transitions is of bounded degree and contains only a single cycle, however the ratio of the maximal and minimal edge weights is exponentially large in the size of the state space. As noted in [17], Aldous’ example, which exhibits pre-cutoff, can be transformed into a sequence of LSRWs on bounded degree graphs (with pre-cutoff). Explicit constructions of such graphs (which were constructed as examples demonstrating that, in general, neither total variation cutoff nor separation cutoff implies the other) can be found in [14].

Until now, as in Pak’s example, every known example in which the product condition does not imply pre-cutoff had unbounded degrees. These examples all share the following behavior. The chain mixes (in some sense) “at once” due to the occurrence of a certain rare event, which occurs before the chain has enough time to get even slightly mixed otherwise. It is plausible that (some concrete formulation of) the aforementioned behavior is necessary in order for pre-cutoff to fail when the product condition holds. Moreover, a-priori, it is not clear whether the mechanism that allows such behavior can be produced in the bounded degree (unweighted) setup.

A question, presented to us by E. Lubetzky (private communication), which naturally arises in light of the above discussion, is whether the product condition is also a sufficient condition for pre-cutoff for a sequence of LSRWs on bounded degree graphs $\{G_n\}_{n \in \mathbb{N}}$. One case in which this holds (as a simple consequence of ℓ_2 -contraction; see Lemma 6.2) is when $\{G_n\}_{n \in \mathbb{N}}$ is a family of bounded degree expanders (that is, $t_{\text{rel}}(G_n) = \Theta(1)$).

Problem 1.4 (E. Lubetzky (private communication)). Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of finite (uniformly) bounded degree graphs satisfying the product condition. Is it always the case that the sequence of lazy simple random walks on $\{G_n\}_{n \in \mathbb{N}}$ exhibits a pre-cutoff?

²Our analysis shows that for all $p \neq q \in (0, 1)$ it is possible to construct an example in which separation cutoff fails for exactly one of the corresponding lazy chains. We do not determine for which of them separation cutoff fails. Determining which of the two chains exhibits separation cutoff requires a more detailed comparison of certain two large deviation rate functions, where for our analysis it suffices to use the fact that the two rate functions are not identical.

Our Theorem 2 provides a negative answer to this question. Our construction may be viewed as a bounded degree (unweighted) version of the aforementioned Pak's example (see § 2 and Remark 5.1 for further details concerning this point).

1.3 Perturbations of edge weights and lumping

Let $G = (V, E)$ be a finite connected simple graph. Given a (weighted) network $(V, E, (c_e)_{e \in E})$, where each edge $\{u, v\} \in E$ is endowed with a conductance $c_{u,v} = c_{v,u} > 0$ (with the convention that $c_{u,v} = 0$ if $\{u, v\} \notin E$), a lazy random walk on $G = (V, E)$ repeatedly does the following: when the current state is $v \in V$, the random walk will stay at v with probability $1/2$ and move to vertex u (such that $\{u, v\} \in E$) with probability $c_{u,v}/(2c_v)$, where $c_v := \sum_w c_{v,w}$. The default choice for $c_{u,v}$ is 1 (in which case, we say that the random walk is *unweighted*), which corresponds to lazy simple random walk on G (in which at each step the walk with equal probability either stays put or moves to a new vertex, chosen from the uniform distribution over the neighbors of the current position). Its stationary distribution is given by $\pi(x) := c_x/c_V$, where $c_V := \sum_{v \in V} c_v = 2 \sum_{e \in E} c_e$. We denote by $t_{\text{mix}, G, (c_e)}(\epsilon)$ the total variation ϵ -mixing time of the lazy random walk on the network induced on the graph G by the edge weights $(c_e)_{e \in E}$. When the walk is unweighted we omit $(c_e)_{e \in E}$ from our notation. As always, when $\epsilon = 1/4$ we omit it from the notation. In the unweighted setup we write $t_{\text{mix}}(G)$ for the $1/4$ -mixing time.

Definition 1.5. Given some network $(V, E, (c_e)_{e \in E})$ with stationary distribution π , consider a partition of V into disjoint sets A_1, \dots, A_k with $\cup_{i=1}^k A_i = V$. The induced network, obtained by lumping together the states belonging to A_i for all $1 \leq i \leq k$ is a network on $[k] := \{1, \dots, k\}$ with transition probabilities $\hat{p}_{i,j} := P_\pi[X_1 \in A_j \mid X_0 \in A_i]$. It can be obtained by collapsing each A_i into a single state i and setting the edge weight of the edge connecting i and j to be $\hat{c}_{i,j} := \sum_{a_i \in A_i, a_j \in A_j} c_{a_i, a_j}$ for all $(i, j) \in [k] \times [k]$ (in particular, it is reversible).

Proposition 4.44 in [2] asserts that several natural parameters of a reversible Markov chain, like the inverses of its spectral gap and cheeger constant, can only decrease as a result of lumping states together. This motivates the following question asked by Aldous and Fill [2, Open Problem 4.45].

Problem 1.6. Is it the case that lumping states together can increase the total variation mixing time of a reversible Markov chain by at most some constant factor K ? Can one take $K = 1$?

Our Theorem 3 (part (c)) gives a negative answer to Problem 1.6.

We say that $(w_e^{(n)})_{e \in E_n}$ is a sequence of **bounded perturbations** (respectively **$o(1)$ -perturbations**) if $w_e^{(n)} \geq 1$ for all $e \in E_n$ for all n and $M((w_e^{(n)})) := \sup_n M_n < \infty$ (resp. $M_n = 1 + o(1)$), where $M_n := \sup_{e \in E_n} w_e^{(n)}$.³

We say that $\{G_n\}_{n \in \mathbb{N}}$ is **robust** if performing a sequence of bounded perturbations, $(w_e^{(n)})_{e \in E_n}$ preserves the mixing-times up to a constant factor of K (independent of n , but which may depend on $M((w_e^{(n)}))$). We say that $\{G_n\}_{n \in \mathbb{N}}$ is **sensitive** if it is not robust. We say that $\{G_n\}_{n \in \mathbb{N}}$ is **$o(1)$ -sensitive** if there exists a sequence of $o(1)$ -perturbations which either increases or decreases the order of the mixing-times.

The girth of a graph G is defined as the minimal length of a cycle in G . In general (even in the weighted case) $\text{girth}(G) \leq 2\text{diameter}(G) + 1 \leq Ct_{\text{mix}}(G)$.

Theorem 2. For every $f(n) = o(\log n / \log \log n)$ such that $\lim_{n \rightarrow \infty} f(n) = \infty$, there exists a sequence of bounded degree graphs $G_n = (V_n, E_n)$ satisfying

(i) $\text{girth}(G_n) = \Theta(t_{\text{mix}}(G_n))$.

³Note that there is no loss of generality in the requirement that $w_e^{(n)} \geq 1$ for all $e \in E_n$, since multiplying all of the edge weights by the same constant has no effect on the distribution of the walk.

- (ii) The corresponding sequence of LSRWs satisfies the product condition but does not exhibit pre-cutoff.
- (iii) There exists a sequence of $o(1)$ -perturbations, $(c_e^{(n)})$, such that the corresponding sequence of lazy walks exhibits a cutoff and satisfies

$$t_{\text{mix}, G_n, (c_e^{(n)})} / t_{\text{mix}}(G_n) = \Theta(f(|V_n|)). \tag{1.4}$$

The following remark explains the significance of the large girth condition.

Remark 1.7. In [5] it was shown that a sequence of lazy random walks on weighted trees exhibits cutoff iff it satisfies the product condition. In [22] it was proved that the mixing time of (possibly weighted) nearest neighbor lazy walks on trees is robust (see [1, Theorem 1] for a recent extension of this result). Combining the two results it follows that for lazy weighted walks on trees the property of exhibiting cutoff is robust. Theorem 2 asserts that the tree assumption in these two results cannot be relaxed to the condition that $\text{girth}(G_n) = \Theta(t_{\text{mix}}(G_n))$ (even in the unweighted setup, and even when considering only pre-cutoff, instead of cutoff).

- Theorem 3.**
- (a) There exists a sequence of bounded degree graphs $G_n = (V_n, E_n)$ satisfying $t_{\text{rel}}(G_n) = \Theta(t_{\text{mix}}(G_n))$ (thus lacking pre-cutoff) such that for every $\epsilon > 0$, increasing the edge weight of some of the edges of G_n to $1 + \epsilon$ increases the mixing time by a factor of $c_\epsilon \log |V_n|$, for some constant $c_\epsilon > 0$ depending only on ϵ . Moreover, the sequence of walks on the perturbed networks exhibits a cutoff.
 - (b) For every $f(n) = o(\log n)$ such that $\lim_{n \rightarrow \infty} f(n) = \infty$ there exists a sequence of graphs $G_n = (V_n, E_n)$ satisfying $t_{\text{rel}}(G_n) = \Theta(t_{\text{mix}}(G_n))$ for which there exists a sequence of $o(1)$ -perturbations which increases the mixing-times by a factor of $\Theta(f(|V_n|))$. Moreover, the sequence of lazy walks on the perturbed networks exhibits a cutoff.
 - (c) There exists a sequence of bounded degree graphs $G_n = (V_n, E_n)$ such that lumping together some pairs of vertices of G_n increases the order of the total variation mixing time by a factor of $\Theta(\log |V_n|)$.

Remark 1.8. We note that the $\log |V_n|$ factor in part (a) of Theorem 3 is optimal. The following general relation holds for lazy reversible chains (e.g. [16] Theorems 12.3-12.4).

$$(t_{\text{rel}} - 1) |\log(2\epsilon)| \leq t_{\text{mix}}(\epsilon) \leq t_{\text{rel}} \left| \log \left(\epsilon \min_x \pi(x) \right) \right|. \tag{1.5}$$

Let $G = (V, E)$ be a graph of maximal degree D . Let $(c_e)_{e \in E}$ satisfy $K^{-1} \leq c_e \leq K$ for all $e \in E$. By the extremal characterization of the spectral gap via the Dirichlet form (e.g. [16, Remark 13.13]), the relaxation-time is robust under bounded perturbations (i.e. a bounded perturbation can change it only by a constant factor). Consequently, by (1.5),

$$1/(C_{D,K} \log |V|) \leq t_{\text{mix}, G, (c_e)} / t_{\text{mix}}(G) \leq C_{D,K} \log |V|, \tag{1.6}$$

for some constant $C_{D,K}$ depending only on D and K . Similarly, since as mentioned earlier, lumping does not increase t_{rel} , also part (c) is optimal, up to a constant factor.

Remark 1.9. We note that in part (a) of Theorem 3 we have that $t_{\text{mix}}(G_n) = \Theta(\log |V_n|)$, and so the mixing time of the perturbed network is $\Theta([t_{\text{mix}}(G_n)]^2)$. This is also optimal by (1.5) and the bound $\text{diam}(G_n) \geq c \log |V_n|$ (for some $c > 0$ depending only on the maximal degree).

Remark 1.10. If the function f in part (b) above is taken to tend to infinity sufficiently slowly, we can have that the $o(1)$ -perturbation from part (b) increases the edge weight by a factor of $1 + \delta_n$ for some $\delta_n = o(1)$ such that $\delta_n \sqrt{t_{\text{mix}}(G_n)}$ tends to infinity arbitrarily slowly. An interesting problem is to determine how small can δ_n be taken in terms of $t_{\text{mix}}(G_n)$ (or to construct such an example with $\delta_n = o([t_{\text{mix}}(G_n)]^{-\alpha})$ for some $\alpha > 1/2$).

Remark 1.11. It is interesting to note that in all of our examples other than the one in the proof of Theorem 1, up to a constant factor, the mixing time of the LSRW is independent of the starting position of the walk (in fact, it is $\Theta(\log |V|)$ in all of our examples).

1.4 Open problems

Let $\{G_n\}_{n \in \mathbb{N}}$ be a sequence of constant degree vertex-transitive graphs.

Problem 1.12 (Re-iterated from [10]). Is $\{G_n\}_{n \in \mathbb{N}}$ robust?

Problem 1.13. Assume that $\{G_n\}_{n \in \mathbb{N}}$ satisfies the product condition. Is it true that the corresponding sequence of LSRWs must exhibit a pre-cutoff? If so, is the pre-cutoff robust?

Similarly, one can consider robustness w.r.t. rough isometries.

1.5 Notation

For every $n \in \mathbb{N}$ we denote $[n] := \{1, 2, \dots, n\}$. For any $a, b \in \mathbb{R}$ we write $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$. Throughout, we use C, C', C_0, C_1, \dots and c, c', c_0, c_1, \dots to denote positive absolute constants that may be different from place to place. Given some parameter, say ϵ , we write C_ϵ and c_ϵ for positive constants which depend only on ϵ . Upper (resp. lower) case letters will be used to denote sufficiently large (resp. small) constants.

2 Related constructions

As mentioned earlier, Ding and Peres have already constructed a sequence of sensitive bounded degree graphs [10]. More precisely, for all $j \in \mathbb{N}$ they constructed a sequence of bounded degree graphs $G_n = (V_n, E_n)$ for which if some of the edge weights are doubled, then the order of the mixing times increases by a multiplicative factor of order $\log |V_n| / \log^{(j)} |V_n|$, where $\log^{(j)}$ is the iterated logarithm of order j (see [10, Remark 2.3])⁴.

Our constructions use a key observation from [10]. Namely, we use the fact that the harmonic measure is sensitive under perturbations and that (as explained below) this can lead to sensitivity of mixing-times. This idea was originally used by Benjamini [6] to study instability of the Liouville property. We note that our construction in the proof of Theorem 2 was greatly influenced by Ding and Peres' construction and is intimately related also to the construction from [13] of a sequence of graphs of uniformly bounded degree whose ℓ_∞ mixing time is sensitive under bounded perturbations.

The first construction of a sequence of finite irreducible lazy reversible chains satisfying the product condition, which does not exhibit pre-cutoff is due to Pak (private communication through Persi Diaconis, see [16, Example 18.7]). Pak's construction gives a general scheme of constructing such sequences of Markov chains. Start with a sequence of lazy reversible chains (Ω_n, P_n, π_n) which exhibits cutoff $(\pi_n P_n = \pi_n)$. Let Π_n be the transition matrix whose rows all equal π_n . Denote $L_n := \sqrt{t_{\text{rel}}^{(n)} t_{\text{mix}}^{(n)}}$. Then the sequence $(\Omega_n, (1 - \frac{1}{L_n})P_n + \frac{1}{L_n}\Pi_n, \pi_n)$ satisfies the product condition but does not exhibit a pre-cutoff.

Loosely speaking, the chain mixes "at once", at a random time having a Geometric distribution with mean L_n , due to the occurrence of one rare event (moving according to Π_n) which (with high probability) occurs before the chain has enough time to get even

⁴The construction is obtained by iterating the construction of the case $j = 2$ and is hence somewhat more involved.

slightly mixed otherwise. At first sight, it is surprising that it is possible to construct an example of bounded degree graphs so that the corresponding sequence of LSRWs imitates this behavior. In our examples, mixing occurs quickly once the chain reaches its “center of mass”, which is an expander. This allows us to reduce the analysis of the mixing time and the occurrence/non-occurrence of cutoff to the easier problem of analyzing the distribution of the hitting time of the center of mass (namely, the mixing time is roughly equal to its mean, and total variation cutoff is equivalent to it being concentrated around its mean).

In the construction from the proof of Theorem 2, the distribution of the hitting time of the “center of mass” (starting from the worst starting state) would be roughly a geometric distribution. Loosely speaking, starting from the worst starting state, until the time the center of mass is reached, the chain looks like a LSRW on a regular tree whose edges were stretched (i.e. replaced by a long path whose length tends to infinity) with some “shortcuts” to the center of mass (see Figures 4-7). The amount and positions of these shortcuts can be chosen so that the center of mass is reached (with high probability) through one of these shortcuts at a random time having roughly a geometric distribution. Moreover, this can be done so that, under a certain perturbation of the edge weights, the harmonic measure is changed in a manner which makes these shortcuts “invisible” to the walk. Hence after the perturbation, the walk (starting from the root) is “trapped” in the tree of stretched edges for a much longer period of time, which results in an increased mixing time.

The idea behind the construction of the graphs G_n from Theorem 2 is simple. Start with an arbitrary sequence of constant degree expanders $H_n = (V(H_n), E(H_n))$ (with $|V(H_n)| \rightarrow \infty$) whose girth is $\ell_n := \Theta(\log |V(H_n)|)$ (see e.g. [18] for the existence of such graphs). We pick some vertex $o \in V(H_n)$ and choose a certain collection of vertices D , all within distance $< \ell_n/2$ from o . Finally, for all $d \in D$ we replace each edge along the shortest path between o and d by a path of length s_n , where $s_n \rightarrow \infty$ and $s_n^2 = o(\ell_n / \log \ell_n)$. With some care, the set D can be chosen (in a canonical manner) so that:

- The asymptotic profile of convergence in total variation of the walk can be understood in terms of the distribution (under P_o) of the escape time from the collection of vertices which are incident to the stretched edges.
- This escape time distribution is “close” to the Geometric distribution with mean $\Theta(\ell_n)$. However, under a certain (canonical) $o(1)$ -perturbation, this escape time distribution becomes concentrated around some time $t_n = \Theta(s_n^2 \ell_n)$.

In [17] Lubetzky and Sly gave an explicit construction of 3-regular expanders, G_n , such that the sequence of lazy simple random walks on G_n exhibits total variation cutoff. Our construction in the proof of Theorem 2 resembles their construction (namely, in [17] they also “stretch” some of the edges of an expander).

3 Proof of Theorem 1

3.1 Preliminaries

Throughout this section we denote the distribution of the associated lazy (resp. continuous-time) chain started from x by P_x (resp. H_x). We denote the transition matrix of the non-lazy (resp. lazy) version of the chain by P (resp. P_L). We denote the separation distance at time t of the continuous-time and lazy chains by $d_{\text{sep},c}(t)$ and $d_{\text{sep},L}(t)$, respectively.

Before presenting the construction for Theorem 1 we first provide some technical machinery, borrowed from [14], which shall assist us in analyzing the asymptotic profile

of convergence in separation. To characterize the separation time, we introduce a notion of “double-hitting time”.

Definition 3.1. Given x, y and z in Ω . We let $T_z^{x,y}$ (resp. $\tau_z^{x,y}$) denote a random variable obtained by taking the sum of two independent realizations of $T_z := \inf\{t : X_t = z\}$, once under P_x and once under P_y (resp. H_x and H_y). More explicitly, we consider realizations of the lazy and continuous-time chains started from x and y , denoted resp., by (X_t^x) and $(X_t^{x,c})$ (resp. (X_t^y) and $(X_t^{y,c})$) defined on the same probability space, so that (X_t^x) and (X_t^y) and also $(X_t^{x,c})$ and $(X_t^{y,c})$ are independent. Denote by $T_u^x := \inf\{t : X_t^x = u\}$ (resp. $\tau_u^x := \inf\{t : X_t^{x,c} = u\}$) the hitting time of state u by (X_t^x) (resp. $(X_t^{x,c})$). Define T_u^y and τ_u^y in an analogous manner. We set

$$T_z^{x,y} := T_z^x + T_z^y, \quad \tau_z^{x,y} := \tau_z^x + \tau_z^y.$$

We define

$$T_{z,y}^x := T_y^x \mathbf{1}_{\{T_y^x \leq T_z^x\}} + (T_z^x + T_z^y) \mathbf{1}_{\{T_y^x > T_z^x\}}$$

$$\tau_{z,y}^x := \tau_y^x \mathbf{1}_{\{\tau_y^x \leq \tau_z^x\}} + (\tau_z^x + \tau_z^y) \mathbf{1}_{\{\tau_y^x > \tau_z^x\}}.$$

Finally, we denote the density function of $\tau_z^{x,y}$ by $f_z^{x,y}$ and that of the sub-distributions of $\tau_y^x \mathbf{1}_{\{\tau_y^x \leq \tau_z^x\}}$ and $(\tau_z^x + \tau_z^y) \mathbf{1}_{\{\tau_y^x > \tau_z^x\}}$ by $g_{z,y}^x$ and $h_{z,y}^x$, respectively.

The following lemma is a slight variation of Lemma 3.5 from [14]. We present its proof in § 6.4 for the sake of completeness.

Lemma 3.2. Let (Ω, P, π) be a finite reversible Markov chain. Consider $x, y, z \in \Omega$.

(i) For all $t \geq 0$ we have that

$$P_L^t(x, y)/\pi(y) = \sum_{k:k \leq t} \mathbb{P}[T_{z,y}^x = k, T_y^x \leq T_z^x] P_L^{t-k}(y, y)/\pi(y)$$

$$+ \sum_{k:k \leq t} \mathbb{P}[T_{z,y}^x = k, T_y^x > T_z^x] P_L^{t-k}(z, z)/\pi(z) \geq \mathbb{P}[T_{z,y}^x \leq t]. \quad (3.1)$$

$$H_t(x, y)/\pi(y) = \int_0^t \left(\frac{g_{z,y}^x(s) H_{t-s}(y, y)}{\pi(y)} + \frac{h_{z,y}^x(s) H_{t-s}(z, z)}{\pi(z)} \right) ds \geq \mathbb{P}[\tau_{z,y}^x \leq t].$$

In particular,

$$P_L^t(x, y)/\pi(y) \geq P_x[T_y \leq t], \quad H_t(x, y)/\pi(y) \geq H_x[T_y < t]. \quad (3.2)$$

(ii) If $P_x[T_z \leq T_y] = 1$ (i.e. if every path from x to y goes through z) then for all $t \geq 0$

$$P_L^t(x, y)/\pi(y) = \sum_{k:k \leq t} \mathbb{P}[T_z^{x,y} = k] P_L^{t-k}(z, z)/\pi(z),$$

$$H_t(x, y)/\pi(y) = \int_0^t f_z^{x,y}(s) H_{t-s}(z, z) ds / \pi(z) \quad (3.3)$$

$$\leq \mathbb{P}[\tau_z^{x,y} \leq t] + t_{\text{rel}} \max_s f_z^{x,y}(s) (1 - \pi(z)) / \pi(z).$$

In particular, if f_y^x is the density of the hitting time of y from x , then

$$H_t(x, y)/\pi(y) \leq H_x[T_y \leq t] + t_{\text{rel}} \max_s f_y^x(s) (1 - \pi(y)) / \pi(y). \quad (3.4)$$

The following example, Example 3.3 (a birth and death chain of size $2n + 1$ with a fixed bias towards its middle point), will serve as a gadget in the construction for Theorem 1. In Example 3.3 the state z serves as the center of mass and has a $\Theta(1)$ stationary probability. The chain from Example 3.3 exhibits cutoff in separation around the expectation of the “double hitting time” of z from a and b .

Example 3.3. Let $\Omega := A \cup B \cup \{z\}$, where $A := \{a_1, a_2, \dots, a_n\}$, $B := \{b_1, b_2, \dots, b_n\}$. Denote $a_0 := z := b_0$, $a := a_n$ and $b := b_n$. Consider the transition matrix $P(a, a_{n-1}) = 1 = P(b, b_{n-1})$, $P(a_i, a_{i-1}) = P(b_i, b_{i-1}) = 2/3 = 2P(a_i, a_{i+1}) = 2P(b_i, b_{i+1})$, for $1 \leq i < n$ and $P(z, a_1) = 1/2 = P(z, b_1)$. Let $\delta > 0$. Define

$$t_\delta = t_\delta^{(n)} := \min\{t : \mathbb{P}[T_z^{a,b} \leq t] \geq 2^{-\delta n}\}, \quad \tau_\delta = \tau_\delta^{(n)} := \min\{t : \mathbb{P}[\tau_z^{a,b} \leq t] \geq 2^{-\delta n}\}. \quad (3.5)$$

Using large deviations theory (cf. [14, Example 3]), it is not hard to show that $\lim_{n \rightarrow \infty} \tau_\delta/n$ and $\lim_{n \rightarrow \infty} t_\delta/n$ both exist. Moreover, $2\tau_\delta < t_\delta$ (for all sufficiently large n) and uniformly in $\delta \in (0, 1/2)$, we have that $\frac{12n-t_\delta}{\sqrt{\delta n}} = \Theta(1) = \frac{6n-\tau_\delta}{\sqrt{\delta n}}$ and

$$\frac{t_\delta - 2\tau_\delta}{\sqrt{\delta n}} = \Theta(1). \quad (3.6)$$

Equation (3.6) shall play a crucial role in the proof of Theorem 1. The proof of (3.6) is deferred to § 6.5.

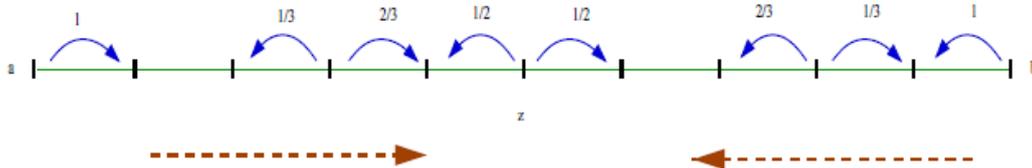


Figure 1: A schematic representation of the transition probabilities for the simple chain from Example 3.3. The endpoints a and b are both of distance n from z . In continuous-time the separation mixing time is twice as large as the total variation mixing time ($3n$ and $6n$, respectively). The transition probabilities (apart from at the special states a , b and z) are $2/3$ in the z direction and $1/3$ in the opposite one.

Remark 3.4. The reader is encouraged to consider a variant of Example 3.3 in which the transition probabilities in are taken to be $1 - e^{-n}$ towards z and e^{-n} away from z (and the definitions of t_δ and τ_δ remain unchanged, as in (3.5)).⁵ In this variant $\frac{4n-t_\delta}{\sqrt{\delta n}} = \Theta(1) = \frac{2n-\tau_\delta}{\sqrt{\delta n}}$ and (as in (3.6)) $\frac{t_\delta-2\tau_\delta}{\sqrt{\delta n}} = \Theta(1)$ (uniformly in $\delta \in (0, 1/2)$), and the proofs of these equations are essentially reduced to a comparison of the large deviation behaviors of the Poisson and of the Binomial distributions. Namely, if $s_1 = s_1(n, \delta) := \inf\{t : \mathbb{P}[\text{Pois}(t) \geq 2n] \geq 2^{-\delta n}\}$ and $s_2 = s_2(n, \delta) := \inf\{t : \mathbb{P}[\text{Bin}(t, 1/2) \geq 2n] \geq 2^{-\delta n}\}$, then (after changing the transition probabilities as above) we have that $\tau_\delta = (1 \pm o(1))s_1$ and $t_\delta = (1 \pm o(1))s_2$. For simplicity, instead of analyzing s_1 and s_2 we now analyze related quantities. Let $X \sim \text{Pois}(n)$ and $Y \sim \text{Bin}(2n, 1/2)$. Pick $t_1 = t_1(n, \delta)$ and $t_2 = t_2(n, \delta)$ such that $\mathbb{P}[X \leq t_1] = \Theta(2^{-\delta n}) = \mathbb{P}[Y \leq t_2]$. Then uniformly in $\delta \in (0, 1/2)$, we have that $\frac{n-t_1}{\sqrt{\delta n}} = \Theta(1) = \frac{n-t_2}{\sqrt{\delta n}}$ and $\frac{t_2-t_1}{\sqrt{\delta n}} = \Theta(1)$. While this can be derived directly via a comparison of the large deviation rate functions of the corresponding distributions, the last equality can also be deduced from the fact that by Poisson thinning if for some $Z \sim \text{Pois}(2n)$ we have that (given Z) $X \sim \text{Bin}(Z, 1/2)$, then $X \sim \text{Pois}(n)$ (for further details cf. the proof of (3.6)).

We now briefly explain how Example 3.3 will be used as a building block in the proof of Theorem 1. As noted in [14, Example 3], by attaching two birth and death chains (“branches”) of length $\Theta(n)$ to z both having the same end-points z, z' with a bias towards z' , we can tune the stationary measure of z to become exponentially small in n .⁶ We

⁵The term e^{-n} can be replaced by any other term which is $o(n)$. Effectively, it is as if the transition towards z equals 1, but if we would have defined it to equal 1 then the chain would be reducible.

⁶We note that in [14, Example 3] the roles of z and z' are interchanged compared to their role here.

pick one of the branches to have a larger average speed than the other. In the example from the proof of Theorem 1 the slower branch is C and the faster is D . For technical reasons, in our construction the end-points of the branches C and D are z and \bar{z} , where \bar{z} is connected to z' through a biased birth and death chain (E) with a fixed bias towards z' (see Figure 2). Using ideas from [14] we can tune simultaneously⁷ both the stationary measure of z (we will take it to be $\Theta(2^{-\delta n})$, for some $\delta \in (0, 1/8)$) and the time it takes to reach z' from z along each branch. For each branch, if we condition on reaching z' through that branch, the (conditional) distribution of the hitting time of z' becomes concentrated.

If the time it takes the chain to reach z' from z along the slow branch is sufficiently small, then the lazy and continuous-time chains would exhibit separation cutoff around the time t_n (resp. τ_n) in which $\mathbb{P}[T_z^{a,b} \leq t_n] = \Theta(\pi(z))$ (resp. $\mathbb{P}[\tau_z^{a,b} \leq \tau_n] = \Theta(\pi(z))$). Indeed this is proven in [14] Example 3, by exploiting the equality in (3.3) above and analyzing the distribution of $T_z^{a,b}$ in the large deviation regime. Note that because we will have that $\pi(z) = \Theta(2^{-\delta n})$, the times t_n and τ_n coincide (up to smaller order terms) with t_δ and τ_δ (respectively) from (3.5).

The significance of (3.6) is that after modifying Example 3.3 as described above so that $\pi(z) = \Theta(2^{-\delta n})$, we get by (3.6) that $\liminf_{n \rightarrow \infty} t_n/\tau_n = \liminf_{n \rightarrow \infty} t_\delta/\tau_\delta > 2$ (where t_n and τ_n are defined in the previous paragraph and t_δ and τ_δ in (3.5)). This allows us to ruin separation cutoff for the continuous-time chain while maintaining it for the lazy chain, by picking the average speed along the two branches wisely, so that the hitting time of z' (for the continuous-time chain), starting from a , is with a constant probability between $\tau_n + \epsilon n$ and $t_n/2$ (for some $\epsilon \in (0, \frac{1}{n}(\frac{t_n}{2} - \tau_n))$) and with the complement probability (up to negligible terms) is smaller than τ_n . In other words, we exploit (3.6) to construct a Markov chain such that the following hold:

- (1) In terms of the separation distance it suffices to consider the case that the initial state is a .
- (2) The state y which up to $o(1)$ terms minimizes $H_t(a, y)/\pi(y)$ and $P_L^{2t}(a, y)/\pi(y)$ is different for t lying in some (sufficiently large) time interval.
- (3) For the continuous-time chain the worst state (i.e. the minimizer from (2)) is z' for every t in the aforementioned time interval. Moreover, we will show that $H_t(a, z') = H_a[T_{z'} > t] \pm o(1)$ for all t . For the sake of simplicity let us assume that the aforementioned time interval is $(0, \infty)$. In this case, the existence of two parallel branches of different average speeds through which z' can be accessed prevents cutoff for the sequence of continuous-time chains.
- (4) For the discrete-time lazy chain we have that $\min_y P_L^t(a, y)/\pi(y) = P_L^t(a, b)/\pi(b) + o(1)$ for all t . Moreover, we will show that $P_L^t(a, b)/\pi(b)$ exhibits an abrupt transition from $o(1)$ to at least $1 - o(1)$ around time t_δ and hence the sequence of lazy chains exhibits cutoff around time t_δ .
- (5) The mechanism which allows us to construct an example with such behavior is that while we always have that $\mathbb{P}_a(T_{z'} > 2t) = H_a(T_{z'} > t) \pm o(1)$, it is not the case that $\frac{1}{2}t_\delta = \tau_\delta(1 \pm o(1))$.

3.2 Proof of Theorem 1

Below we intentionally omit all ceiling and floor signs and suppress the dependence on n of some quantities, for the sake of notational convenience. Fix some $0 < \epsilon < \delta < 1/8$ and $2 \leq s \in \mathbb{N}$ so that

$$3(1 + \frac{\delta}{2}s^2)n + 4\epsilon n \leq \tau_\delta + 2\epsilon n < 3(1 + \frac{\delta}{2}(1 + s^2))n < \frac{1}{2}t_\delta - 2\epsilon n, \quad (3.7)$$

⁷The fact that we can tune them simultaneously is subtle and crucial.

for every sufficiently large n , where t_δ and τ_δ are as in (3.5). Note that by (3.6) we can find such $\epsilon = \epsilon(\delta)$ and $s = s(\delta)$, provided that δ is sufficiently small. We also note that the leftmost inequality in (3.7) is not as important as the other two.⁸

Take $\Omega := A \cup B \cup C \cup D \cup E \cup \{z, \bar{z}, z'\}$, where $A := \{a_1, a_2, \dots, a_n = a\}$, $B := \{b_1, b_2, \dots, b_n = b\}$, $D := \{d_1, \dots, d_{\delta n/2}\}$, $E := \{e_1, \dots, e_{\delta n/2}\}$ and $C := \{c_{i,j} : 1 \leq i \leq \delta n/2, 1 \leq j \leq s\}$ (see Figure 2). Before specifying the transition probabilities we specify some properties that we want the construction to satisfy. The restriction of the chain to $A \cup B \cup \{z\}$ is precisely the chain from Example 3.3, where also here the states a and b are the end-points. The state z' serves as the center of mass of the chain (i.e. $\pi(z') = \Theta(1)$). Below we essentially show that for both the continuous-time and the discrete-time chains it suffices to consider the case that the initial state is a and that started from a the only two other relevant states for mixing in separation are b and z . More precisely, below (combining (3.15)-(3.16) with the analysis of the four cases in the proof of (3.15)) we show that (uniformly in t)

$$d_{\text{sep,L}}^{(n)}(t) = \max\{P_a(T_{z'} > t), 1 - \frac{P_L^t(a,b)}{\pi(b)}\} \pm o(1), \tag{3.8}$$

$$d_{\text{sep,c}}^{(n)}(t) = \max\{H_a(T_{z'} > t), 1 - \frac{H_t(a,b)}{\pi(b)}\} \pm o(1), \tag{3.9}$$

and that $\frac{P_L^t(a,b)}{\pi(b)}$ (resp. $\frac{H_t(a,b)}{\pi(b)}$) exhibits an abrupt transition from $o(1)$ to at least $1 - o(1)$ around time t_δ (resp. τ_δ). Conversely, due to the existence of two parallel branches C and D through which z' can be accessed we have that neither $P_a(T_{z'} > t)$ nor $H_a(T_{z'} > t)$ exhibit an abrupt transition as functions of t . Namely, they decrease from $1 - o(1)$ to $o(1)$ via two drops occurring for the continuous-time chain around times $3(1 + \delta)n$ and $3(1 + \frac{\delta}{2}(1 + s^2))n$ (and for the lazy chain around times $6(1 + \delta)n$ and $6(1 + \frac{\delta}{2}(1 + s^2))n$).

The sets C and D serve as two parallel branches with different average speeds (as in the discussion following Remark 3.4) connecting z to \bar{z} , which in turn is connected to the center of mass, z' , via the segment E (see Figure 2). In order to ensure that the average speed along C is slower than along D we subdivide its edges into paths of length s (see Figure 2).⁹ The term $3(1 + \frac{\delta}{2}(1 + s^2))n$ in (3.7) corresponds to the time around which the hitting time of z' (for the continuous-time chain, started from either a or b) is concentrated, given that z' is reached through the slow branch C (this is explained in more details below). The aforementioned roles of the terms in (3.7) (described over the last two paragraphs, other than that of the term $3(1 + \frac{\delta}{2}s^2)n$, whose significance can be seen from (3.10) below) together with (3.8)-(3.9) motivate (3.7).

Indeed, by (3.7) (and the aforementioned role of the term $3(1 + \frac{\delta}{2}(1 + s^2))n$) we have $P_a(T_{z'} > t_\delta - 2\epsilon n) = o(1)$, and so by (3.8) we have that $d_{\text{sep,L}}^{(n)}(t) = \max\{0, 1 - \frac{P_L^t(a,b)}{\pi(b)}\} \pm o(1)$, which as mentioned below (3.9) exhibits an abrupt change from $1 - o(1)$ to $o(1)$ around time t_δ . Thus the lazy chains exhibit separation cutoff around time t_δ . Similarly, by (3.7)

$$H_a(T_{z'} > \tau_\delta + 2\epsilon n) \geq H_a(z' \text{ is reached through the branch } C) - o(1),$$

$$H_a(T_{z'} > \tau_\delta + \epsilon n) \leq H_a(z' \text{ is reached through the branch } C) + o(1),$$

which (as will be made clear) by construction is bounded away from 0 and 1 (where we say that z' is reached through C if the last state visited before $T_{\bar{z}}$ is in C , see Figure

⁸Even if $3(1 + \frac{\delta}{2}s^2)n \geq \tau_\delta$ only minor changes to the proof below are needed. Namely, instead of arguing that $d_{\text{sep,c}}^{(n)}(\tau_\delta + \epsilon n/2)$ is bounded away from 1, one would have had to argue that $d_{\text{sep,c}}^{(n)}(3(1 + \frac{\delta}{2}s^2)n + \epsilon n/2)$ is bounded away from 1.

⁹We could have simply increased the holding probability along the vertices of the branch C . However, in order to make the details behind Remark 1.1 more transparent we choose to subdivide its edges instead.

2). Hence by (3.9) and the comment following it, for $t \in [\tau_\delta + \epsilon n, \tau_\delta + 2\epsilon n]$ we have that $d_{\text{sep},c}^{(n)}(t) = H_a(T_{z'} > t) \pm o(1)$ is bounded away from 0 and 1. Thus the continuous-time chains do not exhibit separation cutoff.

The lengths of the branches C and D (and also of the interval E) are taken so that $\pi(z) = \Theta(2^{-\delta n})$. Note that this means that t_δ and τ_δ agree (up to negligible terms) with t_n and τ_n , respectively, from the discussion following Remark 3.4. In light of the discussion following Remark 3.4, this explains why (as mentioned above) $\frac{P_L^t(a,b)}{\pi(b)}$ (resp. $\frac{H_t(a,b)}{\pi(b)}$) exhibits an abrupt transition around t_δ (resp. τ_δ).¹⁰ We now specify the edge weights¹¹ and introduce some additional notation (for a schematic representation of the transition probabilities see Figure 2).

- $a := a_n, b := b_n$ (the states a and b have symmetric roles in our construction).
- For all $i, c_{i+1,0} := c_{i,s}$.
- $a_0 = b_0 := z = d_{1+\delta n/2}$ and identify $c_{\delta n/2,s}$ with z .
- $c_{1,0} := \bar{z} =: e_{1+\delta n/2} = d_0$ and $e_0 := z'$.

Consider the following (symmetric) edge weights:

- $w(b_i, b_{i-1}) = 2^{-i} = w(a_i, a_{i-1})$ for all $1 \leq i \leq n$.
- $w(c_{i,j}, c_{i,j-1}) = 2^{\delta n/2-i}$ for all $1 \leq i \leq \delta n/2, 1 \leq j \leq s$.
- $w(d_i, d_{i+1}) = 2^{\delta n/2-i}$ for all $0 \leq i \leq \delta n/2$.
- $w(e_i, e_{i+1}) = 2^{\delta n-i}$ for all $0 \leq i \leq \delta n/2$.

Note that the restriction of the chain to D is a birth and death chain with a bias towards \bar{z} and an average speed of $1/3$ ($1/6$ for the lazy chain), while its restriction to C can be described as follows: first take the same birth and death chain as D and then “stretch” each edge (c_{i+1}, c_i) by a factor s by replacing it by a path of s edges, $(c_{i,s}, c_{i,s-1}), \dots, (c_{i+1,1}, c_{i,0})$, of the same weight as (c_{i+1}, c_i) . It is not hard to see that this results in an average speed of $1/(3s^2)$ along C towards \bar{z} (for the lazy chain the speed is $1/(6s^2)$).

Note that started from a the chain may reach \bar{z} (and thus also z') either through the branch C or D (i.e. the last state to be visited prior to $T_{\bar{z}}$ may be either in D or in C). Started from a , conditioned on taking the branch C (in the above sense), the hitting times of \bar{z} and z' (for the continuous-time chain) are concentrated around $3(1 + s^2\delta/2)n$ and $3(1 + \frac{\delta}{2}(1 + s^2))n$, resp., while conditioned on taking the branch D , they are concentrated around $3(1 + \frac{\delta}{2})n$ and $3(1 + \delta)n$, resp.. From this, we get that there exists some constant¹² $c = c(s) = \Theta(1/s)$ such that for all sufficiently large n

$$\forall y \in C \cup D \cup E \cup \{\bar{z}, z'\}, \quad H_a[T_y \leq \left(3(1 + \frac{\delta}{2}s^2) + \epsilon\right)n] \geq c. \quad (3.10)$$

$$H_a[T_{z'} \leq \left(3(1 + \frac{\delta}{2}(1 + s^2)) - \epsilon\right)n] \leq 1 - c. \quad (3.11)$$

$$\forall y \in C \cup D, \quad \mathbb{P}[T_{\bar{z},y}^a \leq \left(6(1 + \frac{\delta}{2}(1 + s^2)) + \epsilon\right)n] = 1 - o(1). \quad (3.12)$$

$$\forall y \in E \cup \{\bar{z}, z'\}, \quad P_a[T_y \leq \left(6(1 + \frac{\delta}{2}(1 + s^2)) + \epsilon\right)n] = 1 - o(1). \quad (3.13)$$

¹⁰While the abrupt transition of $\frac{P_L^t(a,b)}{\pi(b)}$ around time t_n essentially follows from the analysis in [14, Examples 3 and 5], we will prove it and the abrupt transition of $\frac{H_t(a,b)}{\pi(b)}$ around time τ_n for the sake of completeness.

¹¹We write the edge weights (instead of the transition probabilities, which are given in Figure 2) in order to demonstrate that the chain is indeed reversible and to facilitate the calculation of $\pi(z)$.

¹²Indeed c is bounded away from 0, since s is fixed.

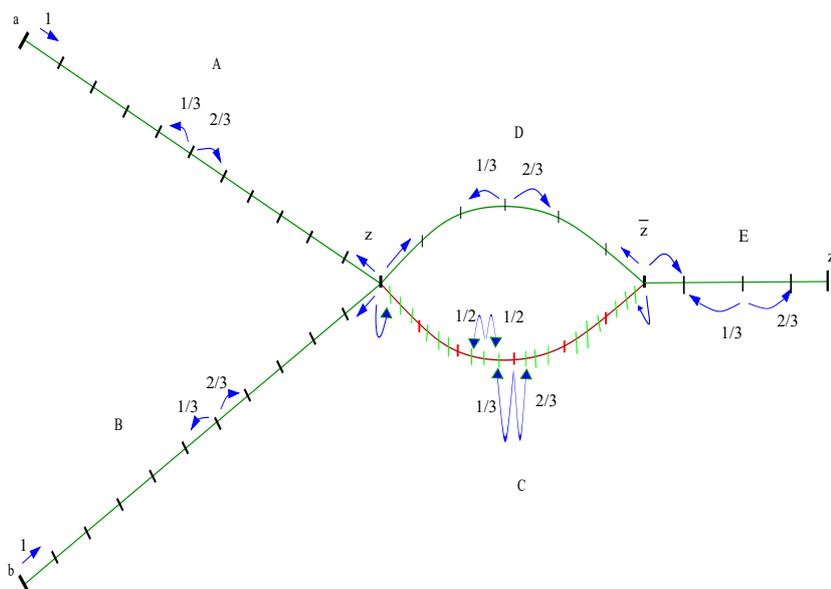


Figure 2: A schematic representation of the transition probabilities for the chain from Theorem 1. The sets A and B are of size n and the sets D and E are of size $\delta n/2$. Along the branch C there are $\delta n/2$ red points (with transitions probabilities $1/3$ to the left and $2/3$ to the right). Between each 2 red points there are $s - 1$ yellow points, from each of which the chain moves to one of its 2 neighbors with equal probability. Also from z and \bar{z} the transitions are of equal probability to each of their neighbors. For a motivation behind the construction see the discussion following Remark 3.4 and the one preceding the specification of the edge weights of the chain.

Note that (since s is fixed)

$$\pi(z) = \Theta(2^{-\delta n}). \tag{3.14}$$

We now argue that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_t |d_{\text{sep},c}^{(n)}(t) - \max\{0, 1 - \min_{y \in \{b, z'\}} H_t(a, y)/\pi(y)\}| &= 0, \\ \limsup_{n \rightarrow \infty} \sup_t |d_{\text{sep},L}^{(n)}(t) - \max\{0, 1 - \min_{y \in \{b, z'\}} P_L^t(a, y)/\pi(y)\}| &= 0. \end{aligned} \tag{3.15}$$

$$\begin{aligned} \lim_{t' \rightarrow \infty} \liminf_{n \rightarrow \infty} H_{\tau_\delta + t'}(a, b)/\pi(b) &= \infty, & \lim_{t' \rightarrow \infty} \limsup_{n \rightarrow \infty} H_{\tau_\delta - t'}(a, b)/\pi(b) &= 0, \\ \lim_{t' \rightarrow \infty} \liminf_{n \rightarrow \infty} P_L^{t_\delta + t'}(a, b)/\pi(b) &= \infty, & \lim_{t' \rightarrow \infty} \limsup_{n \rightarrow \infty} P_L^{t_\delta - t'}(a, b)/\pi(b) &= 0. \end{aligned} \tag{3.16}$$

These equations are essentially borrowed from the analysis of Examples 3 and 5 in [14]. For the sake of completeness we present a sketch of their proofs (all of the missing steps can be found at [14]). We now prove (3.15). We only prove the first line of (3.15) as the second line is proved in a similar fashion. The proof of (3.16) is contained within the analysis of Case 1 below.

By (3.4) with $(x, y) = (a, z')$ (and noting that $t_{\text{rel}}^{(n)} = \Theta(1) = \pi(z')$ while, in the notation of (3.4), $\lim_{n \rightarrow \infty} \max_t f_{z'}^a(t) = 0$ (cf. [14, Remark 3.9]), where $z' = z'(n)$ and $f_{z'}^a = (f_{z'}^a)^{(n)}$)

$$|H_t(a, z')/\pi(z') - H_a[T_{z'} \leq t]| = o(1), \tag{3.17}$$

uniformly in t . Denote $F := C \cup D \cup E \cup \{z, \bar{z}, z'\}$. By symmetry between A and B it suffices to consider each of the following cases: (Case 1) $(x, y) \in A \times B$, (Case 2) $(x, y) \in A \times A$, (Case 3) $(x, y) \in A \times F$ and (Case 4) $(x, y) \in F \times F$ (where (x, y) may depend on n).

For Case 2, if $(x, y) = (a_i, a_j)$ we may assume w.l.o.g. that $i > j$ (since $H_t(x, y)/\pi(y) = H_t(y, x)/\pi(x)$). Using (3.2) we get that for all $t \geq 3n + n^{2/3}$ we have that $1 - H_t(x, y)/\pi(y) \leq \mathbb{H}_x[T_x \leq y] = o(1)$ while by (3.17) we have that $H_{n(3+\delta/2)}(a, z')/\pi(z') = o(1)$ (hence we may neglect Case 2).¹³

For Case 4 it is not hard to see that for $j(n) := n^{\delta/2}(s^2 + 1) + n^{2/3}$ we have

$$\max\{\mathbb{P}[\tau_{\bar{z},y}^x \leq j(n)], \mathbb{P}[\tau_{\bar{z},x}^y \leq j(n)], \mathbb{H}_x[T_y \leq j(n)], \mathbb{H}_y[T_x \leq j(n)]\} = 1 - o(1)$$

(uniformly in all (x, y) as in Case 4). Thus by Lemma 3.2, for all $t \geq j(n)$ we have that $1 - H_t(x, y)/\pi(y) = o(1)$, uniformly in all (x, y) as in Case 4.

We now consider Case 3. For all $x \in A$, if $y \in C \cup D$ then by (3.1) we have that

$$H_t(x, y)/\pi(y) \geq \max\{\mathbb{P}[\tau_{\bar{z},y}^x \leq t], \mathbb{P}[t - 1 \leq \tau_{\bar{z}}^{x,y} \leq t]/[e\pi(z)]\}$$

and if $y \in E \cup \{z, \bar{z}, z'\}$ we have

$$H_t(x, y)/\pi(y) \geq \mathbb{H}_x[T_y \leq t].$$

By (3.4) (as in (3.17)) $|H_t(x, z')/\pi(z') - \mathbb{H}_x[T_{z'} \leq t]| = o(1)$ uniformly for all $x \in A$ and $t \geq 0$. It is thus not hard to see that for each $x \in A$ the worst $y \in F$ (in the sense of minimizing $\min\{H_t(x, y)/\pi(y), 1\}$ for all t , up to $o(1)$ additive terms) is $y = z'$ and that for each fixed t the worst $x \in A$ w.r.t. $y = z'$ (at least up to $o(1)$ additive terms) is $x = a$.

Finally, for Case 1 note that by (3.1) we have that

$$\frac{\mathbb{P}[t - 1 \leq \tau_z^{x,y} \leq t]}{e\pi(z)} \leq H_t(x, y)/\pi(y) \leq \mathbb{P}[\tau_z^{x,y} \leq t]/\pi(z). \tag{3.18}$$

Using the fact that the law of $\tau_z^{x,y}$ is a convolution of Exponential distributions and hence log-concave (cf. the analysis of Example 5 in [14, Section 6.5]), it follows that for all $(x, y) \in A \times B$ we have that $H_t(x, y)/\pi(y)$ exhibits the following behavior around

$$t_{x,y} := \inf\{t : \mathbb{P}[t - 1 \leq \tau_z^{x,y} \leq t] \geq \pi(z)/4\}.$$

For every $\alpha \in (0, 1)$ if $t \leq (1 - \alpha)\mathbb{E}[\tau_z^{x,y}]$, there exists some C_α such that (uniformly in all (x, y) as in Case 1)

$$\mathbb{P}[\tau_z^{x,y} \leq t] \leq C_\alpha e^{-1} \mathbb{P}[t - 1 \leq \tau_z^{x,y} \leq t]$$

for some constant $C_\alpha > 0$ (this uses the aforementioned log-concavity of $\tau_z^{x,y}$ (cf. [14, Section 6.5]) and so by (3.18) (and the fact that $t_{x,y} < (1 - \alpha)\mathbb{E}[\tau_z^{x,y}]$ for some $\alpha > 0$ for all sufficiently large n) there exist a constant $C > 0$ such that

$$\frac{\mathbb{P}[t - 1 \leq \tau_z^{x,y} \leq t]}{e\pi(z)} \leq H_t(x, y)/\pi(y) \leq C \frac{\mathbb{P}[t - 1 \leq \tau_z^{x,y} \leq t]}{e\pi(z)}.$$

Again, using log-concavity and the fact that $t_{x,y}$ by definition is in the large deviation regime of $\tau_z^{x,y}$, such that for some constants $c, C' > 0$ for each fixed $k \in \mathbb{Z}$ we have that $e^{ck} \leq \frac{\mathbb{P}[t_{x,y} + k - 1 \leq \tau_z^{x,y} \leq t_{x,y} + k]}{\mathbb{P}[t_{x,y} - 1 \leq \tau_z^{x,y} \leq t_{x,y}]} \leq e^{C'k}$ for all sufficiently large n (cf. [14, Section 6.5]). Hence

$$\lim_{t' \rightarrow \infty} \liminf_{n \rightarrow \infty} \min_{(x,y) \in A \times B} H_{t_{x,y} + t'}(x, y)/\pi(y) = \infty,$$

¹³The exponent 2/3 can be replaced by an arbitrary exponent in $(1/2, 1)$.

$$\lim_{t' \rightarrow \infty} \limsup_{n \rightarrow \infty} \max_{(x,y) \in A \times B} H_{\max\{t_{x,y}-t',0\}}(x,y)/\pi(y) = 0.$$

This in particular establishes the first line of (3.16) (the proof of the second line is analogous, where Exponential distributions above are replaced by Geometric distributions (see [14, Section 6.5]). Moreover, the same reasoning as in [14, Section 6.5] yields that $H_t(x,y)/\pi(y)$ is increasing in $[0, t^*(x,y)]$, for some $t^*(x,y)$ satisfying $|t^*(x,y) - \mathbb{E}[\tau_z^{x,y}]| \leq C_2\sqrt{n}$, and that for all $t \geq t^*(x,y)$ we have that $H_t(x,y)/\pi(y) \geq 1 - o(1)$ (uniformly in $(x,y) \in A \times B$ and $t \geq t^*(x,y)$). Finally, to conclude the analysis of Case 1 we note that $\max_{(x,y) \in A \times B} t^*(x,y) = t^*(a,b)$.

We are now in a position to conclude the proof. We first argue that the sequence of lazy chains exhibits separation cutoff around time t_δ . Indeed, by (3.1)-(3.2) in conjunction with (3.12)-(3.13) we have $\max_{y \in C \cup D \cup E \cup \{z, \bar{z}, z'\}} \sup_{t \geq (6(1+\frac{\delta}{2}(1+s^2))+\epsilon)n} (1 - \frac{P_L^t(a,y)}{\pi(y)}) = o(1)$. By (3.7) $(6(1+\frac{\delta}{2}(1+s^2))+\epsilon)n \leq t_\delta - \epsilon n$ and thus by combining (3.15) and (3.16), we obtain that $\limsup_{n \rightarrow \infty} \sup_t |d_{\text{sep,L}}^{(n)}(t) - \max(0, 1 - P_L^t(a,b)/\pi(b))| = 0$ and that $\max(0, 1 - P_L^t(a,b)/\pi(b))$ exhibits a sharp transition around time t_δ .

We now consider the continuous-time chain. By (3.15), (3.16), (3.10) combined with (3.1), (3.2) and (3.7) we get that $d_{\text{sep,c}}^{(n)}(\tau_\delta + \epsilon n/2)$ is bounded away from 1 uniformly in n (using the fact that the separation distance is non-increasing in time). Denote $m_n := 3(1+\frac{\delta}{2}(1+s^2))n - \epsilon n$. By (3.7), $\tau_\delta + \epsilon n < m_n$. Thus the proof is concluded by applying (3.17) with $t = m_n$, in order to deduce by (3.11) that

$$d_{\text{sep,c}}^{(n)}(m_n) \geq 1 - H_{m_n}(a, z')/\pi(z') \geq H_a[T_{z'} > m_n] - o(1) \geq c - o(1). \quad \square$$

3.3 Proof of Remark 1.2

We now explain the necessary adaptations for the proof of the assertion of Remark 1.2. The only adjustments are in the choices of δ and s (and possibly, one has to contract some of the s -paths along the slow branch C to a single edge in order to adjust the expected time it takes the walk to cross the slow branch). Fix $\alpha \in (0, 1)$. Consider the α -lazy version of the chain. Denote by $P_x^{(\alpha)}$ the law of the α -lazy version of the chain and by $T_{x',y}^{x,y}(\alpha)$ the version of $T_{x',y}^{x,y}$ corresponding to holding probability α . Note that under the aforementioned modifications to the chain (i.e. adjusting δ and s) the transition probabilities along the A and B segments are unaffected and so (for each fixed α) the law of $T_z^{a,b}(\alpha)$ is also unaffected.

Let $\delta > 0$ to be determined later. The aforementioned modifications can be made so that we still have that $\pi(z) = \Theta(2^{-\delta n})$. Denote the separation distance at time t of the n th α -lazy chain by $d_{\text{sep,\alpha}}^{(n)}(t)$. Let κn be the expected hitting time of z from \bar{z} for the non-lazy chain, conditioned on taking the slow branch C . Similarly to the analysis in the proof of Theorem 1, as long as δ is taken to be sufficiently small, and κ is chosen in an appropriate manner¹⁴, one can replace $d_{\text{sep,\alpha}}^{(n)}(t)$ by $\max\{P_a^{(\alpha)}(T_{z'} > t), 1 - \frac{P_a^{(\alpha)}(X_t=b)}{\pi(b)}\} \pm o(1)$.

Let $t_\delta^{(\alpha)} := \inf\{t : \mathbb{P}[T_z^{a,b}(\alpha) \leq t] \geq 2^{-\delta n}\}$. Similarly to the analysis in the proof of Theorem 1, the quantity $1 - \frac{P_a^{(\alpha)}(X_t=b)}{\pi(b)}$ exhibits a sharp transition around $t_\delta^{(\alpha)}$. Using the aforementioned modifications we can ensure that $P_a^{(\alpha)}(T_{z'} \leq t)$ exhibits two jumps, one around $\frac{3(1+\delta)n}{1-\alpha}$ and a second jump around $s^{(\alpha)} := \frac{3(1+\delta/2+\kappa)n}{1-\alpha}$. It follows that the sequence of α -lazy chains exhibits separation cutoff iff $t_\delta^{(\alpha)} \geq (1 - o(1))s^{(\alpha)}$.

Using the analysis from [14, Example 3 and Lemma 5.1] one can derive the following formula for Ψ_α , the large deviation rate functions of $T_z^{a,b}(\alpha)$, which is given by the

¹⁴More precisely, this is the case whenever $\frac{3(1+\max(\delta,\kappa))n}{1-\alpha} < t_\delta^{(\alpha)} - \epsilon n$ for some $\epsilon > 0$, where $t_\delta^{(\alpha)}$ is defined in the following paragraph. We leave this as an exercise.

following Legendre transform (the exact formula shall be used only to make a certain comment at the end of the analysis)

$$\Psi_\alpha(r) := \sup_{\lambda \in (-\infty, \infty)} [\lambda r - \log \mathbf{F}_\alpha(\lambda)], \tag{3.19}$$

where for $\Delta_\alpha(\lambda) := (e^{-\lambda} - \alpha)^2 - \frac{4(1-\alpha)}{3}$ and λ_α , the smaller solution to $\Delta_\alpha(\lambda) = 0$, we have that

$$\mathbf{F}_\alpha(\lambda) := \begin{cases} \infty & \text{if } \lambda > \lambda_\alpha, \\ \frac{3}{2(1-\alpha)}[(e^{-\lambda} - \alpha) - \sqrt{\Delta_\alpha(\lambda)}] & \text{if } \lambda \leq \lambda_\alpha. \end{cases}$$

Fix $p \neq q \in (0, 1)$. Using the fact that the Legendre transform of a strictly convex smooth function is itself smooth and that the restriction of the Legendre transform to this class of functions is invertible, it is not hard to verify that for every $\epsilon \in (0, 1/100)$ there is some $r \in (3 - \epsilon, 3)$ such that $\Psi_p(\frac{r}{1-p}) \neq \Psi_q(\frac{r}{1-q})$.¹⁵ This implies that for all $p \neq q \in (0, 1)$, we can find $\delta = \delta_{p,q} \in (0, 1/10)$ such that for some $\delta' = \delta'_{p,q} > 0$ either (Case 1:) $(1-p)t_\delta^{(p)} - (1-q)t_\delta^{(q)} \geq \delta'n(1 \pm o(1))$, for all n or (Case 2:) $(1-q)t_\delta^{(q)} - (1-p)t_\delta^{(p)} \geq \delta'n(1 \pm o(1))$ for all n (namely, fixing such r we can pick δ to satisfy $2^{-n\delta} = e^{-2n\Psi_p(\frac{r}{1-p})}$; Case 1 corresponds to $\Psi_p(\frac{r}{1-p}) > \Psi_q(\frac{r}{1-q})$ and Case 2 to $\Psi_p(\frac{r}{1-p}) < \Psi_q(\frac{r}{1-q})$). This implies that we can tune κ such that in Case 1 we have $t_\delta^{(p)} > s^{(p)}$ and $\limsup_{n \rightarrow \infty} t_\delta^{(q)}/s^{(q)} < 1$ and in Case 2 we have $t_\delta^{(q)} > s^{(q)}$ and $\limsup_{n \rightarrow \infty} t_\delta^{(p)}/s^{(p)} < 1$. Thus in the first case the p -lazy chains exhibit separation cutoff while the q -lazy chains do not, and in the second case the q -lazy chains exhibit separation cutoff while the p -lazy chains do not. Note that in the above argument the formulas for Ψ_p and Ψ_q played no role. However, they can be used to distinguish between Cases 1 and 2 for every fixed $p \neq q$.

4 Proof of Theorem 3

4.1 Preliminaries

Before proving Theorem 3 we make several general comments regarding a principle which shall be utilized below repeatedly. We summarize a few different variations of this principle in Fact 4.1 (whose proof is deferred to the appendix § 6.3).

Let $\mathcal{T} = (V, E)$ be an infinite binary tree rooted at o (in practice, we shall work with finite trees; however, it is not hard to show that the “boundary effect” coming from the finiteness of the trees is negligible for our purposes). For any vertex u we distinguish its two children by *left* and *right* child. We denote the collection of all left (resp. right) children in \mathcal{T} by L (resp. R). Denote by \mathcal{L}_n the collection of vertices whose distance from o is n . For any vertex u let $L(u)$, (resp. $R(u)$) be the number of left (resp. right) children along the path from u to the root. Denote $g(u) = L(u) - R(u)$. Let $\tau_k := \sup\{t : X_t \in \mathcal{L}_k\}$.

Fact 4.1. (1 a) Starting from o , for each fixed n we have that $g(X_{T_{\mathcal{L}_n}})$ is distributed like S_n , where $(S_k)_{k \in \mathbb{Z}_+}$ is a SRW on \mathbb{Z} , started at the origin. Moreover, $(g(X_{\tau_k}))_{k \geq 0}$ is distributed like $(S_k)_{k \in \mathbb{Z}_+}$.

(1 b) Let $n \in \mathbb{N}$ and $D \subset \mathcal{L}_n$. Let \mathcal{D} be the event that the walk visits the set D at least once. There exists an absolute constance $c > 0$ such that

$$c \leq \text{P}_o[X_{\tau_n} \in D] / \text{P}_o[\mathcal{D}] \leq 1 \text{ and } c \leq \text{P}_o[X_{T_{\mathcal{L}_n}} \in D] / \text{P}_o[\mathcal{D}] \leq 1. \tag{4.1}$$

(2 a) Fix some $\epsilon > 0$. Consider the network obtained by increasing the edge weight between every u and v such that v is a left child of u to $1 + \epsilon$. Then, $(g(X_{\tau_k}))_{k \geq 0}$ is distributed like a biased nearest-neighbor random walk on \mathbb{Z} , $(\tilde{S}_k)_{k \in \mathbb{Z}_+}$, satisfying $\text{P}[\tilde{S}_1 = 1] = 1 - \text{P}[\tilde{S}_1 = -1] = \frac{\sqrt{1+\epsilon}}{1+\sqrt{1+\epsilon}} = \frac{1}{2} + \epsilon/4 - O(\epsilon^2) > 1/2$.

¹⁵The choice of the constant 1/100 is arbitrary and is made in order to ensure that $3 - r$ is small.

(2 b) Let $n \in \mathbb{N}$ and $D \subset \mathcal{L}_n$. Let \mathcal{D} be the event that the walk visits the set D at least once. There exists an absolute constant c (which is independent also of ϵ) such that also in the perturbed network we have that (4.1) holds.

Fact 4.2 (Local CLT). Let $(S_k)_{k \in \mathbb{Z}_+}$ be simple random walk (SRW) on \mathbb{Z} . Then there exists an absolute constant $C > 1$ such that for all $n \geq 1$ and $1 \leq m \leq \lceil n^{1/4} \rceil$

$$1/C \leq P_0[S_n \geq m\sqrt{n}]m e^{m^2/2} \leq C. \tag{4.2}$$

4.2 Proof of part (a) of Theorem 3

Fix some (large) $k \in \mathbb{N}$. We suppress the dependence on k from the notation (below, $o(\cdot), O(\cdot), \Theta(\cdot)$ and $\Omega(\cdot)$ are taken w.r.t. k). Denote $s := k^3$. We now construct a graph $G = (V, E)$ in three steps (see Figure 3).

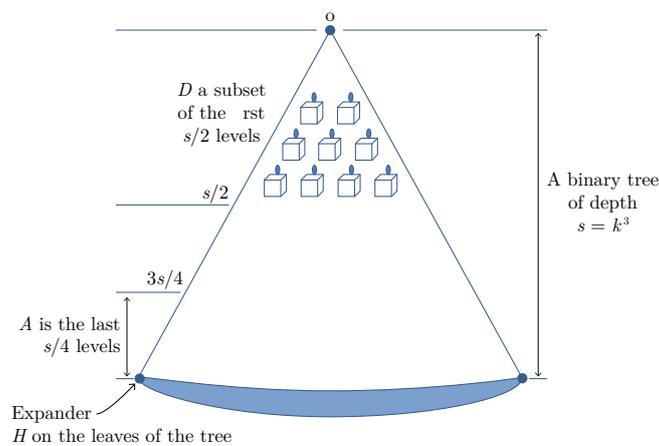


Figure 3: This is schematic representation of the graph G from part (a) of Theorem 3. We start with a binary tree rooted at o of depth $s := k^3$. We then connect its leaves using an expander. We define a set D , contained in the first $s/2$ levels of the tree, which is the collection of “unbalanced” vertices in the sense that the number of left/right turns from them to the root violates the Law of Iterated Logarithm in some strong sense. Finally, $d \in D$ is decorated by a $k \times k \times k$ torus, represented by a square.

Step 1: Start with a binary tree $\mathcal{T} = (V(\mathcal{T}), E(\mathcal{T}))$ of depth s with root o .

Step 2: Form an expander on the leaves of \mathcal{T} : Let $H = (V', E')$ be an arbitrary expander with $|V'| = 2^s$. Identify the leaves of \mathcal{T} with the set V' and connect by an edge every pair of leaves u, v such that $\{u, v\} \in E'$.

Step 3: We pick a set $D \subset V$ strategically as follows and decorate each of its vertices by a 3D torus of side length k (and so of size s):

Denote the collection of vertices belonging to the i -th level of \mathcal{T} by $\mathcal{L}_i(\mathcal{T})$. For any vertex u which is not a leaf of \mathcal{T} , we distinguish its two children by *left* and *right* child. We denote the collection of all left (resp. right) children in \mathcal{T} by L (resp. R). Fix some large integer C to be determined later. We note that one can set $C = 1$, but the analysis is somewhat smoother by taking C to be large. Including the constant C in the construction shall benefit us in the proof of part (b). We denote by D_i the collection of all vertices u belonging to $\mathcal{L}_{Ci}(\mathcal{T})$ such

that if $\gamma_u = (v_0 = u, v_1, \dots, v_{Cj} = o)$ is the path from u to o in \mathcal{T} , then for all $1 \leq j \leq i$

$$|\{v_\ell : 0 \leq \ell \leq Cj\} \cap L| - |\{v_\ell : 0 \leq \ell \leq Cj\} \cap R| \geq 3\sqrt{Cj \log \log(Cj)}.$$

Crucially, above the “base point” v_0 of γ_u was taken to be u itself (rather than o). Let $D_0 := \{o\}$. Denote $D := \bigcup_{i=0}^{\lfloor s/2C \rfloor} D_i$. We decorate each $v \in D$ by a 3D $k \times k \times k$ torus. That is, for each $v \in D$ we attach to v a three dimensional torus, W_v , of side length k , having v as one of its vertices, while the rest of its vertices are disjoint from \mathcal{T} (where W_v and W_u are taken to be disjoint if $v \neq u$). Call the resulting graph $G = (V, E)$.

We argue that

$$t_{\text{mix}}(G) = \Theta(s) \text{ and also } t_{\text{rel}}(G) = \Theta(s). \tag{4.3}$$

By Lemma 6.7 $t_{\text{rel}}(G) = \Omega(s)$. We now explain why indeed $t_{\text{mix}}(G) = \Theta(s)$ (by (1.5) this implies that $t_{\text{rel}}(G) = \Theta(s)$). Let $A := \bigcup_{j \geq 3s/4} \mathcal{L}_j(\mathcal{T})$. The set A is sufficiently far from the set D so that starting from any $u \in A$ the walk mixes in $\Theta(s)$ steps (as if the vertices in D were not decorated by tori). That is, there exists an absolute constant $C_1 \in \mathbb{N}$ such that for every $u \in A$,

$$\|P_u^{C_1 s} - \pi\|_{\text{TV}} = o(1). \tag{4.4}$$

This can be deduced formally using Proposition 6.6. Proposition 6.6 applies because for all $a \in A$, by the tree structure we have that $P_a[T_D \leq C_1 s]$ can be bounded from above by the probability that a LSRW on a binary tree of depth $\lfloor s/4 \rfloor$, started from some leaf, reaches the root by time $C_1 s$ (which occurs with probability of at most $C' e^{-cs}$). Consequently, by Proposition 6.6, there exists some absolute constant C'_2 such that for every $t \geq 0$

$$\begin{aligned} \max_{u \in V} P_u[T_A > t] &\leq d(t) + (1 - \pi(A)) = d(t) + o(1). \\ \max_{u \in V} P_u[T_A > t] &\geq d(t + C'_2 s) - o(1). \end{aligned} \tag{4.5}$$

Let T_1 (resp. T_2) be the total amount of time, prior to time T_A , that the walk spends at $\bigcup_{v \in D} W_v$ (resp. $V \setminus \bigcup_{v \in D} W_v$). Let T_3 be the total number of times the set D was visited prior to time T_A by crossing some edge belonging to \mathcal{T} . That is,

$$T_3 := |\{1 \leq t < T_A : X_t \in D, X_{t-1} \in V(\mathcal{T})\}|.$$

We argue that there exist absolute constants $C_2, C_3, \beta > 0$ such that

$$(1) \quad \lim_{m \rightarrow \infty} \max_{u \in V} P_u[T_3 > m] = 0.$$

$$(2) \quad \max_{u \in V} P_u[T_2 > 5s] = o(1).$$

(3) For every $m, r \in \mathbb{N}$ we have that

$$\max_{u \in V} P_u[T_1 > C_2 m r s \mid T_3 = m] \leq C_3 e^{-\beta m r}.$$

Combining (1)-(3) with (4.5) concludes the proof of the fact that $t_{\text{mix}}(G) = \Theta(s)$.

To see why (1) holds, use part (1 a) of Fact 4.1, in conjunction with the law of the iterated logarithm and the fact that the distribution of the number of visits by time T_A to

each $\mathcal{L}_i(\mathcal{T})$, denoted by N_i , has an exponential tail (along with the fact that $\text{Cov}(N_i, N_{i+j})$ decays exponentially in j for all i, j). We leave the details as an exercise.

For (2), note that the distance from the root of a LSRW on a binary tree behaves like a biased nearest neighbor walk whose average speed is $1/6$.

It is not hard to show that there exist some $C_4, C_5, \beta' > 0$ such that for all $v \in D$

$$\max_{u \in W_v} \mathbb{P}_u[T_{V \setminus W_v} > ms] \leq C_4 e^{-\beta' m}, \text{ for every } m \in \mathbb{N}. \quad (4.6)$$

$$\forall u \in W_v, \quad s/2 \leq \mathbb{E}_u[T_{V \setminus W_v}] \leq C_5 s. \quad (4.7)$$

Thus (3) is obtained as a large deviation estimate.

The bounded perturbation from the assertion of part (a) of Theorem 3 is obtained by increasing the edge weight of the edge between any $v \in \bigcup_{0 \leq i \leq s/2} \mathcal{L}_i(\mathcal{T})$ and its left child to $1 + \epsilon$, for some constant $\epsilon > 0$.

It is easy to show that (4.5) remains valid also in the perturbed network and that (by symmetry) the maximum (in the LHSs of (4.5)) can (still) be taken over the set W_o . To distinguish between the LSRW on G and the lazy random walk on the perturbed network we adopt the convention that when referring to the perturbed network we write \mathbb{P}'_u and \mathbb{E}'_u , instead of \mathbb{P}_u and \mathbb{E}_u .

Let T_1, T_2, T_3 be as above. Using part (2) of Fact 4.1 it is not hard to verify that for any $u \in W_o$ we have that $\mathbb{E}'_u[T_3] \geq c_6(\epsilon)s$ and $\text{Var}'_u[T_3] \leq C_7(\epsilon)s$ for some constants $c_6(\epsilon), C_7(\epsilon) > 0$, depending on ϵ . The fact that $\mathbb{E}'_u[T_3] \geq c_6(\epsilon)s$ (for some $c_6(\epsilon)$) is clear from part (2 a) of Fact 4.1. To see that $\text{Var}'_u[T_3] \leq C_7(\epsilon)s$ use part (2) of Fact 4.1 to deduce that the correlation between the contribution to T_3 from vertices belonging to D_i and D_{i+j} , respectively, decays exponentially in j .

Using these estimates, we now show that the order of the mixing time of the walk on the perturbed network is $\Theta(s^2)$ and that a sequence of walks on such perturbed networks with $k \rightarrow \infty$ exhibits cutoff.

Using (4.6)-(4.7) it is not hard to show that for any $u \in W_o$, we have that $\mathbb{E}'_u[T_1] \geq c_8 s^2$ and that $\text{Var}'_u[T_1] \leq C_9 s^3 = o((\mathbb{E}'_u[T_1])^2)$, for some constants $c_8, C_9 > 0$ depending on ϵ . Moreover, for every $u \in W_o$, $\mathbb{E}'_u[T_o] \leq C_5 s$. Consequently, it follows from Chebyshev's inequality that starting from every $u \in W_o$ we have that T_A is concentrated around $\mathbb{E}'_o[T_1] = \Theta(s^2)$. By (4.5) there exist some constants $c_\epsilon, c'_\epsilon > 0$ such that the mixing time of the walk on the perturbed network is $c'_\epsilon s^2(1 \pm o(1))$ and thus the mixing time increased by a factor of $c_\epsilon \log |V|$. Moreover, a sequence of walks on the perturbed networks with k tending to infinity exhibits cutoff (around time $\mathbb{E}'_o[T_1]$). \square

Remark 4.3. The choice of a 3D torus was made to emphasize the similarity of the construction to the one from [10]; In fact, we could have used any bounded degree graph of size $\Theta(\log |V'|) = \Theta(s)$ which satisfies $\max_{x,y} \mathbb{E}_x[T_y] \leq Cs$.

4.3 Proof of part (b) of Theorem 3

We present two different constructions.

First construction: The first construction is obtained from the construction of the proof of part (a) by replacing the constant C (from step 3) by some $C(k)$ tending to infinity as $k \rightarrow \infty$ such that $C(k) = o(s)$ (where as above $s = k^3$). Call the obtained network $G = (V, E)$. Clearly (4.3) remains valid. This follows from the fact that (4.5) remains valid, and that for all $t > 0$ it is still true that $\max_{u \in V} \mathbb{P}_u[T_A > t] = \max_{u \in W_o} \mathbb{P}_u[T_A > t]$.

Consider a perturbation of the same collection of edges which were perturbed in the proof of part (a), only that now we increase their weights to $1 + K \sqrt{\frac{\log \log(C(k))}{C(k)}}$, where $K > 4$ is some sufficiently large absolute constant to be determined shortly. Similar reasoning as in the proof of part (a) (using part (2 a) of Fact 4.1, together with the law

of the iterated logarithm, here for a biased random walk on \mathbb{Z} , with a fixed bias) shows that if K is sufficiently large, this perturbation increases the order of the mixing time by a factor of order $s/C(k) = \Theta((\log |V|)/C(k))$ and that a sequence of random walks on the perturbed networks with $k \rightarrow \infty$ exhibits a cutoff. By taking $C(k)$ to tend to infinity arbitrarily slowly we can increase the mixing time by any factor $f_k \rightarrow \infty$ such that $f_k = o(\log |V|)$.

Second construction: We now present the second construction. Take sequences of integers k_n, r_n, ℓ_n, m_n tending to infinity as $n \rightarrow \infty$ such that $s_n = k_n^3 = \ell_n r_n$, $m_n \leq \lceil \ell_n^{1/4} \rceil$ and $r_n = \Theta\left(m_n e^{m_n^2/2}\right)$. The first two steps of the construction are taken as in the proof of part (a) with k_n in the role of k . We modify step 3 by changing the definition of the set D as follows.

Step 3': Set $D_0 := \{o\}$. For every $1 \leq i \leq s_n/(2\ell_n) = r_n/2$, we set D_i to be the collection of all vertices $u \in \mathcal{L}_{i\ell_n}(\mathcal{T})$ such that if $(v_0 = u, v_1, \dots, v_{\ell_n})$ is the path in \mathcal{T} between u and its ℓ_n -th ancestor $v_{\ell_n} \in \mathcal{L}_{(i-1)\ell_n}(\mathcal{T})$ then

$$|\{v_j : 0 \leq j \leq \ell_n\} \cap L| - |\{v_j : 0 \leq j \leq \ell_n\} \cap R| \geq \lceil m_n \sqrt{\ell_n} \rceil.$$

By the local CLT (4.2) and our choice of m_n and r_n we have that $|D_1|2^{-\ell_n} = \Theta(r_n^{-1})$. Consequently, for every $i \leq r_n/2 - 1$, for all $u \in \mathcal{L}_{(i-1)\ell_n}(\mathcal{T})$ we have that

$$P_u[T_{D_i} = T_{\mathcal{L}_{i\ell_n}(\mathcal{T})}] = \Theta(r_n^{-1}). \tag{4.8}$$

We define $D := \bigcup_{0 \leq i \leq r_n/2} D_i$. As before, we decorate each $v \in D$ by a 3D $k_n \times k_n \times k_n$ torus, W_v . Call the obtained graph $G_n = (V_n, E_n)$.

As before, let $A = A_n := \bigcup_{j \geq 3s_n/4} \mathcal{L}_j(\mathcal{T})$ and define T_i in an analogous manner to the way they were defined in the proof of part (a) ($i = 1, 2, 3$). By (4.8) we have that $\mathbb{E}[T_3] = \Theta(1)$ (using similar reasoning as in part (1 b) of Fact 4.1). Using this fact it is not hard to verify that (4.3) remains valid also here (with G_n and s_n in the roles of G and s). This follows from the fact that (4.5) remains valid, and that for all $t > 0$ it is still true that $\max_{u \in V_n} P_u[T_A > t] = \max_{u \in W_o} P_u[T_A > t]$.

We now describe the perturbation of the edge weights described in part (b) of the theorem. Let $\delta_n = Km_n/\sqrt{\ell_n}$, for some sufficiently large absolute constant K to be determined later. Perturb the same edges as before by increasing their weights to $1 + \delta_n$.

Much as before, by symmetry, it suffices to consider the case that the initial state of the walk on the perturbed network belongs to W_o . Since $\delta_n \ell_n = Km_n \sqrt{\ell_n}$, if K is taken to be sufficiently large, then by part (2) of Fact 4.1, for every $i \leq r_n/2 - 1$ and all $u \in \mathcal{L}_{(i-1)\ell_n}(\mathcal{T})$ we have that $P'_u[T_{D_i} = T_{\mathcal{L}_{i\ell_n}(\mathcal{T})}] = 1 - o(1)$. This implies that if K is taken to be sufficiently large, then in the perturbed network on G_n , starting from any $u \in W_o$, we have that T_3 is concentrated around some $t_n = \Theta(r_n)$. Consequently, T_A is concentrated around some $t'_n = \Theta(r_n k_n^3)$, which, as before, implies that the walk on the perturbed network exhibits cutoff around time t'_n . In particular, the order of the mixing-times increased by a factor of $\Theta(r_n)$. Setting $m_n = \lceil \ell_n^{1/4} \rceil$, we get that $r_n = \Theta(\log |V_n|/\sqrt{\log \log |V_n|})$. Note however that by taking m_n to tend to infinity arbitrarily slowly we get that s_n/ℓ_n tends to infinity arbitrarily slowly and thus so does $\delta_n \sqrt{t_{\text{mix}}(G_n)}$.

4.4 Proof of part (c) of Theorem 3

Consider the graph G from part (a). Now, stretch all of its edges by a factor 3. The tree \mathcal{T} is replaced by a tree \mathcal{T}' with stretched edges. However, we think of \mathcal{T} as being contained in \mathcal{T}' , with each pair of neighbors in \mathcal{T} being separated by a path of length 3 in \mathcal{T}' . Denote the obtained graph by G' . It is not hard to see that the mixing time of G' can differ from that of G only by a constant factor (the expectation of the hitting

time of the leaf set of \mathcal{T} is delayed by a factor of 3^2). Now, for each path of length 3, (u, w, w', v) connecting some vertex u of \mathcal{T} and its left child v , lump together the pair of its internal vertices w, w' . This has the same effect as replacing the path (u, w, w', v) by a path $(u, z_{u,v}, v)$ with a self loop of weight 2 at $z_{u,v}$. It is easy to see that this results in a bias towards the left children of \mathcal{T} (the bias is the same as when we remove the self-loops at the $z_{u,v}$'s; After removing the self-loops, a standard network reduction, similar to the one in the proof of Fact 4.1, can be used to establish the existence of the aforementioned bias). The analysis can be concluded in a similar manner to the analysis of the perturbed network in part (a). \square

Remark 4.4. It is possible to modify the examples from Theorem 3 so that they also satisfy $\text{girth}(G_n) = \Theta(t_{\text{mix}}(G_n))$. Namely, parts (1)-(2) of the construction can be replaced by starting with some regular expander of logarithmic girth ([19, 18]) as the base graph, like in the construction of Theorem 2. Then, instead of decorating the set D by tori of size $\Theta(\log |V_n|)$, we could decorate them by binary trees of that size.

Remark 4.5. Both Ding and Peres' example [10] and the example from Theorem 2 satisfy the product condition. It is thus natural to ask whether any sequence of bounded degree graphs satisfying $t_{\text{rel}}(G_n) = \Theta(t_{\text{mix}}(G_n))$ must be robust (equivalently, to ask whether the condition $t_{\text{rel}}(G_n) = \Theta(t_{\text{mix}}(G_n))$ is robust). Theorem 3 demonstrates that in fact the condition $t_{\text{rel}}(G_n) = \Theta(t_{\text{mix}}(G_n))$ may be $o(1)$ -sensitive.

Remark 4.6. A non-backtracking random walk on a simple graph $G = (V, E)$ (NBRW) evolves as follows. When at vertex u at some time t , if at time $t - 1$ the walk was at vertex v , the next position of the NBRW is chosen from the uniform distribution over $\{x \in V \setminus \{v\} : \{u, x\} \in E\}$. We may consider the lazy version of a NBRW.

A small variation of the example from part (a) of Theorem 3 shows that for a graph G of bounded degree (with no degree 1 vertices) the total variation mixing time of the lazy NBRW may be larger than that of the LSRW by a factor of $\Theta(\log |V|)$. Namely, one can stretch each edge between a vertex in \mathcal{T} and its right child by a factor 2. As in part (c) of Theorem 3, we think of \mathcal{T} as being contained in the modified tree. One can then define D_i to be the collection of all vertices u belonging to $\mathcal{L}_{C_i}(\mathcal{T})$ such that if $\gamma_u = (v_0 = u, v_1, \dots, v_{C_i} = o)$ is the path from u to o in \mathcal{T} , then for all $1 \leq j \leq i$

$$|\{v_\ell : 0 \leq \ell \leq Cj\} \cap L| - |\{v_\ell : 0 \leq \ell \leq Cj\} \cap R| \leq \epsilon j,$$

for some small $\epsilon > 0$ and, as before, set $D := \bigcup_{i=0}^{\lfloor s/2C \rfloor} D_i$. Finally, as before, decorate each $d \in D$ by a $3D$ torus of side length k and connect the leafs of \mathcal{T} using an expander H . It is not hard to see that if ϵ is taken to be sufficiently small, then due to the bias towards the left children of \mathcal{T} , resulting from stretching the "right edges", w.h.p. the LSRW will visit only a constant number of tori before reaching the $3s/4$ level of \mathcal{T} . However, the harmonic measure of the lazy NBRW is unaffected by the stretched edges (meaning that if v and v' are the children of u in \mathcal{T} , then also in the modified graph, when the lazy NBRW is at u , it has the same probability of reaching either v or v' before the other), which means that w.h.p. it will visit $\Theta(k^3)$ tori before reaching the $3s/4$ level of \mathcal{T} . Since also the lazy NBRW spends at average $\Theta(k^3)$ steps at each torus, this means that the hitting time of the $3s/4$ level of \mathcal{T} is $\Omega(k^6)$, which as before, implies that the mixing time of the lazy NBRW is $\Omega(k^6)$.

5 Proof of Theorem 2

Our construction is obtained by stretching some of the edges of a *Ramanujan Cayley graph*. Let H be a group and $S \subset H$ be a finite symmetric (i.e. $S = S^{-1} := \{s^{-1} : s \in S\}$) set of generators of H (i.e. every $h \in H$ can be written as a finite product of the form

$s_1 s_2 \cdots s_k$ where $s_i \in S$ for all $1 \leq i \leq k$). The **Cayley graph** of H w.r.t. S is defined to be the graph whose vertex set is H and whose edge set is $\{\{h, hs\} : h \in H, s \in S\}$.

Let G be a d -regular connected graph of size n . Denote the transition matrix of simple random walk on G by P . Denote the eigenvalues of P by $\lambda_n \leq \cdots \leq \lambda_2 < \lambda_1 = 1$. We say that G is a **Ramanujan graph** if $|\lambda_i| \in [0, 2d^{-1}\sqrt{d-1}] \cup \{1\}$, for all $i \leq n$.

Let p, q be two distinct prime numbers congruent to 1 modulo 4 such that $q > \sqrt{p}$ and $q \equiv a^2$ modulo p for some integer a . Then there exists a $(p+1)$ -regular Ramanujan Cayley graph $G_{p,q}$ of size $q(q^2-1)/2$ whose girth is of size at least $2 \log_p q$ [18]. We fix $p = 5$ and take an increasing sequence $(q_n)_{n \in \mathbb{N}}$ of such prime numbers and consider $H_n = G_{5,q_n} = (V'_n, E'_n)$. Fix some vertex $o \in V'_n$. Note that up to a distance $\log_5 q_n$ from o the graph H_n looks like a 6-regular tree.

We take some sequences of integers s_n, m_n, b_n tending to infinity such that:

- $\log_5 q_n / 4 \leq s_n^2 m_n b_n \leq \log_5 q_n / 2$.
- $e^{m_n/b_n} = \Theta(s_n^2 b_n)$.

We think of b_n as tending to infinity arbitrarily slowly (compared to s_n). As $e^{m_n/b_n} = \Theta(s_n^2 b_n)$ we think also of $m_n / \log s_n$ and $(\log q_n) / (s_n^2 \log \log q_n)$ as tending to infinity arbitrarily slowly.

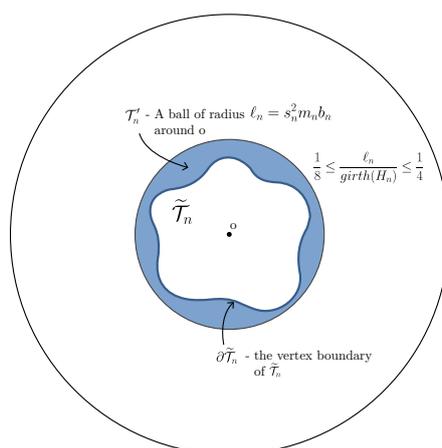


Figure 4: This is a schematic representation of H_n and the trees \mathcal{T}'_n and $\tilde{\mathcal{T}}_n$

Denote the ball of radius $s_n^2 m_n b_n$ around o by $\mathcal{T}'_n = (V(\mathcal{T}'_n), E(\mathcal{T}'_n))$. We think of \mathcal{T}'_n as a 6-regular tree rooted at o . We shall construct a sequence of graphs $G_n = (V_n, E_n)$ by stretching some of the edges of H_n by a factor s_n as follows. We shall pick a certain subtree $\tilde{\mathcal{T}}_n \subset \mathcal{T}'_n$ (which is also rooted at o) and replace each edge $\{v, u\}$ in $\tilde{\mathcal{T}}_n$ by a path $\gamma_{u,v}$ of length s_n whose end-points are u and v . We call the resulting tree $\mathcal{T}_n = (V(\mathcal{T}_n), E(\mathcal{T}_n))$. We identify each vertex of $\tilde{\mathcal{T}}_n = (V(\tilde{\mathcal{T}}_n), E(\tilde{\mathcal{T}}_n))$ with the corresponding vertex of \mathcal{T}_n .

The stretched edges have the effect of significantly “slowing down” the walk while it is confined to \mathcal{T}_n . With some care, we shall choose $\tilde{\mathcal{T}}_n$ so that

- $|\mathcal{T}_n|/|V_n| = o(1/q_n^2)$ and so the walk must escape \mathcal{T}_n before mixing.
- The distribution of the escape time from \mathcal{T}_n starting from the root, is stochastically the largest (compared to all other starting points).
- The distribution of $T_{V(\mathcal{T}_n)^c} / (3s_n^2 m_n b_n)$, starting from o , is “close” to the Exponential distribution with some constant mean (see (5.2) for a precise statement).

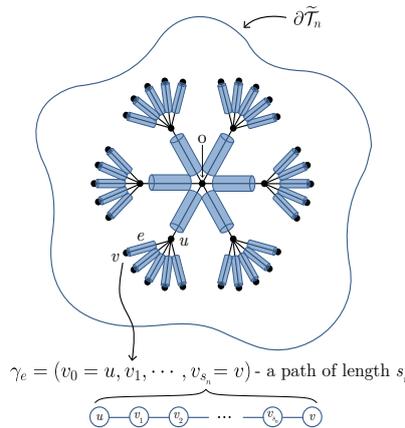


Figure 5: The graph G_n is obtained by stretching each edge of $\tilde{\mathcal{T}}_n$ by a factor of s_n . Each cylinder above represents a path of length s_n , which replaces some edge of $\tilde{\mathcal{T}}_n$.

- Once the walk escapes \mathcal{T}_n it has a negligible chance of crossing any stretched edge by the time it is already extremely mixed. Consequently, there exists some $o(1)$ terms such that the additional amount of time required for the walk to become $\epsilon + o(1)$ mixed, beyond the time required for it to escape \mathcal{T}_n with probability of at least $1 - \epsilon$, can be upper bounded by $t_{\text{mix}, H_n}(o(1)) \leq C \log q_n = \Theta(s_n^2 m_n b_n)$ (more precisely, we derive such a bound using Proposition 6.6). Putting all this together, it follows that

$$\frac{t_{\text{mix}, G_n}(\epsilon)}{|\log \epsilon|} = \Theta(s_n^2 m_n b_n) = \Theta(\log q_n), \text{ uniformly for every } 0 < \epsilon \leq 1/2. \quad (5.1)$$

Hence there is no pre-cutoff, although the product condition holds (by Lemma 6.7 $t_{\text{rel}}(G_n) = O(s_n^2) = o(\log q_n)$).

- Under a certain $o(1)$ -perturbation of some of the edges of \mathcal{T}_n , w.h.p. the walk would remain “trapped” in \mathcal{T}_n until escaping it through the collection of its leaves which have maximal distance from the root. Consequently, as opposed to the situation in the original graph, the escape time from \mathcal{T}_n (starting from the root) is concentrated, but around a time of strictly larger order than $\log q_n$, namely $3s_n^4 m_n b_n = \Theta(s_n^2 \log q_n)$. Thus the walk on the perturbed network exhibits a cutoff around time $3s_n^4 m_n b_n$.

We denote the internal vertex boundary of \mathcal{T}_n w.r.t. G_n and of $\tilde{\mathcal{T}}_n$ w.r.t. \mathcal{T}'_n by

$$\partial \mathcal{T}_n := \{v \in V(\mathcal{T}_n) : \exists u \notin V(\mathcal{T}_n) \text{ such that } \{u, v\} \in E_n\},$$

$$\partial \tilde{\mathcal{T}}_n := \{v \in V(\tilde{\mathcal{T}}_n) : \exists u \notin V(\tilde{\mathcal{T}}_n) \text{ such that } \{u, v\} \in E(\mathcal{T}'_n)\}.$$

By construction $\partial \tilde{\mathcal{T}}_n = \partial \mathcal{T}_n$. As $\tilde{\mathcal{T}}_n$ is rooted at o and $\tilde{\mathcal{T}}_n \subset \mathcal{T}'_n$, in order to define $\tilde{\mathcal{T}}_n$ it suffices to specify its collection of leaves, which is precisely the set $\partial \tilde{\mathcal{T}}_n$. Namely, $V(\tilde{\mathcal{T}}_n)$ is determined by $\partial \tilde{\mathcal{T}}_n$ as the union of the vertices along all paths in \mathcal{T}'_n from the root to $\partial \tilde{\mathcal{T}}_n$. We now describe our procedure for choosing the set $\partial \tilde{\mathcal{T}}_n$ (this concludes the construction).

- (1) Denote the k -th level of \mathcal{T}'_n by $\mathcal{L}_k(\mathcal{T}'_n)$ (this level contains the ks_n -th level of \mathcal{T}_n). Then we shall construct $\partial \tilde{\mathcal{T}}_n$ so that $\partial \tilde{\mathcal{T}}_n \subset \bigcup_{k=2}^{s_n^2 b_n} \mathcal{L}_{km_n}(\mathcal{T}'_n)$. We shall define $D_k :=$

A 5-ary tree \mathcal{T} of depth m_n .

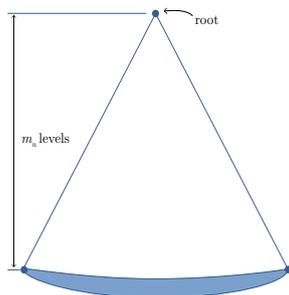


Figure 6: The leaves of \mathcal{T} are partitioned into two parts, A and D , where the set D consist of leaves which are in some sense “unbalanced”. The set D belongs to $\partial\tilde{\mathcal{T}}_n$, while for every $a \in A$ another copy of \mathcal{T} is contained in $\tilde{\mathcal{T}}_n$. We repeat this procedure for $s_n - 1$ iterations (see Figure 7), where in every iteration, the partition of the leaves into A and D is identical.

$\partial\tilde{\mathcal{T}}_n \cap \mathcal{L}_{km_n}(\mathcal{T}'_n)$ recursively starting from $k = 2$ (i.e. we set $\partial\tilde{\mathcal{T}}_n := \bigcup_{k=2}^{s_n^2 b_n} D_k$). Having defined $D_2 = \partial\mathcal{T}_n \cap \mathcal{L}_{2m_n}(\mathcal{T}'_n), \dots, D_k = \partial\mathcal{T}_n \cap \mathcal{L}_{km_n}(\mathcal{T}'_n)$, we define A_{k+1} to be the set of vertices in $\mathcal{L}_{(k+1)m_n}(\mathcal{T}'_n)$ such that the path from them to o in \mathcal{T}'_n does not go through any vertex in $\bigcup_{i=2}^k D_i$. The set D_{k+1} shall be defined to be a certain subset of the vertices in $\mathcal{L}_{(k+1)m_n}(\mathcal{T}'_n)$ which have a vertex in A_k as an ancestor (this is described in (2)-(3) below). To start the construction we set $A_1 := \mathcal{L}_{m_n}(\mathcal{T}'_n)$ and to conclude it we define $D_{s_n^2 b_n} = \partial\mathcal{T}_n \cap \mathcal{L}_{s_n^2 b_n m_n}(\mathcal{T}'_n)$ to equal $A_{s_n^2 b_n}$ (making the last level of \mathcal{T}_n different). We now specify how D_{k+1} is defined in terms of A_k (first qualitatively (2) and then more concretely (3)).

- (2) For every $k \leq m_n b_n s_n^2 - 1$ and every $v \in \mathcal{L}_{km_n}(\mathcal{T}'_n)$ we denote

$$\mathcal{T}_v := \{u \in \bigcup_{i=0}^{m_n} \mathcal{L}_{km_n+i}(\mathcal{T}'_n) : \text{the path in } \mathcal{T}'_n \text{ between } u \text{ and } o \text{ goes through } v\}.$$

Let B_v be the set of leaves of \mathcal{T}_v . For all $2 \leq k \leq s_n^2 b_n - 1$ we will define D_k to be a subset of $\bigcup_{v \in A_{k-1}} B_v$. We shall define $\tilde{\mathcal{T}}_n$ so that for every $1 \leq k_1, k_2 \leq s_n^2 b_n - 2$ and every $v_i \in A_{k_i}$ ($i = 1, 2$) the trivial isomorphism of \mathcal{T}_{v_1} and \mathcal{T}_{v_2} is a bijection from $B_{v_1} \cap \partial\tilde{\mathcal{T}}_n$ onto $B_{v_2} \cap \partial\tilde{\mathcal{T}}_n$. In particular, $|B_v \cap \partial\tilde{\mathcal{T}}_n|/|B_v|$ is some fixed number, which we shall pick to be between b_n^{-1} and $2b_n^{-1}$.

- (3) We now define the sets $D_2, \dots, D_{s_n^2 b_n}$. For every vertex $v \in A_k$ for some $k < s_n^2 b_n - 1$ and every $u \in \mathcal{T}_v \setminus B_v$, we distinguish one of the children of u (w.r.t. the tree \mathcal{T}_v viewed as a rooted tree with root v) as a *left* child. For every $u \in B_v$ let $f(u)$ be the number of left children along the path from u to v in \mathcal{T}_v . We define $F_v = B_v \cap \partial\tilde{\mathcal{T}}_n$ to be the collection of all $u \in B_v$ such that $f(u) \leq m_n/5 - g_n(m_n)$, where g_n (which may depend on n and (b_n, m_n)) is chosen so that $1/b_n \leq |B_v \cap \partial\tilde{\mathcal{T}}_n|/|B_v| \leq 2/b_n$ (for all sufficiently large n). Finally, we set $D_{k+1} := \bigcup_{v \in A_k} F_v$.

We start by describing four properties of G_n which do not depend on the particular choice of $\tilde{\mathcal{T}}_n$.

- (a) $\text{girth}(G_n) \geq \text{girth}(H_n) \geq 2 \log_5 q_n$.
 (b) $|\mathcal{T}_n| \leq s_n |\mathcal{T}'_n| \leq s_n (5^{s_n^2 m_n b_n} + 1) \leq s_n (5^{\frac{1}{2} \log_5 q_n} + 1) = o(\sqrt{q_n \log q_n})$. Consequently, $|V_n| = q_n(q_n^2 - 1)/2 + o(\sqrt{q_n \log q_n})$.

A schematic representation of the recursive construction of $\tilde{\mathcal{T}}_n$

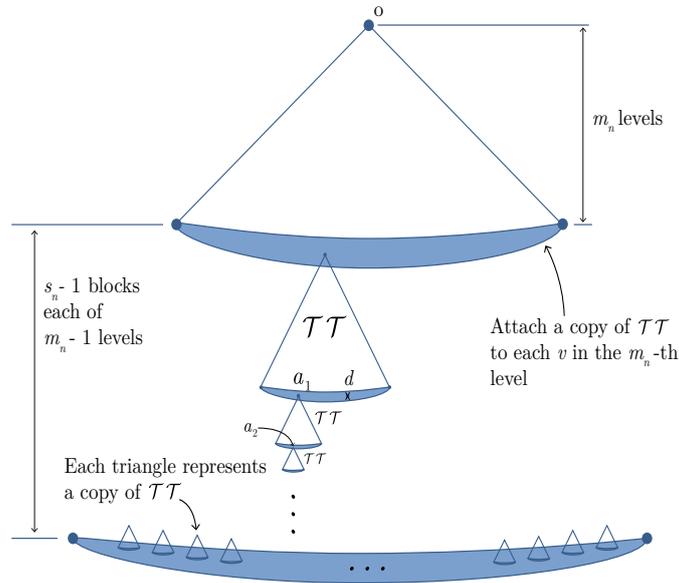


Figure 7: Every triangle represents a copy of $\mathcal{T}\mathcal{T}$. Every $d \in D$ belongs to $\partial\tilde{\mathcal{T}}$, while for every $a \in A$ another copy of $\mathcal{T}\mathcal{T}$ is contained in $\tilde{\mathcal{T}}_n$. We repeat this procedure for $s_n - 1$ iterations, where in every iteration, the partition of the leaves into A and D is identical. The leaves of the copies of $\mathcal{T}\mathcal{T}$ from the last iteration all belong to $\partial\tilde{\mathcal{T}}$. Note that the first iteration is different, in that all of the leaves of the first copy of $\mathcal{T}\mathcal{T}$ (the one rooted at o) are in A .

- (c) $t_{\text{rel}}(G_n) = O(s_n^2)$ (by Lemma 6.7). In particular, $t_{\text{rel}}(G_n) = o(\log q_n)$ (which implies that the product condition holds).
- (d) There exist absolute constants $C_1, C_2 > 0$ such that for every vertex $v \in V_n$ whose distance from \mathcal{T}_n is at least $C_1 \log \log q_n$ we have that $\|\mathbb{P}_v^{[C_2 \log q_n]} - \pi_n\|_{\text{TV}} = o(1)$.
- (e) Let C_1 be as in (d). Denote by J_n the collection of vertices whose distance from \mathcal{T}_n is at least $\lceil C_1 \log \log q_n \rceil$. Then for every $u \in \partial\mathcal{T}_n$ we have that

$$\mathbb{P}_u[T_{J_n} > 5C_1 \log \log q_n] = o(1).$$

(a) and (b) are trivial. For (d) note that for any u in the exterior vertex boundary of \mathcal{T}_n , the intersection of the ball of radius $\lceil C_1 \log \log q_n \rceil$ centered at u with $V_n \setminus V(\mathcal{T}_n)$ is a 5-ary tree. Hence (d) follows from Proposition 6.6.

For (e) observe that for any $v \in \partial\mathcal{T}_n$ the probability that a lazy random walk started from v reaches its parent u w.r.t. $\tilde{\mathcal{T}}_n$ (by crossing the path $\gamma_{u,v}$) before reaching J_n is $o(1)$. Moreover, the probability that the walk would spend at least $C_1 \log \log n$ steps in $\gamma_{u,v}$ before reaching J_n is $o(1)$. Finally, conditioned on hitting J_n before returning to v , the conditional distribution of T_{J_n} is concentrated around $3C_1 \log \log q_n$.

- (f) For any $t \geq 0$,

$$\mathbb{P}_o[T_{\partial\mathcal{T}_n} > t] = \max_{v \in \mathcal{T}_n} \mathbb{P}_v[T_{\partial\mathcal{T}_n} > t].$$

(g) There exists an absolute constant $c_1 > 0$ such that for any $2 \leq k \leq m_n s_n^2 - 1$

$$\begin{aligned} P_o [T_{\partial\mathcal{T}_n} > 3s_n^2 m_n (1 + o(1))k] &\leq (1 - b_n^{-1})^{k-1} + o(1). \\ P_o [T_{\partial\mathcal{T}_n} \leq 3s_n^2 m_n (1 - o(1))k] &\geq 1 - (1 - c_1 b_n^{-1})^{k-1} - o(1). \end{aligned} \tag{5.2}$$

Note that (5.2) together with (d)-(f) imply (5.1).

It is easy to see how (f) follows from the symmetry of the construction together with the fact that the construction of the sets D_i starts only from $k = 2$.

We now explain (g). The function f (from (3) above) could be extended to $V(\mathcal{T}_n)$ by contracting the stretched edges into a single edge (where internal vertices of an edge are assigned the same value as one of the end-points of that stretched edge). Fix some $1 \leq k \leq s_n^2 b_n - 2$ and some $v \in A_k$. Let Y be the last vertex in $\mathcal{L}_{(k+1)m_n}(\tilde{\mathcal{T}}_n)$ visited by the walk prior to $T_{\partial\mathcal{T}_n}$. Then starting from v , conditioned on $T_{\partial\mathcal{T}_n} < T_{\mathcal{L}_{km_n-1}(\tilde{\mathcal{T}}_n)}$ we have that the (conditional) law of $f(Y)$ w.r.t. the walk on $\tilde{\mathcal{T}}_n$ is the same as its (conditional) law w.r.t. the walk on \mathcal{T}_n . This law can be approximated well by that of a sum of m_n i.i.d. Bernoulli(1/5) r.v.'s. Similarly to (4.1), the conditional probability that $X_{T_{\partial\mathcal{T}_n}} \in \mathcal{L}_{(k+1)m_n}(\tilde{\mathcal{T}}_n)$ is at most $c^{-1} P_v[Y \in \partial\mathcal{T}_n \mid T_{\partial\mathcal{T}_n} < T_{\mathcal{L}_{km_n-1}(\tilde{\mathcal{T}}_n)}]$. Using this observation, it is easy to see that the probability that the walk reached $\mathcal{L}_{km_n}(\tilde{\mathcal{T}}_n) = \mathcal{L}_{ks_n^2 m_n}(\mathcal{T}_n)$ without first hitting $\partial\mathcal{T}_n$ is between $(1 - b_n^{-1})^{k-1}$ and $(1 - c_1 b_n^{-1})^{k-1}$. Whence (g) follows from the fact that starting from o the hitting time of $\mathcal{L}_{km_n}(\mathcal{T}'_n)$ (conditioned that it is hit before $\partial\mathcal{T}_n$) is concentrated around $3s_n^2 m_n k$.

To see this, first consider the hitting time of $\mathcal{L}_{km_n}(\mathcal{T}'_n)$ with respect to the non-lazy version of the induced chain on \mathcal{T}'_n (starting from o). Note that it is concentrated around $\frac{3}{2}m_n k$ (its distance from o (w.r.t. \mathcal{T}'_n) is distributed like a biased nearest-neighbor random walk with a fixed bias). Finally, since for a lazy SRW on \mathbb{Z}_+ we have that $\mathbb{E}_0[T_{\{-s_n, s_n\}}] = 2s_n^2$ and $\mathbb{E}_0[T_{\{-s_n, s_n\}}^2] = O(s_n^4)$, the conditional distribution of $T_{\mathcal{L}_{km_n}(\mathcal{T}'_n)}$ (starting from o), given that $T_{\mathcal{L}_{(k-1)m_n}(\mathcal{T}'_n)} > T_{\partial\mathcal{T}_n}$, is extremely close to that of $\sum_{i=1}^T W_i$, where T is distributed like $T_{\mathcal{L}_{km_n}(\mathcal{T}'_n)}$ w.r.t. the aforementioned non-lazy walk on \mathcal{T}'_n and W_1, W_2, \dots are i.i.d. random variables of mean $2s_n^2$ and variance $O(s_n^4)$ (which are also independent of T). Hence (g) follows from the CLT in conjunction with the aforementioned concentration of T around time $\frac{3}{2}m_n k$.

We now describe the $o(1)$ -perturbation described in the assertion of the theorem. For every vertex $v \in A_k$ for some $k < m_n b_n s_n^2 - 1$, we increase the edge weight of every edge belonging to some $\gamma_{u,w}$ such that w is a left child of u in \mathcal{T}_v to $1 + 1/b_n^{1/3}$.

We consider also the perturbed network on $\tilde{\mathcal{T}}_n$ in which we increase the edge weight between each vertex and its left child to $1 + 1/b_n^{1/3}$. Let k, v and Y be as in the paragraph following (5.2). A simple network reduction shows that starting from v , conditioned on $T_{\partial\mathcal{T}_n} < T_{\mathcal{L}_{km_n-1}(\tilde{\mathcal{T}}_n)}$ we have that the (conditional) law of $f(Y)$ w.r.t. the walk on the perturbed network on $\tilde{\mathcal{T}}_n$ is the same as its (conditional) law w.r.t. the perturbed walk on \mathcal{T}_n . Similarly to part (2 a) of Fact 4.1, this law can be approximated well by that of a sum of m_n i.i.d. Bernoulli($1/5 + O(1/b_n^{1/3})$) r.v.'s.

As before, up to constants, we may consider $P_v[Y \in \partial\mathcal{T}_n \mid T_{\partial\mathcal{T}_n} < T_{\mathcal{L}_{km_n-1}(\tilde{\mathcal{T}}_n)}]$ rather than $P_v[X_{T_{\partial\mathcal{T}_n}} \in \mathcal{L}_{(k+1)m_n}(\tilde{\mathcal{T}}_n) \mid T_{\partial\mathcal{T}_n} < T_{\mathcal{L}_{km_n-1}(\tilde{\mathcal{T}}_n)}]$ (here both probabilities are considered w.r.t. the perturbed network). Note that the former event requires a deviation of order $m_n/b_n^{1/3}$ from the mean of $f(Y)$ and thus its conditional probability could be bounded from above by $\exp(-c'_2 m_n b_n^{-2/3})$.

By the above discussion, it is easy to see that also after this perturbation (d), (e) and (f) remain valid and that for all $k < s_n^2 b_n$

$$P_o [T_{\partial\mathcal{T}_n} \geq T_{\mathcal{L}_{(k+1)m_n}(\mathcal{T}'_n)}] \geq 1 - \left(1 - \exp[-c_2 m_n b_n^{-2/3}]\right)^k - o(1),$$

for some constant $c_2 > 0$. In particular, since we took $e^{m_n/b_n} = \Theta(s_n^2 b_n)$ we get that

$$P_o \left[T_{\partial\mathcal{T}_n} \neq T_{\mathcal{L}_{s_n^2 m_n b_n}(\mathcal{T}'_n)} \right] = o(1).$$

As before, it follows that $T_{\partial\mathcal{T}_n}$ is concentrated around $3s_n^4 m_n b_n$. By (d) and (e) this implies that after the perturbation the random walk exhibits cutoff around this time. In particular, the order of the mixing time increased by a factor of s_n^2 . Observe that from our assumption that $e^{m_n/b_n} = \Theta(s_n^2 b_n)$ we get that $\log |V_n| / \log \log |V_n| = O((s_n b_n)^2)$. Note that by taking b_n to tend to infinity arbitrarily slowly, we can increase the order of mixing time by a factor arbitrarily close to $\frac{\log |V_n|}{\log \log |V_n|}$ (as long as it is $o(\log |V_n| / \log \log |V_n|)$). \square

Remark 5.1. Lack of pre-cutoff implies that (along a certain subsequence) the chains mix “very-gradually”. We note that the graphs G_n from Theorem 2 exhibit, in some sense, the most gradual mixing possible. In general, $t_{\text{mix}}(\epsilon) \leq \lceil 2 \log_2 \epsilon \rceil t_{\text{mix}}$, for all $0 < \epsilon < 1/4$. Using the results in [5], it is not hard to show that for reversible chains, under the product condition, there is some $o(1)$ term (depending only on $t_{\text{rel}}^{(n)} / t_{\text{mix}}^{(n)}$) such that

$$\liminf_{n \rightarrow \infty} [t_{\text{mix}}^{(n)}(\epsilon) + t_{\text{mix}}^{(n)}(\delta - o(1))] / t_{\text{mix}}^{(n)}(\epsilon\delta) \geq 1, \text{ for all } 0 < \epsilon, \delta < 1.$$

The graphs G_n from Theorem 2 satisfy that

$$t_{\text{mix}}^{(n)}(\epsilon) \geq c \log \epsilon |t_{\text{mix}}^{(n)}(1/2)|, \text{ for all } 0 < \epsilon < 1/2 \text{ and all } n.$$

We note that (by stretching the stretched edges in the construction by a slightly larger factor) we could have constructed the graphs G_n from Theorem 2 so that we have that $t_{\text{mix}}(G_n) / \text{girth}(G_n)$ tends to infinity arbitrarily slowly (but still $\text{girth}(G_n) = O(\text{diameter}(G_n))$) and so that

$$\lim_{n \rightarrow \infty} [t_{\text{mix}}^{(n)}(\epsilon) + t_{\text{mix}}^{(n)}(\delta)] / t_{\text{mix}}^{(n)}(\epsilon\delta) = 1, \text{ for all } 0 < \epsilon, \delta < 1. \tag{5.3}$$

The probabilistic interpretation of (5.3) is that, loosely speaking, there is some random time τ_n (which can be taken to be a certain hitting time) having roughly a Geometric distribution such that the chain is extremely mixed in time $\tau_n + o(\mathbb{E}[\tau_n])$ (but its distance from π is $1 - o(1)$ at time $\tau_n - o(\mathbb{E}[\tau_n])$).

References

- [1] L. Addario-Berry and M. I. Roberts, Mixing Time Bounds via Bottleneck Sequences, *Journal of Statistical Physics*, (2017): 1–27.
- [2] D. Aldous and J. Fill, Reversible Markov chains and random walks on graphs, (2002) Unfinished manuscript. Available at <http://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [3] N. Alon and V. Milman, λ_1 isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B* 38.1 (1985): 73–88. MR-782626
- [4] N. Alon, λ_1 Eigenvalues and expanders, *Combinatorica* 6.2 (1986): 83–96. MR-875835
- [5] R. Basu, J. Hermon, and Y. Peres, Characterization of cutoff for reversible Markov chains, *Ann. Probab.* 45.3 (2017): 1448–1487. MR-3650406
- [6] I. Benjamini, Instability of the Liouville property for quasi-isometric graphs and manifolds of polynomial volume growth, *Journal of Theoretical Probability* 4.3 (1991): 631–637. MR-1115166
- [7] G. Y. Chen and L. Saloff-Coste, Comparison of cutoffs between lazy walks and Markovian semigroups, *Journal of Applied Probability* 50.4 (2013): 943–959. MR-3161366
- [8] P. Diaconis and L. Saloff-Coste, Logarithmic Sobolev inequalities for finite Markov chains, *The Annals of Applied Probability*, 6(3) (1996): 695–750. MR-1410112

- [9] J. Ding, E. Lubetzky and Y. Peres, Total variation cutoff in birth-and-death chains, *Probability Theory and Related Fields* 146.1-2 (2010): 61–85. MR-2550359
- [10] J. Ding and Y. Peres, Sensitivity of mixing times, *Electronic Communications in Probability* 18.88 (2013): 1–6. MR-3141797
- [11] N. Fountoulakis and B. A. Reed, Faster mixing and small bottlenecks, *Probability Theory and Related Fields* 137.3 (2007): 475–486. MR-2278465
- [12] S. Goel, R. Montenegro and P. Tetali, Mixing time bounds via the spectral profile, *Electron. J. Probab* 11.1 (2006): 1–26. MR-2199053
- [13] J. Hermon, On sensitivity of uniform mixing times, *Annales de l’Institut Henri Poincaré Probabilités et Statistiques* 54.1 (2018): 234–248. MR-3765888
- [14] J. Hermon, H. Lacoïn and Y. Peres, Total Variation and Separation Cutoffs are not equivalent and neither one implies the other, *Electron. J. Probab.*, 21(44) (2016): 1–36. MR-3530321
- [15] G. Kozma, On the precision of the spectral profile, *ALEA Lat. Am. J. Probab. Math. Stat.* 3 (2007): 321–329. MR-2372888
- [16] D. Levin and Y. Peres, *Markov chains and mixing times*, American Mathematical Soc. (2017). With contributions by Elizabeth L. Wilmer and a chapter by James G. Propp and David B. Wilson. MR-3726904
- [17] E. Lubetzky and A. Sly, Explicit expanders with cutoff phenomena, *Electronic Journal Probability* 16.15 (2011): 419–435. MR-2774096
- [18] A. Lubotzky, R. Phillips and P. Sharnak, Ramanujan Graphs, *Combinatorica* 8.3 (1988): 261–277. MR-963118
- [19] M. Morgenstern, Existence and explicit constructions of $q+1$ regular Ramanujan graphs for every prime power q , *Journal of Combinatorial Theory, Series B* 62.1 (1994): 44–62. MR-1290630
- [20] B. Morris and Y. Peres, Evolving sets, mixing and heat kernel bounds, *Probability Theory and Related Fields* 133.2 (2005): 245–266. MR-2198701
- [21] Y. Peres. American Institute of Mathematics (AIM) research workshop “Sharp Thresholds for Mixing Times” (Palo Alto, December 2004). Summary available at <http://www.aimath.org/WWN/mixingtimes>.
- [22] Y. Peres and P. Sousi, Mixing times are hitting times of large sets. *Journal of Theoretical Probability* 28.2 (2013): 1–32. MR-3370663
- [23] C. Pittet and L. Saloff-Coste, On the stability of the behavior of random walks on groups. *Journal of Geometric Analysis*, 10(4) (2000): 713–737. MR-1817783
- [24] A. Sinclair and M. Jerrum, Approximate counting, uniform generation and rapidly mixing Markov chains, *Information and Computation* 82.1 (1989): 93–133. MR-1003059

6 Appendix

6.1 Hitting times connection to mixing times

The aim of this section is to introduce some general theory which shall reduce the analysis of our examples to the analysis of hitting time distributions of certain sets.

Definition 6.1. Let (Ω, P, π) be a finite irreducible reversible Markov chain. For any $f \in \mathbb{R}^\Omega$, let $\mathbb{E}_\pi[f] := \sum_{x \in \Omega} \pi(x)f(x)$ and $\text{Var}_\pi f := \mathbb{E}_\pi[(f - \mathbb{E}_\pi f)^2]$. The inner-product $\langle \cdot, \cdot \rangle_\pi$ and L^p norm are

$$\langle f, g \rangle_\pi := \mathbb{E}_\pi[fg] \text{ and } \|f\|_p := (\mathbb{E}_\pi[|f|^p])^{1/p}, \quad 1 \leq p < \infty,$$

and $\|f\|_\infty := \max_{x \in \Omega} |f(x)|$. For any two distribution μ on Ω and $p \geq 1$ we define

$$\|\mu - \pi\|_{p,\pi} := \|f_\mu - 1\|_p, \quad \text{where } f_\mu(x) := \mu(x)/\pi(x).$$

The following lemma is standard.

Lemma 6.2. *Let (Ω, P, π) be a finite lazy irreducible reversible Markov chain. Let μ be a distribution on Ω and let λ_2 be the second largest eigenvalue of P . Then for all $t \geq 0$*

$$2\|P_\mu^t - \pi\|_{\text{TV}} \leq \|P_\mu^t - \pi\|_{2,\pi} \leq \lambda_2^t \|\mu - \pi\|_{2,\pi}. \quad (6.1)$$

Proof. The first inequality in (6.1) follows from the fact that $2\|P_\mu^t - \pi\|_{\text{TV}} = \|P_\mu^t - \pi\|_{1,\pi} \leq \|P_\mu^t - \pi\|_{2,\pi}$ by the Cauchy-Schwarz inequality. The second inequality in (6.1) is proved using the spectral decomposition in a straightforward manner (e.g. [2, Lemma 3.26]). \square

Definition 6.3. *Let (Ω, P, π) be a finite Markov chain. For a set $A \subset \Omega$ denote $Q(A) := \sum_{x \in A, y \notin A} \pi(x)P(x, y)$. Denote $\Phi(A) := Q(A)/\pi(A)$. We define the **Cheeger constant** of the chain as $\Phi := \min_{A: 0 < \pi(A) \leq 1/2} \Phi(A)$.*

Definition 6.4. *Let $G = (V, E)$ be a finite graph. The edge boundary of a set $S \subset V$ is defined as $\partial_E S := \{\{s, s'\} \in E : s \in S, s' \notin S\}$. The Cheeger constant of lazy simple random walk on G is defined as*

$$\text{ch}_L(G) := \min_{S: 0 < \pi(S) \leq 1/2} \frac{|\partial_E S|}{2 \sum_{v \in S} \deg(v)},$$

which coincides with Definition 6.3 (see e.g. [16, Remark 7.2]). We say that G is a c -lazy expander if $\text{ch}_L(G) > c$. We say that a sequence of finite graphs $(G_n)_{n \geq 1}$ is a family of c -lazy expanders if $\inf_n \text{ch}_L(G_n) > c$.

The following theorem is the well-known discrete analog of Cheeger's inequality [3, 4, 24] (the proof could also be found at [16, Theorem 13.14]).

Theorem 6.5. *Let λ_2 be the second largest eigenvalue of a reversible transition matrix on a finite state space. Let Φ be as in Definition 6.3. Then*

$$\Phi^2/2 \leq 1 - \lambda_2 \leq 2\Phi. \quad (6.2)$$

The following proposition enables us to reduce the problem of bounding $d(t)$ from above in the proofs of Theorems 2 and 3 to the problem of estimating the probability that a certain large set A ("the center of mass of the chain") was not hit by time t .

Proposition 6.6. *Let $G = (V, E)$ be a finite connected graph. Fix some edge weights $(c_e)_{e \in E}$. Assume that $1 \leq \sum_u c_{v,u} \leq D$ for all $v \in V$. Let (V, P, π) be the lazy random walk on the corresponding network.*

Let $0 < \epsilon < 1$. Let $A \subset V$. Let $\partial_V A := \{a \in A : \exists b \notin A \text{ s.t. } \{a, b\} \in E\}$ and $\tilde{A} := A \setminus \partial_V A$. Assume that $\pi(\tilde{A}) \geq 1 - \epsilon/3$. Let $(A, \tilde{P}, \tilde{\pi})$ be lazy random walk on the induced network on A . Denote its Cheeger constant by c . Denote $r := \left\lceil \frac{2}{c^2} \log \left(\frac{3D|A|}{2\epsilon} \right) \right\rceil$. Let $x \in \tilde{A}$ be such that $P_x[T_{\partial_V A} < r] \leq \epsilon/3$. Then

$$\|P_x^r - \pi\|_{\text{TV}} \leq \epsilon.$$

Proof. First note that by (6.1) together with (6.5)

$$\|\tilde{P}^r(x, \cdot) - \tilde{\pi}\|_{\text{TV}} \leq \frac{1}{2} \|\tilde{P}^r(x, \cdot) - \tilde{\pi}\|_{2,\tilde{\pi}} \leq \frac{1}{2\tilde{\pi}(x)} \lambda_2^r \leq \frac{D|A|}{2} e^{-\log(\frac{3D|A|}{2\epsilon})} = \epsilon/3.$$

We also have that $\|\tilde{\pi} - \pi\|_{\text{TV}} \leq 1 - \pi(\tilde{A}) \leq \epsilon/3$. By a straightforward coupling argument $\|P_x^r(\cdot) - \tilde{P}^r(x, \cdot)\|_{\text{TV}} \leq P_x[T_{\partial A} < r] \leq \epsilon/3$. Finally, by the triangle inequality we get that $\|P_x^r - \pi\|_{\text{TV}} \leq 3 \cdot \frac{\epsilon}{3} = \epsilon$. \square

6.2 A useful lemma for bounding the relaxation-time

The following lemma allows us to easily bound the relaxation-time of our examples.

Lemma 6.7. *Let $G_n = (V_n, E_n)$ be a family of c -lazy expanders.*

- (i) *Let H_n be a sequence of graphs obtained by stretching some of the edges in G_n by a factor of s_n . Then $t_{\text{rel}}(H_n) = O(s_n^2)$.*
- (ii) *Let $F_n = (W_n, U_n)$ be a sequence of graphs obtained by decorating some of the vertices of G_n with a 3 dimensional torus of side length k_n . Then $t_{\text{rel}}(H_n) = \Omega(k_n^3)$.*

Proof. Since G_n is an expander and H_n is obtained from it by stretching some of the edges by a factor of s_n , the Cheeger constant of H_n is at least of order s_n^{-1} (e.g. [13, Proposition 2.3]). Hence $t_{\text{rel}}(H_n) = O(s_n^2)$ by (6.2). Part (ii) is obtained from (6.2). \square

6.3 Proof of Fact 4.1

The proof of part (1 a) is easy and hence omitted. We now prove (1 b). The upper bounds in (4.1) are trivial. We now prove that $P_o[X_{\tau_n} \in D], P_o[X_{T_{\mathcal{L}_n}} \in D] \geq cP_o[D]$ (for some $0 < c < 1$, independent of D and n). This follows from the fact that for all $d \in D$, the probability that D was visited and that d is the last (resp. first) vertex in D to be visited is at least $cP_o[X_{\tau_n} = d]$ (resp. $cP_o[X_{T_{\mathcal{L}_n}} = d]$). We leave the details to the reader. The proof of (2 b) is analogous and hence omitted.

We now prove (2 a). Denote the left and right children of o by u and v , respectively. Let \mathcal{T}_v and \mathcal{T}_u be the subtrees rooted at v and u , resp.. We define the left and right trees rooted at o to be the trees obtained by deleting \mathcal{T}_v and \mathcal{T}_u , resp.. Write w, w_L and w_R for the conductance from the root to infinity in the original tree, the left tree and the right tree, respectively. Let A be the event that the last vertex of $\{v, u\}$ which was visited by the walk was u (i.e. the walk got absorbed in \mathcal{T}_u). Using a standard network reduction and the fact that \mathcal{T}_v and \mathcal{T}_u are identical to \mathcal{T} , we get the following relations: $w = w_L + w_R, \frac{1}{w_L} = \frac{1}{1+\epsilon} + \frac{1}{w}, \frac{1}{w_R} = 1 + \frac{1}{w}$ and $P_o[A] = \frac{w_L}{w_L + w_R}$. Solving this system of equations yields that $P_o[A] = \frac{\sqrt{1+\epsilon}}{1+\sqrt{1+\epsilon}}$. Using again the fact that \mathcal{T}_v and \mathcal{T}_u are identical to \mathcal{T} , allows us to repeat the argument and the claim now follows by induction. \square

6.4 Proof of Lemma 3.2

We first prove (3.1). We only prove the discrete-time lazy case, as the continuous-time case is analogous. By reversibility and the Markov property w.r.t. $\min(T_z, T_y)$

$$\begin{aligned} \frac{P_L^t(x, y)}{\pi(y)} &= \sum_{k_1=0}^t \frac{P_x[T_z = k_1 < T_y] P_L^{t-k_1}(z, y)}{\pi(y)} + \frac{P_x[T_y = k_1 < T_z] P_L^{t-k_1}(y, y)}{\pi(y)} \\ &= \sum_{k_1=0}^t \frac{P_x[T_z = k_1 < T_y] P_L^{t-k_1}(y, z)}{\pi(z)} + \sum_{k=0}^t \mathbb{P}[T_{z,y}^x = k, T_y^x \leq T_z^x] P_L^{t-k}(y, y) / \pi(y) \end{aligned} \tag{6.3}$$

Now $\sum_{k_1 \leq t} \frac{P_x[T_z = k_1 < T_y] P_L^{t-k_1}(y, z)}{\pi(z)} = \sum_{k_1, k_2: k_1 + k_2 \leq t} \frac{P_x[T_z = k_1 < T_y] P_y[T_z = k_2] P_L^{t-k_1-k_2}(z, z)}{\pi(z)}$, which equals $\sum_{k: k \leq t} \mathbb{P}[T_{z,y}^x = k, T_y^x > T_z^x] P_L^{t-k}(z, z) / \pi(z)$. Substituting this in (6.3) yields the equality in (3.1). The inequality follows from the fact that for all $a \in \Omega$ we have that $P_L^s(a, a) / \pi(a)$ is decreasing in s and tends to 1 as $s \rightarrow \infty$ which follows from the spectral decomposition and the non-negativity of the eigenvalues of $P_L = \frac{1}{2}(I + P)$. For (3.2) use (3.1) with $z = y$. We now prove (3.3). Here we only prove the continuous-time case.

Denote the densities of T_z under H_x and H_y by f_z^x and f_z^y , resp.. Conditioning on T_z (which is deterministically smaller than T_y under H_x in our current setup), then using

reversibility and finally, conditioning on T_z again (now under H_y), we get that

$$\begin{aligned} \frac{H_t(x, y)}{\pi(y)} &= \int_0^t f_z^x(s) \frac{H_{t-s}(z, y) ds}{\pi(y)} = \int_0^t f_z^x(s) \frac{H_{t-s}(y, z) ds}{\pi(z)} \\ &= \int_{s, r \geq 0: s+r \leq t} f_z^x(s) f_z^y(r) \frac{H_{t-(s+r)}(z, z) dr ds}{\pi(z)} = \int_0^t f_z^{x,y}(s) \frac{H_{t-s}(z, z) ds}{\pi(z)} \\ &= \mathbb{P}[\tau_z^{x,y} \leq t] + \int_0^t f_z^{x,y}(s) \frac{H_{t-s}(z, z) - \pi(z)}{\pi(z)} ds. \end{aligned}$$

Denote $M := \max_s f_z^{x,y}(s)$. Denote the minimal non-zero eigenvalue of $I - P$ by λ . Using the spectral decomposition $\frac{H_s(z, z) - \pi(z)}{\pi(z)} \leq e^{-\lambda s} \frac{H_0(z, z) - \pi(z)}{\pi(z)} = \frac{1 - \pi(z)}{\pi(z)} e^{-\lambda s}$, for all s , thus

$$\int_0^t f_z^{x,y}(s) \frac{H_{t-s}(z, z) - \pi(z) ds}{\pi(z)} \leq M \int_0^\infty \frac{H_s(z, z) - \pi(z) ds}{\pi(z)} \leq M \frac{1 - \pi(z)}{\pi(z)} t_{\text{rel}}.$$

Substituting this above gives (3.3). For (3.4) use (3.3) with $z = y$. □

6.5 Proof of (3.6)

Since $H_t = e^{t(P-I)} = e^{2t(P_L-I)}$, the continuous-time version of the chain is also the continuous-time version of the lazy chain, but run twice faster. Thus, in order to sample $\tau_z^{a,b}$, we may first sample $T_z^{a,b}$ and a Poisson process on \mathbb{R}_+ with rate 2, $(N_t)_{t \geq 0}$, and then set $\tau_z^{a,b} = \inf\{t : N_t \geq T_z^{a,b}\}$. Thus for all s, t we have

$$\mathbb{P}[\tau_z^{a,b} \leq t] \geq \mathbb{P}[N_t \geq s] \mathbb{P}[T_z^{a,b} \leq s]. \tag{6.4}$$

Using (6.4) (for the first inequality in (6.5)), we argue that there exists some sufficiently small absolute constant $c > 0$ such that for all $0 < \delta < 1/2$ we have for all sufficiently large n that

$$\mathbb{P}[\tau_z^{a,b} \leq \frac{1}{2}(t_\delta - cn\sqrt{\delta})] \geq \mathbb{P}[\text{Pois}(t_\delta - cn\sqrt{\delta}) \geq t_\delta + cn\sqrt{\delta}] \mathbb{P}[T_z^{a,b} \leq t_\delta + cn\sqrt{\delta}] \geq 2^{-\delta n}. \tag{6.5}$$

In other words, (when c is sufficiently small) there is an entropy gain in allowing $T_z^{a,b}$ to perform a smaller large deviation (requiring $T_z^{a,b} \leq t_\delta + cn\sqrt{\delta}$ instead of $T_z^{a,b} \leq t_\delta$) while forcing a (relatively) small large deviation in the number of steps performed by the continuous-time chain (requiring that by time $t_\delta - cn\sqrt{\delta}$ it makes at least $t_\delta + cn\sqrt{\delta}$ steps). This behavior is typical in the theory of large deviations. We now justify this claim.

The large deviation behavior of $T_z^{a,b}$ is studied in detail in [14, Example 3 and Lemma 5.1]. In particular, it is shown that it has the same rate function as that of a certain sum i.i.d. r.v.'s with exponential tails (see also § 3.3). Using a second order Taylor expansion for the rate function around the value corresponding to the mean (and the fact that the first derivative of the rate function vanishes at that point), it is also shown that for all $a \in (0, 1/2]$ we have that

$$c_1 a^2 \leq -\frac{1}{n} \log \mathbb{P}[T_z^{a,b} \leq 12n - an] \leq C_1 a^2. \tag{6.6}$$

Using convexity of the rate function (the Legendre transform of a convex function is itself convex) for all $a \in (0, 1/2]$ and $0 < c < a$ we have that

$$-\frac{1}{n} \log \mathbb{P}[T_z^{a,b} \leq 12n - (a+c)n] \geq -\frac{a+c-o(1)}{a} \frac{1}{n} \log \mathbb{P}[T_z^{a,b} \leq 12n - an],$$

and so by (6.6)

$$\frac{1}{n} \log \left(\frac{\mathbb{P}[T_z^{a,b} \leq 12n - an]}{\mathbb{P}[T_z^{a,b} \leq 12n - (a+c)n]} \right) \geq -\frac{c - o(1)}{a} \frac{1}{n} \log \mathbb{P}[T_z^{a,b} \leq 12n - an] \geq c_1 ac - o(1). \quad (6.7)$$

Using (6.7), it is not hard to see that indeed

$$c_0 \sqrt{\delta} n \leq 12n - t_\delta \leq C_0 \sqrt{\delta} n. \quad (6.8)$$

Using the fact that $\mathbb{P}[\text{Pois}(\mu) > \mu(1 + \epsilon)] \geq \exp(-C_2 \epsilon^2 \mu)$ for some absolute constant C_2 for all $\mu > 0$ and $0 < \epsilon < 1$ yields that for some $\epsilon' > 0$, for all $0 < \delta \leq 1/2$ and $0 < c < \epsilon'$ we have

$$\mathbb{P}[\text{Pois}(t_\delta - cn\sqrt{\delta}) \geq t_\delta + cn\sqrt{\delta}] \geq e^{-C_2 c^2 \delta n}.$$

By (6.7)-(6.8) and the definition of t_δ we have that for such c and δ

$$\mathbb{P}[T_z^{a,b} \leq t_\delta + cn\sqrt{\delta}] \geq \mathbb{P}[T_z^{a,b} \leq t_\delta] e^{c_2 c \delta n} \geq 2^{-\delta n} e^{c_2 c \delta n}.$$

The last two inequalities imply (6.5) with $c = c_2/(2C_2)$. □

Acknowledgments. The authors would like to thank the anonymous referee for numerous suggestion and comments which greatly improved the presentation of the paper.