

Total variation and separation cutoffs are not equivalent and neither one implies the other*

Jonathan Hermon[†] Hubert Lacoin[‡] Yuval Peres[§]

Abstract

The cutoff phenomenon describes the case when an abrupt transition occurs in the convergence of a Markov chain to its equilibrium measure. There are various metrics which can be used to measure the distance to equilibrium, each of which corresponding to a different notion of cutoff. The most commonly used are the total-variation and the separation distances. In this note we prove that the cutoff for these two distances are not equivalent by constructing several counterexamples which display cutoff in total-variation but not in separation and with the opposite behavior, including lazy simple random walk on a sequence of uniformly bounded degree expander graphs. These examples give a negative answer to a question of Ding, Lubetzky and Peres

Keywords: Markov chains; mixing time; cutoff; counter example.

AMS MSC 2010: 60J10.

Submitted to EJP on December 18, 2015, final version accepted on May 23, 2016.

Supersedes arXiv:1508.03913.

1 Introduction

Consider an irreducible discrete-time Markov chains $X = (X_t)_{t \geq 0}$, defined on a finite state space Ω (we call a chain finite if Ω is finite). We let P denote its transition matrix. We further assume that X is reversible, that is that there exists a probability measure π which satisfies the detailed balanced equation

$$\forall x, y \in \Omega, \quad \pi(x)P(x, y) = \pi(y)P(y, x).$$

This measure is unique because of irreducibility. Let us assume furthermore that our Markov chain is lazy, meaning that

$$\forall x \in \Omega, \quad P(x, x) \geq 1/2. \tag{1.1}$$

*H.L. acknowledges the support of a productivity grant from the CNPq.

[†]Department of Statistics, UC Berkeley, California, USA. E-mail: jonathan.hermon@berkeley.edu

[‡]Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, Brasil 22460-320. E-mail: lacoin@impa.br

[§]Microsoft Research, Redmond, Washington, USA. E-mail: peres@microsoft.com

A particular important case of such a Markov chain is **lazy simple random walk** (SRW) on a simple graph $G = (V, E)$, in which case $\Omega = V$, $P(x, y) = \frac{1_{\{x=y\}}}{2} + \frac{1_{\{(x,y) \in E\}}}{2 \deg(x)}$ and $\pi(x) = \frac{\deg(x)}{2|E|}$, where $\deg(x) := |\{y : \{x, y\} \in E\}|$ and $|\cdot|$ denotes the cardinality of a set.

It is a classic result of probability theory that for any initial condition the distribution of $X(t)$ converges 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).

We denote by P_x^t (P_x) the distribution of X_t (resp. $(X_t)_{t \geq 0}$), given that $X_0 = x$. For any two distributions μ, ν on Ω , their **total-variation distance** is defined to be

$$\|\mu - \nu\|_{\text{TV}} := \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \sum_{\{x : \mu(x) > \nu(x)\}} \mu(x) - \nu(x) = 1 - \sum_{x \in \Omega} \min(\mu(x), \nu(x)). \quad (1.2)$$

The worst-case total-variation distance at time t is defined as

$$d(t) := \max_{x \in \Omega} d_x(t), \text{ where } d_x(t) := \|P_x(X_t \in \cdot) - \pi\|_{\text{TV}}. \quad (1.3)$$

The (total-variation) ε -**mixing-time** is defined as

$$t_{\text{mix}}(\varepsilon) := \inf \{t : d(t) \leq \varepsilon\}.$$

Similarly, the (worst-case) **separation distance** from stationarity at time t is defined as

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

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

$$t_{\text{sep}}(\varepsilon) := \inf \{t : d_{\text{sep}}(t) \leq \varepsilon\}.$$

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

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:

$$\forall \varepsilon \in (0, 1/2], \quad \lim_{n \rightarrow \infty} t_{\text{mix}}^{(n)}(\varepsilon) / t_{\text{mix}}^{(n)}(1 - \varepsilon) = 1 \quad \left(\text{resp. } \lim_{n \rightarrow \infty} t_{\text{sep}}^{(n)}(\varepsilon) / t_{\text{sep}}^{(n)}(1 - \varepsilon) = 1 \right). \quad (1.4)$$

It is a classical result (e.g. [16, Lemmas 6.13 and 19.3] or (6.9)) that under reversibility the separation and total-variation distances and mixing times can be compared as follows (the second line being an easy consequence of the first)

$$\begin{aligned} \forall t \geq 0, \quad d(t) \leq d_{\text{sep}}(t) \leq 1 - (1 - \min(2d(t/2), 1))^2 \leq 4d(t/2), \\ \forall a \in (0, 1), \quad t_{\text{mix}}(a) \leq t_{\text{sep}}(a) \leq 2t_{\text{mix}}(a/4). \end{aligned} \quad (1.5)$$

Another important family of distances is the family of ℓ_p distances ($1 \leq p \leq \infty$):

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

where $a_{\mu,\nu,\pi}(x) := |\mu(x) - \nu(x)|/\pi(x)$ (observe that $\|\mu - \nu\|_{1,\pi} = 2\|\mu - \nu\|_{\text{TV}}$). Note that the notion of distance to equilibrium and mixing time can be transposed to these distances by replacing $\|\cdot\|_{\text{TV}}$ by $\|\cdot\|_{p,\pi}$ in (1.3). For $a \in (0, \infty)$ we denote the a -th ℓ_p -mixing time by $t_{\ell_p}(a)$. Under reversibility, the ℓ_p distances can be compared as follows (see [6, Proposition 5.1])

$$\begin{aligned} t_{\ell_2}(a) &\leq t_{\ell_p}(a) \leq 2t_{\ell_2}(\sqrt{a}) && \text{for } p \in (2, \infty], \\ \frac{1}{m_p} t_{\ell_2}(a^{m_p}) &\leq t_{\ell_p}(a) \leq t_{\ell_2}(a) && \text{for } p \in (1, 2), \end{aligned} \tag{1.6}$$

where $m_p := \lceil p/(2(p-1)) \rceil$. Hence in some sense, up to a multiplicative constant, the different ℓ_p mixing times ($p \in (1, \infty)$) are equivalent. It turns out that under reversibility the notion of cutoff for these distances are also equivalent.

Theorem A (Chen and Saloff-Coste [6]). Let (Ω_n, P_n, π_n) be a sequence of reversible lazy Markov chains. Let $\lambda_2^{(n)}$ be the second largest eigenvalue of P_n . Then the following assertions are equivalent

- The sequence exhibits ℓ_p -cutoff for some $1 < p \leq \infty$.
- The sequence exhibits ℓ_p -cutoff for all $1 < p \leq \infty$.
- $\lim_{n \rightarrow \infty} (1 - \lambda_2^{(n)}) t_{\ell_2}^{(n)}(1/2) = \infty$.

Observe that under reversibility (for any fixed chain) (1.5) expresses an equivalence between the separation and the total-variation mixing times, parallel to the one, expressed in (1.6), holding between the different ℓ_p mixing times for $p \in (1, \infty]$. Hence a natural question (in light of Theorem A) is whether (under reversibility) there is cutoff in total-variation if and only if there is cutoff in separation. This is Question 5.1 in [10], where an affirmative answer was given for the class of birth and death chains (which are Markov chains for which the set of edges (x, y) with $P(x, y) > 0$ forms a segment). In fact, both cutoffs were shown to be equivalent to the *product condition* (3.2).

Theorem B (Ding, Lubetzky and Peres [10], Diaconis and Saloff-Coste [9]). A sequence of birth and death chains exhibits total variation cutoff iff it exhibits separation cutoff.

In this note we give a negative answer to that question in general by constructing counter-examples.

Theorem 1.1. (i) *Total-variation and separation cutoff are not equivalent for lazy reversible Markov chains and neither one implies the other.*

(ii) *The above statement remains true within the class of lazy simple random walks on graphs of maximal degree at most 7.*

Remark 1.2. We can also produce non-reversible or non-lazy counter-examples by performing artificial modifications in our chains, but this is not a very important point. Non-lazy or non-reversible chains can have very pathological behavior and we want to underline that we are not using “unfair tricks” to produce our counter-examples.

Of course a full proof of this statement only requires two counter-examples as (ii) is a stronger statement than (i). However, we have chosen to include also examples that are not simple random-walks because they are much simpler. We present a total of five counter-examples. Apart from the first one, they are all lazy (weighted nearest-neighbor) random walks on bounded degree graphs, with transition rates which are bounded away from zero. The last two examples, which are a bit more technical to analyze, are lazy SRWs on a sequence of bounded degree graphs $G_n := (V_n, E_n)$ (i.e. $\sup_n \max_{v \in V_n} \deg(v) < \infty$).

Note that for all our counter-examples the graph supporting the transitions contains some cycles. An interesting open problem is to determine whether Theorem B can

extended to the case of lazy weighted nearest-neighbor random walk on trees for which it is already known (cf. [5]) that separation cutoff implies total-variation cutoff.

A sequence of Markov chains is said to display **pre-cutoff** (in total-variation resp. separation) if

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

We call the value of the sup above the **pre-cutoff ratio**. Equation (1.5) implies that

$$\sup_{\varepsilon \in (0, 1/2]} \limsup_{n \rightarrow \infty} t_{\text{sep}}^{(n)}(\varepsilon)/t_{\text{sep}}^{(n)}(1 - \varepsilon) \leq 2 \sup_{\varepsilon \in (0, 1/2]} \limsup_{n \rightarrow \infty} t_{\text{mix}}^{(n)}(\varepsilon)/t_{\text{mix}}^{(n)}(1 - \varepsilon). \quad (1.7)$$

A symmetrized version of this inequality also holds provided that $t_{\text{mix}}^{(n)}$ goes to infinity (the assumption being present just to avoid pathological behavior)

$$\sup_{\varepsilon \in (0, 1/2]} \limsup_{n \rightarrow \infty} t_{\text{mix}}^{(n)}(\varepsilon)/t_{\text{mix}}^{(n)}(1 - \varepsilon) \leq 2 \sup_{\varepsilon \in (0, 1/2]} \limsup_{n \rightarrow \infty} t_{\text{sep}}^{(n)}(\varepsilon)/t_{\text{sep}}^{(n)}(1 - \varepsilon). \quad (1.8)$$

The proof of (1.8) involves more computation than (1.7). We present a complete proof of it in Appendix A.2)

These two inequalities imply that the notion of pre-cutoff is equivalent for the two distances and the pre-cutoff ratio of one is at most twice that of the other. In particular, cutoff in one distance implies pre-cutoff with ratio at most 2 in the other. With our examples, we shall show that this is in fact sharp in some cases:

Remark 1.3. There exists a sequence of lazy reversible Markov chains for which we have cutoff in total-variation and only pre-cutoff with ratio 2 in separation and vice-versa.

Our last point of comparison between total-variation mixing and separation mixing is related to the width of the **cutoff window**. We say that a sequence of chains exhibits total-variation (resp. separation) cutoff with a cutoff window w_n if $w_n = o(t_{\text{mix}}^{(n)})$ and for all $0 < \varepsilon \leq 1/4$ there exists some constant $C_\varepsilon > 0$ (depending only on ε) such that

$$\forall n, \quad t_{\text{mix}}^{(n)}(\varepsilon) - t_{\text{mix}}^{(n)}(1 - \varepsilon) \leq C_\varepsilon w_n \quad (\text{resp. } t_{\text{sep}}^{(n)}(\varepsilon) - t_{\text{sep}}^{(n)}(1 - \varepsilon) \leq C_\varepsilon w_n).$$

Note that the window defined in this manner is not unique, but informally “the” cutoff window is given by the “smallest such w_n ”. Our examples demonstrate that the cutoff windows for total-variation and separation do not have the same behavior.

The following result is due to Chen and Saloff-Coste [7, Theorem 3.4]. We present a much simpler proof in the Appendix.

Theorem C. Let (Ω_n, P_n, π_n) be a sequence of lazy irreducible finite chains which exhibits total-variation cutoff with a cutoff window w_n . Then $w_n = \Omega(\sqrt{t_{\text{mix}}^{(n)}})$.

The bound given by Theorem C is obviously sharp for the biased random walk on a segment. Conversely, some very standard Markov chains like the lazy SRW on the n -dimensional hyper-cube have a cutoff window $w_n \gg \sqrt{t_{\text{mix}}^{(n)}}$ (here $w_n = n$ and $t_{\text{mix}}^{(n)} = (\frac{1}{2} \pm o(1))n \log n$). As indicated in Remark 1.6 the laziness assumption in Theorem C can be replaced by the assumption that $\inf_n \min_{x \in \Omega_n} P_n^2(x, x) > 0$ (as is the case for simple random walk on a sequence of bounded degree graphs).

In light of Theorem C one might expect that whenever separation cutoff occurs for a sequence of discrete-time lazy chains, the width of the separation cutoff window is $\Omega(\sqrt{t_{\text{sep}}^{(n)}})$. We are unaware of any previously analyzed example in which this fails. We find it remarkable that as the following remark asserts, the width of the separation cutoff window for a sequence of discrete-time lazy SRWs on a sequence of bounded degree graphs, can in fact be a constant! This, or more precisely, the mechanism that allows such behavior (see § 2.4 for more on this point) demonstrates that the separation distance can exhibit profoundly different behaviors than the total variation distance.

Our counter-examples show that the cutoff window in one distance can be as small as allowed even if there is no cutoff for the other distance:

Remark 1.4. We will construct sequences of bounded degree graphs such that the corresponding sequences of lazy SRWs exhibit the following behaviors (resp.)

- (i) There is no separation cutoff but there is total-variation cutoff with window $\sqrt{t_{\text{mix}}^{(n)}}$.
- (ii) There is no total-variation cutoff but there is separation cutoff with window 1.

In § 2.4 we refine the statement of (ii) and describe further surprising properties of the relevant example for (ii) above (listed in § 2.4 as properties (i)-(v)).

Remark 1.5. Let $\delta_n \in (0, 1)$. We call a sequence of discrete time chains (Ω_n, P_n, π_n) , δ_n -lazy if for all n , $P_n(x, x) \geq \delta_n$ for all $x \in \Omega_n$. It is not hard to extend the proof of Theorem C and show that if a sequence of δ_n -lazy chains exhibits total-variation cutoff with a window w_n , then $w_n = \Omega(\sqrt{\delta_n(1 - \delta_n)t_{\text{mix}}^{(n)}})$.

Theorem C can also be extended to the continuous time setup, with the additional assumption that the sum of the transition rates from any given state is bounded above by 1 (or by some absolute constant).

Remark 1.6. Let $G_n = (V_n, E_n)$ be a sequence of connected non-bipartite simple graphs of maximal degree d_n . Consider the sequence of (non-lazy) SRWs on G_n . Then $P_n^2(v, v) \geq 1/d_n$, for every $v \in V_n$. By considering P^2 rather than P it follows from the previous remark that if the sequence exhibits total-variation cutoff with a window w_n , then $w_n = \Omega(\sqrt{t_{\text{mix}}^{(n)}/d_n})$. This is in fact sharp by considering a sequence of random d_n -regular graphs of size n for some d_n such that $\lim_{n \rightarrow \infty} d_n = \infty$ and $d_n = o(\frac{\log n}{\log \log n})$ [17, Theorem 3].

1.1 Organization of the note

In § 2 we describe the construction of our examples and our general strategy. We also describe relevant examples due to Aldous and Pak.

In § 3 we introduce a general framework, which under a certain condition, allows to reduce the study of the mixing-time to the study of the hitting time of a special point.

In § 4 we describe two examples of sequences of Markov chains which exhibit total-variation cutoff but do not exhibit separation cutoff. The first example, Example 4.1, demonstrates that (1.7) may be sharp (even when the r.h.s. of (1.7) equals 1). The second example, Example 4.2, is a weighted nearest neighbor random walk on a bounded degree graph with transition probabilities which are bounded away from 0 and 1.

In § 5 we construct an example of a sequence of Markov chains that exhibits separation cutoff but no total-variation cutoff (Example 5.1).

Finally, in § 6 we transform Examples 4.2 and 5.1 into examples of sequences of lazy SRWs on bounded degree Expander graphs. The reason we first describe Examples 4.2 and 5.1 is that the key ideas of our constructions are more transparent in these examples.

2 An overview of the main ideas of our constructions

2.1 A very basic chain with different cutoff times for separation and total variation

In this section we settle with a high-level description of some key ideas. Let us first present a very simple Markov chain which exhibits cutoff in both distances (see Figure 1) but for which the mixing-time in separation is twice as large as that in total variation.

Consider a random walk on a segment a, b of length $2n$ which presents a constant bias towards the middle point which we call z (see Figure 1). Most of the equilibrium measure is concentrated on a small neighborhood of z and for this reason (cf. Proposition 3.3) the total-variation mixing-time corresponds to the time which is needed to hit z (starting from either of the end-points). The system displays cutoff because this hitting time is concentrated around its mean.

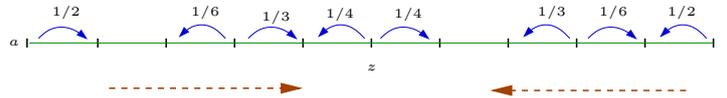


Figure 1: A very simple chain for which the separation mixing-time is twice as large as the total-variation mixing-time ($6n$ and $12n$, respectively). The transition rates (apart from at the special states a, b and z) are $1/3$ in the z direction and $1/6$ in the opposite one (the holding probability is $1/2$), making the chain travel at speed $1/6$ towards z .

The separation mixing-time on the other hand is twice as large. Roughly speaking, this is because for $P^t(a, b)$ to come close to its equilibrium value, “information” has to pass from one end to the other. The time required for this to occur corresponds more or less to the sum of the times needed to reach z from a and b , respectively (see Proposition 3.8).

This scheme with two extremal opposite initial conditions, though not ubiquitous among Markov chains, appears in many natural examples for which cutoff has been proved: e.g. the lazy SRW on the hyper-cube (see [16, Theorem 18.3]), the Ising model at high temperature [19] or the adjacent-transposition shuffle on the segment [15].

2.2 An idea to avoid cutoff in separation while keeping that in total-variation

Our idea to produce counter-examples with total-variation cutoff but only pre-cutoff in separation is to modify the structure (state space and transition rates) of the simple chain above (Figure 1), only on one side (say, the side of b), to break the symmetry. To be precise, in Example 4.2 we first set the holding probabilities on both sides to be $3/4$ (and consider the obtained chain as the “original chain”, as opposed to Example 4.1, for which the chain in Figure 1 serves as the “original chain”) before modifying the b -side. We want to perform our modifications in the following manner:

- We want to keep the property that every path from a to b goes through z , which shall still bear a positive proportion of the equilibrium mass.
- We want a to remain the initial condition from which it takes the longest time to reach equilibrium (equivalently, to hit z). More precisely, we want that also after the modification, the distribution of the hitting time of z , $T_z := \inf\{t : X_t = z\}$, starting from a would still stochastically dominate the distribution of T_z , starting from any other initial state. Moreover, we want the hitting time distribution of z , starting from any state between a and z (including a), to remain un-changed.
- We want the hitting time of z from initial state b , to become non-concentrated, and to remain of the same order of magnitude as the mixing-time of the whole chain. Moreover, we want this hitting time to remain (stochastically) larger than the hitting time of z , starting from any other state which lies between b and z , and to become stochastically dominated by the hitting time distribution of z (from b) in the original chain (which equals the hitting time distribution from a in the modified chain).

In this manner, the hitting time distribution of z under P_a remains un-changed (and in particular, remains concentrated). Moreover, after the modification it is still the case that $d(t) \approx P_a[T_z > t]$, and thus by the aforementioned concentration there is still cutoff in total-variation (see Proposition 3.3). Using Proposition 3.8, we deduce that $d_{\text{sep}}(t_{\text{mix}} + t) \approx P_b[T_z > t]$ and so there is no cutoff in separation as the hitting time distribution of z under P_b in the modified chain is no longer concentrated.

To perform such a modification, we borrow ideas from previous constructions of Pak (for Example 4.1) and Aldous (for Example 4.2), which we present now.

2.3 Related constructions

When the product condition (Definition 3.1) was shown to be a necessary condition for cutoff, it was conjectured that it should also be a sufficient one for “nice” chains. However, two counter-examples constructed, respectively by Aldous and Pak (see [5, Example 8.1], [6] and [16, Chapter 18] for a more detailed description and analysis), show that in general the product condition does not imply cutoff. The mechanisms used to prevent cutoff in those two constructions are of different nature.

- Aldous’ example (Figure 2) locally looks like a biased random walk on a segment, so that most of the equilibrium measure is concentrated on a small neighborhood of the end-point towards which the walk is biased (we call this end of the segment z and the opposite one b). To avoid cutoff, the half of the segment closer to z is split into two distinct parallel branches. The transition rates on these branches are tuned so that there is still a bias towards z but such that one path is slower than the other. Starting furthest away from equilibrium (i.e. at state b) we have two possible scenarios to reach z given by the two distinct branches and the probability of each is bounded away from 0 and 1. As the speed along the two branches is different, the CDF of the hitting time distribution of z starting from b has two abrupt jumps. Consequently, $d^{(n)}(t)$ exhibits two distinct abrupt drops and there is no cutoff.
- Pak’s idea is to start with a sequence of chains which exhibits cutoff and to modify it by adding transitions which are such that with a constant rate (which is chosen to be somewhere between the spectral gap and the inverse of the mixing-time of the original chain, say their geometric mean) the system is brought to equilibrium at once. For the modified Markov chain, the total-variation distance decays (up to a negligible error) exponentially with the rate of the newly added transitions and hence cutoff does not occur, neither pre-cutoff.

In our Example 4.1 (see Figure 3), we adapt Pak’s idea: on the b -side (of the chain from Figure 1) we add transitions from states on the b -side to the center of mass z , and we choose the inverse of the rate to be of the same order as the mixing-time (which is of order of the length of the segment: n). This makes the hitting time of z started from b non-concentrated and (stochastically) smaller than started from a . Moreover, after this modification, all of the properties described in the beginning of § 2.2 are satisfied.

In our Example 4.2, (see Figure 4), we simply replace the b -side by Aldous’ construction, and set the holding probability on the a -side to be $3/4$ (which is the holding probability of the slow branch of the b -side). After this modification, all of the properties described in the beginning of § 2.2 are satisfied.

2.4 An idea to keep cutoff in separation while avoiding that in total-variation

For this part we must rely on a different idea. What we want to alter in our chain is the way the separation distance shrinks to zero. Loosely speaking, in the original chain on the segment, the separation mixing-time is determined by the sum of the hitting

Total variation and separation cutoffs are not equivalent

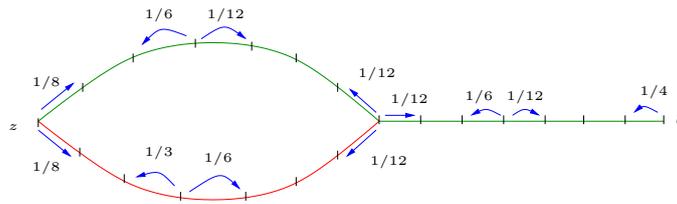


Figure 2: A version of Aldous' example. The walk is always biased towards z but the speed of the walk depends on the branch. On the top branch, as well as on the rest of the segment, the transition rates are $1/6$ in the z direction and $1/12$ in the opposite one (the holding probability is $3/4$) whereas on the bottom branch the (exit) rates are twice as large (and the holding probability is $1/2$), resulting in a larger speed. As a result, two transitions occur for the total-variation distance at times $9n$ and $12n$ respectively, where n denotes the total distance from z to b and the length of each of the two parallel branches is $\lceil n/2 \rceil$ (above $n = 14$). The rates at b , z and at the branching point are not very relevant but we display them for the sake of concreteness.

times of z from a and b since z is the only channel of communication between the two extremities.

Our construction (Example 5.1) relies on the following idea (see Figure 5). We take the length of the line segment to be $2(M + 1)n$ for some large (fixed) integer M .

- We connect the two sides of the segment at a second point z' which is far from the center of mass z . We do so by merging the two states which are of distance n from z (one on the a -side and one on the b -side) into a single state z' . This connection maintains the cutoff in separation. However, it has the effect of shortening the separation cutoff time by some constant factor, while, as we now describe, drastically altering the nature of the abrupt transition of $d_{\text{sep}}^{(n)}(t)$ around the (separation) cutoff time. It follows from our analysis of Example 5.1 and the refined analysis of Example 6.12 in § 6.5, that provided that M is taken to be sufficiently large:

- (i) Also after creating the connection at z' we have that

$$\lim_{n \rightarrow \infty} \sup_t |d_{\text{sep}}^{(n)}(t) - \max(0, 1 - P_n^t(a, b)/\pi_n(b))| = 0.$$

- (ii) Due to the connection of A and B at point z' , up to negligible terms, around the separation cutoff time, $P_n^t(a, b)$ is supported by trajectories which never get much closer to z than z' is, and so are contained in a set whose stationary probability is exponentially small in n .
- (iii) Let $T_{z'}^{a,b}$ (Definition 3.4) be a random variable distributed as a convolution of the hitting time distribution of z' started from a with that started from b (in this case the two distributions are identical). Around the (separation) cutoff time, $P_n^t(a, b)/\pi_n(b)$ can be understood in terms of the behavior of $T_{z'}^{a,b}$ in the large deviation regime (namely, the cutoff occurs around the time t for which $\mathbb{P}[T_{z'}^{a,b} \geq t] \approx \pi_n(z') = \Theta(2^{-n})$).
- (iv) Around $t_{\text{sep}}^{(n)}$, $P_n^t(a, b)/\pi_n(b)$ grows exponentially in $t - t_{\text{sep}}^{(n)}$ for $t \geq t_{\text{sep}}^{(n)}$ (and decays exponentially for $t < t_{\text{sep}}^{(n)}$) and continues to do so for $\Theta(n)$ steps around $t_{\text{sep}}^{(n)}$ (in particular, shortly after $t_{\text{sep}}^{(n)}$, (a, b) no longer minimizes $P_n^t(x, y)/\pi_n(y)$). By (i), it follows that $w_n = 1$ is a (separation) cutoff window (and we can take $C_\varepsilon = C|\log \varepsilon|$, for some absolute constant C , for all $\varepsilon \in (0, 1/4]$).
- (v) $\sup_t P_n^t(a, b)/\pi_n(b) = \Theta(\max_t \mathbb{P}[T_{z'}^{a,b} = t]/\pi_n(z')) = \Theta(2^n/n) \rightarrow \infty$ as $n \rightarrow \infty$.

This behavior (namely, on the one hand having property (i) and on the other having properties (ii), (iv) and (v)) is atypical and quite surprising at first sight.

We are not done yet, as after creating the connection at z' , there are two symmetric parallel distinct branches from z' to the center of mass z , resulting in the hitting time of z from either a or b being concentrated. Consequently, there is still cutoff in total-variation (as by Proposition 3.3, $d(t) \approx P_a[T_z > t] = P_b[T_z > t]$).

- We break the symmetry (between the two branches, but not between a and b) in order to “destroy” the cutoff in total-variation by making the speed along the two paths which link z' to z different as in Aldous’ example (Figure 2). Observe that as opposed to Examples 4.1-4.2, here a and b play symmetric roles (the chain looks the same starting from either one of them).

As one should expect from property (ii) above (provided that M is sufficiently large), breaking the symmetry as described above does not influence the asymptotic pattern of convergence in separation, and (i)-(v) above remain valid. However the quantitative analysis of this example turns out to be more intricate than that of the first two.

2.5 Constructing counter-examples which are lazy SRW on bounded degree graphs

It was observed by Peres and Wilson that the sequence of chains in Aldous’ example could be modified into a sequence of lazy SRWs on bounded degree expander graphs (see Definition 3.6). In [18] Lubetzky and Sly constructed explicit 3-regular expanders with total-variation cutoff.

We use similar ideas to transform our Examples 4.2-5.1 into SRWs on bounded degree graphs (Examples 6.9-6.12). Our constructions includes one new idea: by introducing a sufficient amount of symmetry, (roughly speaking) we are able to reduce the analysis of Examples 6.9-6.12 to that of Examples 4.2-5.1. Consequently, the analysis of the asymptotic convergence profile of $d_n(t)$ is simpler than in [18] (at the cost of having maximal degree ≤ 7 rather than 3).

3 Preliminaries

The aim of this section is to introduce some general theory which shall reduce the analysis of our Examples 4.1-5.1 to the analysis of hitting time distributions of a specific state. The results appearing in this section are later generalized in § 6.1 (these generalizations reduce the analysis of Examples 6.9-6.12 to the analysis of hitting time distributions of a specific set). All proofs are deferred to the appendix. As we shall only prove the more general versions, we now describe the correspondence between the results of this section to the ones from § 6.1: Proposition 6.4 corresponds to Proposition 3.3, Lemma 6.3 to Lemma 3.5 and Proposition 6.5 to Proposition 3.8.

Let us first introduce some notation and standard terminology. Recall that if (Ω, P, π) is a finite irreducible reversible Markov chain, then P is self-adjoint w.r.t. the inner product induced by π on \mathbb{R}^Ω

$$\langle f, g \rangle_\pi := \sum_{x \in \Omega} \pi(x) g(x) f(x). \tag{3.1}$$

Hence it has $|\Omega|$ real eigenvalues satisfying $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} \geq -1$ (where $\lambda_2 < 1$ since the chain is irreducible and if the chain is lazy then $\lambda_{|\Omega|} \geq 0$). Define its **relaxation-time** as $t_{\text{rel}} := (1 - \max(\lambda_2, |\lambda_{|\Omega|}|))^{-1}$. Note that under laziness $t_{\text{rel}} = (1 - \lambda_2)^{-1}$.

Definition 3.1. We say that a family of reversible Markov chains satisfies the **product condition** if

$$\lim_{n \rightarrow \infty} (1 - \max(\lambda_2^{(n)}, |\lambda_{|\Omega|}^{(n)}|)) t_{\text{mix}}^{(n)} = \infty \quad (\text{equivalently, } t_{\text{rel}}^{(n)} = o(t_{\text{mix}}^{(n)})). \quad (3.2)$$

Because of the following well-known fact (e.g. [16, Proposition 18.4]), all our counter-examples satisfy the product condition.

Fact 3.2. For a sequence of irreducible aperiodic reversible Markov chains with relaxation-times $\{t_{\text{rel}}^{(n)}\}$ and mixing-times $\{t_{\text{mix}}^{(n)}\}$, if the sequence exhibits a pre-cutoff (either in total-variation or separation) and $\lim_{n \rightarrow \infty} t_{\text{mix}}^{(n)} = \infty$, then $t_{\text{rel}}^{(n)} = o(t_{\text{mix}}^{(n)})$.

Given $z \in \Omega$ we let

$$T_z := \inf\{t : X_t = z\}$$

denote the hitting time of z . The following result allows us to characterize the mixing-time of the chain in terms of the hitting time of a given point which carries a positive proportion of the mass. As hitting times are sometimes easier to control than mixing-times, it will assist us in determining the total-variation profile of convergence to equilibrium in Examples 4.1-5.1.

Proposition 3.3. Let (Ω_n, P_n, π_n) be a sequence of lazy reversible irreducible finite Markov chains which satisfies the product condition. Let us furthermore assume that there exists $z_n \in \Omega_n$ such that

$$\inf_n \pi_n(z_n) > 0. \quad (3.3)$$

Then setting

$$\tau_n(p) := \inf \left\{ t : \max_{x \in \Omega_n} P_x[T_{z_n} > t] \leq p \right\}, \quad (3.4)$$

we have for any $\varepsilon < \varepsilon' \in (0, 1)$

$$\limsup_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon')}{\tau_n(\varepsilon)} \leq 1 \quad \text{and} \quad \liminf_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{\tau_n(\varepsilon')} \geq 1. \quad (3.5)$$

Note that in particular the result shows that total-variation cutoff occurs if and only if $\tau_n(\cdot)$ displays the following abrupt transition

$$\forall \varepsilon \in (0, 1/2], \quad \lim_{n \rightarrow \infty} \frac{\tau_n(1 - \varepsilon)}{\tau_n(\varepsilon)} = 1. \quad (3.6)$$

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

Definition 3.4. Given x, y and z in Ω . We let $T_z^{x,y}$ denote a random variable obtained by taking the sum of two independent realizations of T_z , once under P_x and once under P_y . That is, $\mathbb{P}[T_z^{x,y} = t] := \sum_{k=0}^t P_x[T_z = k] P_y[T_z = t - k]$.

Lemma 3.5. Let (Ω, P, π) be a finite irreducible lazy reversible Markov chain. Consider $x, y, z \in \Omega$.

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

$$P^t(x, y) / \pi(y) \geq \sum_{k \leq t} \mathbb{P}[T_z^{x,y} = k] P^{t-k}(z, z) / \pi(z) \geq \mathbb{P}[T_z^{x,y} \leq t]. \quad (3.7)$$

In particular,

$$P^t(x, y) / \pi(y) \geq P_x[T_y \leq t] \quad (3.8)$$

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

$$\begin{aligned} P^t(x, y)/\pi(y) &= \sum_{k \leq t} \mathbb{P}[T_z^{x,y} = k] P^{t-k}(z, z)/\pi(z) \\ &\leq \mathbb{P}[T_z^{x,y} \leq t] + \frac{1}{2} t_{\text{rel}} \max_{k \in \mathbb{N}} \mathbb{P}[T_z^{x,y} = k] \sqrt{(1 - \pi(z))/\pi(z)}. \end{aligned} \tag{3.9}$$

All our examples would be of sequences of chains whose spectral gaps are uniformly bounded away from zero, that is, ones satisfying

$$\inf_n (1 - \lambda_2^{(n)}) > 0. \tag{*}$$

Although this is not necessary, working with such chains substantially simplifies the analysis of our examples. To check this condition, we use the notion of the Cheeger constant and the well-known discrete analog of Cheeger's inequality (3.10) [3, 4, 21] (the proof can also be found at [16, Theorem 13.14]).

Definition 3.6. For any (non-empty) set $A \subsetneq \Omega$ we define

$$Q(A) := \sum_{x \in A, y \notin A} \pi(x)P(x, y) \quad \text{and} \quad \Phi(A) := Q(A)/\pi(A).$$

We define the **Cheeger constant** of the chain to be

$$\Phi := \min_{A: 0 < \pi(A) \leq 1/2} \Phi(A).$$

We call a sequence of chains (Ω_n, P_n, π_n) an **expander family** if $\inf_n \Phi_n > 0$.

The following result implies that a sequence of reversible chains satisfies (*) if and only if it is an expander family.

Theorem 3.7. Let λ_2 be the second largest eigenvalue of a reversible transition matrix on a finite state space. Let Φ be as in Definition 3.6. Then

$$\Phi^2/2 \leq 1 - \lambda_2 \leq 2\Phi. \tag{3.10}$$

It is rather straightforward to check in all of our examples that the Cheeger constant is bounded away from zero.

Proposition 3.8. Let (Ω_n, P_n, π_n) be a sequence of lazy reversible irreducible finite Markov chains which satisfies (*). Let us furthermore assume that there exist $z_n \in \Omega_n$, sets $A_n, B_n \subset \Omega_n$, with $A_n \cup B_n = \Omega_n \setminus \{z_n\}$ and $a_n \in A_n, b_n \in B_n$, such that

- (i) $\inf_n \pi_n(z_n) > 0$.
- (ii) For any $x \in A_n$ and $y \in B_n$, $\mathbb{P}_x[T_{z_n} < T_y] = 1$.
- (iii) For all t

$$\max_{x \in A_n \cup B_n} \mathbb{P}_x[T_{z_n} > t] = \mathbb{P}_{a_n}[T_{z_n} > t] \quad \text{and} \quad \max_{y \in B_n} \mathbb{P}_y[T_{z_n} > t] = \mathbb{P}_{b_n}[T_{z_n} > t].$$

(iv)

$$\liminf_{n \rightarrow \infty} \inf_{t \geq 0} \min_{x, y \in A_n} \left(\frac{P_n^t(x, y)}{\pi_n(y)} - \frac{P_n^t(a_n, b_n)}{\pi_n(b_n)} \right) \geq 0. \tag{3.11}$$

(v)

$$\lim_{n \rightarrow \infty} \max_{k \geq 0} \mathbb{P}_{a_n}[T_{z_n} = k] = 0. \tag{3.12}$$

Then

$$\lim_{n \rightarrow \infty} \sup_{t \geq 0} \left| d_{\text{sep}}^{(n)}(t) - \mathbb{P}[T_{z_n}^{a_n, b_n} > t] \right| = 0. \quad (3.13)$$

Proof. We want to show that $P_n^t(x, y)/\pi_n(y)$ achieves its smallest value for $(x, y) = (a_n, b_n)$ up to a negligible correction. According to (iv) we do not need to worry about the case when both x and y lie in A_n . For the other cases, condition (iii) combined with Lemma 3.5 guaranties that

$$\forall t, \forall (x, y) \in \Omega_n^2 \setminus A_n^2, \quad P_n^t(x, y)/\pi_n(y) \geq \mathbb{P}[T_{z_n}^{x, y} \leq t] \geq \mathbb{P}[T_{z_n}^{a_n, b_n} \leq t]. \quad (3.14)$$

Finally, applying Lemma 3.5 again yields that

$$0 \leq \frac{P_n^t(a_n, b_n)}{\pi_n(b_n)} - \mathbb{P}[T_{z_n}^{a_n, b_n} \leq t] \leq \frac{1}{2} t_{\text{rel}}^{(n)} \max_{k \geq 0} \mathbb{P}[T_{z_n}^{a_n, b_n} = k] \sqrt{(1 - \pi_n(z_n))/\pi_n(z_n)}. \quad (3.15)$$

This allows to conclude the proof by noticing that the right-hand side of (3.15) is $o(1)$ (using (i) and (v)). \square

Remark 3.9. We note that for lazy chains condition (v) in Proposition 3.8 follows from the condition $\lim_{n \rightarrow \infty} \text{dist}(a_n, z) = \infty$ (which is satisfied in Examples 4.1-5.1), where $\text{dist}(a_n, z)$ is the minimal k such that $P^k(a_n, z) > 0$. To see this, consider the non-lazy path the chain performed from a_n to z by time T_z , $\gamma = (\gamma_0 = a_n, \gamma_1, \dots, \gamma_\ell = z)$ (i.e. for all $i < \ell$, $\gamma_{i+1} \neq \gamma_i$ and possibly after spending some time at γ_i the chain moved to γ_{i+1}). The conditional law of T_z , given γ , is that of a sum of ℓ independent geometric random variables with parameter $1/2$, and so by the local CLT its mode is at most $C/\sqrt{\ell} \leq C/\sqrt{\text{dist}(a_n, z)}$. Finally, note that the mode of a mixture is at most the maximal mode of a distribution in the mixture.

4 Total-variation cutoff without separation cutoff examples

In this section we describe two similar examples of sequences of reversible chains which exhibit total-variation cutoff but no separation cutoff. The analysis of both examples is extremely similar. We present both examples since while the first demonstrates that (1.7) is indeed sharp, it is much harder to transform it into an example of lazy SRWs on bounded degree expander graphs.

Example 4.1. Given $n \geq 2$, set $\Omega_n := A \cup \{z\} \cup B$ where $A = A_n := \{a = a_n, a_{n-1}, \dots, a_1\}$ and $B = B_n := \{b_1, b_2, \dots, b_{n-1}, b_n = b\}$. For notational convenience we write $a_0 := z =: b_0$. The matrix P_n has positive transition rates on the set of (un-oriented) edges $E = E_A \cup E_B \cup E_{\text{Long}}$, where

$$\begin{aligned} E_A &:= \{e_k^A := \{a_k, a_{k-1}\} : k \in [n]\}, \\ E_B &:= \{e_k^B := \{b_k, b_{k-1}\} : k \in [n]\}, \\ E_{\text{Long}} &:= \{e_k^L := \{z, b_k\} : k \in [n]\}. \end{aligned} \quad (4.1)$$

With a small abuse of notation we define e_1^L and e_1^B to be two distinct parallel edges. To each of these edges, we associate conductances (or weights), as follows

- $w_n(e_k^A) = 2^{-k} = w_n(e_k^B)$, for all $k \in [n]$.
- $w_n(e_k^L) = \frac{w_n(e_k^B) + w_n(e_{k+1}^B)}{n-1} = 3 \cdot \frac{2^{-(k+1)}}{(n-1)}$ for $k \in [n-1]$, and $w_n(e_n^L) = \frac{2^{-n}}{(n-1)}$.

We let P_n be the transition matrix of the $(1/2)$ -lazy random walk on the graph (Ω_n, E) with conductances w_n , i.e. we set

$$P_n(x, x) = 1/2 \text{ for all } x \in \Omega_n,$$

$$P_n(x, y) = \frac{w_n(x, y)1_{x \neq y}}{2w_n(x)}, \tag{4.2}$$

where $w_n(x) := \sum_{y \in \Omega_n} w_n(x, y)$ with the convention that $w_n(z, b_1) = w_n(e_1^L) + w_n(e_1^B)$. This Markov chain is reversible with respect to

$$\pi_n(x) := \frac{w_n(x)}{\sum_{y \in \Omega_n} w_n(y)}.$$

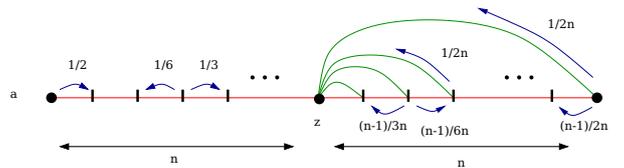


Figure 3: A schematic representation of the transition rates for Example 4.1. On the segments A and B the transition rates away from and towards the center of mass z are equal respectively to $1/6$ and $1/3$ (on the A side) and $(n - 1)/6n$ and $(n - 1)/3n$ (on the B side). The rate for using a green-edge to land on z is equal to $1/2n$. The rates for using green edges in the other direction has a more complicated expression prescribed by reversibility. These rates are described below despite the fact that they play no role in our analysis.

A simple calculation show that

$$w_n(z) = 1 + \frac{3 - 2^{-(n-2)}}{2(n - 1)},$$

$$\sum_{y \in \Omega_n} w_n(y) = 4(1 - 2^{-n}) + \frac{3 - 2^{-(n-2)}}{2(n - 1)}, \tag{4.3}$$

which implies $\lim_{n \rightarrow \infty} \pi_n(z) = 1/4$. The transition matrix obtained from w_n is

- $P_n(x, x) = 1/2$, for all $x \in \Omega_n$.
- $P_n(a_n, a_{n-1}) = 1/2$.
- $2P_n(a_i, a_{i+1}) = 1/3 = P_n(a_i, a_{i-1})$, for all $1 \leq i < n$.
- $P_n(b_i, z) = \frac{1}{2n}$, for $i \geq 2$.
- $P_n(b_i, b_{i-1}) = \frac{1}{3}(1 - \frac{1}{n}) = 2P_n(b_i, b_{i+1})$, for all $2 \leq i \leq n - 1$.
- $P_n(b_n, b_{n-1}) = \frac{1}{2} - \frac{1}{2n}$.
- $P_n(b_1, z) = \frac{1}{2n} + \frac{1}{3}(1 - \frac{1}{n})$ and $P_n(b_1, b_2) = \frac{1}{6}(1 - \frac{1}{n})$.
- $P_n(z, b_1) = \frac{2n+1}{4(2n+1-2^{-(n-2)})} = \frac{1+o(1)}{4}$ and $P_n(z, a_1) = \frac{n-1}{4n+2-2^{-(n-1)}} = \frac{1-o(1)}{4}$.
- $P_n(z, b_k) = \frac{3}{2^{k+1}(2n+1-2^{-(n-2)})} = \frac{3-o(1)}{n2^{k+2}}$, for $2 \leq k \leq n - 1$,
and $P_n(z, b_n) = \frac{1}{2^n(2n+1-2^{-(n-2)})}$.

Note that for this chain, condition (\star) is easily verified using Theorem 3.7. Since under P_{a_n} , T_z is concentrated around time $6n$, to prove total-variation cutoff around time $6n$ for this sequence of chains (using Proposition 3.3), we only need to verify that a_n is the initial state from which T_z is (stochastically) the largest. A crucial fact which shall

assist us in this task is that for all $i \in [n]$ and all t

$$P_{b_i}[T_z > t] \leq P_{a_i}[T_z > t]. \tag{4.4}$$

The reason for this identity is the following: We couple X^A and X^B starting from a_i and b_i (resp.) in the following manner: with probability $1/2$ both stay put, with probability $(\frac{1}{2} - \frac{1}{2n})$ X^A and X^B make “the same move” (+/− 1 (towards/away from z)) with (conditional) probability $1/3$ and $2/3$, resp. (unless the current position of the chain is either a_n or b_n in which case the move has to be −1) and with probability $1/(2n)$, X^B is sent directly to z while X^A moves towards/away from z with probability $2/3$ and $1/3$ (unless it is located at a_n). We do not need to specify how the coupling is defined after X^B has hit z .

A way to describe X_t starting from B before it hits z is the following: at each step it is killed (hits z) with rate $1/(2n)$ and conditionally on not being killed, it performs “the same” random walk as that on A (in terms of the index of its current position) but with holding probability $n/(2n - 1) \geq 1/2$.

Consequently,

$$\begin{aligned} \max_{y \in \Omega_n} P_y[T_z > t] &= \max_{y \in A_n} P_y[T_z > t] = P_{a_n}[T_z > t], \\ \max_{y \in B_n} P_y[T_z > t] &= P_{b_n}[T_z > t]. \end{aligned} \tag{4.5}$$

Moreover, it follows from the above discussion that

$$P_{a_n}[T_z > t] \left(1 - \frac{1}{2n}\right)^t \leq P_{b_n}[T_z > t] \leq \min\left(P_{a_n}[T_z > t], \left(1 - \frac{1}{2n}\right)^t\right). \tag{4.6}$$

We now turn to the task of verifying that there is no cutoff in separation. Note that conditions (i)-(ii) of Proposition 3.8 hold by construction, condition (v) holds by Remark 3.9, while condition (iv) holds by (3.8). Lastly, condition (iii) of Proposition 3.8 follows from (4.5), and so Proposition 3.8 applies. Consequently,

$$\lim_{n \rightarrow \infty} \sup_{t \geq 0} |d_{\text{sep}}^{(n)}(t) - \mathbb{P}[T_z^{a_n, b_n} > t]| = 0. \tag{4.7}$$

Set $m_n := \lceil n^{2/3} \rceil$ (the exponent $2/3$ can be replaced by any number in $(1/2, 1)$). It is standard to check that

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{a_n}[|T_z - 6n| > m_n] &= 0, \\ \lim_{n \rightarrow \infty} \sup_{t \in [0, 6n - 2m_n]} |P_{b_n}(T_z > t + m_n) - P_{b_n}(T_z > t)| &= 0. \end{aligned} \tag{4.8}$$

Hence it follows from (4.7) that for all $c \in (0, 6)$

$$\lim_{n \rightarrow \infty} |d_{\text{sep}}^{(n)}(6n + \lfloor cn \rfloor) - P_{b_n}[T_z > cn]| = 0. \tag{4.9}$$

This and (4.6) yield that for any $0 < \varepsilon \leq 1/4$

$$\lim_{n \rightarrow \infty} d_{\text{sep}}^{(n)}(\lfloor sn \rfloor) = \begin{cases} 1 & \text{if } s \leq 6, \\ e^{-(s-6)/2} & \text{if } s \in [6, 12), \\ 0 & \text{if } s > 12. \end{cases} \tag{4.10}$$

Hence there is no separation cutoff. Moreover,

$$\sup_{0 < \varepsilon < 1/2} \liminf_{n \rightarrow \infty} \frac{t_{\text{sep}}^{(n)}(\varepsilon)}{t_{\text{sep}}^{(n)}(1 - \varepsilon)} = 2.$$

We now describe a variant of the previous example which is a nearest neighbor lazy weighted random walk on a bounded degree graph with bounded transition probabilities.

Example 4.2. Let $\Omega_n = A \cup B \cup C \cup \{z\}$, where

$$\begin{aligned} A = A_n &:= \{a_1, a_2, \dots, a_{2n} = a\}, \\ B = B_n &:= \{b_1, b_2, \dots, b_{2n} = b\}, \\ C = C_n &:= \{c_1, c_2, \dots, c_{n-1}\}. \end{aligned} \tag{4.11}$$

For notational convenience we write $a_0 := z := b_0 = c_0$ and $c_n = b_n$. Consider the following transition matrix

- $P_n(x, x) = 3/4$ for all $x \in \Omega_n \setminus C$ and $P_n(c_i, c_i) = 1/2$ for all $i \in \{1, \dots, n-1\}$,
- $P_n(a_{2n}, a_{2n-1}) = 1/4 = P_n(b_{2n}, b_{2n-1})$,
- $2P_n(a_i, a_{i+1}) = 1/6 = P_n(a_i, a_{i-1})$, for all $1 \leq i < 2n$,
- $2P_n(b_i, b_{i+1}) = 1/6 = P_n(b_i, b_{i-1})$, for all $i \in [2n-1] \setminus \{n\}$,
- $2P_n(c_i, c_{i+1}) = 1/3 = P_n(c_i, c_{i-1})$, for all $1 \leq i \leq n-1$,
- $P_n(b_n, b_{n+1}) = P_n(b_n, c_{n-1}) = P_n(b_n, b_{n+1}) = 1/12$,
- $P_n(z, a_1) = P_n(z, b_1) = P_n(z, c_1) = 1/12$.

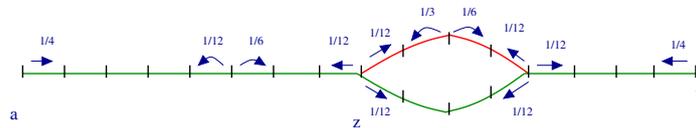


Figure 4: A schematic representation of the transition rates for Example 4.2 for $n = 4$. When at a state of degree two or three (other than z), conditioned on making a non-lazy step, the chain moves away from (resp. towards) z with conditional probability $1/3$ (resp. $2/3$). For vertices of degree 2: along the green edges, rates away from and towards the center of mass z are equal respectively to $1/12$ and $1/6$ and along the red edges they are equal to $1/6$ and $1/3$, respectively. The transitions away from vertices of degree 1 and 3 are given on the figure.

States a_{2n} , b_{2n} and z play here the same respective roles as a_n , b_n and z in the previous example. A simple calculation (similar to (4.3)) yields that

$$\lim_{n \rightarrow \infty} \pi_n(z) = 2/7.$$

We argue that for all $t \geq 0$ and $i \in [2n]$

$$\max(P_{c_i}[T_z > t], P_{b_i}[T_z > t]) \leq P_{a_i}[T_z > t] \leq P_{a_{2n}}[T_z > t] \tag{4.12}$$

In particular,

$$\forall t \geq 0, \quad \max_{x \in \Omega_n} P_x[T_z > t] = P_{a_{2n}}[T_z > t]. \tag{4.13}$$

Since the hitting time of z under $P_{a_{2n}}$ is concentrated around time $t = 24n$, by Proposition 3.3 the sequence exhibits total-variation cutoff around time $24n$.

The last inequality in (4.12) is trivial. For the first one we consider the case where P_n is replaced by P'_n which satisfies $2P'_n(c_i, c_{i+1}) = 1/3 = P'_n(c_i, c_{i-1})$ and $P'_n(c_i, c_i) = 1/2$ for $1 \leq i \leq n-1$ and $P'(x, y) = P(x, y)$ elsewhere. As adding extra laziness increases stochastically the hitting time T_z (as in Remark 3.9 consider the law of γ , the non-lazy path performed by the chain by time T_z ; Clearly it is invariant under this transformation, while the conditional law of T_z , given γ , can only increase, stochastically),

$$P_{b_i}[T_z > t] \leq P'_{b_i}[T_z > t] = P_{a_i}[T_z > t], \tag{4.14}$$

(where P' denotes the distribution of the modified chain with the increased holding probability on C_n) and the same holds when b_i is replaced by c_i .

To prove that b_{2n} is the vertex from which the hitting time of z is the largest, we need to prove the following two inequalities valid for $i \in \{1, \dots, n\}$

$$P_{c_i}[T_z > t] \leq P_{b_i}[T_z > t] \quad \text{and} \quad P_{b_i}[T_z > t] \leq P_{b_{i+n}}[T_z > t]. \quad (4.15)$$

Both can be proved by coupling arguments. For the first one, we can couple the non-lazy path of the chains starting from b_i and c_i until they reach either b_n or z (the second being at position c_j when the first is at position b_j), and then in the case they reach $b_n = c_n$ let them evolve together until they reach z . The larger laziness on the path starting from c_i until the merging time, implies stochastic domination. For the second inequality, the case $i = n$ follows from fact that starting from b_{2n} the chain has to go through b_n before reaching z . For $i < n$, we can couple the chain starting from b_i and b_{i+n} until the pair of chains reaches either (b_n, b_{2n}) or (z, b_n) (the second chain being at position b_{j+n} when the first is at position b_j), and conclude using the case $i = n$.

As in the previous example, we can apply Proposition 3.8. The reason why separation cutoff does not occur is that when starting from b_{2n} , the hitting time T_z is not concentrated. Indeed it is concentrated around $18n$ under the conditioned probability measure $P_{b_{2n}}[\cdot \mid X_{T_z-1} = c_1]$, while it is concentrated around $24n$ under $P_{b_{2n}}[\cdot \mid X_{T_z-1} = b_1]$. As by symmetry

$$P_{b_{2n}}[X_{T_z-1} = c_1] = P_{b_{2n}}[X_{T_z-1} = b_1] = \frac{1}{2},$$

this yields

Lemma 4.3. *We have*

$$\lim_{n \rightarrow \infty} P_{b_{2n}}[T_z \geq sn] = \begin{cases} 1 & \text{if } s < 18, \\ 1/2 & \text{if } s \in (18, 24), \\ 0 & \text{if } s > 24. \end{cases} \quad (4.16)$$

While this result is rather elementary (we use some surgery to compare T_z with a sum of independent variables, and then the law of large number for this sequence), the proof in full detail is long to expose (c.f. [5, Example 8.1]) and we choose to leave it as an exercise. Applying Proposition 3.8 for an adequate choice of sets and states (here $(a_{2n}, b_{2n}, A_n, B_n \cup C_n)$ plays the role of (a_n, b_n, A_n, B_n) from Proposition 3.8) yields

$$\lim_{n \rightarrow \infty} d_{\text{sep}}^{(n)}(sn) = \begin{cases} 1 & \text{if } s < 42, \\ 1/2 & \text{if } s \in (42, 48), \\ 0 & \text{if } s > 48. \end{cases} \quad (4.17)$$

In particular, there is no cutoff in separation.

5 Separation cutoff without total variation cutoff example

In the following example the analysis of the sharp transition of $d_{\text{sep}}^{(n)}(t)$ is reduced to the analysis of the behavior of sum of i.i.d. random variables in the large deviation regime. The analysis below is too coarse for the purpose of determining the width of the cutoff window. We later present a refined analysis for Example 6.12 (which is the bounded degree un-weighted version of Example 5.1) in § 6.5, which shows that in fact $t_{\text{sep}}^{(n)}(\varepsilon) - t_{\text{sep}}^{(n)}(1 - \varepsilon) \leq C|\log \varepsilon|$, for some absolute constant $C > 0$. The analysis in § 6.5 is built upon the analysis of Example 5.1 below, as it relies (in a non-quantitative manner) on the fact that certain large deviation estimates hold uniformly over compact sets (the identity of the large deviation rate function is not important for the analysis in § 6.5).

Example 5.1. Let $M \geq 10$ be a fixed integer whose exact value shall be determined later. Consider the state space $\Omega_n = A \cup B \cup \{z\} \cup C \cup D \cup \{z'\}$, where $A = A_n := \{a_{Mn} = a, a_{Mn-1}, \dots, a_1\}$, $B = B_n := \{b_{Mn} = b, b_{Mn-1}, \dots, b_1\}$, $C = C_n := \{c_1, c_2, \dots, c_{n-1}\}$ and $D = D_n := \{d_1, d_2, \dots, d_{n-1}\}$. We use the following notational convention: $a_0 = b_0 := z' := c_n = d_n$ and $d_0 := z := c_0$. Consider the following transition matrix

- $P_n(i, i) = \begin{cases} 3/4 & i \in C, \\ 1/2 & \text{otherwise.} \end{cases}$
- $P_n(a_{Mn}, a_{Mn-1}) = 1/2 = P_n(b_{Mn}, b_{Mn-1})$.
- $P_n(z, c_1) = 1/4 = P_n(z, d_1)$.
- $P_n(z', c_n) = P_n(z', d_n) = 1/6 = 2P_n(z', a_1) = 2P_n(z', b_1)$.
- $P_n(a_i, a_{i-1}) = P_n(b_i, b_{i-1}) = P_n(d_j, d_{j-1}) = 2P_n(c_j, c_{j-1}) = 1/3$.
 $P_n(a_i, a_{i+1}) = P_n(b_i, b_{i+1}) = P_n(d_j, d_{j+1}) = 2P_n(c_j, c_{j+1}) = 1/6$,
 for all $i \in [Mn - 1]$ and $j \in [n]$.

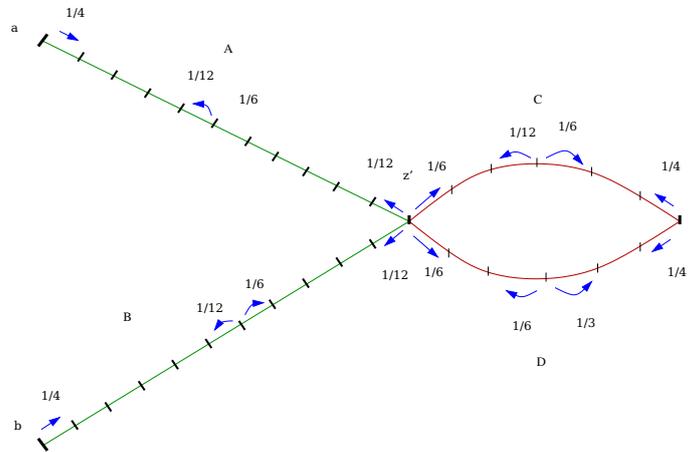


Figure 5: A schematic representation of the transition rates for Example 5.1. When at a state of degree two or four (other than z), conditioned on making a non-lazy step, the chain moves away from (resp. towards) z with conditional probability $1/3$ (resp. $2/3$). The transition rates away from and towards the center of mass z , from degree two states, are equal respectively to $1/6$ and $1/3$, except on the segment C , due to increased holding probability. The transition rates away from the rest of the states are specified in the figure.

This chain is a modification of Aldous' example (which was discussed in § 2). The difference lies in the introduction of an additional branch B to the graph. This branch has no effect on the total-variation profile of the convergence to equilibrium, but crucially modifies the separation profile, as $P_n^t(a, b)/\pi_n(a)$ (recall $a := a_{nM}$ and $b := b_{nM}$) is the quantity that takes the longest time to reach equilibrium (i.e. up to negligible correction $(x, y) = (a, b)$ maximizes $1 - P_n^t(x, y)/\pi_n(y)$ for all relevant t).

A standard calculation yields that

$$\lim_{n \rightarrow \infty} \pi_n(z) = 2/11 \text{ and } \lim_{n \rightarrow \infty} 2^n \pi_n(z') = 6/11. \tag{5.1}$$

By symmetry, the law of T_z starting, resp., from a_i and b_i is identical for all i and by the Markov property, it is stochastically increasing in i (for $i > j$, to reach z from a_i (resp. b_i) the chain must first hit a_j (resp. b_j)). Only minor efforts are necessary to prove rigorously

that a and b are the points in $A \cup B \cup D$ for which the hitting time T_z is stochastically the largest (the coupling arguments are similar to the one developed in the previous section), while for any choice of $M > 1$,

$$\limsup_{n \rightarrow \infty} \sup_t (\max_{c \in C} P_c[T_z > t] - P_a[T_z > t]) \leq 0.$$

Due to the different holding probabilities along the two branches, C, D , the distribution of T_z under P_a is not concentrated around its mean. Thus, by Proposition 3.3, there is no total-variation cutoff, and the total-variation asymptotic profile is given by

$$\lim_{n \rightarrow \infty} d^{(n)}(6sn) := \begin{cases} 0 & \text{if } s < M + 1, \\ 1/2 & \text{if } s \in (M + 1, M + 2), \\ 1 & \text{if } s > M + 2. \end{cases} \quad (5.2)$$

To show that there is separation cutoff, it suffices to prove that

$$\liminf_{n \rightarrow \infty} \inf_t \min_{x, y \in \Omega_n} P_n^t(x, y) / \pi_n(y) - \min(1, P_n^t(a, b) / \pi_n(b)) = 0, \quad (5.3)$$

and to show that $\min(1, P_n^t(a, b) / \pi_n(b))$ displays an abrupt transition. Let us start with the second point. According to Lemma 3.5 (first inequality of (3.7)), we have

$$\mathbb{P}[T_{z'}^{a,b} = t] / \pi_n(z') \leq P_n^t(a, b) / \pi_n(b) \leq \mathbb{P}[T_{z'}^{a,b} \leq t] / \pi_n(z') \quad (5.4)$$

By definition $T_{z'}^{a,b}$ is the sum of two independent hitting times of a biased random walk on a segment of length Mn (from one end-point towards the one towards which there is a bias). We make some efforts to compute the large deviation behavior of this sum.

Lemma 5.2. Consider a lazy random walk $(Z_t)_{t \geq 0}$ on \mathbb{Z}_+ with rates $p(x, x + 1) = 1/3$, $p(x + 1, x) = 1/6$, $x \in \mathbb{Z}_+$. Let \mathcal{T}_N be the first hitting time of N . We have

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}[\mathcal{T}_N = \lfloor sN \rfloor] &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}[\mathcal{T}_N \leq sN] = -\Psi(s), \text{ for } s \in [1, 6] \\ \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}[\mathcal{T}_N = \lfloor sN \rfloor] &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}[\mathcal{T}_N \geq sN] = -\Psi(s), \text{ for } s \geq 6, \end{aligned} \quad (5.5)$$

where Ψ is the following Legendre transform

$$\Psi(s) := \sup_{\lambda \in (-\infty, \infty)} [\lambda s - \log f(\lambda)], \quad (5.6)$$

where

$$f(\lambda) := \begin{cases} \infty & \text{if } \lambda > \log(6/(3 + 2\sqrt{2})), \\ \frac{6e^{-\lambda} - 3}{2} - \frac{\sqrt{(6e^{-\lambda} - 3)^2 - 8}}{2} & \text{if } \lambda \leq \log(6/(3 + 2\sqrt{2})). \end{cases}$$

Moreover, $\Psi(6) = 0 = \Psi'(6)$ and the second derivative $\Psi''(6)$ is positive.

Proof. Let X' be the random walk with the same rates on \mathbb{Z} , and \mathcal{T}'_N be the first hitting time of N for this walk. By the Markov property \mathcal{T}'_N is the sum of N IID copies of \mathcal{T}'_1 and hence we can use Cramér's Theorem (see e.g. [8, Chapter 2]) to obtain the large deviation for \mathcal{T}'_N below its mean. If one decomposes according to the value of X'_1 we notice that the Laplace transform $f(\lambda) := \mathbb{E}[e^{\lambda \mathcal{T}'_1}]$ satisfies

$$f(\lambda) = e^\lambda \left(\frac{1}{3} + \frac{1}{6} f(\lambda)^2 + \frac{1}{2} f(\lambda) \right). \quad (5.7)$$

and we deduce the right value for $f(\lambda)$ from this relation (the fact that $f(0) = 1$ and continuity of f indicates which root to choose in (5.7)). Note that the derivative of $\log f(\lambda)$

at zero is equal to 6 which implies that $\Psi(6) = 0$ (Alternatively, $\mathbb{E}[\mathcal{T}'_1] = 6$, hence by Cramér's Theorem it must be the case that $\Psi(6) = 0$). As Ψ is non-negative (since $\log f(0) = 0$), it must be the case that it attains a global minimum at 6, which implies that $\Psi'(6) = 0$ and $\Psi''(6) > 0$.

Now, note that $\mathcal{T}_i - \mathcal{T}_{i-1}$ are independent variables, which are dominated by \mathcal{T}'_1 and who converge (when i tends to infinity) to \mathcal{T}'_1 in law. In particular, by dominated convergence (and Cesaro's Theorem) we have that for any $\lambda \in (-\infty, \log \frac{6}{3+2\sqrt{2}}]$,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbf{E}[e^{\lambda \mathcal{T}_N}] = f(\lambda). \tag{5.8}$$

and thus in that case the result follows from Gärdner Ellis Theorem [8]. Finally, the local large deviation estimate (the result on $\mathbb{P}[\mathcal{T}_N = \lfloor sn \rfloor]$) can be deduced from the large deviation principle using the fact that due to laziness

$$\frac{\mathbf{P}[\mathcal{T}_N = t + 1]}{\mathbf{P}[\mathcal{T}_N = t]} \geq \frac{1}{2}. \tag{5.9}$$

We leave it as an exercise. Note moreover that the convergence in (5.5) holds uniformly on $s \in K$ for any compact K (it can be deduced e.g. from (5.9)). \square

A consequence of (5.4) and the previous lemma in conjunction with (5.1) and Lemma 3.5 is that if s_M is given by $2Ms^*$, where s^* is the unique solution in $(0, 6)$ of

$$2M\Psi(s) = \log 2, \tag{5.10}$$

then

$$\lim_{n \rightarrow \infty} \frac{P_n^{\lfloor sn \rfloor}(a, b)}{\pi_n(b)} = \begin{cases} 0, & \text{if } s < s_M, \\ \infty, & \text{if } s \in (s_M, 12M]. \end{cases} \tag{5.11}$$

An order 2 Taylor expansion of (5.10) around $s = 6$ readily shows that $6 - s^* = \Theta(1/\sqrt{M})$ (i.e. $12M - s_M = \Theta(\sqrt{M})$) for large M . In particular, $s_M \geq 11M$ for M sufficiently large. What is left to do in order to prove separation cutoff is to check that for any $s \in (s_M, 12M]$ (in fact, by monotonicity it suffices to consider only s arbitrarily close to s_M) we have

$$\liminf_{n \rightarrow \infty} \min_{x, y} P_n^{\lfloor sn \rfloor}(x, y) / \pi_n(y) \geq 1.$$

In what follows we let $s \in (s_M, 12M]$ be fixed.

We first use Lemma 3.5 to reduce to the case of $x = a_i, y = b_j, i, j \geq Mn/2$. Set $E := \{a_i : i \geq \frac{Mn}{2}\} \cup \{b_i : i \geq \frac{Mn}{2}\}$. By (3.7) for any $x \in \Omega_n$ and $y \in \Omega_n \setminus E$ we have

$$\frac{P_n^{\lfloor sn \rfloor}(x, y)}{\pi_n(y)} \geq \mathbb{P}[T_z^{x, y} \leq 11Mn], \tag{5.12}$$

and it is a simple exercise to show that (when M is sufficiently large)

$$\lim_{n \rightarrow \infty} \min_{(x, y) \in \Omega_n \times (\Omega_n \setminus E)} \mathbb{P}[T_z^{x, y} \leq 11Mn] = 1. \tag{5.13}$$

Finally, to treat the case $x = a_i, y = b_j$ (the cases (a_i, a_j) or (b_i, b_j) are treated in the same manner), $i, j \geq Mn/2$, we use again Lemma 3.5 which asserts that

$$\frac{P_n^{\lfloor sn \rfloor}(a_i, b_j)}{\pi_n(b_j)} \geq \max \left(\frac{\mathbb{P}[T_{z'}^{a_i, b_j} = \lfloor sn \rfloor]}{\pi_n(z')}, \mathbb{P}[T_{z'}^{a_i, b_j} \leq \lfloor sn \rfloor] \right) \tag{5.14}$$

Note that $T_{z'}^{a_i, b_j}$ is (cf. the proof of Lemma 5.2) a sum of $i + j$ independent random variables (not identically distributed) and that

$$\lim_{n \rightarrow \infty} \sup_{i, j \geq Mn/2} \left| \frac{1}{i + j} \log \mathbb{E}[e^{\lambda T_{z'}^{a_i, b_j}}] - \log f(\lambda) \right| = 0. \quad (5.15)$$

One deduces from Gärdner Ellis Theorem [8] and the following consequence of laziness

$$\frac{\mathbb{E}[T_{z'}^{a_i, b_j} = t + 1]}{\mathbb{E}[T_{z'}^{a_i, b_j} = t]} \geq 1/2, \quad (5.16)$$

that for any $u \in (1, \infty)$,

$$\lim_{n \rightarrow \infty} \sup_{i, j \geq Mn/2} \left| -\frac{1}{i + j} \log \mathbb{P}[T_{z'}^{a_i, b_j} = \lfloor (i + j)u \rfloor] - \Psi(u) \right| = 0, \quad (5.17)$$

and the convergence holds uniformly on compact sets. Now let us fix η which satisfies

$$\Psi\left(\frac{6s}{s + 6\eta}\right) \leq \frac{\log 2}{4M},$$

Using (5.17), there exists δ such that for all n sufficiently large, for all $i, j \geq Mn/2$,

$$\begin{aligned} \left(i + j \leq \frac{sn}{6} + \eta n\right) &\Rightarrow \mathbb{P}[T_{z'}^{a_i, b_j} \leq \lfloor sn \rfloor] \geq 1 - e^{-\delta n}, \\ \left(i + j \geq \frac{sn}{6} + \eta n\right) &\Rightarrow \log \mathbb{P}[T_{z'}^{a_i, b_j} = \lfloor sn \rfloor] \geq -(i + j) \left[\Psi\left(\frac{sn}{i + j}\right) + \delta \right], \end{aligned} \quad (5.18)$$

Note that the l.h.s. in the second line satisfies

$$(i + j) \left[\Psi\left(\frac{sn}{i + j}\right) + \delta \right] \leq 2Mn \left(\max \left[\Psi\left(\frac{s}{2M}\right), \Psi\left(\frac{6s}{s + 6\eta}\right) \right] + \delta \right). \quad (5.19)$$

As $2M\Psi\left(\frac{s}{2M}\right) < \log 2$ (since $s \in (s_M, 12M]$) and δ can be chosen arbitrarily small, (5.18) (second line) and (5.19) imply that for sufficiently large n , for any i, j satisfying $i + j \geq \frac{sn}{6} + \eta n$, we have

$$\mathbb{P}[T_{z'}^{a_i, b_j} = \lfloor sn \rfloor] \geq 2^{-n(1-\delta)}. \quad (5.20)$$

Combining this with (5.18) (first line) and (5.14) we can conclude that

$$\min_{i, j \geq Mn/2} \frac{P_n^{\lfloor sn \rfloor}(a_i, b_j)}{\pi_n(b_j)} \geq 1 - e^{-\delta n}. \quad (5.21)$$

5.1 Concerning Remark 1.3

Note that by performing a minor modification in the above construction we can bring the pre-cutoff ratio for total-variation to the largest possible value: 2. A way to achieve this is to make one of the branches linking z' to z much faster than the other (instead of only twice faster as in Example 3, we want the ratio of speeds to tend to infinity).

What we can do is to make these branches of length $\lceil \sqrt{n} \rceil$ while A and B are of length n . Furthermore, we choose the speed on one branch to be $1/6$ while that one the other being $1/(6\sqrt{n})$ by increasing the holding probability on this branch (see Figure 6). Using similar reasoning as in the analysis of Example 5.1 one can show that for this construction there is separation cutoff around time $12n$ (note that here $-\log \pi_n(z') = \Theta(\sqrt{n})$, which by (5.4) implies that for $t_n := \lceil (12 - \varepsilon)n \rceil$, $P_n^{t_n}(a, b)/\pi_n(b) \leq \mathbb{P}[T_{z'}^{a, b} \leq t_n]/\pi_n(z') = o(1)$, for every $\varepsilon > 0$).

We can also find a similar example with transition rates bounded uniformly from zero by considering two branches of different lengths, but in that case the analysis turns out to be more intricate.

Total variation and separation cutoffs are not equivalent

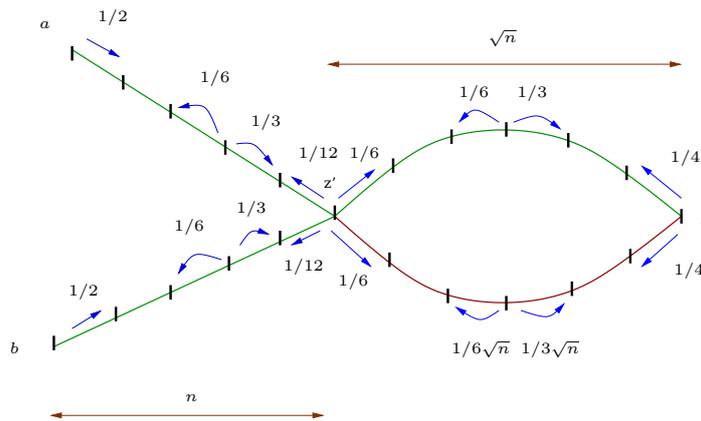


Figure 6: A modification of the graph size and of the holding probability along the slow branch as shown above yields a counter-example with cutoff in separation and the maximal possible pre-cutoff ratio 2 in total-variation.

6 Transforming Examples 4.2 and 5.1 into lazy simple random walk on bounded degree expander graphs

6.1 General comments and preliminaries

In this section we transform Examples 4.2 and 5.1 into lazy SRWs on a sequence of bounded degree expander graphs. For this kind of walk, the equilibrium measure is $\pi(v) = \frac{\deg v}{\sum_u \deg u}$, and thus no particular vertex can have the role of the “center of mass” as in the previous examples. Let us rewrite the definition of the Cheeger constant in this context.

Definition 6.1. Let $G = (V, E)$ be a finite connected graph. For every $S \subset V$ denote $C_S := \sum_{v \in S} \deg v$. For any $S \subset V$ we define its **edge boundary** $\partial_E S$ to be the collection of edges having one vertex in S and the other in $V \setminus S$. The Cheeger constant of lazy simple random walk on G is defined as

$$\text{ch}_{\text{Lazy}}(G) := \min_{S: \pi(S) \leq 1/2} |\partial_E S| / 2C_S,$$

which coincides with Definition 3.6 (see e.g. [16, Remark 7.2]). We say that G is a c -lazy expander if $\text{ch}_{\text{Lazy}}(G) > c$. We say that a sequence of finite graphs $(G_n)_{n \geq 1}$ is a family of c -lazy expanders if $\inf_n \text{ch}_{\text{Lazy}}(G_n) > c$.

In our new context, the center of mass is rather a set which contains a positive fraction of the vertices. We shall relate the mixing-time of the chain to the hitting time of this set. *Mutatis mutandis*, the results of Section 3 and in particular Lemma 3.5 can be adapted to this new context, but only if the set and the starting point satisfy a special relation:

Definition 6.2 (Balanced sets). For any $Z \subset \Omega$ we denote the **hitting time** of Z by $T_Z := \inf\{t : X_t \in Z\}$.

- We say that Z is **balanced** seen from $x \in \Omega$ if for all t such that $P_x[T_Z = t] > 0$,

$$\forall z \in Z, \quad P_x[X_t = z \mid T_Z = t] = \pi_Z(z), \tag{6.1}$$

where $\pi_Z(\cdot) = \frac{1_{\cdot \in Z} \pi(\cdot)}{\pi(Z)}$ is π conditioned on the set Z .

- We say that Z is **balanced** seen from the set A if it is balanced seen from x for all $x \in A$.

- We define $T_Z^{x,y}$ to be a random variable distributed like the sum of two independent realizations of T_Z , once under P_x and once under P_y . That is, for all $t \geq 0$,

$$\mathbb{P}[T_Z^{x,y} = t] := \sum_{k=0}^t P_x[T_Z = k]P_y[T_Z = t - k]. \quad (6.2)$$

Note that sets are not likely to be balanced by “pure luck” and we will be careful to introduce a sufficient amount of symmetry when constructing our graphs, so that our center of mass will be balanced seen from many starting points. However, this property cannot be satisfied for all starting points and we will have to deal with the remaining initial vertices separately (and show that they are irrelevant for determining the worst-case total-variation and separation distances), by using a crude ℓ_2 estimate (Lemma 6.8).

Lemma 6.3. *Let (Ω, P, π) be a finite irreducible lazy reversible Markov chain and consider $x, y \in \Omega$, and $Z \in \Omega$ which is balanced seen from both x and y*

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

$$P^t(x, y)/\pi(y) \geq \sum_{k \leq t} \mathbb{P}[T_Z^{x,y} = k]P_{\pi_Z}^{t-k}(Z)/\pi(Z) \geq \mathbb{P}[T_Z^{x,y} \leq t]. \quad (6.3)$$

- (ii) If $P_x[T_Z < T_y] = 1$ (i.e. if every path from x to y goes through the set Z) then for all $t \geq 0$ we have that

$$\begin{aligned} P^t(x, y)/\pi(y) &= \sum_{k \leq t} \mathbb{P}[T_Z^{x,y} = k]P_{\pi_Z}^{t-k}(Z)/\pi(Z) \\ &\leq \mathbb{P}[T_Z^{x,y} \leq t] + \frac{1}{2}t_{\text{rel}} \max_{k \in \mathbb{N}} \mathbb{P}[T_Z^{x,y} = k] \sqrt{(1 - \pi(Z))/\pi(Z)}. \end{aligned} \quad (6.4)$$

We use this result directly but also to prove the following key propositions whose aim is to replace Propositions 3.3 and 3.8.

Proposition 6.4. *Let (Ω_n, P_n, π_n) be a sequence of lazy reversible irreducible finite chains which satisfies the product condition. Assume that for each n there exist sequences of sets and vertices $I_n, Z_n \subset \Omega_n, a = a(n) \in \Omega_n$ which satisfy*

- (i) $\inf_n \pi_n(Z_n) > 0$.
- (ii) Z_n is balanced seen from I_n for all n .
- (iii) $\limsup_{n \rightarrow \infty} \sup_{t \geq 0} \max_{i \in I_n} P_i[T_{Z_n} > t] - P_a[T_{Z_n} > t] \leq 0$.
- (iv) $\limsup_{n \rightarrow \infty} \sup_{t \geq 0} \max_{x \in \Omega_n \setminus I_n} \|P_x^t - \pi\|_{\text{TV}} - P_a[T_{Z_n} > t] \leq 0$.

Let $\tau_n(p) := \inf\{t : P_a[T_{Z_n} > t] \leq p\}$. Then

$$\limsup_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon')}{\tau_n(\varepsilon)} \leq 1 \text{ and } \liminf_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{\tau_n(\varepsilon')} \geq 1, \text{ for all } 0 < \varepsilon < \varepsilon' < 1. \quad (6.5)$$

In particular, total-variation cutoff occurs if and only if

$$\lim_{n \rightarrow \infty} \frac{\tau_n(\varepsilon)}{\tau_n(1 - \varepsilon)} = 1, \text{ for every } 0 < \varepsilon < 1. \quad (6.6)$$

Proposition 6.5. *Let (Ω_n, P_n, π_n) be a sequence of lazy reversible irreducible finite Markov chains which satisfies (\star) . Assume that there exist sequences of sets and vertices, $A_n, B_n, Z_n \subset \Omega_n, a_n \in A_n, b_n \in B_n$, which satisfy*

- (i) $\inf_n \pi_n(Z_n) > 0$.
- (ii) $\lim_{n \rightarrow \infty} \max_k P_{a_n}[T_{Z_n} = k] = 0$.
- (iii) Z_n is balanced seen from $I_n := A_n \cup B_n$.
- (iv) $P_x[T_{Z_n} < T_{B_n}] = 1$ for all $x \in A_n$.
- (v)

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{t \geq 0} \max_{y \in B_n} P_y[T_{Z_n} > t] - P_{b_n}[T_{Z_n} > t] &= 0, \\ \limsup_{n \rightarrow \infty} \sup_{t \geq 0} \max_{y \in I_n} P_y[T_{Z_n} > t] - P_{a_n}[T_{Z_n} > t] &= 0. \end{aligned} \tag{6.7}$$

(vi)

$$\liminf_{n \rightarrow \infty} \inf_{t \geq 0} \min_{(x,y) \in A_n^2 \cup (\Omega_n^2 \setminus I_n^2)} \frac{P_n^t(x,y)}{\pi_n(y)} - \frac{P_n^t(a_n,b_n)}{\pi_n(b_n)} \geq 0.$$

Then

$$\lim_{n \rightarrow \infty} \sup_{t \geq 0} |d_{\text{sep}}^{(n)}(t) - \mathbb{P}[T_{Z_n}^{a_n,b_n} > t]| = 0. \tag{6.8}$$

In particular, there is separation cutoff if and only if $T_{Z_n}^{a_n,b_n}$ is concentrated around its median.

Remark 6.6. Note that the results presented above are generalizations of those presented in Section 3. Hence we shall only prove the more general versions in the Appendix.

Remark 6.7. Similarly to Remark 3.9, condition (ii) of Proposition 6.5 is satisfied in Examples 6.9 and 6.12 due to laziness and the fact that $\lim_{n \rightarrow \infty} \min\{t : P_n^t(a_n, Z_n) > 0\} = \infty$.

Lemma 6.8. For any reversible Markov chain, (Ω, P, π) and any $x, y \in \Omega$ and $s, t \geq 0$,

$$P^{s+t}(x, y)/\pi(y) \geq (1 - \|P_x^t - P_y^s\|_{\text{TV}})^2. \tag{6.9}$$

In particular, if $d_x(t) + d_y(s) \leq 1$, then

$$P^{s+t}(x, y)/\pi(y) \geq (1 - d_x(t) - d_y(s))^2. \tag{6.10}$$

Proof. Let $f(z) := \sqrt{P^t(x, z)P^s(y, z)}/\pi(z)$. By reversibility and Jensen's inequality,

$$\begin{aligned} \frac{P^{s+t}(x, y)}{\pi(y)} &= \sum_z \pi(z) \frac{P^t(x, z)P^s(y, z)}{\pi(z)\pi(y)} = \sum_z \pi(z) f^2(z) \geq \left(\sum_z \pi(z) f(z) \right)^2 \\ &\geq \left(\sum_z \min(P^t(x, z), P^s(y, z)) \right)^2 = (1 - \|P_x^t - P_y^s\|_{\text{TV}})^2. \end{aligned}$$

(6.10) follows from (6.9) by the triangle inequality. □

6.2 Building blocks of our constructions

Let us now describe the building blocks of our constructions. We assume for simplicity that n is an even integer. To produce the analog of a biased nearest-neighbor random walk, our constructions must include structures which look like regular trees (for which the SRW has a bias towards the leaves). We must also care about adding some extra connections to avoid producing dead-ends on the leaves (which could lead to a small Cheeger constant). Finally, we must introduce extra symmetries to ensure that the center of mass is balanced seen from all vertices which are sufficiently far from it. Finally, we

“stretch” the edges which are far away from the center of mass (that is, replace each such edge by a path of length L , for some fixed large constant L), to ensure that the worst-case total-variation and separation distances are obtained by vertices which are far away from the center of mass (which is balanced, seen from those vertices).

Step 1: Let $\mathcal{T}_a = (V_a, E_a)$ be a binary tree of depth n rooted at a (in the rest of the construction, we keep calling a the **root**, even though the graph will no longer be a tree). Replace each edge between a pair of vertices belonging to the first $n/2$ generations of \mathcal{T}_a by a path of L edges, where L is an integer which does not depend on n . As L shall remain fixed we omit the dependence in L from our notation. In the course of the proof we will have to require L to be sufficiently large for the purpose of applying a certain crude ℓ_2 estimate. We call the obtained graph H_n^1 . It is a tree rooted at a and we denote its set of leaves by

$$\mathcal{L}_n := (u^1, \dots, u^{2^n}),$$

(\mathcal{L}_n stands for the n -th generation of \mathcal{T}_a), where the labels are chosen in an arbitrary fashion.

On H_n^1 the walker starting from a will have a bias towards the set of leaves, which can be considered as the center of mass of these graph, since it contains a positive proportion of the vertices. The parameter L here is present only to make the walk slower (the expected number of steps to cross an L -path is $2L^2$, i.e. if $v \in H_n^1$ is either the root a or a vertex of degree 3 adjacent to three degree 2 vertices $\mathbb{E}_v[\inf\{t : D(X_t, v) = L\}] = 2L^2$ where D denotes the graph distance). This shall assist us in verifying that the worst-case total-variation and separation distances are obtained by vertices which are far away from the center of mass.

The problem of this construction is that seen from a vertex which is not a the set of leaves is not balanced. To cope with this defect, we add n extra “generations” of vertices, which make the center of mass balanced from “many” starting points.

Step 2: For all $1 \leq m \leq n$ we label the vertices of the “ $n + m$ -th generation” (they are at distance $(L + 1)n/2 + m$ from a) as follows

$$\mathcal{L}_{n+m} := \{u_{i_1, \dots, i_m}^k : i_1, \dots, i_m \in [4], k \in [2^{n-m}]\}$$

and we connect them to generation $n + m - 1$ using the following scheme: for all $k \in [2^{n-m}]$ $u_{i_1, \dots, i_{m-1}, 1}^k, u_{i_1, \dots, i_{m-1}, 2}^k, u_{i_1, \dots, i_{m-1}, 3}^k, u_{i_1, \dots, i_{m-1}, 4}^k$ are connected to $u_{i_1, \dots, i_{m-1}}^{2k-1}$ and $u_{i_1, \dots, i_{m-1}}^{2k}$. We call the obtained graph H_n^2 . The “center of mass” of H_n^2 is the set \mathcal{L}_{2n} (it bears roughly half of the total mass of H_n^2), which is balanced seen from any vertex in H_n^1 .

Step 3.1 and 3.2: We now want to plug (attach) to the leaf set of H_n^1 “two paths” with different speeds (to have something similar to the structures present in Examples 4.2 and 5.1). The construction is the following (see Figure 7):

- (i) We start with a rooted binary tree T of depth n (assume $n \geq 4$). And let us call 1 and 2 the two neighbors of the root and T_1 and T_2 the subtrees rooted at 1 and 2, respectively.
- (ii) In T_1 we add edges between any pair of vertices which have a common ancestor and are not leaves.
- (iii) Finally we assign labels to the leaf sets of T_1 and T_2 in a way that the two labeled trees (prior to step (ii) that is) are isomorphic (see e.g. Figure 7) and we merge each leaf of T_1 to the leaf of T_2 with the same label. We let \mathcal{T}_n denote the obtained graph.
- (iv) We let \mathcal{T}'_n denote the graph which is obtained by the same construction, in which we also add edges within T_2 in step (ii) using the same role as for T_1 (see Figure 7).

To each vertex $v \in \mathcal{L}_{2n}$, we glue a copy of \mathcal{T}_n (v is merged with the root of \mathcal{T}_n and we obtain $H_n^{3,1}$). If we glue a copy of \mathcal{T}'_n (to each $v \in \mathcal{L}_{2n}$,) instead of \mathcal{T}_n , we obtain $H_n^{3,2}$. For both graphs we call \mathcal{L}_{3n} the set of vertices at distance $(L + 5)n/2$ (i.e. maximal distance) from a .

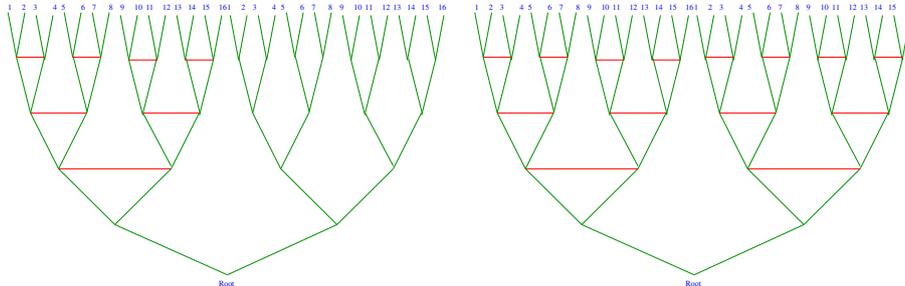


Figure 7: Representations \mathcal{T}_n (on the left) and \mathcal{T}'_n (on the right) for $n = 4$. The red edges are those added in step (ii). On step (iv) leaves with the same label are merged.

Finally we want to link together all the vertices of \mathcal{L}_{3n} in order to avoid dead-ends in the graph. We choose to link them together using an explicit expander (see e.g. [1, 20] for examples of explicit construction of expanders) so that (total-variation) mixing occurs rapidly once \mathcal{L}_{3n} is reached.

Step 4: We let $F_n = (V_n, E_n)$ be a family of explicit 3-regular c -lazy expanders with $V_n = [2^{3n-1}]$. We glue together G_n and $H_n^{3,i}$ ($i = 1, 2$) without adding vertices by identifying V_n with \mathcal{L}_{3n-1} . More precisely, we start with a copy of $H_n^{3,i}$ with root a . We label the vertices of \mathcal{L}_{3n} by $z_1, \dots, z_{2^{3n-1}}$ (the labeling is arbitrary). We then connect z_i with z_j if and only if $\{i, j\} \in E_n$. We call the final result of our construction $H_n^{4,i}$ ($i = 1, 2$). We call a the root of $H_n^{4,i}$ ($i = 1, 2$).

With some efforts and using the tools developed in the following sections, the reader can check that the lazy SRW on $H_n^{4,1}$ exhibits pre-cutoff but not cutoff in total-variation. This is a SRW version of Aldous' counter-example.

6.3 A sequence of lazy SRW on bounded degree expanders with total-variation cutoff and no separation cutoff

The following is a modification of Example 4.2 into a sequence of lazy SRWs on a sequence of bounded degree graphs.

Example 6.9. Take a copy of $H_n^{3,1}$ with root b and a copy of $H_n^{3,2}$ with root a . We glue together the two by merging the vertices of \mathcal{L}_{3n} (of both graphs): we give labels $z_1, \dots, z_{2^{3n-1}}$ to the vertices lying in \mathcal{L}_{3n} of each of the two graphs, and then merge each pair of vertices who share the same label. Finally, we build extra-connections between $z_1, \dots, z_{2^{3n-1}}$ using an expander graph F_n with 2^{3n-1} vertices, like in Step 4. We let $G_n^1 := (V_n^1, E_n^1)$ denote the obtained graph.

In order to apply Propositions 6.4 and 6.5, we need to identify which vertices and sets will play which role.

- The center of mass Z_n is given by the 2^{3n-1} vertices which are linked by the expander.
- a is the vertex which maximizes (stochastically) the hitting time of Z_n .
- The pair of vertices (x, y) which (up to negligible terms) attains the minimum for $P_n^t(x, y)/\pi_n(y)$ (for all $t \geq 0$) is given by (a, b) .

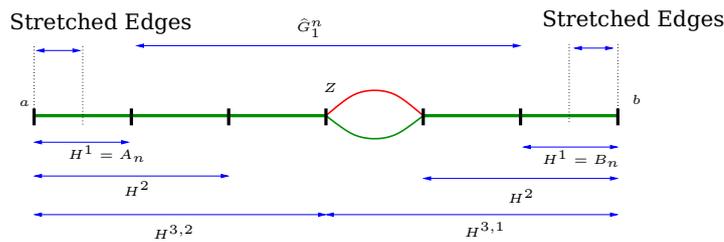


Figure 8: Schematic representation of Example 6.9. In the construction of $H_n^{3,1}$, the asymmetry of \mathcal{T}_n produces two different paths to reach the center of mass Z_n , with different speed. This produce an absence of concentration for the hitting time of Z starting from b .

- The sets A_n and B_n are chosen to be the largest set of points around a and b (resp.) such that Z_n is balanced seen from $I_n := A_n \cup B_n$. Namely, these are the vertices within respective distance $(L + 1)n/2$ from a and b (the vertices of H_n^0 in both $H_n^{3,1}$ and $H_n^{3,2}$). Indeed, due to step 2 of the construction, the set \mathcal{L}_{2n} of $H_n^{3,1}$, respectively, $H_n^{3,2}$ (i.e. the collection of vertices whose distance from a (resp. b) is $(L + 3)n/2$) is balanced seen from A_n , resp. B_n . This implies that the distribution of $X_{T_{Z_n}}$ is uniform on Z_n . Step (iv) of the construction of \mathcal{T}_n is there to guaranty that T_{Z_n} and $X_{T_{Z_n}}$ are independent (and hence that Z_n is balanced seen from A_n and B_n).

It is then not difficult to check (cf. Figure 8) from the construction that assumptions (i) – (iii) resp. (i) – (v) of Propositions 6.4 and 6.5, are satisfied.

Moreover, the hitting time of Z_n from a is concentrated around $(17 + 3L^2)n$, while from b it satisfies that

$$\lim_{n \rightarrow \infty} \text{P}_b [T_{Z_n} \geq sn] = \begin{cases} 1 & \text{if } s < 15 + 3L^2, \\ 1/2 & \text{if } s \in (15 + 3L^2, 17 + 3L^2), \\ 0 & \text{if } s > 17 + 3L^2. \end{cases} \quad (6.11)$$

We want to prove that the system displays cutoff in total-variation around time $(17+3L^2)n$, and that the asymptotic behavior for the separation distance is given by

$$\lim_{n \rightarrow \infty} d_{\text{sep}}^{(n)}(sn) = \lim_{n \rightarrow \infty} \text{P}_b [T_{Z_n} \geq (s - 17 + 3L^2)n] = \begin{cases} 1 & \text{if } s < 32 + 6L^2, \\ 1/2 & \text{if } (32 + 6L^2, 34 + 6L^2), \\ 0 & \text{if } s > 34 + 6L^2. \end{cases} \quad (6.12)$$

The only thing we have to do to prove these statements is to verify condition (iv) in Proposition 6.4 and condition (vi) of Proposition 6.5 (resp.). The only delicate point is to show that for starting points outside of I_n the walk mixes rapidly. I.e. that there exists an absolute constant $C > 0$, which does not depend on L , such that

$$\lim_{n \rightarrow \infty} \max_{v \notin I_n} d_v^{(n)}(\lceil Cn \rceil) = 0. \quad (6.13)$$

Before proving (6.13) let us explain how we use it to verify the remaining conditions. Note that if L is chosen to be sufficiently large (i.e. such that $(17 + 3L^2) > C$) then (6.13) implies condition (iv) of Proposition 6.4.

For condition (vi) of Proposition 6.5, for the case $x \in \Omega_n, y \notin I_n$, we use Lemma 6.8 and the total-variation cutoff result to show that for $t \geq (18 + 3L^2 + C)$

$$P_n^t(x, y)/\pi_n(y) \geq 1 - 2 \left(d_x^{(n)}(18 + 3L^2) + d_y^{(n)}(Cn) \right), \tag{6.14}$$

which is uniformly close to one.

This yields the right condition provided $32 + 6L^2 > 18 + 3L^2 + C$ (which can obviously be fulfilled by picking L to be sufficiently large). We now treat the case where both x and y lie in A_n (whose analysis does not rely on (6.13)). We use Lemma 6.3 with $Z = Z'_n$ chosen to be the set of vertices within distance $(L + 3)/2n$ from a (corresponding to \mathcal{L}_{2n} in the copy of $H_n^{3,2}$). Recall that by construction this set is balanced seen from A_n . By (6.3) we have that

$$P_n^t(x, y)/\pi_n(y) \geq \mathbb{P} \left[T_{Z'_n}^{x,y} \leq t \right]. \tag{6.15}$$

Moreover, for any $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \max_{x, y \in A_n} \mathbb{P} \left[T_{Z'_n}^{x,y} \leq (6L^2 + 18 + \varepsilon)n \right] = 1 \tag{6.16}$$

and this suffices to conclude that condition (vi) of Proposition 6.5 indeed holds.

Now let us prove (6.13). We want to use a simple ℓ_2 bound using the Poincaré inequality (see Lemma A.1). The issue is that the spectral gap of our graph is rather small (of order L^{-2}) due to the presence of stretched edges. However starting outside of I_n the walk has a very small chance to visit the part of the graph where the edges are stretched, before the walk is already extremely mixed. Hence our idea is to apply the ℓ_2 bound for the walk on a smaller graph which corresponds to the vertices which are likely to be visited. This graph will have no stretched edges and a spectral gap which is bounded away from zero and does not depend on L .

We let $\widehat{G}_n^1 = (\widehat{V}_n, \widehat{E}_n)$ denote the graph which is obtained from G_n^1 when all the vertices within distance $Ln/2 + 1$ from a and b have been deleted, together with all edges connected to them. First we observe that the Cheeger constant associated to \widehat{G}_n^1 is large (i.e. it is bounded from below by some positive absolute constant, which is independent also of L), see e.g. Lemma 2.1 in [18] for a proof.

Proposition 6.10. *Let $\kappa := (\min(c/3, 1/18))^2/2$. Then*

$$\text{ch}_{\text{Lazy}}(\widehat{G}_n^1) \geq \sqrt{2\kappa}. \tag{6.17}$$

Consequently, the relaxation-time of the lazy SRW on $\widehat{G}_n^1, \widehat{t}_{\text{rel}}^{(n)}$, satisfies

$$\widehat{t}_{\text{rel}}^{(n)} \leq \kappa^{-1} \tag{6.18}$$

If we let \widehat{P}_x^t and $\widehat{\pi}_n$ refer to the distribution at time t and at equilibrium for the walk on \widehat{G}_n^1 , this implies (by Lemma A.1) that for $x \in \widehat{V}_n^1$, for all $t \geq n\kappa^{-1} \log 9$.

$$\|\widehat{P}_x^t - \widehat{\pi}_n\|_{\text{TV}} \leq \frac{1}{\min_y \widehat{\pi}_n(y)} e^{-\kappa t} \leq \left(\max_{v \in \widehat{V}_n} \deg v \right) |\widehat{V}_n| 9^{-n} \leq 6(8/9)^n. \tag{6.19}$$

What remains to be proven is that if one considers \widehat{V}_n^1 as a subset of V_n^1 , then for any $x \in V_n^1 \setminus I_n$, the distances $\|\widehat{P}_x^t - \widehat{\pi}_n\|_{\text{TV}}$ and $\|P_x^t - \pi_n\|_{\text{TV}}$ are very close. Note that

$$\|P_x^t - \pi_n\|_{\text{TV}} \leq \|P_x^t - \widehat{P}_x^t\|_{\text{TV}} + \|\widehat{P}_x^t - \widehat{\pi}_n\|_{\text{TV}} + \|\pi_n - \widehat{\pi}_n\|_{\text{TV}}, \tag{6.20}$$

The term $\|\pi_n - \widehat{\pi}_n\|_{\text{TV}}$ is exponentially small in n because only an exponentially small fraction of the vertices of G_n^1 lie outside of \widehat{G}_n^1 . Now if one lets $T_{\partial\widehat{V}_n^1}$ denote the hitting time of

$$\partial\widehat{V}_n^1 := \{x \in V_n^1 \setminus \widehat{V}_n^1 : \exists y \in \widehat{V}_n^1, x \sim y\},$$

(recall that V_n^1 is the vertex set of G_n^1) we have (by a standard coupling argument) that

$$\|P_x^t - \widehat{P}_x^t\|_{\text{TV}} \leq P_x[T_{\partial\widehat{V}_n^1} \leq t] \leq \sum_{i=1}^t P_x^i(\partial\widehat{V}_n^1). \quad (6.21)$$

Now if $x \in V_n^1 \setminus I_n$, it lies at distance of at least $n/2$ from $\partial\widehat{V}_n^1$ and has to overcome a drift to reach it. For this reason it should take time which is exponentially large in n . More rigorously, we let Ω_x be the set of vertices $y \in V_n^1$ such that there exists a graph automorphism of G_n^1 preserving a and b which maps x to y (in most cases it is just a pedantic manner to describe the set of points at a fixed distance from a , but we have to introduce this definition due to the lack of symmetry of the b -side). Note that $|\Omega_x|/|\partial\widehat{V}_n^1| \geq 2^{n/2}$ if $x \notin I_n$. Hence we have for all $i > 0$ and $x \notin I_n$ that

$$P_x^i(\partial\widehat{V}_n^1) = \frac{\sum_{y \in \Omega_x} \pi_n(y) P_y^i(\partial\widehat{V}_n^1)}{\pi_n(\Omega_x)} \leq \frac{\pi_n(\partial\widehat{V}_n^1)}{\pi_n(\Omega_x)} \leq \max_{v \in V_n^1} \deg(v) \frac{|\partial\widehat{V}_n^1|}{|\Omega_x|} \leq \frac{6}{2^{n/2}}. \quad (6.22)$$

where in the first inequality we have used the stationarity of π_n ,

$$\sum_{y \in V_n^1} \pi_n(y) P_y^i(\partial\widehat{V}_n^1) = \pi_n(\partial\widehat{V}_n^1).$$

Plugging this in (6.21) we obtain (6.13) or more precisely:

Corollary 6.11. Set $t_n := \lceil n\kappa^{-1} \log 9 \rceil$. Then

$$\lim_{n \rightarrow \infty} \max_{x \in V_n^1 \setminus I_n} \|P_x^{t_n} - \pi_n\|_{\text{TV}} = 0. \quad (6.23)$$

6.4 A sequence of lazy SRW on bounded degree expanders with separation cutoff and no total-variation cutoff

The following is a modification of Example 5.1 into a sequence of lazy SRWs on a sequence of bounded degree graphs.

Example 6.12. Take a copy of $H_n^{4,1}$ with root a and a copy of H_n^1 with root b . We glue them together as follows: we give labels in $[2^{2n}]$ to the vertices in \mathcal{L}_{2n} in the two graphs and merge the vertices which share the same labels. We denote the set of merged vertices by Z'_n (this is the set of vertices of distance $(L+3)n/2$ from a and b). Let G_n^2 denote the obtained graph.

The reader can easily check that here a and b play symmetric roles. We let A_n and B_n denote the vertices within distance $(L+1)n/2$ from a and b , respectively. Moreover,

- The center of mass Z_n is given by the 2^{3n-1} vertices which are linked by the expander (which are the vertices belonging to \mathcal{L}_{3n} of $H_n^{4,1}$).
- Z_n is balanced seen from $A_n \cup B_n$.
- a and b maximize (stochastically) the hitting time of Z_n .

It is then not difficult to check (see Fig.9) from the construction that assumptions (i)–(iii) Proposition 6.4 are satisfied. Assumption (iv) can be showed to be satisfied as in the previous example by using an ℓ_2 bound for the graph in which points within distance $Ln/2$ of a and b have been deleted.

The asymptotic behavior of the hitting time of Z_n from a (or b) is once again given by (6.11) and hence the system does not display cutoff in total-variation.

For cutoff in separation, we cannot use Proposition 6.5. We use instead Lemma 6.3, and the relevant set to hit is Z'_n . This set is balanced seen from $I_n := A_n \cup B_n$ and

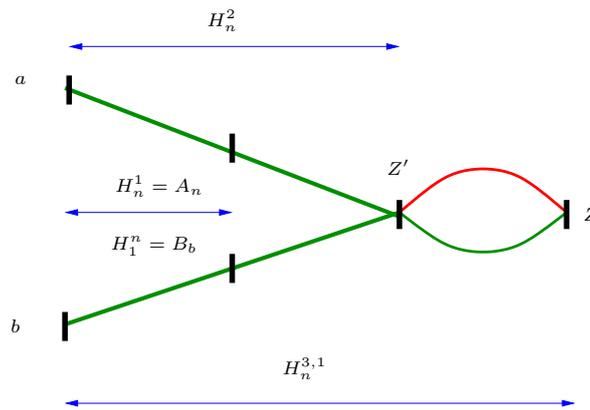


Figure 9: Schematic representation of Example 6.12. Here the asymmetry of \mathcal{T}_n is used to avoid cutoff in total-variation, in a similar manner to what is done in Example 5.1. However, as in Example 5.1, the separation mixing-time is determined by the behavior of $T_{Z'}^{a,b}$ in the large deviation regime. Note that Z' is a set of small equilibrium measure (it has 4^n vertices whereas the full graph has order 8^n vertices).

thus is the relevant one for the purpose of computing the separation mixing time. An analog of the analysis performed for Example 5.1, does the job. To control the quantity $P_n^t(x, y)/\pi_n(y)$ when one of x and y (or both) does not belong to $A_n \cup B_n$ we use an ℓ_2 estimate (in conjunction with Lemma 6.8) for the subgraph \hat{G}_n^2 obtained by deleting the stretched edges in G_n^2 , similarly to what we have done in the analysis of Example 6.9.

6.5 Proof of Remark 1.4

Part (i) follows from the analysis of Example 6.9. We shall prove now that part (ii) is satisfied by Example 6.12.

We denote by $\pi_{Z'}$ the distribution of π_n conditioned on Z' (suppressing the dependence on n). By (6.4) we have that for all t and every $x \in A_n$ and $y \in B_n$

$$P_n^t(x, y)/\pi_n(y) = \sum_{k \leq t} \mathbb{P}[T_{Z'}^{x,y} = k] \mathbb{P}_{\pi_{Z'}}^{t-k}(Z')/\pi_n(Z') \geq \mathbb{P}[T_{Z'}^{x,y} = t]/\pi_n(Z'). \quad (6.24)$$

We know from the previous analysis of Example 6.12 that for the separation distance to equilibrium only $(x, y) \in A_n \times B_n$ matter, or more precisely

$$\lim_{n \rightarrow \infty} \sup_{t \geq 0} |d_{\text{sep}}^{(n)}(t) - \max(0, 1 - \min_{(x,y) \in A_n \times B_n} P_n^t(x, y)/\pi_n(y))| = 0. \quad (6.25)$$

Hence setting

$$t_\eta^n(x, y) := \min\{t : P_n^t(x, y)/\pi_n(y) \geq 1 - \eta\}$$

we prove that cutoff window is constant by proving that, for all $\varepsilon > 0$, there exist some $n_\varepsilon \in \mathbb{N}$ and some absolute constant C_2 such that for all $n \geq n_\varepsilon$ and all $(x, y) \in A_n \times B_n$

$$t_\varepsilon^n(x, y) - t_{1-\varepsilon}^n(x, y) \leq C_2 |\log \varepsilon|. \quad (6.26)$$

$$\forall t \geq t_\varepsilon^n(x, y), \quad P_n^t(x, y)/\pi_n(y) \geq 1 - \varepsilon. \quad (6.27)$$

In what follows for simplicity we drop the dependence in n in the notation $t_\eta(x, y)$. Although this is not used in the analysis below (and hence not proven), we can identify $t_{1/4}(x, y)$ for all $(x, y) \in A_n \times B_n$ as follows:

$$\max\{|t_{1/4}(x, y) - t'(x, y)|, |t_{1/4}(x, y) - \bar{t}(x, y)|\} \leq C_3,$$

where $t'(x, y) := \inf\{t : \mathbb{P}[T_{Z'}^{x,y} \leq t] \geq \pi_n(Z')\}$ and $\bar{t}(x, y) := \inf\{t : \mathbb{P}[T_{Z'}^{x,y} = t] \geq \pi_n(Z')\}$. This follows from the analysis below, together with (6.24) and the exponential decay of $\mathbb{P}_{\pi_{Z'}}^t(Z') - \pi_n(Z')$ as a function of t .

We start by presenting some general machinery which we shall utilize in the proof of Remark 1.4. Let μ be a distribution over \mathbb{Z} . We say that μ is **Unimodal** if for any $z_1 \leq z_2 \leq z_*$ and for any $z_1 \geq z_2 \geq z_*$ we have that $\mu(z_1) \leq \mu(z_2)$, where z_* is the mode of μ (i.e. $\max_z \mu(z) = \mu(z_*)$). We say that μ is **Log-Concave** if $\mu^2(z) \geq \mu(z-1)\mu(z+1)$ for all $z \in \mathbb{Z}$ (equivalently, for all $z_1 < z_2$ ($z_1, z_2 \in \mathbb{Z}$) we have that $\frac{\mu(z_1+1)}{\mu(z_1)} \geq \frac{\mu(z_2+1)}{\mu(z_2)}$, where $0/0$ is interpreted as 0).

Fact 6.13. *Let μ be a log-concave distribution over \mathbb{Z} . Then μ is unimodal.*

Fact 6.14. *The family of Geometric distributions is log-concave.*

Fact 6.15. *The family of log-concave distributions over \mathbb{Z} is closed under convolutions.*

The following representation of hitting times in birth and death chains is due to Karlin and McGregor [13, Equation (45)]. It was later rediscovered by Keilson [14]. The discrete time case of this result was given by Fill [11, Theorem 1.2].

Theorem 6.16. *Let $([n], P, \pi)$ be a lazy birth and death chain (where $[n] := \{1, 2, \dots, n\}$). Let P' be defined by $P'(i, \cdot) = P(i, \cdot)$ if $i \in [n-1]$ and $P'(n, n) = 1$. Denote the non-zero eigenvalues of $I - P'$ by $0 < \beta_1 \leq \dots \leq \beta_{n-1} \leq 1$. Let ξ_1, \dots, ξ_{n-1} be independent random variables such that $\xi_i \sim \text{Geom}(\beta_i)$ for all $i \in [n-1]$. Then the distribution of T_n under \mathbb{P}_1 is the same as the distribution of $\sum_{i \in [n-1]} \xi_i$.*

We are now ready to prove (6.26) and (6.27). For clarity of exposition, we first expose our analysis for the special case $x = a, y = b$. Consider the sequence of graphs G_n^2 from Example 6.12. Let G_n^3 the subgraph of G_n^2 whose set of vertices is given by

$$V_n^3 := \{v : \text{dist}(v, \{a, b\}) \leq (L+3)n/2\},$$

and whose edges are those of E_n^2 for which both ends are in G_n^3 (Note that this graph is connected and includes Z' but not any point further away from $\{a, b\}$)

Let $(Y_t)_{t \in \mathbb{Z}_+}$ be lazy SRW on G_n^3 . Consider the projection $\bar{Y}_t := 1 + \text{dist}(Y_t, \{a, b\})$. Our construction implies that the projection is Markovian and thus $(\bar{Y}_t)_{t \in \mathbb{Z}_+}$ is a lazy birth and death chain on $[1 + (L+3)n/2]$. Consequently, by Theorem 6.16 and Facts 6.14-6.15, the law of $T_{Z'}^{a,b}$, which is a sum of independent hitting time and thus of geometric variables, is log-concave. For any $v \in V_n^3$ the distribution of $T_{Z'}$, given that $Y_0 = v$, is the same as that of $T_{1+(L+3)n/2}$ (for the chain (\bar{Y}_t)), given that $\bar{Y}_0 = 1 + \text{dist}(v, \{a, b\})$. Consequently, by Theorem 6.16 and Facts 6.14-6.15, the law of $T_{Z'}^{a,b}$ is log-concave. Let z_* be the mode of $T_{Z'}^{a,b}$. A standard computation is sufficient to show that

$$|z_* - \mathbb{E}[T_{Z'}^{a,b}]| \leq C_4 \sqrt{\text{Var}(T_{Z'}^{a,b})} \leq C_5 \sqrt{n}, \tag{6.28}$$

(in fact, the first inequality follows from unimodality).

Fix some $\delta > 0$ sufficiently small such that $\mathbb{P}[T_{Z'}^{a,b} \leq z_* - \delta n] \gg 2^{-n}$ (2^{-n} is the order of magnitude of $\pi_n(Z')$). By a large-deviation estimate and log-concavity there is some $\alpha > 1$ such that for all sufficiently large n we have that

$$\alpha^{\delta n} \leq \frac{\mathbb{P}[T_{Z'}^{a,b} = z_*]}{\mathbb{P}[T_{Z'}^{a,b} = z_* - \lfloor \delta n \rfloor]} \leq \left(\frac{\mathbb{P}[T_{Z'}^{a,b} = z_* - \lfloor \delta n \rfloor + 1]}{\mathbb{P}[T_{Z'}^{a,b} = z_* - \lfloor \delta n \rfloor]} \right)^{\lfloor \delta n \rfloor}$$

hence, again by log-concavity,

$$\forall t \leq z_* - \delta n, \quad \frac{\mathbb{P}[T_{Z'}^{a,b} = t + 1]}{\mathbb{P}[T_{Z'}^{a,b} = t]} > \alpha. \tag{6.29}$$

Consequently, by (6.24)

$$\forall t \leq z_* - \delta n, \quad \frac{P_n^{t+1}(a, b)}{\pi_n(b)} \geq \alpha \sum_{k \leq t+1} \mathbb{P}[T_{Z'}^{x,y} = k - 1] \frac{P_n^{t+1-k}(Z')}{\pi_n(Z')} = \alpha \frac{P_n^t(a, b)}{\pi_n(b)}. \quad (6.30)$$

As $T_{Z'}^{a,b}$ is log-concave and hence by Fact 6.13 also unimodal, (6.24) also yields that

$$\forall t \in [z_* - \delta n, z_*), \quad \frac{P_n^t(a, b)}{\pi_n(b)} \leq \sum_{k \leq t} \mathbb{P}[T_{Z'}^{x,y} = k + 1] \frac{P_n^{t-k}(Z')}{\pi_n(Z')} \leq \frac{P_n^{t+1}(a, b)}{\pi_n(b)}, \quad (6.31)$$

and that there exist some absolute constants $c, C_6 > 0, \beta \in (1, 2)$ such that

$$\forall t \in [z_*, z_* + n^{2/3}), \quad \frac{P_n^t(a, b)}{\pi_n(b)} \geq \frac{\mathbb{P}[T_{Z'}^{a,b} = z_* + \lceil n^{2/3} \rceil]}{\pi_n(Z')} \geq c\beta^n. \quad (6.32)$$

$$\forall t \geq z_* + n^{2/3}, \quad 1 - \frac{P_n^t(a, b)}{\pi_n(b)} \leq \mathbb{P}[T_{Z'}^{a,b} > t] \leq C_6/n^{1/3}. \quad (6.33)$$

This concludes the proof of the case $(x, y) = (a, b)$ as (6.30) implies (6.26) with $C_2 := (\log \alpha)^{-1}$ and (6.27) can be deduced from the four other equations. For general $(x, y) \in A_n \times B_n$ we decompose $T_{Z'}^{x,y}$ into a convolution of a log-concave distribution and some other negligible term. Let $(X_t^x)_t$ and $(X_t^y)_t$ be independent realizations of the random walk, started from respective initial vertex x and y , defined on the same probability space. Let $T_{Z'}^x := \inf\{t : X_t^x \in Z'\}$ and $T_{Z'}^y := \inf\{t : X_t^y \in Z'\}$. We define T'_x (and T'_y in an analogous manner, using (X_t^y) and $T_{Z'}^y$) as follows (with the convention $\sup \emptyset = 0$)

$$T'_x := \sup\{t : t < T_{Z'}^x, \text{dist}(X_{t-1}^x, Z') = \text{dist}(x, Z') + 1\}$$

Note that $T'_x, T_{Z'}^x - T'_x, T'_y$ and $T_{Z'}^y, -T'_y$ are independent. We denote $T_1 = T_1(x, y) := (T_{Z'}^x - T'_x) + (T_{Z'}^y - T'_y)$ and $T_2 = T_2(x, y) := T'_x + T'_y$. By Theorem 6.16 and Facts 6.14-6.15 the laws of $T_{Z'}^x - T'_x$ and $T_{Z'}^y - T'_y$ are log-concave (by a similar argument to the one used before using a projection to a birth and death chain), and so T_1 is also log-concave (by Fact 6.15). Observe that $T_1 + T_2$ has the same law as $T_{Z'}^{x,y}$.

Denote the mode of T_1 by $z_* = z_*(x, y)$. Fix some $\delta > 0$ sufficiently small such that $\min_{(x,y) \in A_n \times B_n} \mathbb{P}[T_1(x, y) \leq z_*(x, y) - \delta n] \gg 2^{-n}$. Imitating the proof of (6.30), using a large-deviation estimate on $\frac{\mathbb{P}[T_1(x,y) = z_*(x,y)]}{\mathbb{P}[T_1(x,y) = z_*(x,y) - \lceil \delta n \rceil]}$ which is uniform in (x, y) (the existence of such a uniform large-deviation estimate follows from the analysis of Example 5.1, or alternatively, by [5, Lemma 6.2]), together with log-concavity, we get that if $\alpha > 1$ is chosen sufficiently small, then (6.29) remains valid simultaneously for all choices of x, y , if one replaces $T_{Z'}^{a,b}$ by $T_1(x, y)$ (and z_* with $z_*(x, y)$). We argue that (6.28)-(6.33) can be extended (excluding the middle terms) to all $(x, y) \in A_n \times B_n$ (in the role of (a, b)), with the same choice of constants for all $(x, y) \in A_n \times B_n$. To extend (6.30) and (6.31), note that after conditioning on T_2 we can imitate the above proofs and so the extensions are obtained by averaging over T_2 . For (6.32), note that by unimodality

$$\mathbb{P}[T_{Z'}^{x,y} = z_*(x, y) + \lceil n^{2/3} \rceil] / \pi_n(Z') \geq c_1 2^n \mathbb{P}[T_2(x, y) \leq \lceil n^{2/3} \rceil] \mathbb{P}[T_1(x, y) = z_*(x, y) + \lceil n^{2/3} \rceil].$$

It is not hard to show that there exists some $\gamma < 2$ and $c_2, C_6 > 0$ such that

$$\mathbb{P}[T_1(x, y) = z_*(x, y) + \lceil n^{2/3} \rceil] \geq c_2 \gamma^{-n} \text{ and } \mathbb{P}[T_2(x, y) \leq \lceil n^{2/3} \rceil] \geq 1 - C_6 n^{-2/3}.$$

for all $(x, y) \in A_n \times B_n$ (by Markov inequality and the fact that $\max_{(x,y) \in A_n \times B_n} \mathbb{E}[T_2(x, y)] = O(1)$). For (6.28) use unimodality (first inequality) to show that for all $(x, y) \in A_n \times B_n$

$$|z_*(x, y) - \mathbb{E}[T_1(x, y)]| \leq C_4 \sqrt{\text{Var}(T_1(x, y))} \leq C_4 \sqrt{\text{Var}(T_{Z'}^{a,b})} \leq C_5 \sqrt{n}.$$

Lastly, for (6.33) use (6.28) and Chebyshev's inequality (by noting that $|z_*(x, y) - \mathbb{E}[T_{Z'}^{x,y}]| \leq |z_*(x, y) - \mathbb{E}[T_1(x, y)]| + \mathbb{E}[T_2(x, y)] \leq C_7 \sqrt{n}$). We leave the details to the reader. \square

A Proof of technical results

A.1 Basic ingredients

In our analysis we use various times raw ℓ_2 bounds in order to get estimates on total-variation distance. We cite here a standard result (see e.g. [16, Lemma 12.16]).

Lemma A.1. *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*

$$2\|P_\mu^t - \pi\|_{\text{TV}} \leq \|P_\mu^t - \pi\|_{2,\pi} \leq \lambda_2^t \|\mu - \pi\|_{2,\pi}, \text{ for all } t \geq 0. \quad (\text{A.1})$$

Lemma A.2 (Hitting time from stationary tail estimates). *Let (Ω, P, π) be a finite irreducible reversible Markov chain. Let $A \subset \Omega$. Then for any $t \geq 0$ we have that*

$$P_\pi[T_A > t] \leq (1 - \pi(A)) \exp(-t\pi(A)/t_{\text{rel}}). \quad (\text{A.2})$$

For a proof see [5, Lemma 3.5] (or Proposition 3.21 in conjunction with Theorem 3.33 and Corollary 3.34 in [2]).

A.2 Proof of (1.8)

We can assume that the chain displays pre-cutoff in separation as if not, there is nothing to prove. We know from (1.5) that for every $\varepsilon > 0$, $t_{\text{mix}}^{(n)}(\varepsilon) \leq t_{\text{sep}}^{(n)}(\varepsilon)$. Hence, in our case, it is sufficient to prove that

$$\limsup_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{t_{\text{sep}}^{(n)}(1 - \varepsilon)}{t_{\text{mix}}^{(n)}(1 - 2\sqrt{\varepsilon})} \leq 2, \quad (\text{A.3})$$

as pre-cutoff implies that $\frac{t_{\text{sep}}^{(n)}(1 - 2\sqrt{\varepsilon})}{t_{\text{sep}}^{(n)}(1 - \varepsilon)}$ tends to 1 when n goes to infinity and ε goes to 0 in this order.

We shall show that for all n

$$t_{\text{sep}}^{(n)}(1 - \varepsilon) \leq 2t_{\text{mix}}^{(n)}(1 - 2\sqrt{\varepsilon}) + 2t_{\text{rel}}^{(n)} \log \varepsilon^{-1}. \quad (\text{A.4})$$

This is sufficient to conclude as Fact 3.2 ascertains that the second term is small in comparison to the first.

To prove (A.4), let us introduce the following alternative way of measuring the distance to equilibrium,

$$\bar{d}(t) := \max_{x,y \in \Omega} \|P_x(X_t \in \cdot) - P_y(X_t \in \cdot)\|_{\text{TV}}. \quad (\text{A.5})$$

From [16, Lemma 19.3] (or (6.9)) we know that for every t we have

$$d_{t_{\text{sep}}}(t) \leq 1 - (1 - \bar{d}(t/2))^2. \quad (\text{A.6})$$

Hence if one defines $\bar{t}_{\text{mix}}(\varepsilon)$ to be the first time at which $\bar{d}(t) \leq \varepsilon$, we have for every $\varepsilon > 0$,

$$t_{\text{sep}}(1 - \varepsilon) \leq 2\bar{t}_{\text{mix}}(1 - \sqrt{\varepsilon}). \quad (\text{A.7})$$

Now to conclude the proof of (A.3) we need to show that (under reversibility)

$$\bar{t}_{\text{mix}}(1 - \sqrt{\varepsilon}) \leq t_{\text{mix}}(1 - 2\sqrt{\varepsilon}) + t_{\text{rel}} \log \varepsilon^{-1}. \quad (\text{A.8})$$

Let us set

$$t = t_{\text{mix}}(1 - 2\sqrt{\varepsilon}) \quad \text{and} \quad s = t_{\text{rel}} \log \varepsilon^{-1}. \quad (\text{A.9})$$

For $x \in \Omega$, we set $\mu^x := P_x(X_t \in \cdot)$ and

$$\mu_1^x(z) := \frac{\min(\mu^x(z), \pi(z))}{1 - \|\mu^x - \pi\|_{\text{TV}}} \text{ and } \mu_2^x(z) := \frac{(\mu^x(z) - \pi(z))\mathbf{1}_{\{\mu^x(z) > \pi(z)\}}}{\|\mu^x - \pi\|_{\text{TV}}}. \quad (\text{A.10})$$

Both are probability measures and μ^x can be written as a linear combination of the two

$$\mu^x = (1 - \|\mu^x - \pi\|_{\text{TV}})\mu_1^x + \|\mu^x - \pi\|_{\text{TV}}\mu_2^x.$$

For μ , a distribution on Ω , $f, g \in \mathbb{R}^\Omega$ and $1 \leq p < \infty$, we introduce the notation

$$\mu P^s(x) := \sum_{y \in \Omega} \mu(y) P^s(y, x), \quad P^s f(x) = \sum_{y \in \Omega} P^s(x, y) f(y), \quad \|g\|_p = \left(\sum_y \pi(y) |g(y)|^p \right)^{1/p}.$$

Assuming that x and y maximize the l.h.s. in (A.5) at time $t + s$ and that

$$\|\mu^x - \pi\|_{\text{TV}} \geq \|\mu^y - \pi\|_{\text{TV}}$$

then by the triangular inequality we have that

$$\begin{aligned} \bar{d}(t + s) &= \|\mu^x P^s - \mu^y P^s\|_{\text{TV}} \\ &\leq (1 - \|\mu^x - \pi\|_{\text{TV}}) \|\mu_1^x P^s - \mu_1^y P^s\|_{\text{TV}} + \|\mu^x - \pi\|_{\text{TV}} \|\mu_2^x P^s - \tilde{\mu}_2^y P^s\|_{\text{TV}}, \end{aligned} \quad (\text{A.11})$$

where $\tilde{\mu}_2^y$ is an adequate linear combination of μ_1^y and μ_2^y . Using the definition of t we have $\|\mu^x - \pi\|_{\text{TV}} \leq 1 - \sqrt{\varepsilon}$, and hence we have

$$\bar{d}(t + s) \leq 1 - 2\sqrt{\varepsilon} + \|\mu_1^x P^s - \mu_1^y P^s\|_{\text{TV}}, \quad (\text{A.12})$$

Now to estimate the second term we set $f(z) = \frac{\mu_1^x(z)}{\pi(z)} - \frac{\mu_1^y(z)}{\pi(z)}$. Observe that by reversibility $\mu_1^x P^s - \mu_1^y P^s = P^s f$. Note that by definition $|f(z)| \leq (1 - \|\mu^x - \pi\|_{\text{TV}})^{-1} \leq 2\varepsilon^{-1/2}$. Let $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{|\Omega|}$ be the eigenvalues of P and set $\lambda := \max(\lambda_2, |\lambda_{|\Omega|}|)$. Using the spectral-decomposition of f along with the fact that $\sum_z \pi(z) f(z) = 0$ (and finally, the choice of s) it is standard to show that $\|P^s f\|_2 \leq \lambda^s \|f\|_2 \leq \lambda^s (2\varepsilon^{-1/2}) \leq 2\sqrt{\varepsilon}$. Hence

$$\|\mu_1^x P^s - \mu_1^y P^s\|_{\text{TV}} = \frac{1}{2} \|P^s f\|_1 \leq \frac{1}{2} \|P^s f\|_2 \leq \sqrt{\varepsilon},$$

as desired. □

A.3 Proof of Lemma 6.3

By decomposing over the possible values of T_Z , using the assumption that Z is balanced seen from x and reversibility (which implies that $P_{\pi_Z}^s(y)/\pi(y) = P_y^s(Z)/\pi(Z)$, for all s), we get that

$$\begin{aligned} \frac{P^t(x, y)}{\pi(y)} &= \sum_{k_1 \leq t} P_x[T_Z = k_1] \frac{P_{\pi_Z}^{t-k_1}(y)}{\pi(y)} + \frac{P_x[X_t = y \text{ and } T_Z > t]}{\pi(y)} \\ &\geq \sum_{k_1 \leq t} P_x[T_Z = k_1] \frac{P_{\pi_Z}^{t-k_1}(y)}{\pi(y)} = \sum_{k_1 \leq t} P_x[T_Z = k_1] \frac{P_y^{t-k_1}(Z)}{\pi(Z)} \\ &= \sum_{\substack{0 \leq k_1 \leq t \\ 0 \leq k_2 \leq k_1}} P_x[T_Z = k_2] P_y[T_Z = k_1 - k_2] \frac{P_{\pi_Z}^{t-k_1}(Z)}{\pi(Z)} = \sum_{k \leq t} P[T_Z^{x,y} = k] \frac{P_{\pi_Z}^{t-k}(Z)}{\pi(Z)}. \end{aligned} \quad (\text{A.13})$$

In particular, we have equality if $P_x[T_Z < T_y] = 1$ (i.e. in case (ii)). To conclude, for (i) we note that for lazy irreducible reversible chains that $P_{\pi_Z}^s(Z) \geq \pi(Z)$, which can be

easily verified using the spectral decomposition and the non-negativity of the eigenvalues of P . For (ii) note that by (A.1)

$$P_{\pi_Z}^s(Z) - \pi(Z) \leq \|P_{\pi_Z}^s - \pi\|_{\text{TV}} \leq \frac{1}{2} \|\pi_Z - \pi\|_2 \lambda_2^s \leq \frac{\lambda_2^s}{2} \sqrt{(1 - \pi(Z))/\pi(Z)}.$$

The first inequality in (6.4) is obtained by plugging the last estimate in the second term of (6.4). For the second inequality in (6.4) it follows from the estimate

$$\sum_{k \leq t} \mathbb{P}[T_Z^{x,y} = k] \lambda_2^{t-k} \leq \left(\max_{k \geq 0} \mathbb{P}[T_Z^{x,y} = k] \right) \sum_{k \geq 0} \lambda_2^k = \max_k \mathbb{P}[T_Z^{x,y} = k] / (1 - \lambda_2). \quad \square$$

A.4 Proof of Proposition 6.4

The result is mostly a consequence of the following result which relates the mixing time starting from x to the hitting time of a set Z balanced seen from x .

Lemma A.3. *Let (Ω, P, π) be a finite lazy irreducible reversible Markov chain. Let $Z \subset \Omega$ (we denote its complement by Z^c) and $x \in \Omega$. Given $0 < \varepsilon < 1$. Set*

$$t_{x,Z}(p) := \min\{t : \mathbb{P}_x[T_Z > t] \leq p\},$$

$$s_\varepsilon := \lceil t_{\text{rel}} \log[\pi(Z^c)/\varepsilon] / \pi(Z) \rceil \text{ and } r_\varepsilon := \lceil t_{\text{rel}} \log[\pi(Z^c)/(\pi(Z)\varepsilon^2)] / 2 \rceil.$$

Let $s' := \max(t_{x,Z}(p) - s_\varepsilon, 0)$. Then we have

$$\|P_x^{s'} - \pi\|_{\text{TV}} > p - \varepsilon. \tag{A.14}$$

Moreover if Z is balanced seen from x then we also have that

$$\|P_x^{t_{x,Z}(p)+r_\varepsilon} - \pi\|_{\text{TV}} \leq p + \varepsilon. \tag{A.15}$$

Proof. The first result is proved by coupling the chain with initial distribution $P_x^{k-s_\varepsilon}$ with the stationary chain ($k \geq s_\varepsilon$ to be determine soon). We have

$$\mathbb{P}_x[T_Z \geq k] \leq \|P_x^{k-s_\varepsilon} - \pi\|_{\text{TV}} + \mathbb{P}_\pi[T_Z \geq s_\varepsilon] \leq \|P_x^{k-s_\varepsilon} - \pi\|_{\text{TV}} + \varepsilon. \tag{A.16}$$

where the last inequality is a consequence of (A.2) and the choice of s_ε . Setting $k = t_{x,Z}(p)$ we obtain the result (as if $s' = 0$ there is nothing to prove).

We now prove (A.15). By the assumption that Z is balanced seen from x , for all $\ell \leq t$

$$P_x^t = \mathbb{P}_x[T_Z > \ell] P_x[X_t \in \cdot \mid T_Z > \ell] + \sum_{0 \leq i < \ell} \mathbb{P}_x[T_Z = \ell - i] P_{\pi_Z}^{t-\ell+i}. \tag{A.17}$$

By the triangle inequality and the fact that the distance to π decreases in time, we obtain

$$\begin{aligned} \|P_x^t - \pi\|_{\text{TV}} &\leq \mathbb{P}_x[T_Z > \ell] + \sum_{0 \leq i < \ell} \mathbb{P}_x[T_Z = \ell - i] \|P_{\pi_Z}^{t-\ell+i} - \pi\|_{\text{TV}} \\ &\leq \mathbb{P}_x[T_Z > \ell] + \|P_{\pi_Z}^{t-\ell} - \pi\|_{\text{TV}} \end{aligned}$$

Using this inequality for $\ell := t_{x,Z}(p)$ (and so $t - \ell = r_\varepsilon$) we only have to show that $\|P_{\pi_Z}^{t-\ell} - \pi\|_{\text{TV}} = \|P_{\pi_Z}^{r_\varepsilon} - \pi\|_{\text{TV}} \leq \varepsilon$. Combining (A.1) with the definition of r_ε , we have that

$$\|P_{\pi_Z}^{r_\varepsilon} - \pi\|_{\text{TV}} \leq \lambda_2^{r_\varepsilon} \sqrt{\pi(Z^c)/\pi(Z)} \leq \varepsilon. \tag{A.18}$$

□

We can now proceed to the proof of Proposition 6.4. With our assumptions on t_{rel} and Z_n , Lemma A.3 allows us to show that mixing time starting from x and $t_{x,Z_n}(p)$ are equivalent when Z_n is balanced seen from x (i.e. for $x \in I_n$). Assumption (iv) ensures that what occurs for other initial conditions does not matter and Assumption (iii) establishes that a is the worst initial condition. □

A.5 Proof of Proposition 6.5

From Lemma 6.3 and assumptions (i), (iii) and (v) we know that $P_n^t(x, y)/\pi_n(y)$ and $\mathbb{P}[T_{Z_n}^{x, y} \leq t]$ differ only by a negligible amount, provided that $x \in A_n$ and $y \in B_n$. Assumption (iv) ensures then that

$$\liminf_{n \rightarrow \infty} \inf_{t \geq 0} \min_{(x, y) \in A_n \times B_n} \frac{P_n^t(x, y)}{\pi_n(y)} - \frac{P_n^t(a_n, b_n)}{\pi_n(b_n)} = 0. \quad (\text{A.19})$$

We are left checking the other cases. Assumption (vi) takes care of most of them, and leaves the case where $(x, y) \in B_n \times B_n$, for which Lemma 6.3 implies that $\mathbb{P}[T_{Z_n}^{x, y} \leq t]$ is a lower bound for $P_n^t(x, y)/\pi_n(y)$. Hence the conclusion follows by assumption (v) again. \square

A.6 A short alternative proof of Theorem C

We are going to show that there exists an absolute constant $c > 0$ such that for any lazy chain

$$t_{\text{mix}}(1/4) - t_{\text{mix}}(3/4) \geq c\sqrt{t_{\text{mix}}(3/4)}. \quad (\text{A.20})$$

Indeed set $t := t_{\text{mix}}(1/4)$ and $s := \lfloor c\sqrt{t} \rfloor$. A sample of the distribution of the lazy chain at time t can be generated by running the non-lazy version of the chain for ξ_t steps, where $\xi_t \sim \text{Bin}(t, 1/2)$ and is independent of the non-lazy version of the chain. By the triangle inequality we have (first inequality) and a standard coupling argument (second inequality)

$$\forall t, s \geq 0, \quad d(t) - d(t + s) \leq \max_{x \in \Omega} \|P_x^t - P_x^{t+s}\|_{\text{TV}} \leq \|\xi_t - \xi_{t+s}\|_{\text{TV}}.$$

Moreover, if c is chosen well, we have for every $t \geq 0$ that $\|\xi_t - \xi_{t+\lfloor c\sqrt{t} \rfloor}\|_{\text{TV}} \leq 1/2$. \square

References

- [1] M. Ajtai Recursive construction for 3-regular expanders. *Combinatorica* 14.4 (1994): 379–416. MR-1312869
- [2] D. Aldous, J. Fill. Reversible Markov chains and random walks on graphs. In preparation, <http://www.stat.berkeley.edu/~aldous/RWG/book.html>.
- [3] N. Alon, V. Milman. λ_1 isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory, Series B* 38.1 (1985): 73–88. MR-0782626
- [4] N. Alon. λ_1 Eigenvalues and expanders. *Combinatorica* 6.2 (1986): 83–96. MR-0875835
- [5] R. Basu, J. Hermon, Y. Peres. Characterization of cutoff for reversible Markov chains. preprint: <http://arxiv.org/abs/1409.3250>.
- [6] G. Y. Chen, L. Saloff-Coste. The cutoff phenomenon for ergodic Markov processes. *Electronic Journal of Probability* 13.3 (2008): 26–78. MR-2375599
- [7] G. Y. Chen, L. Saloff-Coste. Comparison of cutoffs between lazy walks and Markovian semi-groups. *Journal of Applied Probability* 50.4 (2013): 943–959. MR-3161366
- [8] A. Dembo, O. Zeitouni. Large Deviation Techniques and Application, 2nd Edition. *Stochastic Modelling and Applied Probability* 38, Springer. MR-2571413
- [9] P. Diaconis and Laurent Saloff-Coste. Separation cutoff for birth and death chains, *The Annals of Applied Probability* 16.4 (2006): 2098–2122. MR-2288715
- [10] J. Ding, E. Lubetzky, Y. Peres. Total variation cutoff in birth-and-death chains. *Probability theory and related fields* 146.1–2 (2010): 61–85. MR-2550359
- [11] J. A. Fill. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. *Journal of Theoretical Probability* 22.3 (2009): 543–557. MR-2530102
- [12] J. Hermon, A technical report on hitting times, mixing and cutoff arXiv:1501.01869 [math.PR].

- [13] S. Karlin and J. McGregor. Coincidence properties of birth and death processes. *Pacific J. Math.* 9 (1959): 1109–1140. MR-0114247
- [14] J. Keilson. Markov chain models, rarity and exponentiality. *Applied Mathematical Sciences*, vol. 28, Springer-Verlag, New York, 1979. MR-0528293
- [15] H. Lacoïn, *Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion*, (preprint) arXiv:1309.3873. MR-3474475
- [16] D. Levin, Y. Peres, E. Wilmer, Markov Chains and Mixing Times. *American Mathematical Society, Providence, RI, (2009)*. MR-2466937
- [17] E. Lubetzky, A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal* 153.3 (2010): 475–510. MR-2667423
- [18] E. Lubetzky, A. Sly. Explicit expanders with cutoff phenomena. *Electronic Journal Probability* 16.15 (2011): 419–435. MR-2774096
- [19] E. Lubetzky, A. Sly. Cutoff for the Ising model on the lattice, *Inventiones Mathematicae* 191 (2013), 719–755. MR-3020173
- [20] O. Reingold, S. Vadhan, A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of mathematics* (2002): 157–187. MR-1888797
- [21] A. Sinclair, M. Jerrum. Approximate counting, uniform generation and rapidly mixing Markov chains. *Information and Computation* 82.1 (1989): 93–133. MR-1003059