Characterization of cutoff for reversible Markov chains

Riddhipratim Basu *  Jonathan Hermon †  Yuval Peres ‡

Abstract

A sequence of Markov chains is said to exhibit (total variation) cutoff if the convergence to stationarity in total variation distance is abrupt. We consider reversible lazy chains. We prove a necessary and sufficient condition for the occurrence of the cutoff phenomena in terms of concentration of hitting time of “worst” (in some sense) sets of stationary measure at least $\alpha$, for some $\alpha \in (0,1)$.

We also give general bounds on the total variation distance of a reversible chain at time $t$ in terms of the probability that some “worst” set of stationary measure at least $\alpha$ was not hit by time $t$. As an application of our techniques we show that a sequence of lazy Markov chains on finite trees exhibits a cutoff iff the ratio of their relaxation-times and their (lazy) mixing-times tends to 0.

Keywords: Cutoff, mixing-time, finite reversible Markov chains, hitting times, trees, maximal inequality.

*Department of Statistics, UC Berkeley, California, USA. E-mail: riddhipratim@stat.berkeley.edu. Supported by UC Berkeley Graduate Fellowship.

†Department of Statistics, UC Berkeley, USA. E-mail: jonathan.heron@stat.berkeley.edu.

‡Microsoft Research, Redmond, Washington, USA. E-mail: peres@microsoft.com.
1 Introduction

In many randomized algorithms, the mixing-time of an underlying Markov chain is the main component of the running-time (see [22]). We obtain a tight bound on $t_{\text{mix}}(\epsilon)$ (up to an absolute constant independent of $\epsilon$) for lazy reversible Markov chains in terms of hitting times of large sets (Proposition 1.6, (1.6)). This refines previous results in the same spirit ([20] and [18], see related work), which gave a less precise characterization of the mixing-time in terms of hitting-times (and were restricted to hitting times of sets whose stationary measure is at most $1/2$).

Loosely speaking, the (total variation) **cutoff phenomenon** occurs when over a negligible period of time, known as the **cutoff window**, the (worst-case) total variation distance (of a certain finite Markov chain from its stationary distribution) drops abruptly from a value close to 1 to near 0.

Our bound on the mixing-time is sufficiently sharp to imply a characterization of cutoff for reversible Markov chains in terms of concentration of hitting times. Though many families of chains are believed to exhibit cutoff, proving the occurrence of this phenomenon is often an extremely challenging task. The cutoff phenomenon was given its name by Aldous and Diaconis in their seminal paper [2] from 1986 in which they suggested the following open problem (re-iterated in [7]), which they refer to as “the most interesting problem”: “Find abstract conditions which ensure that the cutoff phenomenon occurs”.

We use our general characterization of cutoff to give a sharp spectral condition for cutoff in lazy weighted nearest-neighbor random walks on trees (Theorem 1).

Generically, we shall denote the state space of a Markov chain by $\Omega$ and its stationary distribution by $\pi$ (or $\Omega_n$ and $\pi_n$, respectively, for the $n$-th chain in a sequence of chains). We say that the chain is finite, whenever $\Omega$ is finite. Let $(X_t)_{t=0}^\infty$ be an irreducible Markov chain on a finite state space $\Omega$ with transition matrix $P$ and stationary distribution $\pi$. We denote such a chain by $(\Omega, P, \pi)$.

To avoid periodicity and near-periodicity issues, one often considers the lazy version of a discrete time Markov chain. Define the **associated lazy chain** to be $(\Omega, P_L, \pi)$, where $P_L := \frac{1}{2}(I + P)$ and $I$ is the identity matrix. We call a chain lazy if $P(x,x) \geq \frac{1}{2}$, for any $x \in \Omega$ (in which case $P = K_L$, where $K := 2P - I$). In this paper, all discrete-time chains would be assumed to be lazy, unless otherwise is specified. Similarly, one often considers the continuous-time version of the chain. This is a Markov chain whose heat kernel is defined by $H_t(x,y) := \sum_{k=0}^\infty \frac{e^{-tk}}{k!} P^t(x,y)$. We denote the continuous-time version of the chain by $(X_{ct}^t)_{t \geq 0}$.

We denote by $P_t^\mu$ ($P_\mu$) the distribution of $X_t$ (resp. $(X_t)_{t \geq 0}$), given that the initial distribution is $\mu$. We denote by $Q_t^\mu$ ($Q_\mu$) the distribution of $X_{ct}^t$ (resp. $(X_{ct}^t)_{t \geq 0}$), given that the initial distribution is $\mu$. When $\mu = \delta_x$, the Dirac measure on some $x \in \Omega$ (i.e. the chain starts at $x$ with probability 1), we simply write $P_t^x$. For any $x,y \in \Omega$ and $t \in \mathbb{N}$ we write $P_t^x(y) := P_x(X_t = y) = P^t(x,y)$.

A chain $(\Omega, P, \pi)$ is called **reversible** if for any $x,y \in \Omega$,

$$\pi(x)P(x,y) = \pi(y)P(y,x).$$

We denote the set of probability distributions on a (finite) set $B$ by $\mathcal{P}(B)$. For any $\mu, \nu \in \mathcal{P}(B)$, their **total-variation distance** is defined to be $\|\mu - \nu\|_{\text{TV}} := \frac{1}{2} \sum_x |\mu(x) - \nu(x)|$.
the chain is irreducible and λ. Throughout we shall denote them by 1 = π

Recall that if (Ω, P, π) there exists c

d

we refer to the inner product induced by P

For a sequence of irreducible aperiodic reversible Markov chains with relaxation times {t

Fact 1.1. For a sequence of irreducible aperiodic reversible Markov chains with relaxation times \{t_{rel}^{(n)}\} and mixing-times \{t_{mix}^{(n)}\}, if the sequence exhibits a cutoff, then \(t_{rel}^{(n)} = o(t_{mix}^{(n)})\).

In 2004, the third author [19] conjectured that, in many natural classes of chains, the product condition is also sufficient for cutoff. In general, the product condition does not always imply cutoff. Aldous and Pak (private communication via P. Diaconis) have constructed relevant examples (see [16], Chapter 18). This left open the question of characterizing the classes of chains for which the product condition is indeed sufficient.

We now state our main theorem, which generalizes previous results concerning birth and death chains [10]. The relevant setup is weighted nearest neighbor random walks on finite trees. See Section 5 for a formal definition.
Theorem 1. Let \((V, P, \pi)\) be a lazy reversible Markov chain on a tree \(T = (V, E)\) with \(|V| \geq 3\). Then

\[
t_{\text{mix}}(\epsilon) - t_{\text{mix}}(1 - \epsilon) \leq 30\epsilon^{-1}t_{\text{rel}}t_{\text{mix}}^{-1}, \text{ for any } 0 < \epsilon \leq 1/4. 
\]

(1.3)

In particular, if the product condition holds for a sequence of lazy reversible Markov chains \((V_n, P_n, \pi_n)\) on finite trees \(T_n = (V_n, E_n)\), then the sequence exhibits a cutoff with a cutoff window \(w_n = \sqrt{t_{\text{rel}}^{(n)}t_{\text{mix}}^{(n)}}\).

In [8], Diaconis and Saloff-Coste showed that a sequence of birth and death (BD) chains exhibits separation cutoff if and only if \(t_{\text{rel}}^{(n)} = o(t_{\text{mix}}^{(n)})\). In [10], Ding et al. extended this also to the notion of total-variation cutoff and showed that the cutoff window is always at most \(\sqrt{t_{\text{rel}}^{(n)}t_{\text{mix}}^{(n)}}\) and that in some cases this is tight (see Theorem 1 and Section 2.3 ibid). Since BD chains are a particular case of chains on trees, the bound on \(w_n\) in Theorem 1 is also tight.

We note that the bound we get on the rate of convergence ((1.3)) is better than the estimate in [10] (even for BD chains), which is \(t_{\text{mix}}(\epsilon) - t_{\text{mix}}(1 - \epsilon) \leq c\epsilon^{-1}t_{\text{rel}}t_{\text{mix}}^{-1}\) (Theorem 2.2). In fact, in Section 6 we show that under the product condition, \(d(t)\) decays in a sub-Gaussian manner within the cutoff window. More precisely, we show that

\[
t_{\text{mix}}^{(n)}(\epsilon) - t_{\text{mix}}^{(n)}(1 - \epsilon) \leq c\sqrt{t_{\text{rel}}^{(n)}t_{\text{mix}}^{(n)}} \log \epsilon.
\]

This is somewhat similar to Theorem 6.1 in [8], which determines the “shape” of the cutoff and describes a necessary and sufficient spectral condition for the shape to be the density function of the standard normal distribution.

Concentration of hitting times was a key ingredient both in [8] and [10] (as it shall be here). Their proofs relied on several properties which are specific to BD chains. Our proof of Theorem 1 can be adapted to the following setup.

Definition 1.2. For \(n \in \mathbb{N}\) and \(\delta, r > 0\), we call a finite lazy reversible Markov chain, \(([n], P, \pi)\), a \((\delta, r)\)-semi birth and death (SBD) chain if

(i) For any \(i, j \in [n]\) such that \(|i - j| > r\), we have \(P(i, j) = 0\).

(ii) For all \(i, j \in [n]\) such that \(|i - j| = 1\), we have that \(P(i, j) \geq \delta\).

This is a natural generalization of the class of birth and death chains. Conditions (i)-(ii) tie the geometry of the chain to that of the path \([n]\). We have the following theorem.

Theorem 2. Let \(([n_k], P_k, \pi_k)\) be a sequence of \((\delta, r)\)-semi birth and death chains satisfying the product condition for some \(\delta, r > 0\). Then it exhibits a cutoff with a cutoff window

\[w_k := \sqrt{t_{\text{mix}}^{(k)}t_{\text{rel}}^{(k)}}.\]

We now introduce a new notion of mixing, which shall play a key role in this work.

Definition 1.3. Let \((\Omega, P, \pi)\) be an irreducible chain. For any \(x \in \Omega, \alpha, \epsilon \in (0, 1)\) and \(t \geq 0\), define

\[p_x(\alpha, t) := \max_{A \subset \Omega: \pi(A) \geq \alpha} P_x[T_A > t], \text{ where } T_A := \inf\{t : X_t \in A\}\]

is the hitting time of the set \(A\). Set \(p(\alpha, t) := \max_x p_x(\alpha, t)\). We define

\[\text{hit}_{\alpha, x}(\epsilon) := \min\{t : p_x(\alpha, t) \leq \epsilon\} \text{ and } \text{hit}_\alpha(\epsilon) := \min\{t : p(\alpha, t) \leq \epsilon\}.\]

Similarly, we define

\[p_x^c(\alpha, t) := \max_{A \subset \Omega: \pi(A) \geq \alpha} Q_x[T_A > t] \text{ and set } \text{hit}_\alpha^c(\epsilon) := \min\{t : p_x^c(\alpha, t) \leq \epsilon \text{ for all } x \in \Omega\} \text{ (where here } T_A := \inf\{t : X_{t^c} \in A\}\).\]
Definition 1.4. Let \((\Omega_n, P_n, \pi_n)\) be a sequence of irreducible chains and let \(\alpha \in (0, 1)\). We say that the sequence exhibits a \(\text{hit}_\alpha\)-cutoff, if for any \(\epsilon \in (0, 1/4)\)

\[
\text{hit}_\alpha^{(n)}(\epsilon) - \text{hit}_\alpha^{(n)}(1 - \epsilon) = o \left( \text{hit}_\alpha^{(n)}(1/4) \right).
\]

We are now ready to state our main abstract theorem.

Theorem 3. Let \((\Omega_n, P_n, \pi_n)\) be a sequence of lazy reversible irreducible finite chains. The following are equivalent:

1) The sequence exhibits a cutoff.

2) The sequence exhibits a \(\text{hit}_\alpha\)-cutoff for some \(\alpha \in (0, 1)\) and \(t_{\text{rel}}^{(n)} = o(t_{\text{mix}}^{(n)})\).

Remark 1.5. The proof of Theorem 3 can be extended to the continuous-time case. In particular, it follows that a sequence of finite reversible chains exhibits cutoff iff the sequence of the continuous-time versions of these chains exhibits cutoff. This was previously proven in [6] without the assumption of reversibility.

At first glance \(\text{hit}_\alpha(\epsilon)\) may seem like a rather weak notion of mixing compared to \(t_{\text{mix}}(\epsilon)\), especially when \(\alpha\) is close to 1 (say, \(\alpha = 1 - \epsilon\)). The following proposition gives a quantitative version of Theorem 3 (for simplicity we fix \(\alpha = 1/2\) in (1.4) and (1.5)).

Proposition 1.6. For any reversible irreducible finite lazy chain and any \(\epsilon \in (0, \frac{1}{3}]\),

\[
\text{hit}_{1/2}(2\epsilon) - \left[4t_{\text{rel}} \log (2/\epsilon)\right] \leq t_{\text{mix}}(\epsilon) \leq \text{hit}_{1/2}(\epsilon/2) + \left[t_{\text{rel}} \log (4/\epsilon)\right] \quad \text{and} \quad (1.4)
\]

\[
\text{hit}_{1/2}(1 - \epsilon/2) - \left[4t_{\text{rel}} \log (3/\epsilon)\right] \leq t_{\text{mix}}(1 - \epsilon) \leq \text{hit}_{1/2}(1 - 2\epsilon) + \left[t_{\text{rel}}\right]. \quad (1.5)
\]

Moreover,

\[
\max\{\text{hit}_{1-\epsilon/4}(5\epsilon/4), (t_{\text{rel}} - 1) \log (1/2\epsilon)\} \leq t_{\text{mix}}(\epsilon) \leq \text{hit}_{1-\epsilon/4}(3\epsilon/4) + \left[\frac{3t_{\text{rel}}}{2} \log (4/\epsilon)\right]. \quad (1.6)
\]

Finally, if everywhere in (1.4)-(1.6) \(t_{\text{mix}}\) and \(\text{hit}\) are replaced by \(t_{\text{mix}}^{ct}\) and \(\text{hit}^{ct}\), respectively, then (1.4)-(1.6) still hold (and all ceiling signs can be omitted).

Loosely speaking, we show that the mixing of a lazy reversible Markov chain can be partitioned into two stages as follows. The first is the time it takes the chain to escape from some small set with sufficiently large probability. In the second stage, the chain mixes at the fastest possible rate (up to a small constant), which is governed by its relaxation-time.

It follows from Proposition 3.3 that the ratio of the LHS and the RHS of (1.6) is bounded by an absolute constant independent of \(\epsilon\). Moreover, (1.6) bounds \(t_{\text{mix}}(\epsilon)\) in terms of hitting distribution of sets of \(\pi\) measure tending to 1 as \(\epsilon\) tends to 0. In (3.2) we give a version of (1.6) for sets of arbitrary \(\pi\) measure.

Either of the two terms appearing in the sum in RHS of (1.6) may dominate the other. For lazy random walk on two \(n\)-cliques connected by a single edge, the terms in (1.6) involving \(\text{hit}_{1-\epsilon/4}\) are negligible. For a sequence of chains satisfying the product condition, all terms in Proposition 1.6 involving \(t_{\text{rel}}\) are negligible. Hence the assertion of Theorem 3, for \(\alpha = 1/2\), follows easily from (1.4) and (1.5), together with the fact that \(\text{hit}_{1/2}^{(n)}(1/4) = \Theta(t_{\text{mix}}^{(n)})\). In Proposition 3.6, under the assumption that the product condition holds, we prove this fact and show that in fact, if the sequence exhibits \(\text{hit}_\alpha\)-cutoff for some \(\alpha \in (0, 1)\), then it exhibits \(\text{hit}_\beta\)-cutoff for all \(\beta \in (0, 1)\).
1.1 Related work

This work was greatly motivated by the following result of Peres and Sousi ([20] Theorem 1.1) and independently of Oliviera ([18] Theorem 2) which share the general theme of describing mixing-times in terms of hitting-times. Their approach relied on the theory of random times to stationarity combined with a certain “de-randomization” argument which shows that for any lazy reversible irreducible finite chain and any stopping time $T$ such that $X_T \sim \pi$, $t_{\text{mix}} = O(\max_{x \in \Omega} E_x[T])$. As a (somewhat indirect) consequence, they showed that for any $0 < \alpha < 1/2$ (this was extended to $\alpha = 1/2$ in [14]), there exist some constants $c_\alpha, c'_\alpha > 0$ such that for any lazy reversible irreducible finite chain

$$c'_\alpha t_H(\alpha) \leq t_{\text{mix}} \leq c_\alpha t_H(\alpha),$$

where $t_H(\alpha) := \max_{x \in \Omega, A \subset \Omega: \pi(A) \geq \alpha} E_x[T_A]$.

In [15], Lancia et al. established a sufficient condition for cutoff which does not rely on reversibility. However, their condition includes the strong assumption that for some $A_n \subset \Omega_n$ with $\pi_n(A_n) \geq c > 0$, starting from any $x \in A_n$, the $n$-th chain mixes in $\alpha(t_{\text{mix}}^{(n)})$ steps.

1.2 An overview of our techniques

The most important tool we shall utilize is Starr’s $L^2$ maximal inequality (Theorem 2.3). Relating it to the study of mixing-times of reversible Markov chains is one of the main contributions of this work.

**Definition 1.7.** Let $(\Omega, P, \pi)$ be a finite reversible irreducible lazy chain. Let $A \subset \Omega$, $s \geq 0$ and $m > 0$. Denote $\rho(A) := \sqrt{\text{Var}_A T_A} = \sqrt{\pi(A)(1 - \pi(A))}$. Set $\sigma_s := e^{-s/t_{\text{rel}}} \rho(A)$. We define

$$G_s(A, m) := \{y : |P_y^k(A) - \pi(A)| < m\sigma_s \text{ for all } k \geq s\}.$$

(1.7)

We call the set $G_s(A, m)$ the good set for $A$ from time $s$ within $m$ standard-deviations.

As a simple corollary of Starr’s $L^2$ maximal inequality and the $L^2$-contraction lemma we show in Corollary 2.4 that for any non-empty $A \subset \Omega$ and any $m, s \geq 0$ that $\pi(G_s(A, m)) \geq 1 - 8/m^2$. To demonstrate the main idea of our approach we prove the following inequalities.

$$t_{\text{mix}}(2\epsilon) \leq \text{hit}_{1-\epsilon}(\epsilon) + \left[\frac{t_{\text{rel}}}{2} \log \left(\frac{2}{\epsilon^3}\right)\right].$$

(1.8)

$$\text{hit}_{1-\epsilon}(1 - 2\epsilon) \geq t_{\text{mix}}(1 - \epsilon) - \left[\frac{t_{\text{rel}}}{2} \log \left(\frac{8}{\epsilon^2}\right)\right].$$

(1.9)

We first prove (1.8). Fix $A \subset \Omega$ be non-empty. Let $x \in \Omega$. Let $s, t, m \geq 0$ to be defined shortly. Denote $G := G_s(A, m)$. We want this set to be of size at least $1 - \epsilon$. By Corollary 2.4 we know that $\pi(G) \geq 1 - 8/m^2$. Thus we pick $m = \sqrt{8/\epsilon}$. The precision in (1.7) is $m\sigma_s \leq \sqrt{8/\epsilon}(\sqrt{\text{Var}_\pi T_A} e^{-s/t_{\text{rel}}}) \leq \sqrt{2/\epsilon}e^{-s/t_{\text{rel}}}$. We also want $\epsilon$ precision. Hence we pick $s := \left[\frac{\text{Var}_\pi T_A}{2} \log \left(\frac{2}{\epsilon}\right)\right]$.

We seek to bound $|P_x^{t+s}(A) - \pi(A)|$. If $|P_x^{t+s}(A) - \pi(A)| \leq 2\epsilon$, then the chain is “$2\epsilon$-mixed w.r.t. $A$”. This is where we use the set $G$. We now explain that for any $t \geq 0$, hitting $G$ by
In particular,\[ |P_x[X_{t+s} \in A | T_G \leq t] - \pi(A)| \leq \max_{g \in G} \sup_{s' \geq s} |P_g^{s'}(A) - \pi| \leq \epsilon. \]

In particular,\[ |P^{t+s}x(A) - \pi(A)| \leq P_x[T_G > t] + |P_x[X_{t+s} \in A | T_G \leq t] - \pi(A)| \leq P_x[T_G > t] + \epsilon. \tag{1.10} \]

We seek to have the bound \( P_x[T_G > t] \leq \epsilon \). Recall that by our choice of \( m \) we have that \( \pi(G) \geq 1 - \epsilon \). Thus if we pick \( t := \text{hit}_{1-\epsilon}(\epsilon) \), we guarantee that, regardless of the identity of \( A \) and \( x \), we indeed have that \( P_x[T_G > t] \leq \epsilon \). Since \( x \) and \( A \) were arbitrary, plugging this into (1.10) yields (1.8). We now prove (1.9).

We now set \( r := t_{\text{mix}}(1 - \epsilon) - 1 \). Then there exist some \( x \in \Omega \) and \( A \subset \Omega \) such that \( \pi(A) = P^r_x(A) > 1 - \epsilon \). In particular, \( \pi(A) > 1 - \epsilon \). Consider again \( G_2 := G_{s_2}(A, m) \). Since again we seek the size of \( G_2 \) to be at least \( 1 - \epsilon \), we again choose \( m = \sqrt{8/\epsilon} \). The precision in (1.7) is \( m \sigma_{s_2} = \sqrt{8/\epsilon}(\sqrt{\text{Var}}_\pi 1_A e^{-s_2/\text{rel}}) \leq \sqrt{8/\epsilon}(\sqrt{1 - \pi(A)} e^{-s_2/\text{rel}}) \leq \sqrt{8} e^{-s_2/\text{rel}}. \) We again seek \( \epsilon \) precision. Hence we pick \( s_2 := \left\lceil \frac{\log(e^2)}{2} \right\rceil \). As in (1.10) (with \( r - s_2 \) in the role of \( t \) and \( s_2 \) in the role of \( s \)) we have that

\[ P_x[T_{G_2} > r - s_2] \geq \pi(A) - P^r_x(A) - \epsilon > 1 - 2\epsilon. \]

Hence it must be the case that \( \text{hit}_{1-\epsilon}(1 - 2\epsilon) > r - s_2 = t_{\text{mix}}(1 - \epsilon) - 1 - \left\lceil \frac{t_{\text{rel}}}{2} \log\left(\frac{N}{2}\right) \right\rceil \).

## 2 Maximal inequality and applications

In this section we present the machinery that will be utilized in the proof of the main results. Here and in Section 3 we only treat the discrete-time chain. The necessary adaptations for the continuous-time case are explained in Section 4. We start with a few basic definitions and facts.

**Definition 2.1.** Let \((\Omega, P, \pi)\) be a finite reversible chain. For any \( f \in \mathbb{R}^\Omega \), let \( \mathbb{E}_\pi[f] := \sum_{x \in \Omega} \pi(x) f(x) \) and \( \text{Var}_\pi f := \mathbb{E}_\pi[(f - \mathbb{E}_\pi f)^2] \). The inner-product \( \langle \cdot, \cdot \rangle_\pi \) and \( L^p \) norm are

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

We identify the matrix \( P^t \) with the operator \( P^t : L^p(\mathbb{R}^\Omega, \pi) \to L^p(\mathbb{R}^\Omega, \pi) \) defined by \( P^t f(x) := \sum_{y \in \Omega} P^t(x, y) f(y) = \mathbb{E}_\pi f(X_t) \). Then by reversibility \( P^t : L^2 \to L^2 \) is a self-adjoint operator.

The spectral decomposition in discrete time takes the following form. If \( f_1, \ldots, f_{|\Omega|} \) is an orthonormal basis of \( L^2(\mathbb{R}^\Omega, \pi) \) such that \( P f_i := \lambda_i f_i \) for all \( i \), then \( P^t g = \mathbb{E}_\pi P^t g + \sum_{i=2}^{|\Omega|} \langle g, f_i \rangle \pi \lambda_i f_i \), for all \( g \in \mathbb{R}^\Omega \) and \( t \geq 0 \). The following lemma is standard. It is proved using the spectral decomposition in a straightforward manner.

**Lemma 2.2** \((L^2\text{-contraction Lemma})\). Let \((\Omega, P, \pi)\) be a finite lazy reversible irreducible Markov chain. Let \( f \in \mathbb{R}^\Omega \). Then

\[ \text{Var}_\pi P^t f \leq e^{-2t/\text{rel}} \text{Var}_\pi f, \text{ for any } t \geq 0. \]  

(2.1)
We now state a particular case of Starr’s maximal inequality ([23] Theorem 1). It is similar to Stein’s maximal inequality ([24]), but gives the best possible constant. For the sake of completeness we also prove Theorem 2.3 at the end of this section.

**Theorem 2.3** (Maximal inequality). Let \((\Omega, P, \pi)\) be a reversible irreducible Markov chain. Let \(1 < p < \infty\). Then for any \(f \in L^p(\mathbb{R}^\Omega, \pi)\),

\[
\|f^*\|_p \leq \left( \frac{p}{p-1} \right) \|f\|_p,
\]

where \(f^* \in \mathbb{R}^\Omega\) is the corresponding maximal function at even times, defined as

\[
f^*(x) := \sup_{0 \leq k < \infty} \|P^{2k}(f)(x)\| = \sup_{0 \leq k < \infty} \|E_x[f(X_{2k})]\|.
\]

The following corollary follows by combining Lemma 2.2 with Theorem 2.3.

**Corollary 2.4.** Let \((\Omega, P, \pi)\) be a finite reversible irreducible lazy chain. As in Definition 1.7, define \(\rho(A) := \sqrt{\pi(A)(1 - \pi(A))}\), \(\sigma_t := \rho(A)e^{-t/t_{rel}}\) and

\[
G_t(A, m) := \{ y : |P_y^k(A) - \pi(A)| < m\sigma_t \text{ for all } k \geq t \}.
\]

Then

\[
\pi(G_t(A, m)) \geq 1 - 8m^{-2}, \text{ for all } A \subset \Omega, t \geq 0 \text{ and } m > 0.
\]

**Proof.** For any \(t \geq 0\), let \(f_t(x) := P^t(1_A(x) - \pi(A)) = P_x^t(A) - \pi(A)\). Then in the notation of Theorem 2.3,

\[
f_t^*(x) := \sup_{k \geq 0} \|P^{2k}f_t(x)\| = \sup_{k \geq 0} \|P^{2k+t}(A) - \pi(A)\|,
\]

and similarly

\[
(Pf_t)^*(x) = \sup_{k \geq 0} \|P^{2k+1+t}(A) - \pi(A)\|.
\]

Hence \(G_t = \{ x \in \Omega : f_t^*(x), (Pf_t)^*(x) < m\sigma_t \}\). Whence

\[
1 - \pi(G_t) \leq \pi \{ x : f_t^*(x) \geq m\sigma_t \} + \pi \{ x : (Pf_t)^*(x) \geq m\sigma_t \}.
\]

Note that since \(\pi P^t = \pi\) we have that \(E_x(f_t) = E_x(f_0) = E_x(1_A - \pi(A)) = 0\). Since \(\lambda_2 \geq \lambda_{|G|}\), (2.1) implies that

\[
\|Pf_t\|_2^2 \leq \|f_t\|_2^2 = \text{Var}_xP^tf_0 \leq e^{-2t/t_{rel}}\text{Var}_x f_0 = e^{-2t/t_{rel}}\rho^2(A) = \sigma_t^2.
\]

Hence by Markov inequality and (2.2) we have

\[
\pi \{ x : f_t^*(x) \geq m\sigma_t \} = \pi \{ x : (f_t^*(x))^2 \geq m^2\sigma_t^2 \} \leq 4m^{-2},
\]

and similarly, \(\pi \{ x : (Pf_t)^*(x) \geq m\sigma_t \} \leq 4m^{-2}\).

The corollary now follows by substituting the last two bounds in (2.4). \(\square\)
2.1 Proof of Theorem 2.3

As promised, we end this section with the proof of Theorem 2.3.

**Proof of Theorem 2.3.** Let \( p \in (1, \infty) \) and \( f \in L^p(\mathbb{R}^n, \pi) \). Let \( q := \frac{p}{p-1} \) be the conjugate exponent of \( p \). We argue that it suffices to prove the theorem only for \( f \geq 0 \), since for general \( f \), if we denote \( h := |f| \), then \(|f_s| \leq h_s \). Consequently, \( \|f_s\|_p \leq \|h_s\|_p \leq q\|h\|_p = q\|f\|_p \).

Let \((X_n)_{n \geq 0} \) have the distribution of the chain \((\Omega, P, \pi)\) with \( X_0 \sim \pi \). Let \( n \geq 0 \). Let \( 0 \leq f \in L^p(\Omega, \pi) \). By the tower property of conditional expectation (e.g. [11], Theorem 5.1.6.),

\[
P^{2n}f(X_0) := \mathbb{E}[f(X_{2n}) \mid X_0] = \mathbb{E}[\mathbb{E}[f(X_{2n}) \mid X_n] \mid X_0] = \mathbb{E}[R_n \mid X_0], \tag{2.7}
\]

where \( R_n := \mathbb{E}[f(X_{2n}) \mid X_n] \). Since \( X_0 \sim \pi \), by reversibility, \((X_n, X_{n+1}, \ldots, X_{2n})\) and \((X_n, X_{n-1}, \ldots, X_0)\) have the same law. Hence

\[
R_n = \mathbb{E}[f(X_{2n}) \mid X_n] = \mathbb{E}[f(X_0) \mid X_n] = \mathbb{E}[f(X_0) \mid X_n, X_{n+1}, \ldots], \tag{2.8}
\]

where the second equality in (2.8) follows by the Markov property. Fix \( N \geq 0 \). By (2.8) \((R_0)^N_{n=0} \) is a reverse martingale, i.e. \((R_{N-n})^N_{n=0} \) is a martingale. By Doob’s \( L^p \) maximal inequality (e.g. [11], Theorem 5.4.3.)

\[
\| \max_{0 \leq n \leq N} R_n \|_p \leq q \| R_0 \|_p = q \| f(X_0) \|_p. \tag{2.9}
\]

Denote \( h_N := \max_{0 \leq n \leq N} P^{2n}f \). By (2.7),

\[
h_N(X_0) = \max_{0 \leq n \leq N} \mathbb{E}[R_n | X_0] \leq \mathbb{E} \left[ \max_{0 \leq n \leq N} R_n | X_0 \right]. \tag{2.10}
\]

By conditional Jensen inequality \( \| \mathbb{E}[Y | X_0] \|_p \leq \| Y \|_p \) (e.g. [11], Theorem 5.1.4.). So by taking \( L^p \) norms in (2.10), together with (2.9) we get that

\[
\| h_N \|_p \leq \| \max_{0 \leq n \leq N} R_n \|_p \leq q \| f(X_0) \|_p. \tag{2.11}
\]

The proof is concluded using the monotone convergence theorem. \(\square\)

3 Inequalities relating \( t_{\text{mix}}(\epsilon) \) and \( \text{hit}_\alpha(\delta) \)

Our aim in this section is to obtain inequalities relating \( t_{\text{mix}}(\epsilon) \) and \( \text{hit}_\alpha(\delta) \) for suitable values of \( \alpha, \epsilon \) and \( \delta \) using Corollary 2.4.

The following corollary uses the same reasoning as in the proof of (1.8)-(1.9) with a slightly more careful analysis.

**Corollary 3.1.** Let \((\Omega, P, \pi)\) be a lazy reversible irreducible finite chain. Let \( x \in \Omega, \delta, \alpha \in (0, 1), s \geq 0 \) and \( A \subset \Omega \). Denote \( t := \text{hit}_{1-\alpha, x}(\delta) \). Then

\[
P^x_{s+t}[A] \geq (1 - \delta) \left[ \pi(A) - e^{-s/t_{\text{rel}}} \left[ 8\alpha^{-1}\pi(A)(1 - \pi(A)) \right]^{1/2} \right]. \tag{3.1}
\]
Consequently, for any $0 < \epsilon < 1$ we have that

$$
\text{hit}_{1-\alpha}((\alpha+\epsilon)\land 1) \leq \operatorname{t_{mix}}(\epsilon) \quad \text{and} \quad \operatorname{t_{mix}}((\epsilon+\delta)\land 1) \leq \text{hit}_{1-\alpha}(\epsilon) + \left[ \frac{t_{rel}}{2} \log^+ \left( \frac{2(1-\epsilon)^2}{\alpha \epsilon \delta} \right) \right], \quad (3.2)
$$

where $a \land b := \min\{a, b\}$ and $\log^+ x := \max\{\log x, 0\}$. In particular, for any $0 < \epsilon \leq 1/2$,

$$
\text{hit}_{1-\epsilon/4}(5\epsilon/4) \leq \operatorname{t_{mix}}(\epsilon) \leq \text{hit}_{1-\epsilon/4}(3\epsilon/4) + \left[ \frac{3t_{rel}}{2} \log (4/\epsilon) \right], \quad (3.3)
$$

$$
\operatorname{t_{mix}}(\epsilon) \leq \text{hit}_{1/2}(\epsilon/2) + \left[ t_{rel} \log (4/\epsilon) \right] \quad \text{and} \quad \operatorname{t_{mix}}(1-\epsilon/2) \leq \text{hit}_{1/2}(1-\epsilon) + \left[ t_{rel} \right]. \quad (3.4)
$$

Proof. We first prove (3.1). Fix some $x \in \Omega$. Consider the set

$$
G = G_s(A) := \left\{ y : |P_{y}^k(A) - \pi(A)| < e^{-s/t_{rel}} \left( 8\alpha^{-1}\pi(A)(1 - \pi(A)) \right)^{1/2} \right\}.
$$

Then by Corollary 2.4 we have that

$$
\pi(G) \geq 1 - \alpha.
$$

By the Markov property and conditioning on $T_G$ and on $X_{T_G}$ we get that

$$
P^{t+s}_{x}[A \mid T_G \leq t] \geq \pi(A) - e^{-s/t_{rel}} \left[ 8\alpha^{-1}\pi(A)(1 - \pi(A)) \right]^{1/2}.
$$

Since $\pi(G) \geq 1 - \alpha$ we have that $P_x[T_G \leq t] \geq 1 - \delta$ for $t := \text{hit}_{1-\alpha}(\delta)$. Thus

$$
P^{t+s}_{x}[A] \geq P_x[T_G \leq t]P^{t+s}_{x}[A \mid T_G \leq t] \geq (1 - \delta) \left[ \pi(A) - e^{-s/t_{rel}} \left[ 8\alpha^{-1}\pi(A)(1 - \pi(A)) \right]^{1/2} \right],
$$

which concludes the proof of (3.1). We now prove (3.2). The first inequality in (3.2) follows directly from the definition of the total variation distance. To see this, let $A \subset \Omega$ be an arbitrary set with $\pi(A) \geq 1 - \alpha$. Let $t_1 := \operatorname{t_{mix}}(\epsilon)$. Then for any $x \in \Omega$, $P_x[T_A \leq t_1] \geq P_x[X_{t_1} \in A] \geq \pi(A) - \|P^t_x - \pi\|_{TV} \geq 1 - \alpha - \epsilon$. In particular, we get directly from Definition 1.4 that $\text{hit}_{1-\alpha}(\alpha + \epsilon) \leq t_1 = \operatorname{t_{mix}}(\epsilon)$. We now prove the second inequality in (3.2).

Set $t := \text{hit}_{1-\alpha}(\epsilon)$ and $s := \left\lceil \frac{1}{2} t_{rel} \log^+ \left( \frac{2(1-\epsilon)^2}{\alpha \epsilon \delta} \right) \right\rceil$. Let $x \in \Omega$ be such that $d(t + s, x) = d(t + s)$ and set $A := \{ y \in \Omega : \pi(y) > P^{t+s}_{x}(y) \}$. Observe that by the choice of $t, s, x$ and $A$ together with (3.1) we have that

$$
d(t + s) = \pi(A) - P^{t+s}_{x}[A] \leq \epsilon \pi(A) + (1 - \epsilon) e^{-s/t_{rel}} \left[ 8\alpha^{-1}\pi(A)(1 - \pi(A)) \right]^{1/2} \\
\leq \epsilon \left[ \pi(A) + 2\sqrt{\delta/\epsilon} \sqrt{\pi(A)(1 - \pi(A))} \right] \leq \epsilon [1 + (2\sqrt{\delta/\epsilon})^2/4] = \epsilon + \delta,
$$

where in the last inequality we have used the easy fact that for any $c > 0$ and any $x \in [0, 1]$ we have that $x + c\sqrt{x(1-x)} \leq 1 + c^2/4$. Indeed, since $x \in [0, 1]$ it suffices to show that $x + c\sqrt{1-x} \leq 1 + c^2/4$. Write $\sqrt{1-x} = y$ and $c/2 = a$. By subtracting $x$ from both sides, the previous inequality is equivalent to $2ay \leq y^2 + a^2$. This concludes the proof of (3.2).

To get (3.3), apply (3.2) with $(\alpha, \epsilon, \delta)$ being $(\epsilon/4, 3\epsilon/4, \epsilon/4)$. Similarly, to get (3.4) apply (3.2) with $(\alpha, \epsilon, \delta)$ being $(1/2, \epsilon/2, \epsilon/2)$ or $(1/2, 1 - \epsilon, \epsilon/2)$, respectively.
Remark 3.2. Corollary 3.1 holds also in continuous-time case (where everywhere in (3.1)-(3.4) $t_{mix}$ and hit are replaced by $t^{ct}$ and hit$^{ct}$, respectively, and all ceiling signs are omitted). The necessary adaptations are explained in Section 4.

Let $\alpha \in (0, 1)$. Observe that for any $A \subset \Omega$ with $\pi(A) \geq \alpha$, any $x \in \Omega$ and any $t, s \geq 0$ we have that $P_x[T_A > t + s] \leq P_x[T_A > t] \left( \max_t P_x[T_A > s] \right) \leq p(\alpha, t)p(\alpha, s)$. Maximizing over $A$ yields that $p(\alpha, t + s) \leq p(\alpha, t)p(\alpha, s)$, from which the following proposition follows.

Proposition 3.3. For any $\alpha, \epsilon, \delta \in (0, 1)$ we have that

$$\text{hit}_\alpha(\epsilon \delta) \leq \text{hit}_\alpha(\epsilon) + \text{hit}_\alpha(\delta). \quad (3.6)$$

In the next corollary, we establish inequalities between hit$\alpha(\delta)$ and hit$\beta(\delta')$ for appropriate values of $\alpha, \beta, \delta$ and $\delta'$.

Corollary 3.4. For any reversible irreducible finite chain and $0 < \epsilon < \delta < 1$,

$$\text{hit}_\beta(\delta) \leq \text{hit}_\alpha(\delta) \leq \text{hit}_\beta(\delta - \epsilon) + \left[ \alpha^{-1} t_{rel} \log \left( \frac{1 - \alpha}{(1 - \beta)\epsilon} \right) \right], \text{ for any } 0 < \alpha \leq \beta < 1. \quad (3.7)$$

The general idea behind Corollary 3.4 is as follows. Loosely speaking, we show that any (not too small) set $A \subset \Omega$ has a “blow-up” set $H(A)$ (of large $\pi$-measure), such that starting from any $x \in H(A)$, the set $A$ is hit “quickly” (in time proportional to $t_{rel}$ times a constant depending on the size of $A$) with large probability.

In order to establish the existence of such a blow-up, it turns out that it suffices to consider the hitting time of $A$, starting from the initial distribution $\pi$, which is well-understood.

Lemma 3.5. Let $(\Omega, P, \pi)$ be a finite irreducible reversible Markov chain. Let $A \subset \Omega$ be non-empty. Let $\alpha > 0$ and $w \geq 0$. Let $B(A, w, \alpha) := \left\{ y : P_y \left[ T_A > \left[ \frac{t_{rel}w}{\pi(A)} \right] \right] \geq \alpha \right\}$. Then

$$P_\pi[T_A > t] \leq \pi(A^c) \left( 1 - \frac{\pi(A)}{t_{rel}} \right)^t \leq \pi(A^c) \exp \left( \frac{-t\pi(A)}{t_{rel}} \right), \text{ for any } t \geq 0. \quad (3.8)$$

In particular,

$$\pi(B(A, w, \alpha)) \leq \pi(A^c) e^{-w} \alpha^{-1} \text{ and } \pi(A) \mathbb{E}_\pi[T_A] \leq t_{rel} \pi(A^c). \quad (3.9)$$

The proof of Lemma 3.5 is deferred to the end of this section.

Proof of Corollary 3.4. Denote $s = s_{\alpha, \beta, \epsilon} := \left[ \alpha^{-1} t_{rel} \log \left( \frac{1 - \alpha}{(1 - \beta)\epsilon} \right) \right]$. Let $A \subset \Omega$ be an arbitrary set such that $\pi(A) \geq \alpha$. Consider the set

$$H_1 = H_1(A, \alpha, \beta, \epsilon) := \left\{ y \in \Omega : P_y[T_A \leq s] \geq 1 - \epsilon \right\}.$$

Then by (3.9)

$$\pi(H_1) \geq 1 - (1 - (1 - \epsilon))^{-1}(1 - \pi(A)) \exp \left[ -\frac{s\pi(A)}{t_{rel}} \right] \geq 1 - \epsilon^{-1}(1 - \alpha) \exp \left[ -\log \left( \frac{1 - \alpha}{(1 - \beta)\epsilon} \right) \right] = \beta.$$
By the definition of $H_1$ together with the Markov property and the fact that $\pi(H_1) \geq \beta$, for any $t \geq 0$ and $x \in \Omega$,

$$
P_x[T_A \leq t+s] \geq P_x[T_{H_1} \leq t, T_A \leq t+s] \geq (1-\epsilon)P_x[T_{H_1} \leq t] \\
\geq (1-\epsilon)(1-p_x(\beta, t)) \geq 1 - \epsilon - \max_{y \in \Omega} p_y(\beta, t).
$$

(3.10)

Taking $t := \text{hit}_\beta(\delta - \epsilon)$ and minimizing the LHS of (3.10) over $A$ and $x$ gives the second inequality in (3.7). The first inequality in (3.7) is trivial because $\alpha \leq \beta$.

### 3.1 Proofs of Proposition 1.6 and Theorem 3

Now we are ready to prove our main abstract results.

**Proof of Proposition 1.6.** First note that (1.6) follows from (3.3) and the first inequality in (1.2). Moreover, in light of (3.4) we only need to prove the first inequalities in (1.4) and (1.5). Fix some $0 < \epsilon \leq 1/4$. By the first inequality in (3.2) and (3.7) (with $(\alpha, \beta, \epsilon, \delta)$ in (3.7) being $(1/2, 1-\epsilon/2, \epsilon/2, \epsilon)$ and $(1/2, 1-\epsilon/3, \epsilon/6, 1-\epsilon/2)$, respectively)

$$
t^{(n)}_{\text{mix}}(\epsilon) \geq \text{hit}^{(n)}_{1-\epsilon/2}(3\epsilon/2) \geq \text{hit}^{(n)}_{1/2}(2\epsilon) - \left[ 4t^{(n)}_{\text{rel}} \log (2/\epsilon) \right] \text{ and similarly }
$$

$$
t^{(n)}_{\text{mix}}(1-\epsilon) \geq \text{hit}^{(n)}_{1-\epsilon/3}(1-2\epsilon/3) \geq \text{hit}^{(n)}_{1/2}(1-\epsilon/2) - \left[ 4t^{(n)}_{\text{rel}} \log (3/\epsilon) \right].
$$

(3.11)

Before completing the proof of Theorem 3, we prove that under the product condition if a sequence of reversible chains exhibits hit$_\alpha$-cutoff for some $\alpha \in (0, 1)$, then it exhibits hit$_\alpha$-cutoff for all $\alpha \in (0, 1)$.

**Proposition 3.6.** Let $(\Omega_n, P_n, \pi_n)$ be a sequence of lazy finite irreducible reversible chains. Assume that the product condition holds. Then (1) and (2) below are equivalent:

1. There exists $\alpha \in (0, 1)$ for which the sequence exhibits a hit$_\alpha$-cutoff.

2. The sequence exhibits a hit$_\alpha$-cutoff for any $\alpha \in (0, 1)$.

Moreover, for any $\alpha \in (0, 1)$,

$$
\text{hit}^{(n)}_\alpha(1/4) = \Theta(t^{(n)}_{\text{mix}}), \text{ for any } \alpha \in (0, 1).
$$

(3.11)

Furthermore, if (2) holds then

$$
\lim_{n \to \infty} \text{hit}^{(n)}_\alpha(1/4)/\text{hit}^{(n)}_{1/2}(1/4) = 1, \text{ for any } \alpha \in (0, 1).
$$

(3.12)

**Proof.** We start by proving (3.11). Assume that the product condition holds. Fix some $\alpha \in (0, 1)$. Note that we have

$$
\text{hit}^{(n)}_\alpha(1/4) \leq 4\alpha^{-1}\text{hit}^{(n)}_\alpha \left( 1 - \frac{3\alpha}{4} \right) \leq 4\alpha^{-1}t^{(n)}_{\text{mix}} \left( \frac{\alpha}{4} \right) \leq 4\alpha^{-1}(3 + \log_2(1/\alpha))t^{(n)}_{\text{mix}}.
$$
Let $k, t, \alpha, \epsilon, \delta$ in (3.2) with (4.24) and Lemma 4.12). Conversely, by (3.6) (second inequality) and the second inequality in (3.2) with $(\alpha, \epsilon, \delta)$ here being $(1 - \alpha, 1/8, 1/8)$ (first inequality)

$$\frac{t^{(n)}_{\text{mix}}}{2} - \left[\frac{t^{(n)}_{\text{rel}}}{4} \log \left(\frac{100}{1 - \alpha}\right)\right] \leq \frac{\text{hit}^{(n)}_{\alpha}(1/8)}{2} \leq \text{hit}^{(n)}_{\alpha}(1/4).$$

This concludes the proof of (3.11). We now prove the equivalence between (1) and (2) under the product condition. It suffices to show that $(1) \implies (2)$, as the reversed implication is trivial. Fix $0 < \alpha < \beta < 1$. It suffices to show that hit$_{\alpha}$-cutoff occurs iff hit$_{\beta}$-cutoff occurs.

Fix $\epsilon \in (0, 1/8)$. Denote $s_n = s_n(\alpha, \beta, \epsilon) := \left[\frac{t^{(n)}_{\text{rel}}}{\alpha^{-1}} \log \left(\frac{1 - \alpha}{(1 - \beta)\epsilon}\right)\right]$. By the second inequality in Corollary 3.4

$$\text{hit}^{(n)}_{\alpha}(1 - \epsilon) \leq \text{hit}^{(n)}_{\beta}(1 - 2\epsilon) + s_n \quad \text{and} \quad \text{hit}^{(n)}_{\alpha}(2\epsilon) \leq \text{hit}^{(n)}_{\beta}(\epsilon) + s_n. \quad (3.13)$$

By the first inequality in Corollary 3.4

$$\text{hit}^{(n)}_{\beta}(2\epsilon) \leq \text{hit}^{(n)}_{\alpha}(2\epsilon) \leq \text{hit}^{(n)}_{\alpha}(\epsilon) \quad \text{and} \quad \text{hit}^{(n)}_{\beta}(1 - \epsilon) \leq \text{hit}^{(n)}_{\beta}(1 - 2\epsilon) \leq \text{hit}^{(n)}_{\alpha}(1 - 2\epsilon). \quad (3.14)$$

Hence

$$\text{hit}^{(n)}_{\beta}(2\epsilon) - \text{hit}^{(n)}_{\beta}(1 - 2\epsilon) \leq \text{hit}^{(n)}_{\alpha}(\epsilon) - \text{hit}^{(n)}_{\alpha}(1 - \epsilon) + s_n, \quad (3.15)$$

$$\text{hit}^{(n)}_{\alpha}(2\epsilon) - \text{hit}^{(n)}_{\alpha}(1 - 2\epsilon) \leq \text{hit}^{(n)}_{\beta}(\epsilon) - \text{hit}^{(n)}_{\beta}(1 - \epsilon) + s_n.$$

Note that by the assumption that the product condition holds, we have that $s_n = o(t^{(n)}_{\text{mix}})$. Assume that the sequence exhibits hit$_{\alpha}$-cutoff. Then by (3.11) the RHS of the first line of (3.15) is $o(t^{(n)}_{\text{mix}})$. Again by (3.11), this implies that the RHS of the first line of (3.15) is $o(\text{hit}^{(n)}_{\beta}(1/4))$ and so the sequence exhibits hit$_{\beta}$-cutoff. Applying the same reasoning, using the second line of (3.15), shows that if the sequence exhibits hit$_{\beta}$-cutoff, then it also exhibits hit$_{\alpha}$-cutoff.

We now prove (3.12). Let $a \in (0, 1)$. Denote $\alpha := \min\{a, 1/2\}$ and $\beta := \max\{a, 1/2\}$. Let $s_n = s_n(\alpha, \beta, \epsilon)$ be as before. By the second inequality in Corollary 3.4

$$\text{hit}^{(n)}_{\alpha}(1/4 + \epsilon) - s_n \leq \text{hit}^{(n)}_{\beta}(1/4) \leq \text{hit}^{(n)}_{\alpha}(1/4). \quad (3.16)$$

By assumption (2) together with the product condition and (3.11), the LHS of (3.16) is at least $(1 - o(1))\text{hit}^{(n)}_{\alpha}(1/4)$, which by (3.16), implies (3.12).

**Proof of Theorem 3.** By Proposition 3.6 it suffices to show that the sequence exhibits cutoff iff it exhibits hit$_{1/2}$-cutoff. This follows at once from (1.4), (1.5) and (3.11).
3.2 Proof of Lemma 3.5

Now we prove Lemma 3.5. As mentioned before, the hitting time of a set $A$ starting from stationary initial distribution is well-understood (see [13]; for the continuous-time analog see [3], Chapter 3 Sections 5 and 6.5 or [5]). Assuming that the chain is lazy, it follows from the theory of complete monotonicity together with some linear-algebra that this distribution is dominated by a distribution which gives mass $\pi(A)$ to 0, and conditionally on being positive, is distributed as the Geometric distribution with parameter $\frac{\pi(A)}{t_{\text{rel}}}$. Since the existing literature lacks simple treatment of this fact (especially for the discrete-time case), for the sake of completeness, we now prove it. We shall prove this fact without assuming laziness. Although without assuming laziness the distribution of $T_A$ under $P_\pi$ need not be completely monotone, the proof is essentially identical as in the lazy case.

For any non-empty $A \subseteq \Omega$, we write $\pi_A$ for the distribution of $\pi$ conditioned on $A$. That is, $\pi_A(\cdot) := \frac{\pi(\cdot \mid 1 \in A)}{\pi(A)}$.

**Lemma 3.7.** Let $(\Omega, P, \pi)$ be a reversible irreducible finite chain. Let $A \subsetneq \Omega$ be non-empty. Denote its complement by $\overline{A}$ and write $\overline{B} = |B|$. Consider the sub-stochastic matrix $P_B$, which is the restriction of $P$ to $B$. That is $P_B(x,y) := P(x,y)$ for $x,y \in B$. Assume that $P_B$ is irreducible, that is, for any $x,y \in B$, exists some $t \geq 0$ such that $P_B^t(x,y) > 0$. Then

(i) $P_B$ has $k$ real eigenvalues $1 - \frac{\pi(A)}{t_{\text{rel}}} \geq \gamma_1 > \gamma_2 \geq \cdots \geq \gamma_k \geq -\gamma_1$.

(ii) There exist some non-negative $a_1, \ldots, a_k$ such that for any $t \geq 0$ we have that

$$P_{\pi_B}[T_A > t] = \sum_{i=1}^k a_i \gamma_i^t. \quad (3.17)$$

(iii) $$P_{\pi_B}[T_A > t] \leq \left(1 - \frac{\pi(A)}{t_{\text{rel}}} \right)^t \leq \exp \left(-\frac{t\pi(A)}{t_{\text{rel}}} \right), \text{ for any } t \geq 0. \quad (3.18)$$

**Proof.** We first note that (3.18) follows immediately from (3.17) and (i). Indeed, plugging $t = 0$ in (3.17) yields that $\sum_{i} a_i = 1$. Since by (i), $|\gamma_i| \leq \gamma_1 \leq 1 - \frac{\pi(A)}{t_{\text{rel}}}$ for all $i$, (3.17) implies that $P_{\pi_B}[T_A > t] \leq \gamma_1^t \leq \left(1 - \frac{\pi(A)}{t_{\text{rel}}} \right)^t$ for all $t \geq 0$.

We now prove (i). Consider the following inner-product on $\mathbb{R}^B$ defined by $\langle f,g \rangle_{\pi_B} := \sum_{x \in B} \pi_B(x)f(x)g(x)$. Since $P$ is reversible, $P_B$ is self-adjoint w.r.t. this inner-product. Hence indeed $P_B$ has $k$ real eigenvalues $\gamma_1 > \gamma_2 \geq \cdots \geq \gamma_k$ and there is a basis of $\mathbb{R}^B$, $g_1, \ldots, g_k$ of orthonormal vectors w.r.t. the aforementioned inner-product, such that $P_B g_i = \gamma_i g_i$ ($i \in [k]$).

By the Perron-Frobenius Theorem $\gamma_1 > 0$ and $\gamma_1 \geq -\gamma_k$. Since $P_B$ is strictly sub-stochastic, $\gamma_1 < 1$.

The claim that $1 - \gamma_1 \geq \frac{\pi(A)}{t_{\text{rel}}}$ follows by the Courant-Fischer characterization of the spectral gap and comparing the Dirichlet forms of $\langle \cdot, \cdot \rangle_{\pi_B}$ and $\langle \cdot, \cdot \rangle_{\pi}$ (c.f. Lemma 2.7 in [10] or Theorem 3.3 and Corollary 3.4 in Section 6.5 of Chapter 3 in [3]). This concludes the proof of part (i). We now prove part (ii).
By summing over all paths of length $t$ which are contained in $B$ we get that
\begin{equation}
    P_{\pi_B}[T_A > t] = \sum_{x,y \in B} \pi_B(x) P_B^t(x,y).
\end{equation}

By the spectral representation (c.f. Lemma 12.2 in [16] and Section 4 of Chapter 3 in [3]) for any $x, y \in B$ and $t \in \mathbb{N}$ we have that $P_B^t(x,y) = \sum_{i=1}^{k} \pi_B(x) g_i(x) g_i(y) \gamma_i^t$. So by (3.19)
\begin{equation}
    P_{\pi_B}[T_A > t] = \sum_{x,y \in B} \pi_B(x) \sum_{i=1}^{k} \pi_B(y) g_i(x) g_i(y) \gamma_i^t = \sum_{i=1}^{k} \left( \sum_{x \in B} \pi_B(x) g_i(x) \right)^2 \gamma_i^t.
\end{equation}

**Proof of Lemma 3.5.** We first note that (3.9) follows easily from (3.8). For the first inequality in (3.9) denote $B := B(A,w,\alpha) = \{ y : P_y \left[ T_A > \left[ \frac{t_{rel}}{\pi(A)} \right] \right] \geq \alpha \}$ and $t = t(A,w) := \left[ \frac{t_{rel}}{\pi(A)} \right]$. Then by (3.8)
\begin{align*}
    \alpha \pi(B) &\leq \pi(B) P_{\pi_B}[T_A > t] \leq P_\pi[T_A > t] \leq \pi(A^c) \exp \left( - \frac{t \pi(A)}{t_{rel}} \right) \leq \pi(A^c) e^{-w}.
\end{align*}

We now prove (3.8). Denote the connected components of $A^c := \Omega \setminus A$ by $\{C_1, \ldots, C_k\}$. Denote the complement of $C_i$ by $C_i^c$. By (3.18) we have that
\begin{align*}
    P_\pi[T_A > t] &\leq \sum_{i=1}^{k} \pi(C_i) P_{\pi_{C_i}}[T_A > t] = \sum_{i=1}^{k} \pi(C_i) P_{\pi_{C_i}}[T_{C_i^c} > t] \leq \\
    \sum_{i=1}^{k} \pi(C_i) \exp \left( - \frac{t \pi(C_i^c)}{t_{rel}} \right) &\leq \sum_{i=1}^{k} \pi(C_i) \exp \left( - \frac{t \pi(A)}{t_{rel}} \right) = \pi(A^c) \exp \left( - \frac{t \pi(A)}{t_{rel}} \right). \quad \square
\end{align*}

4 Continuous-time

In this section we explain the necessary adaptations in the proof of Proposition 1.6 for the continuous-time case. We fix some finite, irreducible, reversible chain $(\Omega, P, \pi)$. For notational convenience, exclusively for this section, we shall denote the transition-matrix of the non-lazy version of the discrete-time chain, by $P$, and that of the lazy version of the chain by $P_L := (P + I)/2$.

We denote the eigenvalues of $P$ by $1 = \lambda_1^{ct} > \lambda_2^{ct} \geq \cdots \geq \lambda_\ell^{ct} \geq -1$ and that of $P_L$ by $1 = \lambda_1^{ct} > \lambda_2^{ct} \geq \cdots \geq \lambda_{\ell/2}^{ct} \geq -1$ (where $1 + \lambda_1^{ct} = 2\lambda_2^{ct}$). We denote $t_{rel}^{ct} := (1 - \lambda_2^{ct})^{-1}$ and $t_{rel}^L := (1 - \lambda_2^L)^{-1}$. We identify $H_t$ with the operator $H_t : L^2(\mathbb{R}_+, \pi) \rightarrow L^2(\mathbb{R}_+, \pi)$, defined by $H_t f(x) = \mathbb{E}_x[f(X^c_t)]$. The spectral decomposition in continuous time takes the following form. If $f_1, \ldots, f_{\ell/2}$ is an orthonormal basis such that $P f_i := \lambda_i^{ct} f_i$ for all $i$, then
Lemma takes the following form in continuous-time:

\[ H_t g = E_\pi H_t g + \sum_{i=2}^{[\Omega]} \langle g, f_i \rangle \pi e^{-(1-\lambda_i t)} f_i, \]  
for all \( g \in \mathbb{R}^\Omega \) and \( t \geq 0 \). Thus the \( L^2 \)-contraction

\[
\text{Var}_\pi H_t f \leq e^{-2t/\tau_{\text{rel}}} \text{Var}_\pi f, \]  
for any \( f \in \mathbb{R}^\Omega \), for any \( t \geq 0 \). (4.1)

Starr’s inequality holds also in continuous-time ([23] Proposition 3) and takes the following form. Let \( f \in \mathbb{R}^\Omega \). Define the continuous-time maximal function as \( f_{\text{ct}}^*(x) := \sup_{t \geq 0} |H_t f(x)| \). Then

\[
\|f_{\text{ct}}^*\|_2 \leq 2\|f\|_2. \]  
(4.2)

We note that our proof of Theorem 2.3 can easily be adapted to the continuous-time case.

For any \( A \subseteq \Omega \) and \( s \in \mathbb{R}_+ \), set \( \rho(A) := \sqrt{\pi(A)(1 - \pi(A))} \) and \( \sigma_s^{\text{ct}} := \rho(A)e^{s/\tau_{\text{rel}}} \). Define

\[
G_{\text{ct}}^s(A, m) := \{ y : |Q_y^k(A) - \pi(A)| < m\sigma_s^{\text{ct}} \text{ for all } k \geq s \},
\]

where \( Q_y^k(A) = \sum_{a \in A} H_k(y, a) \). Then similarly to Corollary 2.4, combining (4.1) and (4.2) (in continuous-time there is no need to treat odd and even times separately) yields

\[
\pi(G_{\text{ct}}^s(A, m)) \geq 1 - 4/m^2, \]  
for all \( A \subseteq \Omega, s \geq 0 \) and \( m > 0 \). (4.3)

The proof of Corollary 3.1 carries over to the continuous-time case (where everywhere in (3.1)-(3.4), \( t_{\text{mix}} \) and \( \text{hit} \) are replaced by \( t_{\text{mix}}^{\text{ct}} \) and \( \text{hit}^{\text{ct}} \), respectively, and all ceiling signs are omitted), using (4.3) rather than (2.3) as in the discrete-time case.

In Lemma 3.7, we showed that for any non-empty \( A \subseteq \Omega \) such that \( P_{A^c} \) is irreducible, \( P_{A^c} \) has \( k \) real eigenvalues \( 1 - \pi(A)/\tau_{\text{rel}} \geq \gamma_1 > \gamma_2 \geq \cdots \geq \gamma_k \geq -\gamma_1 \) and that there exists some convex combination \( a_1, \ldots, a_k \) such that \( P_{\pi A^c}[T_A > t] = \sum_{i=1}^k a_i \gamma_i^t \leq (1 - \pi(A)/\tau_{\text{rel}})^t \leq e^{-\pi(A)/\tau_{\text{rel}}} \), for any \( t \geq 0 \). Repeating the argument while using the spectral decomposition of \( (H_{A^c}) \) (the restriction of \( H_t \) to \( A^c \)) in continuous-time, rather than the discrete-time spectral decomposition, yields that \( Q_{\pi A^c}[T_A > t] = \sum_{i=1}^k a_i e^{-(1-\gamma_i)t} \leq e^{-\pi(A)/\tau_{\text{rel}}} \), for any \( t \geq 0 \). Consequently, as in Lemma 3.5, \( B_{\text{ct}}(A, w, \alpha) := \{ y : Q_y \left[ T_A \geq \frac{\tau_{\text{ct}} w}{\pi(A)} \right] \geq \alpha \} \) satisfies that

\[
\pi(B_{\text{ct}}(A, w, \alpha)) \leq \pi(A^c) e^{-w} \alpha^{-1} \text{ for all } w \geq 0 \text{ and } 0 < \alpha \leq 1. \]  
(4.4)

Using (4.4) rather than (3.9), Corollary 3.4 is extended to the continuous-time case. Namely, for any reversible irreducible finite chain and any \( 0 < \epsilon < \delta < 1 \),

\[
\text{hit}_{\beta}^{\text{ct}}(\delta) \leq \text{hit}_{\alpha}^{\text{ct}}(\delta) \leq \text{hit}_{\beta}^{\text{ct}}(\delta - \epsilon) + \alpha^{-1} t_{\text{rel}}^{\text{ct}} \log \left( \frac{1 - \alpha}{1 - \beta \epsilon} \right), \]  
for any \( 0 < \alpha \leq \beta < 1 \). (4.5)

Finally, using (4.5), rather than (3.7) as in the discrete-time case, together with the version of Corollary 3.1 for the continuous-time chain, the proof of Proposition 1.6 for the continuous-time case is concluded in the same manner as the proof in the discrete-time case.
5 Trees

We start with a few definitions. Let \( T := (V,E) \) be a finite tree. Throughout the section we fix some lazy Markov chain, \( (V,P,\pi) \), on a finite tree \( T := (V,E) \). That is, a chain with stationary distribution \( \pi \) and state space \( V \) such that \( P(x,y) > 0 \) iff \( \{x,y\} \in E \) or \( y = x \) (in which case, \( P(x,x) \geq 1/2 \)). Then \( P \) is reversible by Kolmogorov’s cycle condition.

Following [20], we call a vertex \( v \in V \) a central-vertex if each connected component of \( T \setminus \{v\} \) has stationary probability at most 1/2. A central-vertex always exists (and there may be at most two central-vertices). Throughout, we fix a central-vertex \( o \) and call it the root of the tree. We denote a (weighted) tree with root \( o \) by \((T,o)\).

Loosely speaking, the analysis below shows that a chain on a tree satisfies the product condition iff it has a global bias towards \( o \). A non-intuitive result is that one can construct such unweighed trees [21].

The root induces a partial order \( \prec \) on \( V \), as follows. For every \( u \in V \), we denote the shortest path between \( u \) and \( o \) by \( \ell(u) = (u_0 = u, u_1, \ldots, u_k = o) \). We call \( f_u := u_1 \) the parent of \( u \) and denote \( \mu_u := P(u,f_u) \). We say that \( u' \prec u \) if \( u' \in \ell(u) \). Denote \( W_u := \{ v : u \in \ell(v) \} \). Recall that for any \( \emptyset \neq A \subset V \), we write \( \pi_A \) for the distribution of \( \pi \) conditioned on \( A \), \( \pi_A(\cdot) := \frac{\pi(\cdot)1_{\epsilon \leq A}}{\pi(A)} \).

**Definition 5.1.** We call \( x \in V \) a \( \beta \)-vertex if \( \pi(W_x) \geq \beta \) (\( \beta \in [0,1] \)). For any vertex \( x \) define \( y_\beta(x) \) to be the \( \beta \)-vertex closest to \( x \) in \( \ell(x) \) (if \( \pi(W_x) \geq \beta \), \( y_\beta(x) := x \)). For any \( p \in (0,1) \) we set \( \tau_\beta(p) := \min\{t : P_x[T_{y_\beta(x)} > t] \leq p\} \). In simple word, \( \tau_\beta(p) \) is the minimal time by which for every \( x \), the chain started from \( x \) hits the nearest \( \beta \)-vertex to \( x \) with probability at least \( 1 - p \).

Note that for any \( x \neq o \), necessarily \( \pi(W_x) \leq 1/2 \), as \( W_x \) is contained in one of the connected components of \( T \setminus \{o\} \) (while \( \pi(W_o) = 1 \)). Thus \( \pi(V \setminus W_x) \geq 1/2 \geq \pi(W_x) \).

Fix some \( 0 < \epsilon \leq 1/2 \). A key observation is that if \( y \) is such that \( \pi(W_y) \geq \epsilon/2 \), then the chain started from \( y \) mixes rapidly (this follows implicitly from the following lemma). More precisely, there exists an absolute constants \( 0 < c_1 < c_2 \) such that \( \text{hit}_{1-\epsilon/2,y}(\epsilon/2) \leq c_1 t_{\text{rel}} \log(1/\epsilon) =: t_1 \), where for any \( \nu \in \mathcal{P}(\Omega) \) and \( \alpha, \delta \in (0,1) \), \( \text{hit}_{\alpha,\nu}(\delta) := \min\{t : \max_{A \in \Omega} \pi_A[T_A > t] \leq \delta \} \). Consequently, \( d(c_2 t_{\text{rel}} \log(1/\epsilon), y) \leq 1 - \epsilon \). Whence if \( y \prec x \), and \( A \) is an arbitrary set such that \( \pi(A) \geq 1 - \epsilon/2 \), and if \( P_x[T_y \leq t_2] \geq 1 - \epsilon/2 \), then \( P_x[T_A > t_1 + t_2] \leq P_x[T_y > t_2] + P_y[T_A > t_1] \leq \epsilon \). Thus one can replace \( \text{hit}_{1-\epsilon}(\cdot) \) by the somewhat simpler expression \( \tau_\epsilon(\cdot) \).

The following lemma makes the above observations precise and shows that \( \tau_\epsilon(\cdot) \) satisfies similar relations with \( t_{\text{mix}} \) as the ones described between \( \text{hit}_{1-\epsilon}(\cdot) \) and \( t_{\text{mix}} \) in Corollary 3.1. In other words, one can use the tree structure to bound the \( \epsilon \) and \( 1 - \epsilon \)-mixing times, in terms of \( t_{\text{mix}} \), \( \min_{y} P_x[T_{W_y} \geq t] : x \in W_y : \pi(W_y) \geq 1 - \epsilon/2 \).

**Lemma 5.2.** Let \( \beta \in (0,1/4) \). Set \( s = s_\beta := \lceil 4 t_{\text{rel}} \log(6/\beta) \rceil \) and \( s' = s'_\beta := \lceil 3 \log(4/\beta) \rceil \).

Then \( t_{\text{mix}}(1 - \beta) \geq \tau_{\beta/2}(1 - \beta/2) \) and \( t_{\text{mix}}(\beta) \leq \tau_{\beta/2}(\beta/2) + s_\beta + s'_\beta \). (5.1)

**Proof.** The first inequality holds since \( t_{\text{mix}}(1 - \beta) \geq \text{hit}_{1-\beta/2}(1 - \beta/2) \geq \tau_{\beta/2}(1 - \beta/2) \). We now prove the second inequality. Let \( x \in V \). If \( x \neq o \) set \( A = A_x := W_{y_{\beta/2}(x)} \) and \( \beta' := \pi(A) \).
If \( x = o \), pick some \( y \sim o \) such that \( \pi(W_y) = \max_{y \sim o} \pi(W_o) := \beta' \leq 1/2 \) and set \( A := W_y \).

Let \( B := A^c \). Let \( C \subseteq V \) be an arbitrary set such that \( \pi(C) \geq 1 - \beta'/2 \). Denote \( D := C \cap B \).

By the Markov property and the tree structure (namely, the fact that in order for the chain with initial distribution \( \pi_A \) to hit \( D \), it must first hit \( y_{\beta/2}(x) \)), for any \( t \geq 0 \) we have that

\[
P_{\pi_A}[T_D \leq t] = \sum_{i=0}^{t} P_{\pi_A}[T_{y_{\beta/2}(x)} = i]P_{y_{\beta/2}(x)}[T_D \leq t - i]
\]

(5.2)

By (3.8) and the fact that \( \pi(D) \geq 1 - 3\beta'/2 \geq 1/4 \),

\[
\beta' P_{\pi_A}[T_D > s] = P_{\pi_A}[T_D > s] \pi(A) \leq P_x[T_D > s] \leq (1 - \pi(D)) e^{-\log(6/\beta)} \leq \frac{3\beta'}{2} \beta/6 = \beta' \beta/4.
\]

By (5.2) \( P_{y_{\beta/2}(x)}[T_C > s] \leq P_{\pi_A}[T_D > s] \leq \beta/4 \). Since \( C \) was arbitrary, it follows that \( \text{hit}_{1-\beta/4,y_{\beta/2}(x)}(\beta/4) \leq s_\beta \).

It now follows by the Markov property that

\[
\text{hit}_{1-\beta/4,x}(3\beta/4) \leq \tau_{\beta/2}(\beta/2) + \text{hit}_{1-\beta/4,y_{\beta/2}(x)}(\beta/4) \leq \tau_{\beta/2}(\beta/2) + s_\beta.
\]

(5.3)

The right inequality in (3.3) together with (5.3) imply that \( t_{\text{mix}}(\beta) \leq \tau_{\beta/2}(\beta/2) + s_\beta + s'_\beta \). \( \square \)

In light of Lemma 5.2, in order to show that in the setup of Theorem 1 (under the product condition) cutoff occurs is to show that \( \tau_{\beta/2}(\beta/2) - \tau_{\beta/2}(1 - \beta/2) = o(t_{\text{mix}}^{(n)}(\beta)), \) for any \( \beta \in (0, 1/2) \). We actually show more than that. Instead of identifying the “worst” couple \((x,y_{\beta/2}(x))\) and prove that \( T_{y_{\beta/2}(x)} \) is concentrated under \( P_x \), we shall show that for any \( x, y \in V_n \) such that \( y \prec x \) and \( E_x[T_y] = \Theta(t_{\text{mix}}^{(n)}), T_y \) is concentrated under \( P_x \), around \( E_x[T_y] \), with deviations of order \( \sqrt{t_{\text{rel}}^{(n)} t_{\text{mix}}^{(n)}} \). This shall follow from Chebyshev inequality, once we establish that \( \Var_y[T_A] \leq 4t_{\text{rel}} E_x[T_y] \).

Let \( (v_0 = x, v_1, \ldots, v_k = y) \) be the path from \( x \) to \( y \) \((y \prec x)\). Define \( \tau_i := T_{v_i} - T_{v_{i-1}} \). Then by the tree structure, under \( P_x \) we have that \( T_y = \sum_{i=1}^{k} \tau_i \) and that \( \tau_1, \ldots, \tau_k \) are independent. This reduces the task of bounding \( \Var_x[T_y] \) from above, to the task of estimating \( \Var_{v_i}[T_{v_{i+1}}] = \Var_{v_i}[T_{v_i}] \) from above for each \( i \).

**Lemma 5.3.** Let \( \beta \in (0, 1/2) \). For any vertex \( u \) such that \( \pi(W_u) \leq \beta \) we have that

\[
t_u := E_u[T_{f_u}] = \frac{\pi(W_u)}{\pi(u)} \mu_u \quad \text{and} \quad r_u := E_u[T_{f_u}^2] = 2t_u E_{\pi u}[T_{f_u}] - t_u \leq \frac{2t_u t_{\text{rel}}}{1 - \beta}.
\]

(5.4)

The assertion of Lemma 5.3 follows as a particular case of Proposition 5.8 at the end of this section.

**Corollary 5.4.** Let \( x \in V \). Let \( 0 \leq \beta \leq 1/2 \) and \( c \geq 0 \). Denote \( \sigma_{x,\beta} := \sqrt{2E_x[T_{y_{\beta/2}(x)}] t_{\text{rel}}}/(1 - \beta) \).

Then

\[
\Var_x[T_{y_{\beta/2}(x)}] \leq \sigma_{x,\beta}^2.
\]

(5.5)
and

\[ P_x[T_{y_\beta(x)} \geq \mathbb{E}_x[T_{y_\beta(x)}] + c\sigma_{x,\beta}] \leq \frac{1}{1 + c^2} \text{ and } P_x[T_{y_\beta(x)} \leq \mathbb{E}_x[T_{y_\beta(x)}] - c\sigma_{x,\beta}] \leq \frac{1}{1 + c^2}. \quad (5.6) \]

In particular, if \((V_n, P_n, \pi_n)\) is a sequence of lazy Markov chains on trees \((T_n, o_n)\) which satisfies the product condition, and \(x_n, y_n \in V_n\) satisfy that \(y_n < x_n\) and \(\mathbb{E}_{x_n}[T_{y_n}] / t_{rel}^{(n)} \to \infty\), then for any \(\epsilon > 0\) we have that

\[ \lim_{n \to \infty} P_x[x_n \cdot (T_{y_n} - \mathbb{E}_{x_n}[T_{y_n}]) \geq \epsilon \mathbb{E}_{x_n}[T_{y_n}]] = 0. \quad (5.7) \]

Proof. We first note that (5.6) follows from (5.5) by the one-sided Chebyshev inequality. Also, (5.7) follows immediately from (5.6). We now prove (5.5). Let \((v_0 = x, v_1, \ldots, v_k = y_\beta(x))\) be the path from \(x\) to \(y_\beta(x)\). Define \(\tau_i := T_{v_i} - T_{v_{i-1}}\). Then by the tree structure, under \(P_x\), we have that \(T_{y_\beta(x)} = \sum_{i=1}^k \tau_i\) and that \(\tau_1, \ldots, \tau_k\) are independent. Whence, by (5.4) we get that

\[ \text{Var}_x[T_{y_\beta(x)}] = \sum_{i=1}^k \text{Var}_x[\tau_i] = \sum_{i=1}^k \text{Var}_{v_{i-1}}[T_{v_i}] \leq \sum_{i=1}^k \mathbb{E}_{v_{i-1}}[T_{v_i}^2] \leq \frac{2t_{rel}}{1 - \beta} \sum_{i=1}^k \mathbb{E}_{v_{i-1}}[T_{v_i}] = \sigma_{x,\beta}^2. \]

Lemma 5.5. Let \((\Omega, P, \pi)\) be a finite lazy irreducible Markov chain. Then for any \(x \in \Omega\) and any \(A \subset \Omega\), such that \(\pi(A) \geq 3/4\) we have that \(\mathbb{E}_x[T_A] \leq 2t_{mix}\). In particular, if \((V, P, \pi)\) is a lazy chain on a (weighted) tree \((T, o)\), then

\[ \mathbb{E}_x[T_{y_\beta(x)}] \leq 2t_{mix}, \text{ for any } x \in V \text{ and any } 0 \leq \beta \leq 1/4. \quad (5.8) \]

Proof. Let \(A \subset V\) be such that \(\pi(A) \geq 3/4\). Consider \(\tau_A := \inf\{k \in \mathbb{N} : X_{kt_{mix}}^x \in A\}\). Then \(T_A \leq \tau_A t_{mix}\). By the Markov property and the definition of the total variation distance, the distribution of \(\tau_A\) is stochastically dominated by the Geometric distribution with parameter \(3/4 - 1/4 = 1/2\). Hence \(\mathbb{E}_x[T_A] \leq t_{mix} \mathbb{E}_x[\tau_A] \leq 2t_{mix}\).

This implies (5.8), since if \(B := \{z \in V \setminus \{y_\beta(x)\} : y_\beta(x) < z\}\) and \(A = B^c\), then \(\pi(A) \geq 1 - \beta \geq 3/4\), and by the tree structure \(T_{y_{\beta}(x)} = T_A\), under \(P_x\).

Corollary 5.6. In the setup of Lemma 5.3, let \(\beta \in (0, 1/4]\). For any \(x \in V\) denote \(t_{x,\beta} := \mathbb{E}_x[T_{y_\beta(x)}]\). Denote

\[ \rho_\beta := \max_{x \in V} t_{x,\beta}, \text{ and } \kappa_\beta := \sqrt{2\beta^{-1}\rho_\beta t_{rel}}, \text{ then } \]

\[ \rho_\beta \leq 2t_{mix}, \tau_\beta(1 - \beta) \geq \rho_\beta - \kappa_\beta \text{ and } \tau_\beta(\beta) < \rho_\beta + \kappa_\beta. \quad (5.9) \]

In particular,

\[ \tau_{\beta/2}(\beta/2) - \tau_{\beta/2}(1 - \beta/2) \leq 2\kappa_{\beta/2} \leq 4\sqrt{\beta^{-1}\rho_\beta t_{mix}}. \quad (5.10) \]

Proof. By (5.8) \(\rho_\beta \leq 2t_{mix}\). Denote \(\sigma_\beta := \sqrt{\frac{2\rho_\beta t_{rel}}{1 - \beta}}\) and \(c_\beta := \sqrt{\beta^{-1}} - 1\). Let \(x \in V\) be such that \(\pi(W_x) < \beta\). By (5.5) \(\sigma^2_{x,\beta} := \text{Var}_x[T_{y_\beta}] \leq \sigma^2_\beta\). The assertion of the corollary now follows from (5.6) by noting that \(c_\beta \sigma_\beta = \kappa_\beta\).
Finally, the next lemma bounds the term $s_\beta + s'_\beta$.

**Lemma 5.7.** Assume that $|V| \geq 3$. Let $0 < \beta \leq 1/4$. Then in the setup of Lemma 5.2 we have that $t_{rel} \leq 6 t_{mix}$ and $s_\beta + s'_\beta \leq 23 \sqrt{\beta^{-1} t_{rel} t_{mix}}$.

**Proof.** For any irreducible Markov chain on $n > 1$ states we have that $\lambda_2 \geq - \frac{1}{n-1}$ ([3], Chapter 3 Proposition 3.18). Hence for a lazy chain with at least 3 states we have that $t_{rel} \geq 4/3$ and so by (1.2) $t_{rel} \leq 6(t_{rel} - 1) \log 2 \leq 6 t_{mix}$. Hence

$$s_\beta = [4 t_{rel} \log(6/\beta)] \leq [8 \sqrt{6 t_{rel} t_{mix}} \log(\sqrt{6/\beta})] \leq [8 \sqrt{6 t_{rel} t_{mix}}(\frac{\sqrt{6/\beta}}{3})] \leq 16 \sqrt{\beta^{-1} t_{rel} t_{mix}} + 1,$$

where we have used the fact that $\log x \leq x/3$, for any $x \geq \sqrt{24}$. Similarly,

$$s'_\beta = [3 t_{rel} \log(2/\beta^{1/2})] \leq 3 \sqrt{6 t_{rel} t_{mix}} \log(2/\sqrt{\beta}) + 1 \leq 6 \sqrt{\beta^{-1} t_{rel} t_{mix}} + 1.$$

Finally, since $\beta^{-1} t_{rel} t_{mix} > 4$, we get that $s_\beta + s'_\beta \leq 23 \sqrt{\beta^{-1} t_{rel} t_{mix}}$. 

These lemmata put together establish Theorem 1.

**Proof of Theorem 1.** The proof follows from (5.1) together with Lemma 5.7 and (5.10). 

As promised earlier, the following proposition implies the assertion of Lemma 5.3. For any set $A \subset \Omega$, we define $\psi_{A^c} \in \mathcal{P}(A^c)$ as $\psi_{A^c}(y) := P_{\pi_A}[X_1 = y \mid X_1 \in A^c]$. For $A \subset \Omega$, we denote $T_A^+ := \inf\{t \geq 1 : X_t \in A\}$ and $\Phi(A) := \sum_{a \in A, b \in A^c} \frac{\pi(a) P(a, b)}{\pi(A)} = P_{\pi_A}[X_1 \notin A]$. Note that

$$\pi(A) \Phi(A) = \sum_{a \in A, b \in A^c} \pi(a) P(a, b) = \sum_{a \in A, b \in A^c} \pi(b) P(b, a) = \pi(A^c) \Phi(A^c). \quad (5.11)$$

This is true even without reversibility, since the second term (resp. third term) is the asymptotic frequency of transitions from $A$ to $A^c$ (resp. from $A^c$ to $A$).

**Proposition 5.8.** Let $(\Omega, P, \pi)$ be a finite irreducible reversible Markov chain. Let $A \subset \Omega$ be non-empty. Denote the complement of $A$ by $B$. Then

$$P_{\pi_B}[T_A = t]/\Phi(B) = P_{\psi_B}[T_A \geq t], \text{ for any } t \geq 1. \quad (5.12)$$

Consequently,

$$E_{\psi_B}[T_A] = \frac{1}{\Phi(B)} \text{ and } E_{\psi_B}[T_A^2] = E_{\psi_B}[T_A] (2E_{\pi_B}[T_A] - 1) \leq \frac{2E_{\psi_B}[T_A] t_{rel}}{\pi(A)}. \quad (5.13)$$

**Proof.** We first note that the inequality $2E_{\psi_B}[T_A] E_{\psi_B}[T_A] \leq \frac{2E_{\psi_B}[T_A] t_{rel}}{\pi(A)}$ follows from the second inequality in (3.9) (this is the only part of the proposition which relies upon reversibility).

Summing (5.12) over $t$ yields the first equation in (5.13). Multiplying both sides of (5.12) by $2t - 1$ and summing over $t$ yields the second equation in (5.13). We now prove (5.12). Let $t \geq 1$. Then

$$\pi(B) P_{\pi_B}[T_A = t] = P_{\pi}[T_A = t] = P_{\pi}[T_A^+ = t + 1] = P_{\pi}[X_1 \notin A, \ldots, X_t \notin A, X_{t+1} \in A]$$

$$= P_{\pi}[X_1 \notin A, \ldots, X_t \notin A] - P_{\pi}[X_1 \notin A, \ldots, X_t \notin A, X_{t+1} \notin A]$$

$$= P_{\pi}[X_1 \notin A, \ldots, X_t \notin A] - P_{\pi}[X_0 \notin A, \ldots, X_t \notin A] = P_{\pi}[X_0 \in A, X_1 \notin A, \ldots, X_t \notin A]$$

$$= \pi(A) \Phi(A) P_{\psi_B}[X_0 \notin A, \ldots, X_{t-1} \notin A] = \pi(A) \Phi(A) P_{\psi_B}[T_A \geq t],$$

which by (5.11) implies (5.12). 

\square
6 Refining the bound for trees

The purpose of this section is to improve the concentration estimate (5.6). As a motivating example, consider a lazy nearest neighbor random walk on a path of length $n$ with some fixed bias to the right. For concreteness, say, $\Omega := \{1, 2, \ldots, n\}$, $P_n(i, i) = 1/2$, $P_n(i, i-1) = 1/8$ and $P_n(i, i+1) = 3/8$ for all $1 < i < n$. Then $t_{\text{mix}}^{(n)} = 4n(1 + o(1))$ and $t_{\text{rel}}^{(n)} = \Theta(1)$.

In this case, there exists some constant $c_1 > 0$ such that for any $\lambda > 0$ we have that $P_1[|T_n - 4n| \geq \lambda \sqrt{n}] \leq 2e^{-c_1 \lambda^2}$. Observe that $\sqrt{t_{\text{mix}}^{(n)} t_{\text{rel}}^{(n)}} = \Theta(\sqrt{n})$. Hence there exists some constant $c_2$ such that $P_1 \left[ |T_n - 4n| \geq \lambda \sqrt{t_{\text{mix}}^{(n)} t_{\text{rel}}^{(n)}} \right] \leq 2e^{-c_2 \lambda^2}$. Using Proposition 1.6, it is not hard to show that this implies that $t_{\text{mix}}^{(n)} \leq 4n + c_3 \sqrt{\log(1/\epsilon) t_{\text{mix}}^{(n)} t_{\text{rel}}^{(n)}}$ and that $t_{\text{mix}}^{(n)} (1 - \epsilon) \geq 4n - c_4 \sqrt{\log(1/\epsilon) t_{\text{mix}}^{(n)} t_{\text{rel}}^{(n)}}$.

In Lemma 6.2 we show that for any lazy Markov chain on a tree $T = (V, E, o)$ and any $x \in V$ and $0 \leq \epsilon \leq 1/2$, we have that $P_x[|T_{y_k} - \mathbb{E}_x[T_{y_k}]| \geq \lambda \sqrt{\mathbb{E}_x[T_{y_k}] t_{\text{rel}}^{(n)}}] \leq 2e^{-c_3 \lambda^2}$. Besides being of independent interest, using Proposition 1.6, one can deduce from Lemma 6.2 that under the product condition,

$$
t_{\text{mix}}^{(n)}(1 - \epsilon) = O(1), \text{ for any } 0 < \epsilon \leq 1/4.
$$

The details of the derivation of (6.1) from Lemma 6.2 are left to the reader.

**Proposition 6.1.** Let $(\Omega, P, \pi)$ be a finite irreducible reversible Markov chain. Let $0 < \epsilon < 1$. Let $A \subseteq \Omega$ be such that $\pi(A) \geq 1 - \epsilon$. Denote the complement of $A$ by $B$. Denote $p := 1 - \frac{1}{t_{\text{rel}}}$ and $a := \mathbb{E}_B[T_A]$. Let $z > 1$ be such that $2p(z - 1) \leq 1 - p$. Then

$$\max(\mathbb{E}_B[z^{T_A - \mathbb{E}_B[T_A]}], \mathbb{E}_B[z^{\mathbb{E}_B[T_A] - T_A}]) \leq \exp \left[ \frac{2a(z - 1)^2}{1 - p} \right].$$

**Proof.** By (5.12) and (3.8)

$$\mathbb{E}_B[z^{T_A}] = \sum_{k \geq 1} z^k P_{\pi_B}[T_A = k] = 1 + (z - 1) \sum_{k \geq 1} z^{k-1} P_{\pi_B}[T_A \geq k]$$

$$= 1 + (z - 1)a \sum_{k \geq 1} z^{k-1} P_{\pi_B}[T_A = k] \leq 1 + (z - 1)a \sum_{k \geq 1} (1 - p)(pz)^k$$

$$\leq 1 + (z - 1)a \left(1 + \frac{2p(z - 1)}{1 - p}\right) \leq \exp[a(z - 1) + \frac{2ap(z - 1)^2}{1 - p}],$$

where in the penultimate inequality we have used the assumption that $2p(z - 1) \leq 1 - p$. We also have that

$$z^{-\mathbb{E}_B[T_A]} \leq \left(1 - (z - 1) + (z - 1)^2\right)^a \leq \exp[-a(z - 1) + a(z - 1)^2].$$

21
Denote \( t \) the cutoff for a sequence of \( \{x, \epsilon \} \).

In this section we prove Theorem 2 and establish that product condition is sufficient for

\[
\text{Proof. Lemma 6.2. Let } E = \sum_{k \geq 1} z^{-k} P_{\psi B}[T_A = k] = 1 - (1 - z^{-1}) \sum_{k \geq 1} z^{-(k-1)} P_{\psi B}[T_A \geq k] 
\]

\[
= 1 - (1 - z^{-1})a \sum_{k \geq 1} z^{-(k-1)} P_{\pi B}[T_A = k] = 1 - (1 - z^{-1})a \sum_{k \geq 1} (1 - p)(p/z)^{k-1} 
\]

\[
= 1 - (1 - z^{-1})a \left(1 - \frac{p/(1-z^{-1})}{1 - p/(1-z^{-1})}\right) \leq 1 - (1 - z^{-1})a \left(1 - \frac{2p(1-z^{-1})}{1 - p}\right) 
\]

\[
\leq \exp \left[ -a(1 - z^{-1}) + \frac{2ap(z - 1)^2}{1 - p} \right].
\]

We also have that \( z^E \psi_B[T_A] \leq (1 + (z - 1))^a \leq e^{a(z-1)}. \) Note that \( a(z-1) - a(1 - z^{-1}) = a(z - 1)^2 / z \leq a(z - 1)^2. \) Hence \( E \psi_B[z^E \psi_B[T_A] - T_A] \leq \exp \left[ a(z - 1)^2 \left(1 + \frac{2p}{1 - p}\right) \right] \leq \exp \left[ \frac{2a(z - 1)^2}{1 - p} \right]. \]

\[ \square \]

\textbf{Lemma 6.2. Let } (V, P, \pi) \text { be a Markov chain on a tree } (T, o). \text { Let } x \in V. \text { Let } 0 \leq \epsilon \leq 1/2. \text { Denote } t_{x, \epsilon} := E_x[T_{y_i(x)}] \text { and } b = b_{x, \epsilon} := \sqrt{t_{x, \epsilon}/t_{rel}}. \text { Then}

\[
P_x[T_{y_i(x)} - t_{x, \epsilon} \geq cb] \forall P_x[t_{x, \epsilon} - T_{y_i(x)} \geq cb] \leq e^{-c^2/20}, \text { for any } 0 \leq c \leq 5(1 - \epsilon) \sqrt{t_{x, \epsilon}/t_{rel}}. \quad \text{(6.6)}
\]

\textbf{Proof. Let } (v_0 = x, v_1, \ldots, v_k = y_i(x)) \text { be the path from } x \text { to } y_i(x). \text { Define } \tau_i := T_{v_i} - T_{v_{i-1}}. \text { Then by the tree structure, under } P_x, \text { we have that } T_{y_i(x)} = \sum_{i=1}^k \tau_i \text { and that } \tau_1, \ldots, \tau_k \text { are independent. Denote } p := 1 - \frac{1}{t_{rel}}. \text { Denote } a_i := E_x[\tau_i] \text { and that } 0 \leq c \leq 5(1 - \epsilon) \sqrt{t_{x, \epsilon}/t_{rel}}. \text { Set } z_c = z_{c, x, \epsilon} := 1 + \frac{c}{10b}. \text { Note that } 2p(z_c - 1) \leq \frac{c}{5b} \leq \frac{1}{t_{rel}} = 1 - p. \text { Then by (6.2) }

\[
P_x[T_{y_i(x)} - t_{x, \epsilon} \geq cb] = P_x[z_{c, x, \epsilon} - t_{x, \epsilon} \geq z_{c, x, \epsilon} cb] \leq E_x[z_{c, x, \epsilon} - t_{x, \epsilon} - z_{c, x, \epsilon} cb] = z_{c, x, \epsilon} cb \prod_{i=1}^k E_x[\tau_i - a_i] 
\]

\[
\leq \exp[(-z_{c, x, \epsilon} - 1 + (z_{c, x, \epsilon} - 1)^2)cb] \prod_{i=1}^k \exp \left[ \frac{2a_i(z_{c, x, \epsilon} - 1)^2}{1 - p} \right] \quad \text{(6.7)}
\]

\[
= \exp \left[ -\frac{c^2}{10} + \frac{c^3}{100b} \right] \exp \left[ \frac{2t_{rel}t_{x, \epsilon}c^2}{100b^2(1 - \epsilon)} \right] \leq \exp \left[ -\frac{c^2}{10} + \frac{c^3}{100b} + \frac{c^2}{25} \right] \leq e^{-c^2/20}. 
\]

The inequality \( P_x[t_{x, \epsilon} - T_{y_i(x)} \geq cb] \leq e^{-c^2/20} \) is proved in an analogous manner. \( \square \)

\section{Weighted random walks on the interval with bounded jumps}

In this section we prove Theorem 2 and establish that product condition is sufficient for cutoff for a sequence of \( (\delta, r) \)-SBD chains. Although we think of \( \delta \) as being bounded away from 0, and of \( r \) as a constant integer, it will be clear that our analysis remains valid as
long as $\delta$ does not tend to 0, nor does $r$ to infinity, too rapidly in terms of some functions of $t_{\text{rel}}/t_{\text{mix}}$.

Throughout the section, we use $C_1, C_2, \ldots$ to describe positive constants which depend only on $\delta$ and $r$. We call a state $i \in [n]$ a central-vertex if $\pi([i-1]) \cup \pi([n] \setminus [i]) \leq 1/2$. As opposed to the setting of Section 5, the sets $[i-1]$ and $[n] \setminus [i]$ need not be connected components of $[n] \setminus \{(i)\}$ w.r.t. the chain, in the sense that it might be possible for the chain to get from $[i-1]$ to $[n] \setminus [i]$ without first hitting $i$ (skipping over $i$). We pick a central-vertex $o$ and call it the root. As before, the root induces a partial order $\prec$, where $j \prec k$ if either $o \leq j \leq k$ or $k \leq j \leq o$.

We denote $W_j := \{ k \in [n] : j \prec k \}$. Let $0 \leq \epsilon \leq 1$. We say that $j \in [n]$ is an $\epsilon$-vertex if $\pi(W_j) \geq \epsilon$. For any $v \in V$ we define $y_v(v)$ to be the $\epsilon$-vertex closest to $v$ (i.e. $y_v(v) := \arg\min\{|y-v|^2 : y \text{ is an } \epsilon\text{-vertex}\}$). We note that $y_v(v)$ might be $v$ itself or might equal $o$, even when $\epsilon < 1/2$.

Recall that in Section 5 we exploited the tree structure to argue that if $y \prec x$ and $A \cap W_y$ is empty, then under both $P_x$ and $P_{\pi y}$ we have that $T_y \leq T_A$. This used to show that if $y$ is an $\epsilon$-vertex, then the chain started from $y$ mixes rapidly. This was also used to show that $\text{hit}_{1-\epsilon/2}(\epsilon/2)$ could be replaced by $\max_x \min\{t : P_x[T_{y_{\pi y}}(x) > t] \leq \epsilon/2\}$. Under condition (ii), Definition 1.2, the following lemma shows that if $y \prec x$ and $A \cap W_y$ is empty, then $E_x[T_A] \vee E_{\pi y}[T_A] \geq \delta^r E_y[T_A]$. We soon show that this allows one to extend the aforementioned facts which were used in the proof of Theorem 1 to our current setup.

**Lemma 7.1.** In the above setup, let $I := \{ v, v+1, \ldots, v+r-1 \} \subset [n]$. Let $\mu \in \mathcal{P}(I)$. Then

$$E_{\mu}[T_A] \leq \max_{y \in I} E_y[T_A] \leq \delta^{-r} \min_{x \in I} E_x[T_A], \text{ for any } A \subset \Omega \setminus I. \quad (7.1)$$

Consequently, for any $i \in I$ and $A \subset [v-1]$ (resp. $A \subset [n] \setminus [v+r-1]$) we have that

$$E_i[T_A] \leq \delta^{-r} E_{\pi_{[v-1]}}[T_A], \quad (\text{resp. } E_i[T_A] \leq \delta^{-r} E_{\pi_{[v+r-1]}}[T_A]). \quad (7.2)$$

**Proof.** We first note that (7.2) follows from (7.1). Indeed, by condition (i) of the definition of a $(\delta, r)$-SBD chain, if $A \subset [v-1]$ (resp. $A \subset [n] \setminus [v+r-1]$), then under $P_{\pi_{[v-1]}}$ (resp. under $P_{\pi_{[v+r-1]}}$), $T_I \leq T_A$. Thus (7.2) follows from (7.1) by averaging over $X_{T_i}$. We now prove (7.1).

Fix some $A$ such that $A \subset [n] \setminus I$. Fix some distinct $x, y \in I$. Let $B_1$ be the event that $T_y \leq T_A$. One way in which $B_1$ can occur is that the chain would move from $x$ to $y$ in $|y-x|$ steps such that $|X_k - X_{k-1}| = 1$ for all $1 \leq k \leq |y-x|$. Denote the last event by $B_2$. Then

$$E_x[T_A] \geq E_x[T_A 1_{B_2}] \geq P[B_2] E_y[T_A] \geq \delta^r E_y[T_A].$$

Minimizing over $x$ yields that for any $y \in I$ we have that $E_y[T_A] \leq \delta^{-r} \min_{x \in I} E_x[T_A]$, from which (7.1) follows easily.

**Lemma 7.2.** Let $(\Omega, P, \pi)$ be an irreducible finite lazy reversible Markov chain. Let $0 < \beta \leq 1/2$. Let $A, B \subset \Omega$ be disjoint sets such that $\pi(A) \geq \beta$ and $\pi(B) \geq 1 - 2\beta$. Then $E_{\pi A}[T_B] \leq 2t_{\text{rel}}$.

**Proof.** By (3.9) we have that

$$E_{\pi A}[T_B]/2 \leq \pi_B(A) E_{\pi A}[T_B] = E_{\pi B}(T_B 1_{X_0 \in A}) \leq E_{\pi B}[T_B] = \pi(B^c) E_{\pi}[T_B] \leq t_{\text{rel}}. \quad \Box$$
Corollary 7.3. Let $v \in [n]$ and $0 < \epsilon, \beta \leq 1/2$. Assume that $\pi(W_v) \geq \beta$. Let $C \subset [n]$ be such that $\pi(C) \geq 1 - \beta$. Then $\mathbb{E}_o[T_C] \leq 2\delta r t_{rel}$. In particular, $\text{hit}_{1-\beta,v}(\epsilon) \leq 2\epsilon^{-1}\delta^{-r}t_{rel}$ and consequently $d(3\epsilon^{-1}\delta^{-r}t_{rel}, v) \leq 2\epsilon$.

Proof. If $v \neq o$, set $A := W_v$. If $v = o$ and $\pi(C \cap \{o\}) \leq 1/3$, set $A := \{o\}$. If $v = o$ and $\pi(C \cap \{o\}) > 1/3$, set $A := [n] \setminus \{o\}$. Let $D := A^c$. Denote $B := C \cap D$. Then by Lemma 7.2, $\mathbb{E}_{\pi_A}[T_B] \leq 2t_{rel}$. By (7.2), $\mathbb{E}_v[T_C] \leq \mathbb{E}_o[T_B] \leq \delta^{-r}\mathbb{E}_{\pi_A}[T_B] \leq 2\delta^{-r}t_{rel}$. By Markov inequality, $\mathbb{P}_v[T_C \geq 2\epsilon^{-1}\delta^{-r}t_{rel}] \leq \epsilon$. Since $C$ was arbitrary, $\text{hit}_{1-\beta,v}(\epsilon) \leq 2\epsilon^{-1}\delta^{-r}t_{rel}$, which by the proof of (3.2) implies that $d(3\epsilon^{-1}\delta^{-r}t_{rel}, v) \leq 2\epsilon$.

Divide $[n]$ into $m := \lceil n/r \rceil$ consecutive disjoint intervals, $I_1, \ldots, I_m$ each of size $r$, apart from perhaps $I_m$. We call each such interval a block. Denote by $I_o$ the unique block such that the root $o$ belongs to it. For any $k \in [m]$ denote $A_k := \{x : \exists y \in I_k : y < x\}$ and $B_k := [n] \setminus A_k$. Fix some $0 \leq \epsilon \leq 1$. We call $k \in [m]$ an $\epsilon$-block if $\pi(A_k) \geq \epsilon$.

As Lemma 5.2 served to reduce the problem of establishing cutoff to establishing concentration of hitting times of $\epsilon$-vertices, Corollary 7.3 also reduces the proof of Theorem 2 to the problem of establishing concentration of hitting times of $\epsilon$-blocks. We omit the proof of the last statement to avoid repetitions.

Observe that the root induces a partial order also on the blocks. We say that $I_j \prec I_k$ if for any $x \in I_j$ and $y \in I_k$ we have that $x < y$. Let $j \in [m]$ and $0 \leq \epsilon \leq 1/2$. We define $b_k(j)$ to be the index of the $\epsilon$-block $I_k$ such that $k$ minimizes $|j - k|$ out of all $\epsilon$-blocks $I_k$ such that $I_k \prec I_j$. For any $j \in [m]$ we define the parent block of $j$ in the obvious manner and denote its index by $f_j$. We define

$$T(j) := T_{I_j} \text{ and } \bar{T}_j := T(f_j) - T(j).$$

To establish concentration of the hitting time of $I_{b_k(j)}$ from any $x \in I_j$, for $j$ such that $t_{rel} = o \left( \min_{x \in I_j} \mathbb{E}_x[T_{I_{b_k(j)}}] \right)$, we will bound $\text{Var}_x[\sum \bar{T}_j]$, where $x \in I_j$ is arbitrary, and the sum is taken over blocks between $I_j$ and $I_{b_k(j)}$. As opposed to the situation in Section 5, the terms in the sum are no longer independent. We now show that the correlation between them decays exponentially (Lemma 7.5) and that for all $\ell$ we have that $\text{Var}_x[\bar{T}_j] \leq C_2 t_{rel} \mathbb{E}_x[\bar{T}_\ell]$ (Lemma 7.6). The aforementioned concentration result follows from combining these two lemmata in a similar manner to the derivation of Corollary 5.4 (using the one-sided Chebyshev inequality). We omit the details to avoid repetitions.

Lemma 7.4. In the above setup, let $v \in [m] \setminus \{o\}$ Let $(v_0 = v, v_1, \ldots, v_s)$ be indices of consecutive blocks. Let $\mu_1, \mu_2 \in \mathcal{P}(I_v)$. Let $k \in [s]$. Denote by $\nu_k^{(j)} (j = 1, 2)$ the hitting distribution of $I_{v_k}$ starting from initial distribution $\mu_j$ (i.e. $\nu_k^{(j)}(z) := P_{\mu_j}[X_{T(v_k)} = z]$). Then $\|\nu_k^{(1)} - \nu_k^{(2)}\|_{TV} \leq (1 - \delta^r)^k$.

Proof. It suffices to prove the case $k = 1$ as the general case follows by induction using the Markov property. The case $k = 1$ follows from coupling the chain with the two different starting distributions in a way that with probability at least $\delta^r$ there exists some $z_v \in I_v$ such that both chains hit $z_v$ before hitting $I_{v_1}$ and from that moment on they follow the same trajectory. The fact that the hitting time of $z_v$ might be different for the two chains makes no difference. We now describe this coupling more precisely.
Let \( \mu_1, \mu_2 \in \mathcal{P}(I_v) \). There exists a coupling \((X_t^{(1)}, X_t^{(2)})_{t \geq 0}\) in which \((X_t^{(i)})_{t \geq 0}\) is distributed as the chain \((\Omega, P, \pi)\) with initial distribution \(\mu_i (i = 1, 2)\), such that \(P_{\mu_1, \mu_2}[S] \geq \delta\), where \(P_{\mu_1, \mu_2}\) is the corresponding probability measure and the event \(S\) is defined as follows. Let \(R := \min\{t : X_t^{(1)} = X_0^{(2)}\}\) and \(L_i := \min\{t : X_t^{(i)} \in I_{f_i}\}\). Let \(S\) denote the event: \(R \leq L_1\) and \(X_t^{(1)} = X_t^{(2)}\) for any \(t \geq 0\). Note that on \(S\), \(X_{L_1}^{(1)} = X_{L_2}^{(2)}\). Hence for any \(D \subset I_{v_i}\),

\[
\nu_1^{(1)}(D) - \nu_1^{(2)}(D) = P_{\mu_1, \mu_2}[X_{L_1}^{(1)} \in D] - P_{\mu_1, \mu_2}[X_{L_2}^{(2)} \in D] \\
\leq P_{\mu_1, \mu_2}[X_{L_1}^{(1)} \in D, X_{L_2}^{(2)} \notin D] \leq 1 - P_{\mu_1, \mu_2}[S] \leq 1 - \delta^r.
\]

\(\square\)

**Lemma 7.5.** In the setup of Lemma 7.4, let \(0 \leq i < j < s\). Let \(\mu \in \mathcal{P}(I_v)\). Write \(\tau_i := \bar{\tau}_{v_i}\) and \(\tau_j := \bar{\tau}_{v_j}\). Then

\[
E_\mu[\tau_i \tau_j] \leq E_\mu[\tau_i] E_\mu[\tau_j] \left(1 + (1 - \delta^r)^{j-i-1} \delta^{-r}\right).
\]

**Proof.** Let \(\mu_{i+1}\) and \(\mu_j\) be the hitting distributions of \(I_{v_{i+1}}\) and of \(I_{v_j}\), respectively, of the chain with initial distribution \(\mu\). Note that \(E_\mu[\tau_j] = E_{\mu_{i+1}}[\tau_j] = E_{\mu_j}[\tau_j]\). Clearly

\[
E_\mu[\tau_i \tau_j] \leq E_\mu[\tau_i] \max_{y \in I_{v_{i+1}}} E_y[\tau_j]. \tag{7.3}
\]

Let \(y^* \in I_{v_{i+1}}\) be the state achieving the maximum in the RHS above. By Lemma 7.4 we can couple successfully the hitting distribution of \(I_{v_j}\) of the chain started from \(y^*\) with that of the chain starting from initial distribution \(\mu_{i+1}\) with probability at least \(1 - (1 - \delta^r)^{j-i-1}\). The latter distribution is simply \(\mu_j\). If the coupling fails, then by (7.1) we can upper bound the conditional expectation of \(\tau_j\) by \(\delta^{-r} E_\mu[\tau_j]\). Hence

\[
E_{y^*}[\tau_j] \leq E_{\mu_j}[\tau_j] + (1 - \delta)^{j-i-1} \delta^{-r} E_\mu[\tau_j] = E_\mu[\tau_j] \left(1 + (1 - \delta^r)^{j-i-1} \delta^{-r}\right).
\]

The assertion of the lemma follows by plugging this estimate in (7.3). \(\square\)

**Lemma 7.6.** Let \(j \in [m] \setminus \{o\}\). Let \(\nu \in \mathcal{P}([n])\). Then there exists some \(C_1, C_2 > 0\) such that \(E_\nu[\bar{\tau}_j^2] \leq C_1 t_{rel} \Phi(A_j) \leq C_2 t_{rel} E_\nu[\bar{\tau}_j]\).

**Proof.** Let \(\mu := \psi_{A_j}\). By condition (i) in the definition of a \((\delta, r)\)-SBD chain, \(\mu \in \mathcal{P}(I_j)\). By (5.13), \(E_\mu[\bar{\tau}_j^2] \leq C_3 t_{rel} \Phi(A_j) \leq C_4 t_{rel} E_\mu[\bar{\tau}_j]\). The proof is concluded using the same reasoning as in the proof of (7.1) to argue that the first and second moments of \(\bar{\tau}_j\) w.r.t. different initial distributions change by at most some multiplicative constant. \(\square\)

8 Aldous’ example

We now present a small variation of Aldous’ example (see [16], Chapter 18) of a sequence of chains which satisfies the product condition but does not exhibit cutoff. This example demonstrates that Theorem 2 may fail if condition (ii) in the definition of a \((\delta, r)\)-semi birth and death chain is not satisfied. The main point in the construction is that the hitting times of worst sets are not concentrated.

25
ends and a path of length $2\pi$ bounded from below. In particular, the product condition holds. As 

\[ P_{n}(x,x) = \frac{1}{2} \text{ for } x \text{ even and } P_{n}(x,x) = \frac{3}{4} \text{ for } x \text{ odd. } \]

For $n=0$, $P_{n}(0,1) = \frac{1}{5}$, $P_{n}(0,-2) = \frac{1}{10}$, $P_{n}(-10n,10n+2) = \frac{1}{2}$, $P_{n}(2n+1,2n) = P_{n}(2n+1,2n-1) = \frac{1}{8}$.

All other transition probabilities are given by: $P_{n}(2i, \min\{2i+2,2n+1\}) = \frac{1}{5}$, $P_{n}(2i,2i-2) = P_{n}(2i-1,2i+1) = \frac{1}{6}$, $P_{n}(2i-1,\max\{2i-3,0\}) = \frac{1}{12}$.

**Example 8.1.** Consider the chain $(\Omega_{n}, P_{n}, \pi_{n})$, where $\Omega_{n} := \{-10n,-10n+2, \ldots , -2,0\} \cup [2n+1]$. Think of $\Omega$ as two paths (we call them branches) of length $n$ joined together at the ends and a path of length $5n$ joined to them at $0$ (see Figure 1). Set $P_{n}(x,x) = \frac{1}{2}$ if $x$ is even, $P_{n}(x,x) = \frac{3}{4}$ if $x$ is odd and $x < 2n+1$ and $P_{n}(2n+1,2n+1) = \frac{9}{10}$. Conditionally on not making a lazy step the walk moves with a fixed bias towards $2n+1$ (apart from at the states $-10n,0,2n+1$):

\[ P_{n}(2i, \min\{2i+2,2n+1\}) = 2P_{n}(2i,2i-2) = 2P_{n}(2i-1,2i+1) = 4P_{n}(2i-1, \max\{2i-3,0\}) = \frac{1}{3}. \]

Finally, we set $P_{n}(-10n,-10n+2) = \frac{1}{2}$, $P_{n}(0,2) = P_{n}(0,1) = P_{n}(0,-2) = \frac{1}{5}$ and $P_{n}(2n+1,2n) = P_{n}(2n+1,2n-1) = \frac{1}{20}$. It is easy to check that this chain is indeed reversible.

By Cheeger inequality (e.g. [16], Theorem 13.14), $\mathcal{L}_{\text{rel}}^{(n)} = O(1)$, as the bottleneck-ratio is bounded from below. In particular, the product condition holds. As $\pi_{n}(2n+1) > 1/2$, there is hit$_{1/2}$-cutoff iiff starting from $-10n$, the hitting-time of $2n+1$ is concentrated. We now explain why this is not the case. In particular, by Theorem 3, there is no cutoff.

Let $Y$ denote the last step away from 0 before $T_{2n+1}$. Observe that if $Y = 2$ (respectively,
If $Y = 1$, then the chain had to reach $2n + 1$ through the path $(2, 4, \ldots, 2n)$ ($(1, 3, \ldots, 2n - 1)$, respectively). Denote, $Z_i := T_{2n} 1_{Y = i}$, $i = 1, 2$. Then on $Y = i$, $T_{2n} = Z_i$, and its conditional distribution is concentrated around $42n$ for $i = 1$ and around $36n$ for $i = 2$, with deviations of order $\sqrt{n}$. Since both $Y = 1$ and $Y = 2$ have probability bounded away from 0, it follows that $d_n(37n)$ and $d_n(41n)$ are both bounded away from 0 and 1 (see Figure 2). In particular, the product condition holds but there is no cutoff.

Acknowledgements

We are grateful to David Aldous, Allan Sly, Perla Sousi and Prasad Tetali for many helpful suggestions.

References


