The Annals of Probability
2017, Vol. 45, No. 3, 1448–1487
DOI: 10.1214/16-AOP1090

CHARACTERIZATION OF CUTOFF FOR REVERSIBLE MARKOV CHAINS[1]

BY RIDDHIPRATIM BASU, JONATHAN HERMON AND YUVAL PERES

Stanford University, University of California, Berkeley and Microsoft Research, Redmond

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 α, 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 α 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 product of their spectral gaps and their (lazy) mixing-times tends to ∞.

1. Introduction. We obtain a tight bound on the mixing-time $t_{\mathrm{mix}}(\varepsilon)$ (up to an absolute constant independent of ε) for lazy reversible Markov chains in terms of hitting times of large sets [Proposition 1.8, (1.6)]. This refines previous results in the same spirit ([24] and [21], 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. In other words, one should run the nth chain until the cutoff point for it to even slightly mix in total variation, whereas running it any further is essentially redundant.

Though many families of chains are believed to exhibit cutoff, proving the occurrence of this phenomenon is often an extremely challenging task. Although drawing much attention, the progress made in the investigation of the cutoff phenomenon was done mostly through understanding examples and the field suffers from a lack of general theory. The cutoff phenomenon was given its name by Aldous and Diaconis in their seminal paper [1] from 1986 in which they suggested

Received December 2014; revised November 2015.

[1]An extended abstract of this paper appeared in the conference proceedings of *ACM-SIAM Symposium on Discrete Algorithms* (SODA), 2015.

MSC2010 subject classifications. 60J10.

Key words and phrases. Cutoff, mixing-time, finite reversible Markov chains, hitting times, trees, maximal inequality.

the following open problem (reiterated in [9]), which they refer to as "the most interesting problem": "*Find abstract conditions which ensure that the cutoff phenomenon occurs.*" 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.

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 Ω and its stationary distribution by π (or Ω_n and π_n, respectively, for the nth chain in a sequence of chains). Let $(X_t)_{t=0}^{\infty}$ be an irreducible Markov chain on a finite state space Ω with transition matrix P and stationary distribution π. We denote such a chain by (Ω, P, π). We say that the chain is finite, whenever Ω is finite. We say the chain is *reversible* if $\pi(x)P(x,y) = \pi(y)P(y,x)$, for any $x, y \in \Omega$.

We call a chain *lazy* if $P(x,x) \geq 1/2$, for all x. In this paper, all discrete-time chains would be assumed to be lazy, unless otherwise is specified. To avoid periodicity and near-periodicity issues, one often considers the lazy version of the chain, defined by replacing P with $P_{\mathrm{L}} := (P+I)/2$. Another way to avoid periodicity issues is to consider the continuous-time version of the chain, $(X_t^{\mathrm{ct}})_{t\geq 0}$, which is a continuous-time Markov chain whose heat kernel is defined by $H_t(x,y) := \sum_{k=o}^{\infty} \frac{e^{-t}t^k}{k!} P^k(x,y)$.

We denote by P_{μ}^t (P_{μ}) the distribution of X_t [resp., $(X_t)_{t\geq 0}$], given that the initial distribution is μ. We denote by H_{μ}^t (H_{μ}) the distribution of X_t^{ct} [resp. $(X_t^{\mathrm{ct}})_{t\geq 0}$], given that the initial distribution is μ. 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_x^t (P_x) and H_x^t (H_x). For any $x, y \in \Omega$ and $t \in \mathbb{N}$ we write $\mathrm{P}_x^t(y) := \mathrm{P}_x(X_t = y) = P^t(x,y)$.

We denote the set of probability distributions on a (finite) set B by $\mathscr{P}(B)$. For any $\mu, \nu \in \mathscr{P}(B)$, their *total-variation distance* is defined to be $\|\mu - \nu\|_{\mathrm{TV}} := \frac{1}{2}\sum_x |\mu(x) - \nu(x)| = \sum_{x \in B:\mu(x)>\nu(x)} \mu(x) - \nu(x)$. The worst-case total variation distance at time t is defined as

$$d(t) := \max_{x\in\Omega} d_x(t) \qquad \text{where for any } x \in \Omega, d_x(t) := \|\mathrm{P}_x(X_t \in \cdot) - \pi\|_{\mathrm{TV}}.$$

The *ε-mixing-time* is defined as

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

Similarly, let $d_{\mathrm{ct}}(t) := \max_{x\in\Omega} \|\mathrm{H}_x^t - \pi\|_{\mathrm{TV}}$ and let $t_{\mathrm{mix}}^{\mathrm{ct}}(\varepsilon) := \inf\{t : d_{\mathrm{ct}}(t) \leq \varepsilon\}$.

When $\varepsilon = 1/4$ we omit it from the above notation. Next, consider a sequence of such chains, $((\Omega_n, P_n, \pi_n) : n \in \mathbb{N})$, each with its corresponding worst-distance from stationarity $d^{(n)}(t)$, its mixing-time $t_{\mathrm{mix}}^{(n)}$, etc. We say that the sequence exhibits a *cutoff* if the following sharp transition in its convergence to stationarity occurs:

$$\lim_{n\to\infty} \frac{t_{\mathrm{mix}}^{(n)}(\varepsilon)}{t_{\mathrm{mix}}^{(n)}(1-\varepsilon)} = 1 \qquad \text{for any } 0 < \varepsilon < 1.$$

We say that the sequence has a *cutoff window* w_n, if $w_n = o(t_{\mathrm{mix}}^{(n)})$ and for any $\varepsilon \in (0, 1)$ there exists $c_\varepsilon > 0$ such that for all n

$$t_{\mathrm{mix}}^{(n)}(\varepsilon) - t_{\mathrm{mix}}^{(n)}(1-\varepsilon) \le c_\varepsilon w_n. \tag{1.1}$$

Recall that if (Ω, P, π) is a finite reversible irreducible lazy chain, then P is self-adjoint w.r.t. the inner product induced by π (see Definition 2.1), and hence has $|\Omega|$ real eigenvalues. Throughout we shall denote them by $1 = \lambda_1 > \lambda_2 \ge \cdots \ge \lambda_{|\Omega|} \ge 0$ (where $\lambda_2 < 1$ since the chain is irreducible and $\lambda_{|\Omega|} \ge 0$ by laziness). Define the *relaxation-time* of P as $t_{\mathrm{rel}} := (1-\lambda_2)^{-1}$. The following general relation holds for lazy chains (see [18] Theorems 12.3 and 12.4)

$$(t_{\mathrm{rel}} - 1)\log\left(\frac{1}{2\varepsilon}\right) \le t_{\mathrm{mix}}(\varepsilon) \le t_{\mathrm{rel}} \log\left(\frac{1}{\varepsilon \min_x \pi(x)}\right). \tag{1.2}$$

We say that a family of chains satisfies the *product condition* if $(1-\lambda_2^{(n)})t_{\mathrm{mix}}^{(n)} \to \infty$ as $n \to \infty$ [or equivalently, $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$]. The following well-known fact follows easily from the first inequality in (1.2) (cf. [18], Proposition 18.4).

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

In 2004, the third author [22] 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 [18], 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 [11]. The relevant setup is weighted nearest neighbor random walks on finite trees. See Section 5 for a formal definition.

THEOREM 1. *Let (V, P, π) be a lazy reversible Markov chain on a tree $T = (V, E)$ with $|V| \ge 3$. Then*

$$t_{\mathrm{mix}}(\varepsilon) - t_{\mathrm{mix}}(1-\varepsilon) \le 35\sqrt{\varepsilon^{-1} t_{\mathrm{rel}} t_{\mathrm{mix}}} \qquad \textit{for any } 0 < \varepsilon \le 1/4. \tag{1.3}$$

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

In [10], Diaconis and Saloff-Coste showed that a sequence of birth and death (BD) chains exhibits separation cutoff if and only if $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$. In [11], Ding et al. extended this also to the notion of total-variation cutoff and showed that the cutoff window is always at most $\sqrt{t_{\mathrm{rel}}^{(n)} t_{\mathrm{mix}}^{(n)}}$ 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 [11] (even for BD chains), which is $t_{\mathrm{mix}}(\varepsilon) - t_{\mathrm{mix}}(1-\varepsilon) \leq c\varepsilon^{-1}\sqrt{t_{\mathrm{rel}} t_{\mathrm{mix}}}$ (Theorem 2.2). In fact, in Section 5.1 we show that under the product condition, $d(t)$ decays in a sub-Gaussian manner within the cutoff window. More precisely, we show that $t_{\mathrm{mix}}^{(n)}(\varepsilon) - t_{\mathrm{mix}}^{(n)}(1-\varepsilon) \leq c\sqrt{t_{\mathrm{rel}}^{(n)} t_{\mathrm{mix}}^{(n)} |\log \varepsilon|}$. This is somewhat similar to Theorem 6.1 in [10], 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 [10] and [11] (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. Denote $[n] := \{1, 2, \ldots, n\}$.

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

(i) For all $i, j \in [n]$ 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* (δ, r)*-semi birth and death chains, for some* $\delta, r > 0$, *satisfying the product condition. Then it exhibits a cutoff with a cutoff window* $w_k := \sqrt{t_{\mathrm{mix}}^{(k)} t_{\mathrm{rel}}^{(k)}}$.

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

DEFINITION 1.3. Let (Ω, P, π) be an irreducible chain. For any $x \in \Omega$, $\alpha, \varepsilon \in (0, 1)$ and $t \geq 0$, define $p_x(\alpha, t) := \max_{A \subset \Omega: \pi(A) \geq \alpha} \mathrm{P}_x[T_A > t]$, 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

$$\mathrm{hit}_{\alpha,x}(\varepsilon) := \min\{t : p_x(\alpha, t) \leq \varepsilon\} \quad \text{and} \quad \mathrm{hit}_{\alpha}(\varepsilon) := \min\{t : p(\alpha, t) \leq \varepsilon\}.$$

Similarly, we define $p_x^{\mathrm{ct}}(\alpha, t) := \max_{A \subset \Omega: \pi(A) \geq \alpha} \mathrm{H}_x[T_A^{\mathrm{ct}} > t]$ (where $T_A^{\mathrm{ct}} := \inf\{t : X_t^{\mathrm{ct}} \in A\}$) and set $\mathrm{hit}_{\alpha}^{\mathrm{ct}}(\varepsilon) := \min\{t : p_x^{\mathrm{ct}}(\alpha, t) \leq \varepsilon \text{ for all } x \in \Omega\}$.

DEFINITION 1.4. Let (Ω_n, P_n, π_n) be a sequence of irreducible chains and let $\alpha \in (0,1)$. We say that the sequence exhibits a hit_α-cutoff, if for any $\varepsilon \in (0, 1/4)$

$$\mathrm{hit}_\alpha^{(n)}(\varepsilon) - \mathrm{hit}_\alpha^{(n)}(1-\varepsilon) = o\big(\mathrm{hit}_\alpha^{(n)}(1/4)\big).$$

We are now ready to state our main abstract theorem.

THEOREM 3. *Let (Ω_n, P_n, π_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* hit_α*-cutoff for some* $\alpha \in (0, 1/2]$.
(3) *The sequence exhibits a* hit_α*-cutoff for some* $\alpha \in (1/2, 1)$ *and* $t_{\mathrm{rel}}^{(n)} = o(t_{\mathrm{mix}}^{(n)})$.

REMARK 1.5. In Example 7.2, we show that there exists a sequence of lazy reversible irreducible finite Markov chains, (Ω_n, P_n, π_n), such that the product condition fails, yet for all $1/2 < \alpha < 1$ there is hit_α-cutoff. Thus, the assertion of Theorem 3 is sharp.

REMARK 1.6. The proof of Theorem 3 can be extended to the continuous-time case (the necessary adaptations are sketched in Section 4). In particular, it follows that a sequence of finite lazy reversible chains exhibits cutoff iff the sequence of the continuous-time versions of these chains exhibits cutoff. This was previously proven in [7] without the assumption of reversibility.

REMARK 1.7. Using somewhat similar techniques as in this work, it was shown in [16] that under reversibility the sequence of associated continuous-time chains exhibits a cutoff around time t_n iff the same holds for the sequence of *associated averaged* ("averaged at two consecutive time steps") *chains*, defined by replacing P^k by $A_k := (P^k + P^{k+1})/2$. This result and its connections with the results and techniques of this paper are discussed in more details in the related work section.

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

PROPOSITION 1.8. *For any reversible irreducible finite lazy chain and any* $\varepsilon \in (0, \frac{1}{4}]$,

$$\mathrm{hit}_{1/2}(3\varepsilon/2) - \lceil 2t_{\mathrm{rel}}|\log\varepsilon|\rceil \le t_{\mathrm{mix}}(\varepsilon) \le \mathrm{hit}_{1/2}(\varepsilon/2) + \lceil t_{\mathrm{rel}}|\log(\varepsilon/4)|\rceil \tag{1.4}$$

and

$$\begin{aligned}\text{hit}_{1/2}(1-\varepsilon/2) - \lceil 2t_{\text{rel}}|\log\varepsilon|\rceil &\le t_{\text{mix}}(1-\varepsilon)\\ &\le \text{hit}_{1/2}(1-2\varepsilon) + 1_{\varepsilon>1/18}\left\lceil \frac{1}{2}t_{\text{rel}}\log 8\right\rceil.\end{aligned} \tag{1.5}$$

Moreover,

$$\begin{aligned}&\max\{\text{hit}_{1-\varepsilon/4}(5\varepsilon/4), (t_{\text{rel}}-1)|\log(2\varepsilon)|\}\\ &\quad\le t_{\text{mix}}(\varepsilon)\\ &\quad\le \text{hit}_{1-\varepsilon/4}(3\varepsilon/4) + \left\lceil \frac{3t_{\text{rel}}}{2}|\log(\varepsilon/4)|\right\rceil.\end{aligned} \tag{1.6}$$

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

REMARK 1.9. Define $t_{\text{rel}}^{\text{absolute}} := \max\{(1-\lambda_2)^{-1}, (1-|\lambda_{|\Omega|}|)^{-1}\}$. Our only use of the laziness assumption is to argue that $t_{\text{rel}} = t_{\text{rel}}^{\text{absolute}}$. In particular, Proposition 1.8 holds also without the laziness assumption if one replaces t_{rel} by $t_{\text{rel}}^{\text{absolute}}$. Similarly, without the laziness assumption the assertion of Theorem 3 should be transformed as follows. A sequence of finite irreducible aperiodic reversible Markov chains exhibits cutoff iff $(t_{\text{rel}}^{\text{absolute}})^{(n)} = o(t_{\text{mix}}^{(n)})$ and there exists some $0 < \alpha < 1$ such that the sequence exhibits hit_α-cutoff.

Note that for any finite irreducible reversible chain, (Ω, P, π), it suffices to consider a δ-lazy version of the chain, $P_\delta := (1-\delta)P + \delta I$, for some $\delta \ge \frac{1-\max\{\lambda_2,0\}}{2}$, to ensure that $t_{\text{rel}} = t_{\text{rel}}^{\text{absolute}}$ (which by the previous paragraph, guarantees that all near-periodicity issues are completely avoided).

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 a rate which is governed by its relaxation-time. This estimate is sharp is some cases (i.e., there are examples in which the above description is accurate and the rate of convergence in the "second stage" is also lower bounded by the 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 ε. Moreover, (1.6) bounds $t_{\text{mix}}(\varepsilon)$ in terms of hitting distribution of sets of π measure tending to 1 as ε tends to 0. In (3.2) we give a version of (1.6) for sets of arbitrary π measure.

Either of the two terms appearing in the sum in RHS of (1.6) may dominate the other. For lazy simple random walk on two n-cliques connected by a single edge, the terms in (1.6) involving $\text{hit}_{1-\varepsilon/4}$ are negligible. For a sequence of chains satisfying the product condition, all terms in Proposition 1.8 involving t_{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 $\mathrm{hit}_{1/2}^{(n)}(1/4) = \Theta(t_{\mathrm{mix}}^{(n)})$. 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 hit_α-cutoff for some $\alpha \in (0, 1)$, then it exhibits hit_β-cutoff for all $\beta \in (0, 1)$.

1.1. *Related work.* The idea that expected hitting times of sets which are "worst in expectation" [in the sense of (1.7) below] could be related to the mixing time is quite old and goes back to Aldous' 1982 paper [3]. A similar result was obtained later by Lovász and Winkler ([19], Proposition 4.8).

This aforementioned connection was substantially refined recently by Peres and Sousi ([24], Theorem 1.1) and independently by Oliveira ([21], Theorem 2). In [24], Peres and Sousi considered the mixing times of the associated lazy and "averaged" chains [recall from Remark 1.7 that the distribution at time t of the latter is obtained by replacing P^t by $A_t := (P^t + P^{t+1})/2$] denoted, respectively, by $t_{\mathrm{L}} := \inf\{t : \max_x \|P_{\mathrm{L}}^t(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \le 1/4\}$ and $t_{\mathrm{ave}} := t_{\mathrm{ave}}(1/4)$, where $t_{\mathrm{ave}}(\varepsilon) := \inf\{t + 1 : d_{\mathrm{ave}}(t) \le \varepsilon\}$, and $d_{\mathrm{ave}}(t) := \max_x \|A_t(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}}$. They proved that under reversibility t_{L} and t_{ave} are equivalent to each other (i.e., that for some universal constants, $0 < c < C$, $c \le t_{\mathrm{L}}/t_{\mathrm{ave}} \le C$ for all reversible chains) and also to various other mixing parameters, including $t_{\mathrm{stop}} := \max_{x\in\Omega, T \text{ stopping time}: X_T \sim \pi} \mathbb{E}_x[T]$. Their approach relied on the theory of random times to stationarity combined with a certain complicated "de-randomization" argument which shows that (under reversibility) $t_{\mathrm{ave}} \le C t_{\mathrm{stop}}$. As a (somewhat indirect) consequence, they deduced 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

$$c'_\alpha t_{\mathrm{H}}(\alpha) \le t_{\mathrm{mix}} \le c_\alpha t_{\mathrm{H}}(\alpha) \qquad \text{where} \tag{1.7}$$
$$t_{\mathrm{H}}(\alpha) := \max_{x\in\Omega} t_{\mathrm{H},x}(\alpha) \text{ and } t_{\mathrm{H},x}(\alpha) := \max_{A\subset\Omega:\pi(A)\ge\alpha} \mathbb{E}_x[T_A].$$

This work was greatly motivated by the aforementioned results. It is natural to ask whether (1.7) could be further refined so that the cutoff phenomenon could be characterized in terms of concentration of the hitting times of a sequence of sets $A_n \subset \Omega_n$ which attain the maximum in the definition of $t_{\mathrm{H}}^{(n)}(1/2)$ (starting from the worst initial states). Corollary 1.5 in [15] asserts that this is indeed the case in the transitive setup. More generally, Theorem 2 in [15] asserts that this is indeed the case for any fixed sequence of initial states $x_n \in \Omega_n$ if one replaces $t_{\mathrm{H}}^{(n)}(1/2)$ and $d^{(n)}(t)$ by $t_{\mathrm{H},x_n}^{(n)}(1/2)$ and $d_{x_n}^{(n)}(t)$ (i.e., when the hitting times and the mixing times are defined only w.r.t. these starting states). Alas, Proposition 1.6 in [15] asserts that in general cutoff could not be characterized in this manner.

In [17], 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 $\pi_n(A_n) \ge c > 0$, starting from any $x \in A_n$, the nth chain mixes in $o(t_{\mathrm{mix}}^{(n)})$ steps.

Very recently, Chen and Saloff-Coste [8] obtained a detailed criterion for cutoff (both in total variation and separation distance) for the class of birth and death chains using concentration of hitting times. They also obtained formulae for the cutoff time as well as the cutoff window in terms of the moments of certain hitting times.

The most important tool we shall utilize is Starr's L^2 maximal inequality (Theorem 2.3), which as we demonstrate in Section 2, can become extremely powerful when combined with simple spectral techniques (e.g., the L^2-contraction Lemma). Its central role in our approach is explained in the following section. Relating it to the study of mixing-times of reversible Markov chains is one of the main contributions of this work. It is the belief of the authors that this technique can be applied to other theoretical problems concerning Markov chains. Maximal inequalities were the main tool used in two other recent works [16, 20] which resolved long lasting open problems related to mixing times of reversible chains. In [20] Starr's L^p maximal inequality was used to prove (under reversibility) the inequality $\sum_{y \in \Omega} \sup_t P^t(x,y) \le 2e(1 \vee |\log \pi(x)|)$. We note that for their application they had to take $p \approx 1 + \pi(x)$.

In [16], the second and third authors substantially refined the aforementioned equivalence of t_{L} and t_{ave}, established by Peres and Sousi, by showing that $d_{\mathrm{ave}}(t + \lceil Mt \rceil) \le \max_x \|\mathrm{H}_x^t - \pi\|_{\mathrm{TV}} + C\sqrt{(1 \vee \log M)/M}$ and $\max_x \|\mathrm{H}_x^{t+M\sqrt{t}} - \pi\|_{\mathrm{TV}} \le d_{\mathrm{ave}}(t) + e^{-cM^2}$, for all $t, M > 0$ (for some absolute constants $C, c > 0$). The main tool used in [16] is a certain L^2 maximal inequality involving the discrete derivative of the transition matrix. These quantitative relations resolve a conjecture of Aldous and Fill [2], Open Problem 4.17. Moreover, it is shown in [16] that these inequalities not only imply the equivalence of cutoffs for the sequences of associated (resp.) continuous-time and averaged chains, but also allows one to express the (optimal) cutoff window of one in terms of that of the other.

1.2. *An overview of our techniques.*

DEFINITION 1.10. Let (Ω, P, π) be a finite reversible irreducible lazy chain. Let $A \subset \Omega$, $s \ge 0$ and $m > 0$. Denote $\rho(A) := \sqrt{\mathbf{Var}_\pi 1_A} = \sqrt{\pi(A)(1-\pi(A))}$. Set $\sigma_s := e^{-s/t_{\mathrm{rel}}}\rho(A)$. We define

$$G_s(A,m) := \{y : |\mathrm{P}_y^k(A) - \pi(A)| < m\sigma_s \text{ for all } k \ge s\}. \tag{1.8}$$

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 nonempty $A \subset \Omega$ 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 now prove the following inequalities:

$$t_{\text{mix}}(2\varepsilon) \leq \text{hit}_{1-\varepsilon}(\varepsilon) + \left\lceil \frac{t_{\text{rel}}}{2} \log\left(\frac{2}{\varepsilon^3}\right) \right\rceil, \tag{1.9}$$

$$\text{hit}_{1-\varepsilon}(1 - 2\varepsilon) \geq t_{\text{mix}}(1 - \varepsilon) - \left\lceil \frac{t_{\text{rel}}}{2} \log\left(\frac{8}{\varepsilon^2}\right) \right\rceil. \tag{1.10}$$

We first prove (1.9). Let $A \subset \Omega$ be nonempty. 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 $1 - \varepsilon$. By Corollary 2.4, we know that $\pi(G) \geq 1 - 8/m^2$. Thus, we pick $m = \sqrt{8/\varepsilon}$. The precision in (1.8) is $m\sigma_s \leq \sqrt{8/\varepsilon}(\sqrt{\mathbf{Var}_\pi\, 1_A} e^{-s/t_{\text{rel}}}) \leq \sqrt{2/\varepsilon} e^{-s/t_{\text{rel}}}$. As we want to have ε precision, we pick $s := \lceil \frac{t_{\text{rel}}}{2} \log(\frac{2}{\varepsilon^3}) \rceil$.

We seek to bound $|\mathrm{P}_x^{t+s}(A) - \pi(A)|$. If $|\mathrm{P}_x^{t+s}(A) - \pi(A)| \leq 2\varepsilon$, then the chain is "$2\varepsilon$-mixed w.r.t. A". This is where we use the set G. We now demonstrate that for any $t \geq 0$, hitting G by time t serves as a "certificate" that the chain is ε-mixed w.r.t. A at time $t + s$. Indeed, from the Markov property and the definition of G,

$$\left|\mathrm{P}_x[X_{t+s} \in A \mid T_G \leq t] - \pi(A)\right| \leq \max_{g \in G} \sup_{s' \geq s} \left|\mathrm{P}_g^{s'}(A) - \pi(A)\right| \leq \varepsilon.$$

In particular,

$$\begin{aligned} \left|\mathrm{P}_x^{t+s}(A) - \pi(A)\right| &\leq \mathrm{P}_x[T_G > t] + \left|\mathrm{P}_x[X_{t+s} \in A \mid T_G \leq t] - \pi(A)\right| \\ &\leq \mathrm{P}_x[T_G > t] + \varepsilon. \end{aligned} \tag{1.11}$$

We seek to have the bound $\mathrm{P}_x[T_G > t] \leq \varepsilon$. Recall that by our choice of m we have that $\pi(G) \geq 1 - \varepsilon$. Thus if we pick $t := \text{hit}_{1-\varepsilon}(\varepsilon)$, we guarantee that, regardless of the identity of A and x, we indeed have that $\mathrm{P}_x[T_G > t] \leq \varepsilon$. Since x and A were arbitrary, plugging this into (1.11) yields (1.9). We now prove (1.10).

We now set $r := t_{\text{mix}}(1 - \varepsilon) - 1$. Then there exist some $x \in \Omega$ and $A \subset \Omega$ such that $\pi(A) - \mathrm{P}_x^r(A) > 1 - \varepsilon$. In particular, $\pi(A) > 1 - \varepsilon$. Consider again $G_2 := G_{s_2}(A, m)$. Since again we seek the size of G_2 to be at least $1 - \varepsilon$, we again choose $m = \sqrt{8/\varepsilon}$. The precision in (1.8) is $m\sigma_{s_2} \leq \sqrt{8/\varepsilon}(\sqrt{\mathbf{Var}_\pi\, 1_A} e^{-s_2/t_{\text{rel}}}) \leq \sqrt{8/\varepsilon}(\sqrt{1 - \pi(A)} e^{-s_2/t_{\text{rel}}}) \leq \sqrt{8} e^{-s_2/t_{\text{rel}}}$. We again seek ε precision. Hence, we pick $s_2 := \lceil \frac{t_{\text{rel}}}{2} \log(\frac{8}{\varepsilon^2}) \rceil$. As in (1.11) (with $r - s_2$ in the role of t and s_2 in the role of s) we have that

$$\mathrm{P}_x[T_{G_2} > r - s_2] \geq \pi(A) - \mathrm{P}_x^r(A) - \varepsilon > 1 - 2\varepsilon.$$

Hence, it must be the case that $\text{hit}_{1-\varepsilon}(1 - 2\varepsilon) > r - s_2 = t_{\text{mix}}(1 - \varepsilon) - 1 - \lceil \frac{t_{\text{rel}}}{2} \log(\frac{8}{\varepsilon^2}) \rceil$.

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 (Ω, P, π) 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 $\mathbf{Var}_{\pi} f := \mathbb{E}_{\pi}[(f - \mathbb{E}_{\pi} f)^2]$. The inner-product $\langle \cdot, \cdot \rangle_{\pi}$ and L^p norm are

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

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 $P^t f(x) := \sum_{y \in \Omega} P^t(x, y) f(y) = \mathbb{E}_x[f(X_t)]$. 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 $f_1, \dots, f_{|\Omega|}$ is an orthonormal basis of $L^2(\mathbb{R}^{\Omega}, \pi)$ such that $P f_i := \lambda_i f_i$ for all i, then $P^t g = \mathbb{E}_{\pi} P^t g + \sum_{i=2}^{|\Omega|} \langle g, f_i \rangle_{\pi} \lambda_i^t f_i$, for all $g \in \mathbb{R}^{\Omega}$ and $t \ge 0$. The following lemma is standard. It is proved using the spectral decomposition in a straightforward manner.

LEMMA 2.2 (L^2-contraction lemma). *Let (Ω, P, π) be a finite lazy reversible irreducible Markov chain. Let $f \in \mathbb{R}^{\Omega}$. Then*

$$\mathbf{Var}_{\pi} P^t f \le e^{-2t/t_{\mathrm{rel}}} \mathbf{Var}_{\pi} f \qquad \text{for any } t \ge 0. \tag{2.1}$$

We now state a particular case of Starr's maximal inequality ([25], Theorem 1). It is similar to Stein's maximal inequality ([26]), 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 (Ω, P, π) be a reversible irreducible Markov chain. Let $1 < p < \infty$ and $p^* := p/(p-1)$ be its conjugate exponent. Then for any $f \in L^p(\mathbb{R}^{\Omega}, \pi)$,*

$$\|f^*\|_p \le p^* \|f\|_p, \tag{2.2}$$

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

$$f^*(x) := \sup_{0 \le k < \infty} \left| P^{2k}(f)(x) \right| = \sup_{0 \le k < \infty} \left| \mathbb{E}_x[f(X_{2k})] \right|.$$

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

COROLLARY 2.4. *Let (Ω, P, π) be a finite reversible irreducible lazy chain. As in Definition 1.10, define $\rho(A) := \sqrt{\pi(A)(1-\pi(A))}$, $\sigma_t := \rho(A)e^{-t/t_{\mathrm{rel}}}$ and*

$$G_t(A,m) := \{y : |\mathrm{P}_y^k(A) - \pi(A)| < m\sigma_t \text{ for all } k \geq t\}.$$

Then

$$\pi(G_t(A,m)) \geq 1 - 8m^{-2} \qquad \text{for all } A \subset \Omega, t \geq 0 \text{ and } m > 0. \tag{2.3}$$

PROOF. For any $t \geq 0$, let $f_t(x) := P^t(1_A - \pi(A))(x) = \mathrm{P}_x^t(A) - \pi(A)$. Then in the notation of Theorem 2.3,

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

and similarly

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

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

$$1 - \pi(G_t) \leq \pi\{x : f_t^*(x) \geq m\sigma_t\} + \pi\{x : (Pf_t)^*(x) \geq m\sigma_t\}. \tag{2.4}$$

Note that since $\pi P^t = \pi$ we have that $\mathbb{E}_\pi(f_t) = \mathbb{E}_\pi(f_0) = \mathbb{E}_\pi(1_A - \pi(A)) = 0$. Now (2.1) implies that

$$\|Pf_t\|_2^2 \leq \|f_t\|_2^2 = \mathbf{Var}_\pi\, P^t f_0 \leq e^{-2t/t_{\mathrm{rel}}}\, \mathbf{Var}_\pi\, f_0 = e^{-2t/t_{\mathrm{rel}}}\rho^2(A) = \sigma_t^2. \tag{2.5}$$

Hence, by Markov inequality and (2.2) we have

$$\pi\{x : f_t^*(x) \geq m\sigma_t\} = \pi\{x : (f_t^*(x))^2 \geq m^2\sigma_t^2\} \leq 4m^{-2}, \tag{2.6}$$

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). □

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 $f \in L^p(\mathbb{R}^\Omega, \pi)$. 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 $|f^*| \leq h^*$. Consequently, $\|f^*\|_p \leq \|h^*\|_p \leq q\|h\|_p = q\|f\|_p$.

Let $(X_n)_{n\geq 0}$ have the distribution of the chain (Ω, P, π) 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., [12], 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, $(X_n, X_{n+1}, \dots, X_{2n})$ and $(X_n, X_{n-1}, \dots, X_0)$ have the same law. Hence,

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

where the third equality in (2.8) follows by the Markov property. Fix $N \geq 0$. By (2.8) $(R_n)_{n=0}^N$ is a reverse martingale, that is, $(R_{N-n})_{n=0}^N$ is a martingale. By Doob's L^p maximal inequality (e.g., [12], Theorem 5.4.3.)

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

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

$$h_N(X_0) = \max_{0 \leq n \leq N} \mathbb{E}[R_n \mid X_0] \leq \mathbb{E}\Big[\max_{0 \leq n \leq N} R_n \mid X_0\Big]. \tag{2.10}$$

By conditional Jensen inequality $\|\mathbb{E}[Y \mid X_0]\|_p \leq \|Y\|_p$ (e.g., [12], Theorem 5.1.4.). So by taking L^p norms in (2.10), together with (2.9) we get that

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

The proof is complete using the monotone convergence theorem. □

3. Inequalities relating $t_{\mathrm{mix}}(\varepsilon)$ and $\mathrm{hit}_\alpha(\delta)$. Our aim in this section is to obtain inequalities relating $t_{\mathrm{mix}}(\varepsilon)$ and $\mathrm{hit}_\alpha(\delta)$ for suitable values of α, ε and δ using Corollary 2.4.

The following corollary uses the same reasoning as in the proof of (1.9)–(1.10) with a slightly more careful analysis.

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

$$\mathrm{P}_x^{t+s}(A) \geq (1-\delta)\big[\pi(A) - e^{-s/t_{\mathrm{rel}}}\big[8\alpha^{-1}\pi(A)\big(1-\pi(A)\big)\big]^{1/2}\big]. \tag{3.1}$$

Consequently, for any $0 < \varepsilon < 1$ we have that

$$\begin{aligned} &\mathrm{hit}_{1-\alpha}\big((\alpha+\varepsilon) \wedge 1\big) \leq t_{\mathrm{mix}}(\varepsilon) \quad \textit{and} \\ &t_{\mathrm{mix}}\big((\varepsilon+\delta) \wedge 1\big) \leq \mathrm{hit}_{1-\alpha}(\varepsilon) + \left\lceil \frac{t_{\mathrm{rel}}}{2} \log^+\left(\frac{2(1-\varepsilon)^2}{\alpha\varepsilon\delta}\right)\right\rceil, \end{aligned} \tag{3.2}$$

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

$$\mathrm{hit}_{1-\varepsilon/4}(5\varepsilon/4) \leq t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}_{1-\varepsilon/4}(3\varepsilon/4) + \left\lceil \frac{3t_{\mathrm{rel}}}{2} \log(4/\varepsilon)\right\rceil, \tag{3.3}$$

$$\begin{aligned} &t_{\mathrm{mix}}(\varepsilon) \leq \mathrm{hit}_{1/2}(\varepsilon/2) + \lceil t_{\mathrm{rel}} \log(4/\varepsilon)\rceil \quad \textit{and} \\ &t_{\mathrm{mix}}(1-\varepsilon/2) \leq \mathrm{hit}_{1/2}(1-\varepsilon) + 1_{\varepsilon > 1/9}\left\lceil \frac{1}{2} t_{\mathrm{rel}} \log 8\right\rceil. \end{aligned} \tag{3.4}$$

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

$$G = G_s(A) := \{y : |\mathrm{P}_y^k(A) - \pi(A)| < e^{-s/t_{\mathrm{rel}}}(8\alpha^{-1}\pi(A)(1-\pi(A)))^{1/2} \text{ for all } k \geq s\}.$$

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

$$\mathrm{P}_x[X_{t+s} \in A \mid T_G \leq t] \geq \pi(A) - e^{-s/t_{\mathrm{rel}}}[8\alpha^{-1}\pi(A)(1-\pi(A))]^{1/2}.$$

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

$$\begin{aligned}\mathrm{P}_x^{t+s}(A) &\geq \mathrm{P}_x[T_G \leq t]\mathrm{P}_x[X_{t+s} \in A \mid T_G \leq t] \\ &\geq (1-\delta)[\pi(A) - e^{-s/t_{\mathrm{rel}}}[8\alpha^{-1}\pi(A)(1-\pi(A))]^{1/2}],\end{aligned}$$

which completes 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 $t_1 := t_{\mathrm{mix}}(\varepsilon)$. Then for any $x \in \Omega$,

$$\mathrm{P}_x[T_A \leq t_1] \geq \mathrm{P}_x[X_{t_1} \in A] \geq \pi(A) - \|\mathrm{P}_x^{t_1} - \pi\|_{\mathrm{TV}} \geq 1 - \alpha - \varepsilon.$$

In particular, we get directly from Definition 1.3 that $\mathrm{hit}_{1-\alpha}(\alpha + \varepsilon) \leq t_1 = t_{\mathrm{mix}}(\varepsilon)$. We now prove the second inequality in (3.2).

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

$$\begin{aligned}d(t+s) &= \pi(A) - \mathrm{P}_x^{t+s}(A) \\ &\leq \varepsilon\pi(A) + (1-\varepsilon)e^{-s/t_{\mathrm{rel}}}[8\alpha^{-1}\pi(A)(1-\pi(A))]^{1/2} \\ &\leq \varepsilon[\pi(A) + 2\sqrt{\delta/\varepsilon}\sqrt{\pi(A)(1-\pi(A))}] \\ &\leq \varepsilon[1 + (2\sqrt{\delta/\varepsilon})^2/4] = \varepsilon + \delta,\end{aligned} \tag{3.5}$$

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 completes the proof of (3.2).

For the second inequality of (3.3), apply (3.2) with $(\alpha, \varepsilon, \delta)$ being $(\varepsilon/4, 3\varepsilon/4, \varepsilon/4)$. Similarly, to get (3.4) apply (3.2) with $(\alpha, \varepsilon, \delta)$ being $(1/2, \varepsilon/2, \varepsilon/2)$ or $(1/2, 1-\varepsilon, \varepsilon/2)$, 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 $t_{\mathrm{mix}}^{\mathrm{ct}}$ and $\mathrm{hit}^{\mathrm{ct}}$, 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 $t, s \geq 0$, by the Markov property we have that $\mathrm{P}_x[T_A > t+s] \leq \mathrm{P}_x[T_A > t](\max_z \mathrm{P}_z[T_A > s]) \leq p(\alpha,t)p(\alpha,s)$. Maximizing over x and 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* $\alpha, \varepsilon, \delta \in (0,1)$, *we have that*

$$\mathrm{hit}_\alpha(\varepsilon\delta) \leq \mathrm{hit}_\alpha(\varepsilon) + \mathrm{hit}_\alpha(\delta). \tag{3.6}$$

In the next corollary, we establish inequalities between $\mathrm{hit}_\alpha(\delta)$ and $\mathrm{hit}_\beta(\delta')$ for appropriate values of α, β, δ and δ'.

COROLLARY 3.4. *For any reversible irreducible finite chain and* $0 < \varepsilon < \delta < 1$,

$$\begin{aligned} \mathrm{hit}_\beta(\delta) &\leq \mathrm{hit}_\alpha(\delta) \\ &\leq \mathrm{hit}_\beta(\delta - \varepsilon) \\ &\quad + \left\lceil \alpha^{-1} t_{\mathrm{rel}} \log\left(\frac{1-\alpha}{(1-\beta)\varepsilon}\right)\right\rceil \qquad \textit{for any } 0 < \alpha \leq \beta < 1. \end{aligned} \tag{3.7}$$

The general idea behind Corollary 3.4 is as follows. Loosely speaking, we show that any set $A \subset \Omega$ has a "blow-up" set $H(A)$ (of large π-measure), such that starting from any $x \in H(A)$, the set A is hit "quickly" [in time proportional to $t_{\mathrm{rel}}/\pi(A)$] 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 π, which is well understood.

LEMMA 3.5. *Let* (Ω, P, π) *be a finite irreducible reversible Markov chain. Let* $A \subsetneq \Omega$ *be non-empty. Let* $\alpha > 0$ *and* $w \geq 0$. *Let* $B(A, w, \alpha) := \{y : \mathrm{P}_y[T_A > \lceil \frac{t_{\mathrm{rel}} w}{\pi(A)} \rceil] \geq \alpha\}$. *Then*

$$\begin{aligned} \mathrm{P}_\pi[T_A > t] &\leq \pi(A^c)\left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t \\ &\leq \pi(A^c)\exp\left(-\frac{t\pi(A)}{t_{\mathrm{rel}}}\right) \qquad \textit{for any } t \geq 0. \end{aligned} \tag{3.8}$$

In particular,

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

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

PROOF OF COROLLARY 3.4. Denote $s = s_{\alpha,\beta,\varepsilon} := \lceil \alpha^{-1} t_{\mathrm{rel}} \log(\frac{1-\alpha}{(1-\beta)\varepsilon}) \rceil$. Let $A \subset \Omega$ be an arbitrary set such that $\pi(A) \geq \alpha$. Consider the set

$$H_1 = H_1(A, \alpha, \beta, \varepsilon) := \{ y \in \Omega : \mathrm{P}_y[T_A \leq s] \geq 1 - \varepsilon \}.$$

Then by (3.9)

$$\begin{aligned}\pi(H_1) &\geq 1 - \big(1 - (1-\varepsilon)\big)^{-1}\big(1 - \pi(A)\big) \exp\left[-\frac{s\pi(A)}{t_{\mathrm{rel}}}\right] \\ &\geq 1 - \varepsilon^{-1}(1-\alpha) \exp\left[-\log\left(\frac{1-\alpha}{(1-\beta)\varepsilon}\right)\right] = \beta.\end{aligned}$$

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 $x \in \Omega$,

$$\begin{aligned}\mathrm{P}_x[T_A \leq t + s] &\geq \mathrm{P}_x[T_{H_1} \leq t, T_A \leq t + s] \geq (1-\varepsilon)\mathrm{P}_x[T_{H_1} \leq t] \\ &\geq (1-\varepsilon)\big(1 - p_x(\beta, t)\big) \geq 1 - \varepsilon - \max_{y \in \Omega} p_y(\beta, t).\end{aligned} \tag{3.10}$$

Taking $t := \mathrm{hit}_\beta(\delta - \varepsilon)$ 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* 1.8 *and Theorem* 3. Now we are ready to prove our main abstract results.

PROOF OF PROPOSITION 1.8. 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 $0 < \varepsilon \leq 1/4$ and $t \geq 0$. Take any set A with $\pi(A) \geq \frac{1}{2}$ and $x \in \Omega$. Denote $s_\varepsilon := \lceil 2t_{\mathrm{rel}} |\log \varepsilon| \rceil$. Consider a coupling $(\mathbb{P}, (Y_k, Z_k)_{k \geq 0})$ of the chain $(Y_k)_{k \geq 0}$ with initial distribution $Y_0 \sim \mathrm{P}_x^t$ with the stationary chain $(Z_k)_{k \geq 0}$ so that $\mathbb{P}[(Y_k)_{k \geq 0} \neq (Z_k)_{k \geq 0}] = d_x(t)$ (cf. the proofs of Proposition 4.7 and of Theorem 5.2 in [18] for the existence of such a coupling). By the Markov property,

$$\begin{aligned}\mathrm{P}_x[T_A > t + s_\varepsilon] &\leq \mathrm{P}_x[X_k \notin A \text{ for all } t \leq k \leq t + s_\varepsilon] \\ &= \mathbb{P}[Y_k \notin A \text{ for all } k \leq s_\varepsilon] \\ &\leq \mathbb{P}\big[(Y_k)_{k \geq 0} \neq (Z_k)_{k \geq 0}\big] + \mathbb{P}[Z_k \notin A \text{ for all } k \leq s_\varepsilon] \\ &= d_x(t) + \mathrm{P}_\pi[T_A > s_\varepsilon].\end{aligned}$$

Hence, by (3.8)

$$\mathrm{P}_x[T_A > t + s_\varepsilon] \leq d_x(t) + \frac{1}{2} e^{-s_\varepsilon / 2t_{\mathrm{rel}}} \leq d(t) + \frac{\varepsilon}{2}.$$

Putting $t = t_{\mathrm{mix}}(\varepsilon)$ and $t = t_{\mathrm{mix}}(1-\varepsilon)$ successively in the above equation and maximizing over $x \in \Omega$ and A such that $\pi(A) \geq \frac{1}{2}$ gives

$$\mathrm{hit}_{1/2}(3\varepsilon/2) \leq t_{\mathrm{mix}}(\varepsilon) + s_\varepsilon \quad \text{and} \quad \mathrm{hit}_{1/2}(1-\varepsilon/2) \leq t_{\mathrm{mix}}(1-\varepsilon) + s_\varepsilon,$$

which completes the proof. □

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

PROPOSITION 3.6. *Let (Ω_n, P_n, π_n) be a sequence of lazy finite irreducible reversible chains for which 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_α*-cutoff.*
(2) *The sequence exhibits a* hit_α*-cutoff for any $\alpha \in (0,1)$.*

Moreover,

$$\mathrm{hit}_\alpha^{(n)}(1/4) = \Theta\big(t_{\mathrm{mix}}^{(n)}\big) \qquad \text{for any } \alpha \in (0,1). \tag{3.11}$$

Furthermore, if (2) *holds then*

$$\lim_{n\to\infty} \mathrm{hit}_\alpha^{(n)}(1/4)/\mathrm{hit}_{1/2}^{(n)}(1/4) = 1 \qquad \text{for any } \alpha \in (0,1). \tag{3.12}$$

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

$$\begin{aligned}\mathrm{hit}_\alpha^{(n)}(1/4) &\leq 4\alpha^{-1}\,\mathrm{hit}_\alpha^{(n)}\left(1-\frac{3\alpha}{4}\right) \leq 4\alpha^{-1} t_{\mathrm{mix}}^{(n)}\left(\frac{\alpha}{4}\right) \\ &\leq 4\alpha^{-1}\big(2+\lceil \log_2(1/\alpha)\rceil\big) t_{\mathrm{mix}}^{(n)}.\end{aligned}$$

The first inequality above follows from (3.6) and the fact that $(1-3\alpha/4)^{4\alpha^{-1}-1} \leq 4e^{-3} \leq 1/4$. The second one follows from (3.2) (first inequality). The final inequality above is a consequence of the sub-multiplicativity property: for any $k, t \geq 0$, $d(kt) \leq (2d(t))^k$ (e.g., [18], (4.24) and Lemma 4.12).

Conversely, by (3.6) (second inequality) and the second inequality in (3.2) with $(\alpha, \varepsilon, \delta)$ here being $(1-\alpha, 1/8, 1/8)$ (first inequality)

$$\frac{t_{\mathrm{mix}}^{(n)}}{2} - \left\lceil \frac{t_{\mathrm{rel}}^{(n)}}{4} \log\left(\frac{100}{1-\alpha}\right)\right\rceil \leq \frac{\mathrm{hit}_\alpha^{(n)}(1/8)}{2} \leq \mathrm{hit}_\alpha^{(n)}(1/4).$$

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

Fix $\varepsilon \in (0, 1/8)$. Denote $s_n = s_n(\alpha, \beta, \varepsilon) := \lceil t_{\mathrm{rel}}^{(n)} \alpha^{-1} \log(\frac{1-\alpha}{(1-\beta)\varepsilon}) \rceil$. By the second inequality in Corollary 3.4,

$$\mathrm{hit}_\alpha^{(n)}(1-\varepsilon) \le \mathrm{hit}_\beta^{(n)}(1-2\varepsilon) + s_n \quad \text{and} \quad \mathrm{hit}_\alpha^{(n)}(2\varepsilon) \le \mathrm{hit}_\beta^{(n)}(\varepsilon) + s_n. \tag{3.13}$$

By the first inequality in Corollary 3.4,

$$\begin{aligned} \mathrm{hit}_\beta^{(n)}(2\varepsilon) \le \mathrm{hit}_\alpha^{(n)}(2\varepsilon) \le \mathrm{hit}_\alpha^{(n)}(\varepsilon) \quad \text{and} \\ \mathrm{hit}_\beta^{(n)}(1-\varepsilon) \le \mathrm{hit}_\beta^{(n)}(1-2\varepsilon) \le \mathrm{hit}_\alpha^{(n)}(1-2\varepsilon). \end{aligned} \tag{3.14}$$

Hence

$$\begin{aligned} \mathrm{hit}_\beta^{(n)}(2\varepsilon) - \mathrm{hit}_\beta^{(n)}(1-2\varepsilon) \le \mathrm{hit}_\alpha^{(n)}(\varepsilon) - \mathrm{hit}_\alpha^{(n)}(1-\varepsilon) + s_n, \\ \mathrm{hit}_\alpha^{(n)}(2\varepsilon) - \mathrm{hit}_\alpha^{(n)}(1-2\varepsilon) \le \mathrm{hit}_\beta^{(n)}(\varepsilon) - \mathrm{hit}_\beta^{(n)}(1-\varepsilon) + s_n. \end{aligned} \tag{3.15}$$

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

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

$$\mathrm{hit}_\alpha^{(n)}(1/4+\varepsilon) - s_n \le \mathrm{hit}_\beta^{(n)}(1/4) \le \mathrm{hit}_\alpha^{(n)}(1/4). \tag{3.16}$$

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

The following proposition shows that for all $\alpha \le 1/2$ the occurrence of hit_α-cutoff implies that the product condition holds. In particular, this implies the equivalence of (2) and (3) in Theorem 3.

PROPOSITION 3.7. *Let (Ω_n, P_n, π_n) be a sequence of lazy finite irreducible reversible chains. Assume that the product condition fails. Then for any $\alpha \le 1/2$ the sequence does not exhibit* hit_α*-cutoff.*

Before providing the proof of Proposition 3.7, we complete the proof of Theorem 3.

PROOF OF THEOREM 3. By Fact 1.1 and Proposition 3.7, it suffices to consider the case in which the product condition holds. By Propositions 3.6, it suffices to consider the case $\alpha = 1/2$ (i.e., it suffices to show that under the product condition the sequence exhibits cutoff iff it exhibits $\text{hit}_{1/2}$-cutoff). This follows at once from (1.4), (1.5) and (3.12). □

PROOF OF PROPOSITION 3.7. Fix some $0 < \alpha \le 1/2$. We first argue that for all $n, k \ge 1$

$$\text{hit}_{\alpha}^{(n)}\big([1-\alpha/2]^k\big) \le k\big\lceil|\log_2(\alpha/2)|\big\rceil t_{\text{mix}}^{(n)}. \tag{3.17}$$

By the sub-multiplicativity property (3.6), it suffices to verify (3.17) only for $k = 1$. As in the proof of Proposition 3.6, by the sub-multiplicativity property $d(mt) \le (2d(t))^m$, together with (3.2), we have that $\text{hit}_{\alpha}^{(n)}(1-\alpha/2) \le t_{\text{mix}}^{(n)}(\alpha/2) \le \lceil|\log_2(\alpha/2)|\rceil) t_{\text{mix}}^{(n)}$.

Conversely, by the laziness assumption, we have that for all n,

$$\text{hit}_{\alpha}^{(n)}(\varepsilon/2) \ge |\log_2 \varepsilon| \qquad \text{for all } 0 < \varepsilon < 1. \tag{3.18}$$

To see this, consider the case that $X_0^{(n)} = y_n$, for some $y_n \in \Omega_n$ such that $\pi_n(y_n) \le 1/2 \le 1 - \alpha$, and that the first $\lfloor|\log_2 \varepsilon|\rfloor$ steps of the chain are lazy (i.e., $y_n = X_1^{(n)} = \cdots = X_{\lfloor|\log_2 \varepsilon|\rfloor}$).

By (3.17) in conjunction with (3.18), we may assume that $\lim_{n\to\infty} t_{\text{mix}}^{(n)} = \infty$, as otherwise there cannot be hit_α-cutoff. By passing to a subsequence, we may assume further that there exists some $C > 0$ such that $t_{\text{mix}}^{(n)} < Ct_{\text{rel}}^{(n)}$. In particular, $\lim_{n\to\infty} t_{\text{rel}}^{(n)} = \infty$ and we may assume without loss of generality that $(\lambda_2^{(n)})^{t_{\text{mix}}^{(n)}} \ge e^{-C}$ for all n, where $\lambda_2^{(n)}$ is the second largest eigenvalue of P_n.

For notational convenience, we now suppress the dependence on n from our notation. Let $f_2 \in \mathbb{R}^{\Omega}$ be a nonzero vector satisfying that $Pf_2 = \lambda_2 f_2$. By considering $-f_2$ if necessary, we may assume that $A := \{x \in \Omega : f_2 \le 0\}$ satisfies $\pi(A) \ge 1/2$. Let $x \in \Omega$ be such that $f_2(x) = \max_{y\in\Omega} f_2(y) =: L$. Note that $L > 0$ since $\mathbb{E}_\pi[f_2] = 0$.

Consider $N_k := \lambda_2^{-k} f_2(X_k)$ and $M_k := N_{k\wedge T_A}$, where $X_0 = x$. Observe that $(N_k)_{k\ge 0}$ is a martingale, and hence so is $(M_k)_{k\ge 0}$ (w.r.t. the natural filtration induced by the chain). As $M_k \le 0$ on $\{T_A \le k\}$ and $M_k \le \lambda_2^{-k} L$ on $\{T_A > k\}$, we get that for all $k > 0$, $M_k \le \lambda_2^{-k} L 1_{T_A > k}$, and so

$$L = \mathbb{E}_x[M_0] = \mathbb{E}_x[M_k] \le \mathbb{E}_x\big[\lambda_2^{-k} L 1_{T_A>k}\big] = \lambda_2^{-k} L \mathrm{P}_x[T_A > k]. \tag{3.19}$$

Thus, $\mathrm{P}_x[T_A > k] \ge \lambda_2^k$, for all k. Consequently, for all $a > 0$,

$$\mathrm{P}_x[T_A > at_{\text{mix}}] \ge \lambda_2^{at_{\text{mix}}} \ge e^{-aC}. \tag{3.20}$$

Thus

$$\text{hit}_\alpha(\varepsilon/2) \ge \text{hit}_{1/2}(\varepsilon/2) \ge C^{-1} t_{\text{mix}} |\log \varepsilon| \qquad \text{for any } 0 < \varepsilon < 1.$$

This, in conjunction with (3.17), implies that $\frac{\text{hit}_\alpha(\varepsilon)}{\text{hit}_\alpha(1-\varepsilon)} \ge \frac{|\log \varepsilon|}{C\lceil \log_2(\alpha/2)\rceil}$, for all $0 < \varepsilon \le \alpha/2$. Consequently, there is no hit_α-cutoff. □

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 [2], Chapter 3, Sections 5 and 6.5 or [4]). 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 $\pi(A)/t_{\mathrm{rel}}$. Since the existing literature lacks simple treatment of this fact (especially for the discrete-time case), we now prove it for the sake of completeness. We shall prove this fact without assuming laziness. Although without assuming laziness, the distribution of T_A under P_π need not be completely monotone, the proof is essentially identical as in the lazy case.

For any nonempty $A \subset \Omega$, we write π_A for the distribution of π conditioned on A. That is, $\pi_A(\cdot) := \frac{\pi(\cdot)1_{\cdot \in A}}{\pi(A)}$. For any matrix P and $f \in \mathbb{R}^\Omega$ we denote $\mathcal{E}_P(f) := \langle (I-P)f, f\rangle_\pi$.

LEMMA 3.8. *Let (Ω, P, π) be a reversible irreducible finite chain. Let $A \subsetneq \Omega$ be nonempty. Denote its complement by B and write $k = |B|$. 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 $t \ge 0$ such that $P_B^t(x,y) > 0$. Then*:

(i) *P_B has k real eigenvalues $1 - \pi(A)/t_{\mathrm{rel}} \ge \gamma_1 > \gamma_2 \ge \cdots \ge \gamma_k \ge -\gamma_1$.*

(ii) *There exist some nonnegative $a_1, \dots, a_k$ satisfying $\sum_{i=1}^k a_i = 1$ such that for any $t \ge 0$,*

$$\mathrm{P}_{\pi_B}[T_A > t] = \sum_{i=1}^k a_i \gamma_i^t. \tag{3.21}$$

(iii)

$$\mathrm{P}_{\pi_B}[T_A > t] \le \left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t \le \exp\left(-\frac{t\pi(A)}{t_{\mathrm{rel}}}\right) \qquad \textit{for all } t \ge 0. \tag{3.22}$$

PROOF. We first note that (3.22) follows immediately from (3.21) and (i). Indeed, by (i), $|\gamma_i| \le \gamma_1 \le 1 - \frac{\pi(A)}{t_{\mathrm{rel}}}$ for all i, and so (3.21) implies that $\mathrm{P}_{\pi_B}[T_A > t] \le \gamma_1^t \le (1 - \frac{\pi(A)}{t_{\mathrm{rel}}})^t$ for all $t \ge 0$.

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

By the Courant–Fischer variational characterization of eigenvalues, we have

$$1-\gamma_1=\inf\left\{\frac{\mathcal{E}_P(g)}{\langle g,g\rangle_\pi}:g\geq 0, g=0 \text{ on } A, g \text{ nonconstant}\right\}. \tag{3.23}$$

Also observe that for all $g\geq 0$ such that $g=0$ on A we have by the Cauchy–Schwarz inequality that $\mathbb{E}_{\pi_B} g^2 \geq (\mathbb{E}_{\pi_B} g)^2$ [where for $f\in\mathbb{R}^\Omega$ we denote $\mathbb{E}_{\pi_B} f:=\sum_b \pi_B(b)f(b)$] which rearranges to

$$\mathbf{Var}_\pi\, g=\langle g-\mathbb{E}_\pi g, g-\mathbb{E}_\pi g\rangle_\pi\geq\pi(A)\langle g,g\rangle_\pi.$$

Thus, by (3.23) $1-\gamma_1\geq\pi(A)\inf\{\mathcal{E}_P(g)/\mathbf{Var}_\pi\, g: g\geq 0, g=0 \text{ on } A, g$ nonconstant$\}$, which in comparison with the variational characterization of t_{rel} (e.g., [18], Remark 13.13)

$$1/t_{\mathrm{rel}}=\inf\{\mathcal{E}_P(g)/\mathbf{Var}_\pi\, g: g \text{ nonconstant}\},$$

yields that $1-\gamma_1\geq\pi(A)/t_{\mathrm{rel}}$. This completes 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

$$\mathrm{P}_{\pi_B}[T_A>t]=\sum_{x,y\in B}\pi_B(x)P_B^t(x,y). \tag{3.24}$$

By the spectral representation (cf. [18], Lemma 12.2, and Section 4 of Chapter 3 in [2]) for any $x,y\in B$ and $t\in\mathbb{N}$ we have that $P_B^t(x,y)=\sum_{i=1}^k\pi_B(y)g_i(x)g_i(y)\gamma_i^t$. So by (3.24)

$$\mathrm{P}_{\pi_B}[T_A>t]=\sum_{x,y\in B}\pi_B(x)\sum_{i=1}^k\pi_B(y)g_i(x)g_i(y)\gamma_i^t=\sum_{i=1}^k a_i\gamma_i^t,$$

where $a_i:=(\sum_{x\in B}\pi_B(x)g_i(x))^2$. Plugging $t=0$ shows that indeed $\sum_{i=1}^k a_i=1$, as desired. $\square$

Using the same argument for the continuous-time setup, it follows that

$$\begin{aligned}\mathrm{H}_{\pi_B}[T_A^{\mathrm{ct}}>t]&=\sum_{x,y\in B}\pi_B(x)\sum_{i=1}^k\pi_B(y)g_i(x)g_i(y)e^{-(1-\gamma_i)t}\\&=\sum_{i=1}^k a_ie^{-(1-\gamma_i)t}\leq e^{-t\pi(A)/t_{\mathrm{rel}}}.\end{aligned}$$

We now present an alternative argument, which does not require reversibility (which allows one to extend Corollary 3.4 and Lemma 3.5 to the nonreversible continuous-time setup).

REMARK 3.9. When P is nonreversible, one can consider its additive symmetrization $S = (P + P^*)/2$ (which is reversible), where $P^*(x, y) := \pi(y)P(y, x)/\pi(x)$ is the time-reversal of P (and the dual operator). Define $t_{\mathrm{rel}} := 1/(1 - \lambda_2(S))$, where $\lambda_2(S)$ is the second largest eigenvalue of S. Let $A \subsetneq \Omega$ be an irreducible set. Denote its complement by B. We argue that with this notation, it is still the case that $\mathrm{H}_{\pi_B}[T_A^{\mathrm{ct}} > t] \le e^{-t\pi(A)/t_{\mathrm{rel}}}$.

As above, consider the sub-stochastic matrices P_B, P_B^*, H_B^t and S_B, which are the restrictions of P, P^*, H_t and S (resp.) to B. For any $f, g \in \mathbb{R}^B$ we denote $\mathbb{E}_{\pi_B}[f] := \sum_{x \in B} \pi_B(x) f(x)$, $\langle f, g\rangle_{\pi_B} := \mathbb{E}_{\pi_B}[fg]$ and $\|f\|_{B,p}^p := \mathbb{E}_{\pi_B}[|f|^p]$. Let $h_t(x) := H_B^t 1_B(x) = \mathrm{H}_x[T_A^{\mathrm{ct}} > t]$. Then

$$\mathrm{H}_{\pi_B}[T_A^{\mathrm{ct}} > t] = \mathbb{E}_{\pi_B}[h_t] = \|h_t\|_{B,1} \le \|h_t\|_{B,2}.$$

Let $f \in \mathbb{R}^B$. For any linear operator $Q : \mathbb{R}^B \to \mathbb{R}^B$ denote $\mathcal{E}_Q(f) := \langle (I - Q)f, f\rangle_{\pi_B}$. Since $\mathcal{E}_{P_B}(f) = \mathcal{E}_{P_B^*}(f)$ it is also the case that $\mathcal{E}_{P_B}(f) = \mathcal{E}_{S_B}(f)$. An elementary calculation shows that

$$-\frac{d}{dt}\|H_B^t f\|_{B,2}^2 = 2\mathcal{E}_{P_B}(H_B^t f) = 2\mathcal{E}_{S_B}(H_B^t f).$$

Moreover, if f is nonnegative and nonzero, then as in (3.23) we have that

$$\mathcal{E}_{S_B}(H_B^t f) \ge \pi(A)\|H_B^t f\|_{B,2}^2/t_{\mathrm{rel}}.$$

Hence,

$$\frac{d}{dt}\|H_B^t f\|_{B,2}^2 \le -2\pi(A)\|H_B^t f\|_{B,2}^2/t_{\mathrm{rel}}.$$

Substituting $f = 1_B$ yields that $\mathrm{H}_{\pi_B}[T_A^{\mathrm{ct}} > t] \le \|h_t\|_{B,2} \le \exp[-t\pi(A)/t_{\mathrm{rel}}]$, as desired.

PROOF OF LEMMA 3.5. We first note that (3.9) follows easily from (3.8). For the first inequality in (3.9) set $t = t(A, w) := \lceil t_{\mathrm{rel}} w/\pi(A)\rceil$ and $B := B(A, w, \alpha) = \{y : \mathrm{P}_y[T_A > t] \ge \alpha\}$. Then by (3.8)

$$\begin{aligned}\alpha\pi(B) &\le \pi(B)\mathrm{P}_{\pi_B}[T_A > t] \le \mathrm{P}_\pi[T_A > t] \\ &\le \pi(A^c)\exp(-t\pi(A)/t_{\mathrm{rel}}) \le \pi(A^c)e^{-w}.\end{aligned}$$

For the first inequality in (3.9), recall that $\mathbb{E}_\pi[T_A] = \sum_{t>0} \mathrm{P}_\pi[T_A > t]$ and apply (3.8).

We now prove (3.8). Denote the connected components of $A^c := \Omega \setminus A$ by $\{C_1, \dots, C_k\}$. Denote the complement of C_i by C_i^c. By (3.22), we have that

$$\begin{aligned} \mathrm{P}_\pi[T_A > t] &= \sum_{i=1}^{k} \pi(C_i)\mathrm{P}_{\pi_{C_i}}[T_A > t] = \sum_{i=1}^{k} \pi(C_i)\mathrm{P}_{\pi_{C_i}}[T_{C_i^c} > t] \\ &\le \sum_{i=1}^{k} \pi(C_i)\left(1 - \frac{\pi(C_i^c)}{t_{\mathrm{rel}}}\right)^t \le \sum_{i=1}^{k} \pi(C_i)\left(1 - \frac{\pi(A)}{t_{\mathrm{rel}}}\right)^t \\ &= \pi(A^c)\exp\left(-\frac{t\pi(A)}{t_{\mathrm{rel}}}\right). \end{aligned}$$

□

4. Continuous-time. In this section, we explain the necessary adaptations in the proof of Proposition 1.8 for the continuous-time case. We also explain the necessary adaptations in the proof of the continuous time analogue of Theorem 3. More details can be found at [15]. We fix some finite, irreducible, reversible chain (Ω, P, π). For notational convenience, exclusively for this section, we shall denote the transition-matrix of $(X_k^{\mathrm{NL}})_{k \ge 0}$, the nonlazy 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 $1 = \lambda_1^{\mathrm{ct}} > \lambda_2^{\mathrm{ct}} \ge \cdots \ge \lambda_{|\Omega|}^{\mathrm{ct}} \ge -1$ and that of P_L by $1 = \lambda_1^L > \lambda_2^L \ge \cdots \ge \lambda_{|\Omega|}^L \ge -1$ (where $1 + \lambda_i^{\mathrm{ct}} = 2\lambda_i^L$). We denote $t_{\mathrm{rel}}^{\mathrm{ct}} := (1 - \lambda_2^{\mathrm{ct}})^{-1}$ and $t_{\mathrm{rel}}^L := (1 - \lambda_2^L)^{-1}$. We identify H_t with the operator $H_t : L^2(\mathbb{R}^\Omega, \pi) \to L^2(\mathbb{R}^\Omega, \pi)$, defined by $H_t f(x) = \mathbb{E}_x[f(X_t^{\mathrm{ct}})]$. The spectral decomposition in continuous-time takes the following form. If $f_1, \dots, f_{|\Omega|}$ is an orthonormal basis such that $Pf_i := \lambda_i^{\mathrm{ct}} f_i$ for all i, then $H_t g = \mathbb{E}_\pi H_t g + \sum_{i=2}^{|\Omega|} \langle g, f_i \rangle_\pi e^{-(1-\lambda_i^{\mathrm{ct}})t} f_i$, for all $g \in \mathbb{R}^\Omega$ and $t \ge 0$. Thus, the L^2-contraction lemma takes the following form in continuous-time (see, e.g., [18], Lemma 20.5):

$$\mathbf{Var}_\pi\, H_t f \le e^{-2t/t_{\mathrm{rel}}^{\mathrm{ct}}}\,\mathbf{Var}_\pi\, f \qquad \text{for any } f \in \mathbb{R}^\Omega, \text{ for any } t \ge 0. \tag{4.1}$$

Starr's inequality holds also in continuous-time ([25], Proposition 3) and takes the following form. Let $f \in \mathbb{R}^\Omega$. Define the *continuous-time maximal function* as $f_{\mathrm{ct}}^*(x) := \sup_{t \ge 0} |H_t f(x)|$. Then

$$\|f_{\mathrm{ct}}^*\|_2 \le 2\|f\|_2. \tag{4.2}$$

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

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

$$G_s^{\mathrm{ct}}(A, m) := \{y : |H_t(1_A)(y) - \pi(A)| < m\sigma_s^{\mathrm{ct}} \text{ for all } t \ge s\}.$$

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

$$\pi(G_s^{\mathrm{ct}}(A, m)) \ge 1 - 4/m^2 \qquad \text{for all } A \subset \Omega, s \ge 0 \text{ and } m > 0. \tag{4.3}$$

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

Finally, the proof of Proposition 1.8 in the continuous-time case is concluded by noting that the coupling argument from the discrete-time case carries over to the continuous-time case, with (3.8) replaced by the inequality $\mathrm{H}_\pi[T_A^{\mathrm{ct}} > t] \le \pi(A^c)\exp(-t\pi(A)/t_{\mathrm{rel}}^{\mathrm{ct}})$ (see Remark 3.9).

We now explain the required adaptations for proving the continuous-time analog of Theorem 3. By the last inequality, as in Lemma 3.5, $B_{\mathrm{ct}}(A, w, \alpha) := \{y : \mathrm{H}_y[T_A^{\mathrm{ct}} \ge w t_{\mathrm{rel}}^{\mathrm{ct}}/\pi(A)] \ge \alpha\}$ satisfies

$$\pi\big(B_{\mathrm{ct}}(A, w, \alpha)\big) \le \pi\big(A^c\big)e^{-w}\alpha^{-1} \qquad \text{for all } w \ge 0 \text{ and } 0 < \alpha \le 1. \tag{4.4}$$

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

$$\begin{aligned} \mathrm{hit}_\beta^{\mathrm{ct}}(\delta) \le \mathrm{hit}_\alpha^{\mathrm{ct}}(\delta) \le \mathrm{hit}_\beta^{\mathrm{ct}}(\delta - \varepsilon) + \alpha^{-1} t_{\mathrm{rel}}^{\mathrm{ct}} \log\left(\frac{1-\alpha}{(1-\beta)\varepsilon}\right) \\ \text{for any } 0 < \alpha \le \beta < 1. \end{aligned} \tag{4.5}$$

Using (4.5) and the continuous-time analog of Proposition 1.8 and of (3.6), one can obtain a continuous-time analog of Proposition 3.6 by imitating the discrete time proof.

Finally, in order to obtain a the continuous-time analog of Proposition 3.7 replace the discrete time martingale $\lambda_2^{-k} f_2(X_k)$ by the continuous-time martingale $e^{(1-\lambda_2)t} f_2(X_t)$.

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, π), on a finite tree $T := (V, E)$. That is, a chain with stationary distribution π and state space V such that $P(x, y) > 0$ iff $\{x, y\} \in E$ or $y = x$ [in which case, $P(x, x) \ge 1/2$]. Then P is reversible by Kolmogorov's cycle condition.

Following [24], 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" toward o. A nonintuitive result is that one can construct such unweighed trees [23].

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 $\ell(u) = (u_0 = u, u_1, \dots, u_k = o)$. 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 $\varnothing \neq A \subset V$, we write π_A for the distribution of π conditioned on A, $\pi_A(\cdot) := \frac{\pi(\cdot)1_{\cdot \in A}}{\pi(A)}$.

A key observation is that starting from the central vertex o, the chain mixes rapidly (this follows implicitly from the following analysis). Let T_o denote the hitting time of the central vertex. We define the mixing parameter $\tau(\varepsilon)$ for $\varepsilon \in (0, 1)$ by

$$\tau_o(\varepsilon) := \min\{t : \mathrm{P}_x[T_o > t] \le \varepsilon \ \forall x \in \Omega\}.$$

We show that up to terms of the order of the relaxation-time (which are negligible under the product condition) $\tau_o(\cdot)$ approximates $\mathrm{hit}_{1/2}(\cdot)$ and then using Proposition 3, the question of cutoff is reduced to showing concentration for the hitting time of the central vertex. Below we make this precise.

LEMMA 5.1. *Denote* $s_\delta := \lceil 4t_{\mathrm{rel}} |\log(4\delta/9)|\rceil$. *Then*

$$\tau_o(\varepsilon) \le \mathrm{hit}_{1/2}(\varepsilon) \le \tau_o(\varepsilon - \delta) + s_\delta \qquad \text{for every } 0 < \delta < \varepsilon < 1. \tag{5.1}$$

PROOF. First observe that by the definition of central vertex, for any $x \in V$, $x \neq o$ there exists a set A with $\pi(A) \ge \frac{1}{2}$ such that the chain starting at x cannot hit A without first hitting o. Indeed, we can take A to be the union of $\{o\}$ and all components of $T \setminus \{o\}$ not containing x. The first inequality in (5.1) follows trivially from this.

To establish the other inequality, fix $A \subseteq V$ with $\pi(A) \ge \frac{1}{2}$, $x \in V$ and some $0 < \delta < \varepsilon < 1$. It follows using Markov property and the definition of $\tau_o(\varepsilon - \delta)$ that

$$\begin{aligned}\mathrm{P}_x[T_A > \tau_o(\varepsilon - \delta) + s_\delta] &\le \mathrm{P}_x[T_o > \tau_o(\varepsilon - \delta)] + \mathrm{P}_o[T_A > s_\delta] \\ &\le \varepsilon - \delta + \mathrm{P}_o[T_A > s_\delta].\end{aligned}$$

Hence, it suffices to show that $\mathrm{P}_o[T_A > s_\delta] \le \delta$. If $o \in A$ then $\mathrm{P}_o[T_A > s_\delta] = 0$, so without loss of generality assume $o \notin A$. It is easy to see that we can partition $T \setminus \{o\} = T_1 \cup T_2$ such that both T_1 and T_2 are unions of components of $T \setminus \{o\}$ and $\pi(T_1), \pi(T_2) \le 2/3$. For $i = 1, 2$, let $A_i := A \cap T_i$ and without loss of generality let us assume $\pi(A_1) \ge \frac{1}{4}$. Let $B = T_2 \cup \{o\}$. Clearly, the chain started at any $x \in B$ must hit o before hitting A_1. Hence,

$$\begin{aligned}\mathrm{P}_o[T_A > s_\delta] &\le \mathrm{P}_o[T_{A_1} > s_\delta] \\ &\le \mathrm{P}_{\pi_B}[T_{A_1} > s_\delta] \\ &\le \pi(B)^{-1}\mathrm{P}_\pi[T_{A_1} > s_\delta].\end{aligned} \tag{5.2}$$

Using $\pi(A_1) \ge \frac{1}{4}$, $\pi(B) \ge \frac{1}{3}$ it follows from (3.8) that $\pi(B)^{-1}\mathrm{P}_\pi[T_{A_1} > s_\delta] \le \delta$. $\square$

In light of Lemma 5.1 and Proposition 1.8, in order to show that in the setup of Theorem 1 (under the product condition) cutoff occurs it suffices to show that $\tau_{o_n}^{(n)}(\varepsilon) - \tau_{o_n}^{(n)}(1-\varepsilon) = o(t_{\mathrm{mix}}^{(n)})$, for any $\varepsilon \in (0, 1/4]$. We actually show more than that. Instead of identifying the "worst" starting position x and proving that T_o is concentrated under P_x, we shall show that for any $x, y \in V_n$ such that $y \prec x$ and $\mathbb{E}_x[T_y] = \Theta(t_{\mathrm{mix}}^{(n)})$, T_y is concentrated under P_x, around $\mathbb{E}_x[T_y]$, with deviations of order $\sqrt{t_{\mathrm{rel}}^{(n)} t_{\mathrm{mix}}^{(n)}}$. This shall follow from Chebyshev inequality, once we establish that $\mathbf{Var}_x[T_y] \le 4 t_{\mathrm{rel}} \mathbb{E}_x[T_y]$.

Let $(v_0 = x, v_1, \dots, v_k = y)$ 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 $\tau_1, \dots, \tau_k$ are independent. This reduces the task of bounding $\mathbf{Var}_x[T_y]$ from above, to the task of estimating $\mathbf{Var}_{v_i}[T_{v_{i+1}}] = \mathbf{Var}_{v_i}[T_{f_{v_i}}]$ from above for each i.

LEMMA 5.2. *For any vertex $u \neq o$, we have that*

$$t_u := \mathbb{E}_u[T_{f_u}] = \frac{\pi(W_u)}{\pi(u)\mu_u} \quad \textit{and}$$
$$r_u := \mathbb{E}_u[T_{f_u}^2] = 2 t_u \mathbb{E}_{\pi_{W_u}}[T_{f_u}] - t_u \le 4 t_u t_{\mathrm{rel}}. \tag{5.3}$$

The assertion of Lemma 5.2 follows as a particular case of Proposition 5.6 at the end of this section.

COROLLARY 5.3. *Let $x, y \in V$ be such that $y \preceq x$ and $c \ge 0$. Denote $\sigma_{x,y} := \sqrt{4\mathbb{E}_x[T_y] t_{\mathrm{rel}}}$. Then*

$$\mathbf{Var}_x[T_y] \le \sigma_{x,y}^2, \tag{5.4}$$

$$\mathrm{P}_x\big[T_y \ge \mathbb{E}_x[T_y] + c\sigma_{x,y}\big] \le \frac{1}{1+c^2} \quad \textit{and}$$
$$\mathrm{P}_x\big[T_y \le \mathbb{E}_x[T_y] - c\sigma_{x,y}\big] \le \frac{1}{1+c^2}. \tag{5.5}$$

In particular, if (V_n, P_n, π_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 $y_n \prec x_n$ and $\mathbb{E}_{x_n}[T_{y_n}]/t_{\mathrm{rel}}^{(n)} \to \infty$, then for any $\varepsilon > 0$ we have that

$$\lim_{n\to\infty} \mathrm{P}_{x_n}\big[\big|T_{y_n} - \mathbb{E}_{x_n}[T_{y_n}]\big| \ge \varepsilon \mathbb{E}_{x_n}[T_{y_n}]\big] = 0. \tag{5.6}$$

PROOF. We first note that (5.5) follows from (5.4) by the one-sided Chebyshev inequality. Also, (5.6) follows immediately from (5.5). We now prove (5.4). Let $(v_0 = x, v_1, \dots, v_k = y)$ be the path from x to y. 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 $\tau_1, \dots, \tau_k$ are independent. Whence, by (5.3) we get that

$$\mathbf{Var}_x[T_y] = \sum_{i=1}^{k} \mathbf{Var}_{v_{i-1}}[T_{v_i}] \le \sum_{i=1}^{k} \mathbb{E}_{v_{i-1}}[T_{v_i}^2] \le 4t_{\mathrm{rel}} \sum_{i=1}^{k} \mathbb{E}_{v_{i-1}}[T_{v_i}] = \sigma_{x,y}^2.$$

This completes the proof. □

LEMMA 5.4. *If* (V, P, π) *is a lazy chain on a* (*weighted*) *tree* (T, o) *then*

$$\mathbb{E}_x[T_o] \le 4t_{\mathrm{mix}} \qquad \text{for all } x \in V. \tag{5.7}$$

PROOF. Fix some $x \in V$. Let C_x be the component of $T \setminus \{o\}$ containing x. Denote $B := V \setminus C_x$. Consider $\tau_B := \inf\{k \in \mathbb{N} : X_{kt_{\mathrm{mix}}} \in B\}$. Clearly, $T_o \le \tau_B t_{\mathrm{mix}}$. Since $\pi(B) \ge 1/2$, by the Markov property and the definition of the total variation distance, the distribution of τ_B is stochastically dominated by the geometric distribution with parameter $1/2 - 1/4 = 1/4$. Hence, $\mathbb{E}_x[T_o] = \mathbb{E}_x[T_B] \le t_{\mathrm{mix}}\mathbb{E}_x[\tau_B] \le 4t_{\mathrm{mix}}$. □

COROLLARY 5.5. *In the setup of Lemma* 5.2, *for any* $x \in V$ *denote* $t_x := \mathbb{E}_x[T_o]$. *Fix* $\varepsilon \in (0, \frac{1}{4}]$, *Denote*

$$\rho := \max_{x \in V} t_x, \quad \text{and} \quad \kappa_\varepsilon := \sqrt{4\varepsilon^{-1}\rho t_{\mathrm{rel}}}, \qquad \text{then}$$

$$\rho \le 4t_{\mathrm{mix}}, \qquad \tau_o(1-\varepsilon) \ge \rho - \kappa_\varepsilon \quad \text{and} \quad \tau_o(\varepsilon) < \rho + \kappa_\varepsilon. \tag{5.8}$$

PROOF. By (5.7) $\rho \le 4t_{\mathrm{mix}}$. Denote $\sigma := \sqrt{4\rho t_{\mathrm{rel}}}$ and $c_\varepsilon := \sqrt{\varepsilon^{-1} - 1}$. Take $x \in V \setminus \{o\}$. By (5.4) $\sigma_{x,o}^2 := \mathbf{Var}_x[T_o] \le \sigma^2$. The assertion of the corollary now follows from (5.5) by noting that $c_\varepsilon \sigma \le \kappa_\varepsilon$. □

Now we are ready to prove Theorem 1.

PROOF OF THEOREM 1. Fix $\varepsilon \in (0, \frac{1}{4}]$. It follows from (1.4) and (1.5) that

$$\begin{aligned} &t_{\mathrm{mix}}(\varepsilon) - t_{\mathrm{mix}}(1-\varepsilon) \\ &\quad \le \mathrm{hit}_{1/2}(\varepsilon/2) - \mathrm{hit}_{1/2}(1-\varepsilon/2) + t_{\mathrm{rel}}\big(3|\log \varepsilon| + \log 4\big) + 2. \end{aligned} \tag{5.9}$$

Using Lemma 5.1 with (ε, δ) there replaced by $(\varepsilon/2, \varepsilon/4)$, it follows that

$$\mathrm{hit}_{1/2}(\varepsilon/2) - \mathrm{hit}_{1/2}(1-\varepsilon/2) \le \tau_o(\varepsilon/4) - \tau_o(1-\varepsilon/2) + s_{\varepsilon/4}, \tag{5.10}$$

where $s_{\varepsilon/4}$ is as in Lemma 5.1. It follows from (5.9), (5.10) and (5.8) that

$$\begin{aligned} &t_{\mathrm{mix}}(\varepsilon) - t_{\mathrm{mix}}(1-\varepsilon) \\ &\quad \le \kappa_{\varepsilon/4} + \kappa_{\varepsilon/2} + t_{\mathrm{rel}}\big(7|\log \varepsilon| + 4\log 9 - 3\log 4\big) + 3. \end{aligned} \tag{5.11}$$

It follows from (5.8) that $\kappa_{\varepsilon/4}+\kappa_{\varepsilon/2} \le 14\sqrt{\varepsilon^{-1}t_{\mathrm{rel}}t_{\mathrm{mix}}}$. For any irreducible Markov chain on $n>1$ states, we have that $\lambda_2 \ge -\frac{1}{n-1}$ ([2], Chapter 3, Proposition 3.18). As any lazy chain is a lazy version of some chain, it follows that for a lazy chain with at least 3 states we have that $\lambda_2 \ge (1+(-\frac{1}{2}))/2$ and so $t_{\mathrm{rel}} \ge 4/3$. Thus, by (1.2) $t_{\mathrm{rel}} \le 6(t_{\mathrm{rel}}-1)\log 2 \le 6t_{\mathrm{mix}}$. Using the fact that $|\log\varepsilon| \le \frac{2}{e\sqrt{\varepsilon}}$ for every $0<\varepsilon\le 1$ ($h(x)=2\sqrt{x}/e-\log x$ attains its minimum in $[1,\infty)$ at e^2 and $h(e^2)=0$), it follows that $7t_{\mathrm{rel}}|\log\varepsilon| \le \frac{14\sqrt{6}}{e}\sqrt{\varepsilon^{-1}t_{\mathrm{rel}}t_{\mathrm{mix}}} \le 13\sqrt{\varepsilon^{-1}t_{\mathrm{rel}}t_{\mathrm{mix}}}$. As $\sqrt{6}(4\log 9-3\log 4)<12$, $\sqrt{\varepsilon^{-1}}\ge 2$ and $4\sqrt{t_{\mathrm{rel}}t_{\mathrm{mix}}} \ge 4\sqrt{4/3}\ge 3$; we also have that $t_{\mathrm{rel}}(4\log 9-3\log 4)+3 \le 8\sqrt{\varepsilon^{-1}t_{\mathrm{rel}}t_{\mathrm{mix}}}$. Plugging these estimates in (5.11) completes the proof of the theorem. □

As promised earlier, the following proposition implies the assertion of Lemma 5.2. For any set $A\subset\Omega$, we define $\psi_{A^c}\in\mathscr{P}(A^c)$ as $\psi_{A^c}(y):=\mathrm{P}_{\pi_A}[X_1=y \mid X_1\in A^c]$. For $A\subset\Omega$, we denote $T_A^+ := \inf\{t\ge 1 : X_t\in A\}$ and $\Phi(A) := \frac{\sum_{a\in A,b\in A^c}\pi(a)P(a,b)}{\pi(A)} = \mathrm{P}_{\pi_A}[X_1\notin A]$. Note that

$$\pi(A)\Phi(A) = \sum_{a\in A,b\in A^c}\pi(a)P(a,b) = \sum_{a\in A,b\in A^c}\pi(b)P(b,a) = \pi(A^c)\Phi(A^c). \tag{5.12}$$

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 5.6. *Let (Ω,P,π) be a finite irreducible reversible Markov chain. Let $A\subsetneq\Omega$ be nonempty. Denote the complement of A by B. Then*

$$\mathrm{P}_{\pi_B}[T_A=t]/\Phi(B) = \mathrm{P}_{\psi_B}[T_A\ge t] \qquad \text{for any } t\ge 1. \tag{5.13}$$

Consequently,

$$\mathbb{E}_{\psi_B}[T_A] = \frac{1}{\Phi(B)} \quad \text{and} \quad \mathbb{E}_{\psi_B}[T_A^2] = \mathbb{E}_{\psi_B}[T_A]\bigl(2\mathbb{E}_{\pi_B}[T_A]-1\bigr) \le \frac{2\mathbb{E}_{\psi_B}[T_A]t_{\mathrm{rel}}}{\pi(A)}. \tag{5.14}$$

PROOF. We first note that the inequality $\mathbb{E}_{\psi_B}[T_A](2\mathbb{E}_{\pi_B}[T_A]-1) \le 2\mathbb{E}_{\psi_B}[T_A]t_{\mathrm{rel}}/\pi(A)$ follows from the second inequality in (3.8) (this is the only part of the proposition which relies upon reversibility).

Summing (5.13) over t yields the first equation in (5.14). Multiplying both sides of (5.13) by $2t-1$ and summing over t yields the second equation in (5.14). We now prove (5.13). Let $t\ge 1$. As $\{T_A=t\}=\{X_0\notin A,\dots,X_{t-1}\notin A, X_t\in$

$A\}$, $\{T_A^+ = t+1\} = \{X_1 \notin A, \dots, X_t \notin A, X_{t+1} \in A\}$ we have by stationarity that $\mathrm{P}_\pi[T_A = t] = \mathrm{P}_\pi[T_A^+ = t+1]$. Thus,

$$\begin{aligned}
&\pi(B)\mathrm{P}_{\pi_B}[T_A = t] \\
&\quad = \mathrm{P}_\pi[T_A = t] = \mathrm{P}_\pi\big[T_A^+ = t+1\big] \\
&\quad = \mathrm{P}_\pi[X_1 \notin A, \dots, X_t \notin A, X_{t+1} \in A] \\
&\quad = \mathrm{P}_\pi[X_1 \notin A, \dots, X_t \notin A] - \mathrm{P}_\pi[X_1 \notin A, \dots, X_t \notin A, X_{t+1} \notin A] \\
&\quad = \mathrm{P}_\pi[X_1 \notin A, \dots, X_t \notin A] - \mathrm{P}_\pi[X_0 \notin A, \dots, X_t \notin A] \\
&\quad = \mathrm{P}_\pi[X_0 \in A, X_1 \notin A, \dots, X_t \notin A] \\
&\quad = \pi(A)\Phi(A)\mathrm{P}_{\psi_B}[X_0 \notin A, \dots, X_{t-1} \notin A] \\
&\quad = \pi(A)\Phi(A)\mathrm{P}_{\psi_B}[T_A \geq t],
\end{aligned}$$

which by [5.12] implies (5.13). □

5.1. *Refining the bound for trees.* The purpose of this section is to improve the concentration estimate (5.5). 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, $\Omega_n := \{1, 2, \dots, n\}$, $P_n(i,i) = 1/2$, $P_n(i, i-1) = 1/8$ and $P_n(i, i+1) = 3/8$ for all $1 < i < n$. Then $t_{\mathrm{mix}}^{(n)} = 4n(1+o(1))$ and $t_{\mathrm{rel}}^{(n)} = \Theta(1)$.

In this case, there exists some constant $c_1 > 0$ such that for any $\lambda > 0$ we have that $\mathrm{P}_1[|T_n - 4n| \geq \lambda\sqrt{n}] \leq 2e^{-c_1\lambda^2}$. Observe that $\sqrt{t_{\mathrm{mix}}^{(n)} t_{\mathrm{rel}}^{(n)}} = \Theta(\sqrt{n})$. Hence, there exists some constant c_2 such that $\mathrm{P}_1[|T_n - 4n| \geq \lambda\sqrt{t_{\mathrm{mix}}^{(n)} t_{\mathrm{rel}}^{(n)}}] \leq 2e^{-c_2\lambda^2}$. Using Proposition 1.8, it is not hard to show that this implies that $t_{\mathrm{mix}}^{(n)}(\varepsilon) \leq t_{\mathrm{mix}}^{(n)} + c_3\sqrt{t_{\mathrm{mix}}^{(n)} t_{\mathrm{rel}}^{(n)} |\log \varepsilon|}$ and that $t_{\mathrm{mix}}^{(n)}(1-\varepsilon) \geq t_{\mathrm{mix}}^{(n)} - c_3\sqrt{t_{\mathrm{mix}}^{(n)} t_{\mathrm{rel}}^{(n)} |\log \varepsilon|}$. It is also not hard to verify that in this case (and also in many other examples of birth and death chains) this is sharp.

In Lemma 5.8, we show that for any lazy Markov chain on a tree $T = (V, E, o)$ and any $x \in V$, we have that $\mathrm{P}_x[|T_o - \mathbb{E}_x[T_o]| \geq \lambda\sqrt{\mathbb{E}_x[T_o] t_{\mathrm{rel}}}] \leq 2e^{-c_4\lambda^2}$. Besides being of independent interest, using Proposition 1.8, one can deduce from Lemma 5.8 that under the product condition,

$$\frac{t_{\mathrm{mix}}^{(n)}(\varepsilon) - t_{\mathrm{mix}}^{(n)}(1-\varepsilon)}{\sqrt{t_{\mathrm{mix}}^{(n)} t_{\mathrm{rel}}^{(n)} |\log \varepsilon|}} = O(1) \qquad \text{for any } 0 < \varepsilon \leq 1/4. \tag{5.15}$$

The details of the derivation of (5.15) from Lemma 5.8 are left to the reader.

PROPOSITION 5.7. *Let (Ω, P, π) be a finite irreducible reversible lazy Markov chain. Let $0 < \varepsilon < 1$. Let $A \subsetneq \Omega$ be such that $\pi(A) \geq 1 - \varepsilon$. Denote the complement of A by B. Denote $p := 1 - \frac{1-\varepsilon}{t_{\mathrm{rel}}}$ and $a := \mathbb{E}_{\psi_B}[T_A]$. Let*

$1 < z < \max(1/p, 2)$ *be such that* $q := p(z-1)/(1-p) \leq 1/2$. *Then*

$$\mathbb{E}_{\psi_B}[z^{T_A - a}] \leq \exp\left[\frac{2a(z-1)^2}{(1-p)}\right] \quad \text{and} \quad \mathbb{E}_{\psi_B}[z^{a - T_A}] \leq \exp\left[\frac{a(z-1)^2}{(1-p)}\right]. \tag{5.16}$$

PROOF. Let $q := \frac{p(z-1)}{1-p}$ and $0 \leq \gamma \leq p$. Then by our assumption that $pz < 1$, we have that

$$\begin{aligned}\sum_{k \geq 1}(1-\gamma)(\gamma z)^{k-1} &= (1-\gamma)/(1-\gamma z) \leq (1-p)/(1-pz) \\ &= 1 + \frac{q}{1-q} \leq 1 + 2q\end{aligned}$$

[alternatively, this follows by noting that (for a fixed z) z^k is a monotone increasing function and that the Geometric distribution with parameter $(1-p)$ stochastically dominates the Geometric distribution with parameter $(1-\gamma)$]. By Lemma 3.8 [parts (i)–(ii)] and laziness, there exist $\sum_{i=1}^{\ell} a_i = 1$ and $0 \leq \gamma_i \leq p$ such that

$$\begin{aligned}\sum_{k \geq 1} z^{k-1} \mathrm{P}_{\pi_B}[T_A = k] &= \sum_{i=1}^{\ell} a_i \sum_{k \geq 1}(1-\gamma_i)(\gamma_i z)^{k-1} \\ &\leq \sum_{j=1}^{\ell} a_i(1+2q) = 1 + 2q.\end{aligned}$$

Hence, by (5.13)–(5.14) (third equality) we have that

$$\begin{aligned}\mathbb{E}_{\psi_B}[z^{T_A}] &= \sum_{k \geq 1} z^k \mathrm{P}_{\psi_B}[T_A = k] \\ &= 1 + (z-1)\sum_{k \geq 1} z^{k-1} \mathrm{P}_{\psi_B}[T_A \geq k] \\ &= 1 + (z-1)a \sum_{k \geq 1} z^{k-1} \mathrm{P}_{\pi_B}[T_A = k] \\ &\leq 1 + (z-1)a(1+2q) \leq \exp[a(z-1)(1+2q)].\end{aligned} \tag{5.17}$$

As for any $1 \leq x \leq 2$, we have that $0 \leq \log x - (x-1) + (x-1)^2/2$, we also have that

$$z^{-a} \leq \exp[-a(z-1) + a(z-1)^2/2]. \tag{5.18}$$

Thus, $\mathbb{E}_{\psi_B}[z^{T_A - a}] \leq \exp[a(z-1)^2(1/2 + 2p/(1-p))] \leq \exp[\frac{2a(z-1)^2}{(1-p)}]$, as desired. We now turn to the task of bounding $\mathbb{E}_{\psi_B}[z^{-T_A}]$. Let $y := \frac{p(1-z^{-1})}{1-p}$. In the above notation, much as before, we have

$$\sum_{k\geq 1} z^{-(k-1)}\mathrm{P}_{\pi_B}[T_A=k] \geq \sum_{k\geq 1}(1-p)(p/z)^k$$
$$= (1-p)/(1-p/z) = \frac{1}{1+y} \geq 1-y,$$

$$\begin{aligned}\mathbb{E}_{\psi_B}[z^{-T_A}] &= \sum_{k\geq 1} z^{-k}\mathrm{P}_{\psi_B}[T_A=k] \\ &= 1-(1-z^{-1})\sum_{k\geq 1} z^{-(k-1)}\mathrm{P}_{\psi_B}[T_A\geq k] \\ &= 1-(1-z^{-1})a\sum_{k\geq 1} z^{-(k-1)}\mathrm{P}_{\pi_B}[T_A=k] \\ &\leq 1-(1-z^{-1})a(1-y) \\ &\leq \exp[-a(1-z^{-1})(1-y)].\end{aligned} \tag{5.19}$$

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

LEMMA 5.8. *Let (V,P,π) be a lazy Markov chain on a tree (T,o). Let $x,y\in V$ be such that $y\prec x$. Denote $t_{x,y} := \mathbb{E}_x[T_y]$ and $b=b_{x,y} := \sqrt{t_{x,y}t_{\mathrm{rel}}}$. Then*

$$\mathrm{P}_x[T_y - t_{x,y} \geq cb] \vee \mathrm{P}_x[t_{x,y}-T_y\geq cb] \leq e^{-3c^2/64} \quad \text{for any } 0\leq c\leq 2\sqrt{t_{x,y}/t_{\mathrm{rel}}}. \tag{5.20}$$

PROOF. Let $(v_0=x,v_1,\dots,v_k=y)$ be the path from x to y. 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 $\tau_1,\dots,\tau_k$ are independent. Denote $p := 1-\frac{1}{2t_{\mathrm{rel}}}$. Denote $a_i := \mathbb{E}_x[\tau_i]$. Fix some $0\leq c\leq 2\sqrt{t_{x,y}/t_{\mathrm{rel}}}$. Set $z_c=z_{c,x} := 1+\frac{c}{8b}$. Note that $2p(z_c-1)\leq \frac{c}{4b}\leq \frac{1}{2t_{\mathrm{rel}}} = 1-p$ (i.e., $q<1/2$) and $z_c<\max(1/p,2)$. Then by (5.16)

$$\begin{aligned}&\mathrm{P}_x[T_y-t_{x,y}\geq cb] \\ &\quad = \mathrm{P}_x[z_c^{T_y-t_{x,y}} \geq z_c^{cb}] \\ &\quad \leq \mathbb{E}_x[z_c^{T_y-t_{x,y}}]z_c^{-cb} = z_c^{-cb}\prod_{i=1}^k \mathbb{E}_x[z_c^{\tau_i-a_i}] \\ &\quad \leq \exp[(-(z_c-1)+(z_c-1)^2/2)cb]\prod_{i=1}^k \exp\left[\frac{2a_i(z_c-1)^2}{1-p}\right]\end{aligned} \tag{5.21}$$

$$= \exp\left[-\frac{c^2}{8} + \frac{c^3}{128b}\right]\exp\left[\frac{4t_{\mathrm{rel}}t_{x,y}c^2}{64b^2}\right]$$
$$\le \exp\left[-\frac{c^2}{8} + \frac{c^3}{128b} + \frac{c^2}{16}\right] \le e^{-3c^2/64}.$$

The inequality $\mathrm{P}_x[t_{x,y} - T_y \ge cb] \le e^{-3c^2/64}$ is proved in an analogous manner. □

6. 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 (δ, r)-SBD chains. Although we think of δ 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 δ does not tend to 0, nor does r to infinity, too rapidly in terms of some functions of $t_{\mathrm{rel}}/t_{\mathrm{mix}}$.

Throughout the section, we use $C_1, C_2, \dots$ to describe positive constants which depend only on δ and r. Consider a (δ, r)-SBD chain on $([n], P, \pi)$. We call a state $i \in [n]$ a central-vertex if $\pi([i-1]) \vee \pi([n] \setminus [i]) \le 1/2$. 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.

Divide $[n]$ into $m := \lceil n/r \rceil$ consecutive disjoint intervals, $I_1, \dots, I_m$ each of size r, apart from perhaps I_m. We call each such interval a *block*. Denote by $I_{\tilde{o}}$ the unique block such that the root o belongs to it. Since we are assuming the product condition (and thus $t_{\mathrm{mix}}^{(n)} \to \infty$), in the setup of Theorem 2 we can assume that $n \gg r$, and hence $I_{\tilde{o}} \neq [n]$ (it is not hard to show that $t_{\mathrm{mix}}^{(n)}$ can be bounded from above in terms of n and δ, and thus we must have $n \to \infty$). Observe the following. Consider some $v \notin I_{\tilde{o}}$ and $u \in I_{\tilde{o}}$ such that $|u - v| = 1$. Then by the definition of a (δ, r)-SBD chain, we have for all $v' \in I_{\tilde{o}}$, $\pi(v) \ge \delta^r \pi(v')$. Hence, $\pi(I_{\tilde{o}}) \le \frac{r}{r+\delta^r}$. For the rest of this section, let us fix $\alpha = \alpha(\delta, r) := 1 - \frac{\delta^r}{4(r+\delta^r)}$.

Recall that in Section 5 we exploited the tree structure to reduce the problem of showing cutoff to showing the concentration of the hitting time of the central vertex by showing that starting from the central vertex the chain hits any large set (with large probability) quickly. We argue similarly in this case with the central vertex replaced by the central block. First, we need the following lemma.

LEMMA 6.1. *In the above setup, let* $I := \{v, v+1, \dots, v+r-1\} \subset [n]$. *Let* $\mu \in \mathscr{P}(I)$. *Then*

$$\mathbb{E}_\mu[T_A] \le \max_{y \in I} \mathbb{E}_y[T_A] \le \delta^{-r} \min_{x \in I} \mathbb{E}_x[T_A] \qquad \text{for any } A \subset [n] \setminus I. \tag{6.1}$$

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

$$\mathbb{E}_i[T_A] \le \delta^{-r}\mathbb{E}_{\pi_{[n]\setminus[v-1]}}[T_A], \qquad \big(\textit{resp. } \mathbb{E}_i[T_A] \le \delta^{-r}\mathbb{E}_{\pi_{[v+r-1]}}[T_A]\big). \tag{6.2}$$

PROOF. We first note that (6.2) follows from (6.1). Indeed, by condition (i) of the definition of a (δ, r)-SBD chain, if $A \subset [v-1]$ (resp. $A \subset [n] \setminus [v+r-1]$), then under $\mathrm{P}_{\pi_{[n]\setminus[v-1]}}$ (resp., under $\mathrm{P}_{\pi_{[v+r-1]}}$), $T_I \le T_A$. Thus, (6.2) follows from (6.1) by averaging over X_{T_I}. We now prove (6.1).

Fix some A such that $A \subset [n] \setminus I$. Fix some distinct $x, y \in I$. Let B_1 be the event that $T_y \le T_A$. 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 $1 \le k \le |y-x|$. Denote the last event by B_2. Then

$$\mathbb{E}_x[T_A] \ge \mathbb{E}_x[T_A 1_{B_2}] \ge \mathrm{P}_x[B_2]\mathbb{E}_y[T_A] \ge \delta^r \mathbb{E}_y[T_A].$$

Minimizing over x yields that for any $y \in I$ we have that $\mathbb{E}_y[T_A] \le \delta^{-r} \min_{x \in I} \mathbb{E}_x[T_A]$, from which (6.1) follows easily. □

The next proposition reduces the question of proving cutoff for a sequence of (δ, r)-SBD chains under the product condition to that of showing an appropriate concentration for the hitting time of the central block. The argument is analogous to the one in Section 5, and hence we only provide a sketch to avoid repetitions. As in Section 5, for $\varepsilon \in (0,1)$ let $\tau_C(\varepsilon) := \min\{t : \mathrm{P}_x[T_{I_{\tilde{o}}} > t] \le \varepsilon, \forall x \in [n]\}$. As always, we write $\tau_C^{(k)}(\cdot)$ to indicate that this parameter is taken w.r.t. the kth chain in a sequence of (δ, r)-SBD chains.

PROPOSITION 6.2. *Let* $([n_k], P_k, \pi_k)$ *be a sequence of* (δ, r)*-SBD chains. Suppose that there exist constants* C_ε *for* $\varepsilon \in (0, \frac{1}{8})$ *and a some sequence* $(w_k)_{k=1}^{\infty}$ *of numbers such that for all* k

$$\tau_C^{(k)}(\varepsilon) - \tau_C^{(k)}(1-\varepsilon) \le C_\varepsilon w_k \qquad \textit{for all } 0 < \varepsilon < 1/8. \tag{6.3}$$

Then there exist some constants $C'_\varepsilon, C''_\varepsilon$ *such that for all* k *and all* $\varepsilon \in (0, 1/8)$

$$\mathrm{hit}_{1/2}^{(k)}(3\varepsilon/2) - \mathrm{hit}_{1/2}^{(k)}(1-3\varepsilon/2) \le C_\varepsilon w_k + C'_\varepsilon t_{\mathrm{rel}}^{(k)} \quad \textit{and} \tag{6.4}$$

$$t_{\mathrm{mix}}^{(k)}(2\varepsilon) - t_{\mathrm{mix}}^{(k)}(1-2\varepsilon) \le C_\varepsilon w_k + C''_\varepsilon t_{\mathrm{rel}}^{(k)}. \tag{6.5}$$

PROOF. Observe that (6.5) follows from (6.4) using Proposition 1.8 and Corollary 3.4. To deduce (6.4) from (6.3), we argue as in Lemma 5.1 using Lemma 6.3 below, which shows that starting from any vertex in $I_{\tilde{o}}$ the chain hits any set of π-measure at least α in time proportional to t_{rel} with large probability. We omit the details. □

LEMMA 6.3. *Let* $v \in I_{\tilde{o}}$. *Let* $D \subset [n]$ *be such that* $\pi(D) \geq \frac{1+\alpha}{2}$. *Then* $\mathbb{E}_v[T_D] \leq C(\alpha)\delta^{-r} t_{\mathrm{rel}}$ *for some constant* $C(\alpha)$. *In particular, by Markov inequality* $\mathrm{hit}_{\alpha,v}(\varepsilon) \leq \varepsilon^{-1} C(\alpha)\delta^{-r} t_{\mathrm{rel}}$.

PROOF. Let $I_{\tilde{o}} = \{v_1, v_1+1, \dots, v_2\}$. Set $A_1 = [v_1 - 1]$ and $A_2 = [n] \setminus [v_2]$. For $i = 1, 2$, let $D_i = D \cap A_i$. Using the definition of α, without loss of generality let $\pi(D_1) \geq \frac{1-\alpha}{2}$. Set $A = A_2 \cup I_{\tilde{o}}$. By (6.2) and the fact that $\pi(A) \geq \frac{1}{2}$

$$\mathbb{E}_v[T_D] \leq \mathbb{E}_v[T_{D_1}] \leq \delta^{-r}\mathbb{E}_{\pi_A}[T_{D_1}] \leq 2\delta^{-r}\mathbb{E}_{\pi}[T_{D_1}].$$

The proof is completed using Lemma 3.5. □

Observe that, arguing as in Corollary 5.5, it follows using Chebyshev inequality that (6.3) holds for some constants C_ε if we take $w_n = \max_{x \in [n]} \sqrt{\mathbf{Var}_x[T_{I_{\tilde{o}}}]}$. Theorem 2 therefore follows at once from Proposition 6.2 provided we establish $\mathbf{Var}_x[T_{I_{\tilde{o}}}] \leq C_1 \mathbb{E}_x[T_{I_{\tilde{o}}}] t_{\mathrm{rel}}$ for all $x \notin I_{\tilde{o}}$ [since as in (5.8) (first inequality) $\mathbb{E}_x[T_{I_{\tilde{o}}}] = O(t_{\mathrm{mix}})$, alternatively, this follows by (1.7)]. This is what we shall do.

Observe that the root induces a partial order on the blocks. We say that $I_j \prec I_k$ if I_j is a block between I_k and $I_{\tilde{o}}$. For $j \in [m]$, $I_j \neq I_{\tilde{o}}$, we define the parent block of I_j in the obvious manner and denote its index by f_j. We define

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

Consider some arbitrary $x \in [n]$ and $j \in [m] \setminus \{\tilde{o}\}$ such that $x \in I_j$. Denote $k := |j - \tilde{o}|$, $j_0 = j$ and $j_{i+1} = f_j$ for all $0 < i < k$. Observe that starting from x we have that $T_{I_{\tilde{o}}} = \sum_{\ell=0}^{k-1} \bar{\tau}_{j_\ell}$. As mentioned above, we will bound $\mathbf{Var}_x[\sum_{\ell=0}^{k-1} \bar{\tau}_{j_\ell}]$. 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 6.5) and that for all ℓ we have that $\mathbf{Var}_x[\bar{\tau}_{j_\ell}] \leq C_2 t_{\mathrm{rel}} \mathbb{E}_x[\bar{\tau}_{j_\ell}]$ (Lemma 6.6). The desired variance estimate, $\mathbf{Var}_x[\sum_{\ell=0}^{k-1} \bar{\tau}_{j_\ell}] \leq C_1 \mathbb{E}_x[T_{I_{\tilde{o}}}] t_{\mathrm{rel}}$, follows by combining these two lemmata. We omit the details.

LEMMA 6.4. *In the above setup, let* $v \in [m] \setminus \{\tilde{o}\}$. *Let* $(v_0 = v, v_1, \dots, v_s)$ *be indices of consecutive blocks. Let* $\mu_1, \mu_2 \in \mathscr{P}(I_v)$. *Let* $k \in [s]$. *Denote by* $\nu_k^{(j)}$ ($j = 1, 2$) *the hitting distribution of* I_{v_k} *starting from initial distribution* μ_j (*i.e.*, $\nu_k^{(j)}(z) := \mathrm{P}_{\mu_j}[X_{T(v_k)} = z]$). *Then* $\|\nu_k^{(1)} - \nu_k^{(2)}\|_{\mathrm{TV}} \leq (1 - \delta^r)^k$.

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 δ^r there exists some $z_v \in I_v$ such that both chains hit z_v before hitting I_{f_v} (not necessarily at the same time) and from that moment on (which may occur at different times for the two chains) they follow the same trajectory. The fact that the hitting time of z_v (and thus also of I_{f_v}) might be different for the two chains makes no difference [as regardless of the hitting time of I_{f_v} w.r.t. the two chains, this coupling is also a coupling of $(\nu_1^{(1)}, \nu_1^{(2)})$, having the desired property]. We now describe this coupling more precisely.

Let $\mu_1, \mu_2 \in \mathscr{P}(I_v)$. Let $(X_t^{(1)})_{t\ge 0}$ and $(X_t^{(2)})_{t\ge 0}$ be independent Markov chains where $(X_t^{(i)})_{t\ge 0}$ is distributed as the chain (Ω, P, π) with initial distribution μ_i $(i = 1, 2)$ as follows. Pick $v_1 \sim \mu_1$ and $v_2 \sim \mu_2$, respectively. Run the chain $X_t^{(1)}$ started from v_1. Let $R := \min\{t : X_t^{(1)} = X_0^{(2)}\}$ and $L_i := \min\{t : X_t^{(i)} \in I_{f_v}\}$. Let S denote the event: $R \le L_1$. On S, define $Y_t^{(1)}$ by setting $Y_t^{(1)} = X_t^{(1)}$ for $t < R$ and $Y_{R+t}^{(1)} = X_t^{(2)}$ for any $t \ge 0$, and on S^c, define $Y_t^{(1)} = X_t^{(1)}$ for all t. Denote the joint law of $(Y_t^{(1)}, X_t^{(2)})$ by P_{μ_1,μ_2} and of $(X_t^{(1)}, Y_t^{(1)}, X_t^{(2)})$ by $\mathrm{P}_{\mu_1,\mu_1,\mu_2}$. Clearly, P_{μ_1,μ_2} is a coupling with the correct marginals and $\mathrm{P}_{\mu_1,\mu_1,\mu_2}[S] \ge \delta^r$. Let L_2 be as above and $\bar{L}_1 := \min\{t : Y_t^{(1)} \in I_{f_v}\}$. Note that on S, $X_{L_2}^{(2)} = Y_{\bar{L}_1}^{(1)}$. Hence, for any $D \subset I_{v_k}$,

$$\begin{aligned}\nu_1^{(1)}(D) - \nu_1^{(2)}(D) &= \mathrm{P}_{\mu_1,\mu_2}\big[Y_{\bar{L}_1}^{(1)} \in D\big] - \mathrm{P}_{\mu_1,\mu_2}\big[X_{L_2}^{(2)} \in D\big] \\ &\le \mathrm{P}_{\mu_1,\mu_2}\big[Y_{\bar{L}_1}^{(1)} \in D, X_{L_2}^{(2)} \notin D\big] \\ &\le 1 - \mathrm{P}_{\mu_1,\mu_1,\mu_2}[S] \le 1 - \delta^r. \end{aligned}$$ □

LEMMA 6.5. *In the setup of Lemma 6.4, let $0 \le i < j < s$. Let $\mu \in \mathscr{P}(I_v)$. Write $\tau_i := \bar{\tau}_{v_i}$ and $\tau_j := \bar{\tau}_{v_j}$. Then*

$$\mathbb{E}_\mu[\tau_i\tau_j] \le \mathbb{E}_\mu[\tau_i]\mathbb{E}_\mu[\tau_j]\big(1 + (1-\delta^r)^{j-i-1}\delta^{-r}\big).$$

PROOF. Let μ_{i+1} and μ_j be the hitting distributions of $I_{v_{i+1}}$ and of I_{v_j}, respectively, of the chain with initial distribution μ. By the Markov property, the hitting distribution of I_{v_j} for the chain started with initial distribution either μ or μ_{i+1} is simply μ_j. Again by the Markov property $\mathbb{E}_\mu[\tau_j] = \mathbb{E}_{\mu_{i+1}}[\tau_j] = \mathbb{E}_{\mu_j}[\tau_j]$ and

$$\mathbb{E}_\mu[\tau_i\tau_j] \le \mathbb{E}_\mu[\tau_i] \max_{y\in I_{v_{i+1}}} \mathbb{E}_y[\tau_j]. \tag{6.6}$$

Let $y^* \in I_{v_{i+1}}$ be the state achieving the maximum in the RHS above. By Lemma 6.4, we can couple successfully the hitting distribution of I_{v_j} (and thus also τ_j) of the chain started from y^* with that of the chain starting from initial distribution μ_{i+1} with probability at least $1 - (1-\delta^r)^{j-i-1}$. If the coupling fails, then by (6.1) we can upper bound the conditional expectation of τ_j by $\delta^{-r}\,\mathbb{E}_\mu[\tau_j]$. Hence,

$$\mathbb{E}_{y^*}[\tau_j] \le \mathbb{E}_{\mu_j}[\tau_j] + (1-\delta)^{j-i-1}\delta^{-r}\mathbb{E}_\mu[\tau_j] = \mathbb{E}_\mu[\tau_j]\big(1 + (1-\delta^r)^{j-i-1}\delta^{-r}\big).$$

The assertion of the lemma follows by plugging this estimate in (6.6). □

LEMMA 6.6. *Let* $j \in [m] \setminus \{o\}$. *Let* $\nu \in \mathscr{P}([n])$. *Then there exists some* $C_1, C_2 > 0$ (*depending on* δ *and* r) *such that* $\mathbb{E}_\nu[\bar{\tau}_j^2] \le C_1 t_{\mathrm{rel}} \Phi(I_j)^{-1} \le C_2 t_{\mathrm{rel}} \mathbb{E}_\nu[\bar{\tau}_j]$.

PROOF. Let $\mu := \psi_{I_j}$. By condition (i) in the definition of a (δ, r)-SBD chain, $\mu \in \mathscr{P}(I_j)$. By (5.14), $\mathbb{E}_\mu[\bar{\tau}_j^2] \le C_3 t_{\mathrm{rel}} \Phi(I_j)^{-1} \le C_4 t_{\mathrm{rel}} \mathbb{E}_\mu[\bar{\tau}_j]$ for constants C_3, C_4 depending on δ and r. The proof is complete using the same reasoning as in the proof of (6.1) to argue that the first and second moments of $\bar{\tau}_j$ w.r.t. different initial distributions can change by at most some multiplicative constant. $\square$

7. Examples.

7.1. *Aldous' example*. We now present a small variation of Aldous' example of a sequence of chains which satisfies the product condition but does not exhibit cutoff. The reason we present this example is that it demonstrates that Theorem 2 may fail if condition (ii) in the definition of a (δ, r)-semi birth and death chain is not satisfied. Loosely speaking, the main point in the construction is that the set of stationary measure at least $1/2$ which is hardest to hit (by time t for all t) can be taken to be a certain singleton, $\{2n+1\}$, and that there is a state, $-10n$, satisfying $\lim_{n\to\infty} \sup_t |\mathrm{P}_{-10n}[T_{2n+1} > t] - p^{(n)}(1/2, t)| = 0$ such that T_{2n+1} is not concentrated under P_{-10n}. In particular, there is no $\mathrm{hit}_{1/2}$-cutoff. Because this example is classic and was analyzed in details in [5, 6], we shall only give a sketch of the proof of the above claims.

EXAMPLE 7.1. Consider the chain (Ω_n, P_n, π_n), where $\Omega_n := \{-10n, -10n+2, \dots, -2, 0\} \cup [2n+1]$. Think of Ω_n 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 $P_n(x,x) = 1/2$ if x is even, $P_n(x,x) = 3/4$ if x is odd and $x < 2n+1$ and $P_n(2n+1, 2n+1) = 9/10$.

Conditionally on making a nonlazy step the walk moves with a fixed bias toward $2n+1$ (apart from at the states $-10n, 0, 2n+1$):

$$P_n(2i, \min\{2i+2, 2n+1\}) = 2P_n(2i, 2i-2) = 2P_n(2i-1, 2i+1)$$
$$= 4P_n(2i-1, \max\{2i-3, 0\}) = \tfrac{1}{3}.$$

Finally, we set $P_n(-10n, -10n+2) = 1/2$, $P_n(0,2) = P_n(0,1) = 2P_n(0,-2) = \frac{1}{5}$ 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., [18], Theorem 13.14), $t_{\mathrm{rel}}^{(n)} = O(1)$, as the bottleneck-ratio is bounded from below. In particular, the product condition holds. As $\pi_n(2n+1) > 1/2$, there is $\mathrm{hit}_{1/2}$-cutoff [and hence by Proposition 3.6 there is hit_α-cutoff for all $\alpha \in (0,1)$] iff starting from $-10n$, the hitting-time of $2n+1$ is

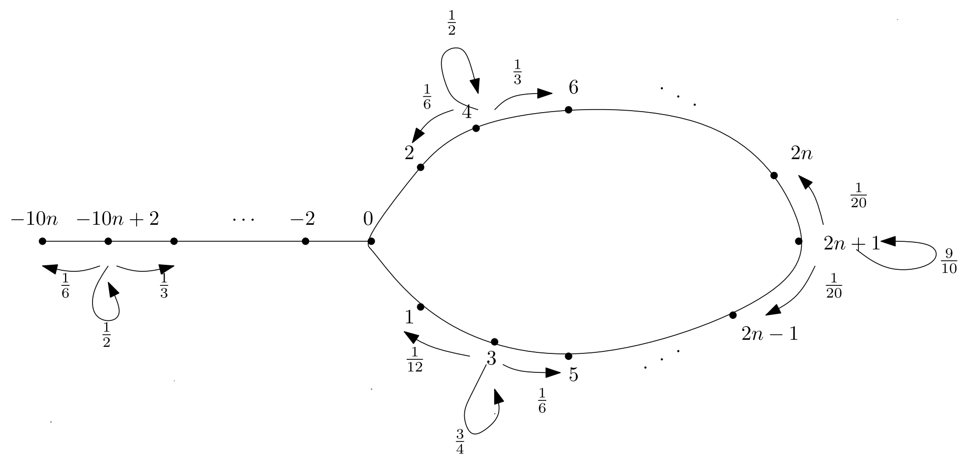

FIG. 1. *We consider a Markov chain on the above graph with the following transition probabilities*: $P_n(x,x) = 1/2$ *for* x *even and* $P_n(x,x) = 3/4$ *for* $x < 2n+1$ *and odd*. $P_n(0,2) = P_n(0,1) = \frac{1}{5}, P_n(0,-2) = \frac{1}{10}$, $P_n(-10n,-10n+2) = 1/2$, $P_n(2n+1,2n) = P_n(2n+1,2n-1) = \frac{1}{20}$, *and hence* $P_n(2n+1,2n+1) = \frac{9}{10}$. *All other transition probabilities are given by*: $P_n(2i, \min\{2i+2, 2n+1\}) = \frac{1}{3}$, $P_n(2i, 2i-2) = P_n(2i-1, 2i+1) = \frac{1}{6}$, $P_n(2i-1, \max\{2i-3, 0\}) = \frac{1}{12}$.

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 $Y=2$ (resp., $Y=1$), then the chain had to reach $2n+1$ through the path $(2,4,\dots,2n)$ $[(1,3,\dots,2n-1)$, resp.]. Denote, $Z_i := T_{2n+1}1_{Y=i}$, $i=1,2$. Then on $Y=i$, $T_{2n+1} = Z_i$, 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.

7.2. *Sharpness of Theorem* 3. Now we give an example to show that in Proposition 3.7 (and hence in Theorem 3) the value $\frac{1}{2}$ cannot be replaced by any larger value.

EXAMPLE 7.2. Let (Ω_n, P_n, π_n) be the nearest-neighbor weighted random walk from Figure 3. Then $t_{\mathrm{rel}}^{(n)} = \Theta(t_{\mathrm{mix}}^{(n)})$, yet for every $1/2 < \alpha < 1$, the sequence exhibits hit$_\alpha$-cutoff.

PROOF. Let $\Phi_n := \min_{A \subset \Omega_n : 0 < \pi_n(A) \le 1/2} \Phi_n(A)$ be the Cheeger constant of the nth chain, where $\Phi_n(A) := \frac{\sum_{a \in A, b \in A^c} \pi_n(a) P_n(a,b)}{\pi_n(A)}$. Then by taking A to be either

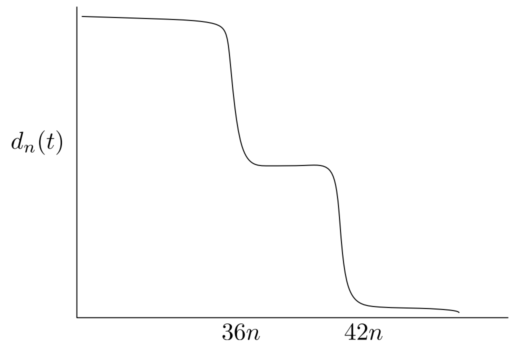

FIG. 2. *Decay in total variation distance for Aldous' example: it does not have cutoff.*

A_1 or A_2, by Cheeger inequality (e.g., [18], Theorem 13.14), we have that $t_{\mathrm{rel}}^{(n)} \geq \frac{1}{2\Phi_n} \geq c_1 n^2 \geq c_2 t_{\mathrm{mix}}^{(n)}$ [it is easy to show that by (1.7) and the fact that $\pi_n(A_i) = 1/2 - o(1)$ for $i = 1, 2$ we have that $t_{\mathrm{mix}}^{(n)} \leq Cn^2$]. By (1.2), indeed $t_{\mathrm{rel}}^{(n)} = \Theta(t_{\mathrm{mix}}^{(n)})$ and it follows from Fact 1.1 that there is no cutoff.

Fix some $1/2 < \alpha < 1$. Let $B \subset \Omega_n$ be such that $\pi_n(B) \geq \alpha$. Denote the set of vertices belonging to the path, but not to A_1 by D. Then $\pi_n(D) = O(n^{-2}) = o(1)$. Consequently, $\pi_n(A_i \cap B) \geq \alpha - 1/2 - o(1)$, for $i = 1, 2$. Using this observation,

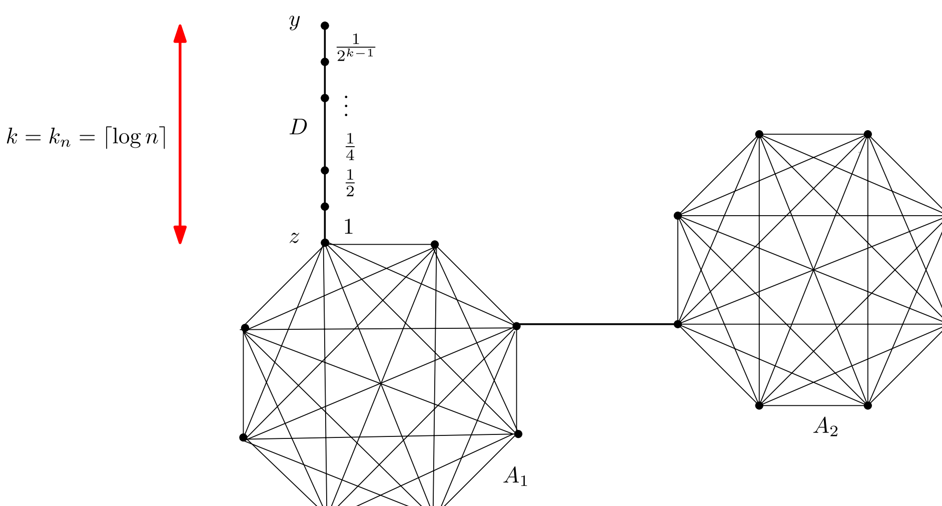

FIG. 3. *We consider a lazy weighted nearest-neighbor random walk on the above graph consisting of two disjoint cliques A_1 and A_2 of size n connected by a single edge and a path of length $k_n = \lceil \log n \rceil$ connected to A_1. The edge weights of all edges incident to vertices in $A_1 \cup A_2$ is 1, while those belonging to the path are indicated in the figure. Inside the path, the walk has a fixed bias towards the clique.*

it is easy to verify that for all $x \in A_1 \cup A_2$ we have that

$$\text{hit}_{\alpha,x}(\varepsilon) \le c_\alpha \log(1/\varepsilon) \qquad \text{for any } 0 < \varepsilon < 1, \tag{7.1}$$

for some constant c_α independent of n.

Let y be the endpoint of the path which does not lie in A_1. Let z be the other endpoint of the path. The hitting time of z under P_y is concentrated around time $6 \log n$. Then by (7.1), together with the Markov property (using the same reasoning as in the proof of Lemma 5.1) for all sufficiently large n we have that for any $0 < \varepsilon \le 1/4$

$$\begin{aligned} \text{hit}_{\alpha,y}^{(n)}(2\varepsilon) &\le \big(6 + o(1)\big) \log n + \text{hit}_{\alpha,z}^{(n)}(\varepsilon) = \big(6 + o(1)\big) \log n, \\ \text{hit}_{\alpha,y}^{(n)}(1-\varepsilon) &\ge \big(6 - o(1)\big) \log n. \end{aligned} \tag{7.2}$$

Similarly to the proof of Lemma 5.1, for any $B \subset \Omega_n$ and any $x \in D$, we have that $\mathrm{P}_y[T_{B \setminus D} > t] \ge \mathrm{P}_x[T_B > t]$, for all t. Since $\pi_n(D) = o(1)$, this implies that for all sufficiently large n, for any $1/2 < \alpha < 1$, there exists some $1/2 < \alpha' < \alpha$ (α' depends on α but not on n), such that for any $x \in D$ we have that $\text{hit}_{\alpha,y}^{(n)}(\varepsilon) \ge \text{hit}_{\alpha',x}^{(n)}(\varepsilon)$, for all $0 < \varepsilon < 1$. This, together with (7.1) and the fact that the leftmost terms in both lines of (7.2) are up to negligible terms independent of α and ε, implies that the sequence of chains exhibits hit_α-cutoff for all $1/2 < \alpha < 1$. □

REMARK 7.3. One can modify the sequence from Example 7.2 into a sequence of lazy simple nearest-neighbor random walks on a graph. Construct the nth graph in the sequence as follows. Start with a binary tree T of depth n. Denote its root by y, the set of its leaves by A_1 and $D := T \setminus A_1$. Turn A_1 into a clique by connecting every two leaves of T by an edge. Take another disjoint complete graph of size $|A_1| = 2^n$ and denote its vertices by A_2. Finally, connect A_1 and A_2 by a single edge. Since the number of edges which are incident to D is at most 2^{n+2}, while the total number of edges of the graph is greater than 2^{2n}, we have that $\pi_n(D) = o(1)$. The analysis above can be extended to this example with minor adaptations (although a rigorous analysis of this example is somewhat more tedious).

Acknowledgments. We are grateful to David Aldous, Allan Sly, Perla Sousi and Prasad Tetali for many helpful suggestions. We also thank anonymous referees for many useful comments and suggestions that helped improve the paper. This work was completed while the first author was a graduate student at the Department of Statistics, University of California, Berkeley, supported by UC Berkeley Graduate Fellowship.

REFERENCES

[1] ALDOUS, D. and DIACONIS, P. (1986). Shuffling cards and stopping times. *Amer. Math. Monthly* **93** 333–348. MR0841111

[2] ALDOUS, D. and FILL, J. A. (2000). *Reversible Markov Chains And Random Walks On Graphs*. Book in preparation. Available at http://www.stat.berkeley.edu/users/aldous.

[3] ALDOUS, D. J. (1982). Some inequalities for reversible Markov chains. *J. Lond. Math. Soc.* (2) **2** 564–576. MR0657512

[4] BROWN, M. (1999). Interlacing eigenvalues in time reversible Markov chains. *Math. Oper. Res.* **24** 847–864. MR1854356

[5] CHEN, G.-Y. (2006). The cutoff phenomenon for finite Markov chains Ph.D. thesis, Cornell University.

[6] CHEN, G.-Y. and SALOFF-COSTE, L. (2008). The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.* **13** 26–78. MR2375599

[7] CHEN, G.-Y. and SALOFF-COSTE, L. (2013). Comparison of cutoffs between lazy walks and Markovian semigroups. *J. Appl. Probab.* **50** 943–959. MR3161366

[8] CHEN, G.-Y. and SALOFF-COSTE, L. (2015). Computing cutoff times of birth and death chains. *Electron. J. Probab.* **20** 1–47. MR3371435

[9] DIACONIS, P. (1996). The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* **93** 1659–1664. MR1374011

[10] DIACONIS, P. and SALOFF-COSTE, L. (2006). Separation cut-offs for birth and death chains. *Ann. Appl. Probab.* **16** 2098–2122. MR2288715

[11] DING, J., LUBETZKY, E. and PERES, Y. (2010). Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields* **146** 61–85. MR2550359

[12] DURRETT, R. (2010). *Probability: Theory and Examples*, 4th ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge Univ. Press, Cambridge. MR2722836

[13] FILL, J. A. and LYZINSKI, V. (2012). Hitting times and interlacing eigenvalues: A stochastic approach using intertwinings. *J. Theoret. Probab.* **27** 954–981.

[14] GRIFFITHS, S., KANG, R. J., OLIVEIRA, R. I. and PATEL, V. (2012). Tight inequalities among set hitting times in Markov chains. Preprint. Available at arXiv:1209.0039.

[15] HERMON, J. (2015). A technical report on hitting times, mixing and cutoff. Preprint. Available at arXiv:1501.01869.

[16] HERMON, J. and PERES, Y. (2015). The power of averaging at two consecutive time steps: Proof of a mixing conjecture by Aldous and Fill. Preprint. Available at arXiv:1508.04836.

[17] LANCIA, C., NARDI, F. R. and SCOPPOLA, B. (2012). Entropy-driven cutoff phenomena. *J. Stat. Phys.* **149** 108–141. MR2981642

[18] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. With a chapter by James G. Propp and David B. Wilson. MR2466937

[19] LOVÁSZ, L. and WINKLER, P. (1998). Mixing times. In *Microsurveys in Discrete Probability* (*Princeton*, *NJ*, 1997). *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* **41** 85–134. Amer. Math. Soc., Providence, RI. MR1630411

[20] NORRIS, J., PERES, Y. and ZHAI, A. (2015). Surprise probabilities in Markov chains. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms* 1759–1773. SIAM, Philadelphia, PA.

[21] OLIVEIRA, R. I. (2012). Mixing and hitting times for finite Markov chains. *Electron. J. Probab.* **17** 1–12. MR2968677

[22] PERES, Y. American Institute of Mathematics (AIM) research workshop “Sharp Thresholds for Mixing Times” (Palo Alto, December 2004). Summary available at http://www.Aimath.Org/WWN/mixingtimes.

[23] PERES, Y. and SOUSI, P. (2013). Total variation cutoff in a tree. Preprint. Available at arXiv:1307.2887.

[24] PERES, Y. and SOUSI, P. (2015). Mixing times are hitting times of large sets. *J. Theoret. Probab.* **28** 488–519. MR3370663

[25] STARR, N. (1966). Operator limit theorems. *Trans. Amer. Math. Soc.* **121** 90–115. MR0190757

[26] STEIN, E. M. (1961). On the maximal ergodic theorem. *Proc. Natl. Acad. Sci. USA* **47** 1894–1897. MR0131517

R. BASU
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
450 SERRA MALL, BUILDING 380
STANFORD, CALIFORNIA 94305
USA
E-MAIL: rbasu@stanford.edu

J. HERMON
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY
365 EVANS HALL
BERKELEY, CALIFORNIA 94720
USA
E-MAIL: jonathan.hermon@stat.berkeley.edu

Y. PERES
MICROSOFT RESEARCH
ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052
USA
E-MAIL: peres@microsoft.com