

# ANNALES DE LA FACULTÉ DES SCIENCES DE TOULOUSE Mathématiques

YUVAL PERES AND PERLA SOUSI

*Total variation cutoff in a tree*

Tome XXIV, n° 4 (2015), p. 763-779.

[http://afst.cedram.org/item?id=AFST\\_2015\\_6\\_24\\_4\\_763\\_0](http://afst.cedram.org/item?id=AFST_2015_6_24_4_763_0)

© Université Paul Sabatier, Toulouse, 2015, tous droits réservés.

L'accès aux articles de la revue « Annales de la faculté des sciences de Toulouse Mathématiques » (<http://afst.cedram.org/>), implique l'accord avec les conditions générales d'utilisation (<http://afst.cedram.org/legal/>). Toute reproduction en tout ou partie de cet article sous quelque forme que ce soit pour tout usage autre que l'utilisation à fin strictement personnelle du copiste est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

cedram

Article mis en ligne dans le cadre du  
Centre de diffusion des revues académiques de mathématiques  
<http://www.cedram.org/>

## Total variation cutoff in a tree

YUVAL PERES<sup>(1)</sup>, PERLA SOUSI<sup>(2)</sup>

**RÉSUMÉ.** — On construit une famille d'arbres sur laquelle la marche aléatoire paresseuse présente un phénomène de transition abrupte à l'équilibre au sens de la variation totale. L'idée principale de la construction est le fait que les temps d'atteinte de grands sous-ensembles doivent se concentrer autour de leurs moyennes. Pour cette famille d'arbres, nous calculons le temps de mélange, le temps de relaxation et la fenêtre de transition abrupte.

**ABSTRACT.** — We construct a family of trees on which a lazy simple random walk exhibits total variation cutoff. The main idea behind the construction is that hitting times of large sets should be concentrated around their means. For this sequence of trees we compute the mixing time, the relaxation time and the cutoff window.

### 1. Introduction

Let  $X$  be an irreducible aperiodic Markov chain on a finite state space with stationary distribution  $\pi$  and transition matrix  $P$ . The lazy version of  $X$  is a Markov chain with transition matrix  $(P + I)/2$ . Let  $\varepsilon > 0$ . The  $\varepsilon$ -total variation mixing time is defined to be

$$t_{\text{mix}}(\varepsilon) = \min\{t \geq 0 : \max_x \|P^t(x, \cdot) - \pi\| \leq \varepsilon\},$$

where  $\|\mu - \nu\| = \sup_A |\mu(A) - \nu(A)|$  is the total variation distance between the measures  $\mu$  and  $\nu$ . When  $\varepsilon = 1/4$ , we simply write  $t_{\text{mix}} = t_{\text{mix}}(1/4)$ .

Suppose  $P$  is reversible. Let  $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3, \dots$  be its eigenvalues in decreasing order. Let  $\lambda_* = \max_{i \geq 2} |\lambda_i|$  and define the relaxation time  $t_{\text{rel}} = (1 - \lambda_*)^{-1}$ . Note that for a lazy chain  $\lambda_* = \lambda_2$ .

<sup>(1)</sup> Microsoft Research, Redmond, Washington, USA  
peres@microsoft.com

<sup>(2)</sup> University of Cambridge, Cambridge, UK  
p.sousi@statslab.cam.ac.uk

We say that a sequence of chains  $X^n$  exhibits total variation cutoff if for all  $0 < \varepsilon < 1$

$$\lim_{n \rightarrow \infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(1 - \varepsilon)} = 1.$$

We say that a sequence  $w_n$  is a cutoff window for a family of chains  $X^n$  if  $w_n = o(t_{\text{mix}})$  and for all  $\varepsilon > 0$  there exists a positive constant  $c_\varepsilon$  such that for all  $n$

$$t_{\text{mix}}(\varepsilon) - t_{\text{mix}}(1 - \varepsilon) \leq c_\varepsilon w_n.$$

Loosely speaking cutoff occurs when over a negligible period of time the total variation distance from stationarity drops abruptly from near 1 to near 0. It is standard that if  $t_{\text{rel}}$  and  $t_{\text{mix}}$  are of the same order, then there is no cutoff (see for instance [5, Proposition 18.4]). From that it follows that a lazy simple random walk on the interval  $[0, n]$  or a lazy simple random walk on a finite binary tree on  $n$  vertices do not exhibit cutoff, since in both cases  $t_{\text{rel}} \asymp t_{\text{mix}}$  (see for instance [5, Examples 18.5 and 18.6]).

Although the above two extreme types of trees do not exhibit cutoff, in this paper we construct a sequence of trees, where a lazy simple random walk exhibits total variation cutoff. We start by describing the tree and then state the results concerning the mixing and the relaxation time of the lazy simple random walk on it.

Let  $n_j = 2^{2^j}$  for  $j \in \mathbb{N}$ . We construct the tree  $\mathcal{T}$  of Figure 1 by placing a binary tree at the origin consisting of  $N = n_k^3$  vertices. Then for all  $j \in \{[k/2], \dots, k\}$  we place a binary tree at distance  $n_j$  from the origin consisting of  $N/n_j$  vertices.

For each  $j$  we call  $\mathcal{T}_j$  the binary tree attached at distance  $n_j$  and  $\mathcal{T}_0$  the binary tree at 0. We abuse notation and denote by  $n_j$  the root of  $\mathcal{T}_j$  and by 0 the root of  $\mathcal{T}_0$ .

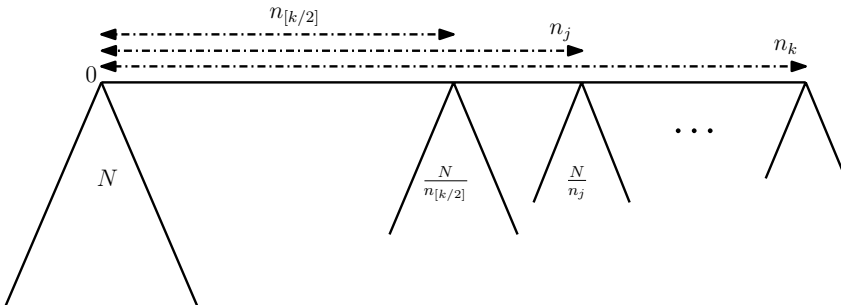


Figure 1. — The tree  $\mathcal{T}$  (not drawn to scale)

THEOREM 1.1. — *The lazy simple random walk on the tree  $\mathcal{T}$  exhibits total variation cutoff and for all  $\varepsilon$*

$$t_{\text{mix}}(\varepsilon) \sim 2Nk.$$

*Further, the cutoff window is of size  $N\sqrt{k}$ , i.e. for all  $0 < \varepsilon < 1$*

$$t_{\text{mix}}(\varepsilon) - t_{\text{mix}}(1 - \varepsilon) \leq c_\varepsilon N\sqrt{k},$$

*where  $c_\varepsilon$  is a positive constant.*

By Chen and Saloff-Coste [2] cutoff also holds for the continuous time random walk on  $\mathcal{T}$ .

The main ingredient in the proof of Theorem 1.1 is to establish the concentration of the first hitting time of 0 starting from  $n_k$ . Once this has been completed, cutoff follows easily. In Section 2, we prove the concentration result of the hitting time, which then gives a lower bound on the mixing time. Then in Section 3 we describe the coupling that will yield the matching upper bound on the mixing time.

*Remark 1.2.* — We note that the same idea of showing concentration of hitting times was used in [3] in order to establish cutoff for birth and death chains satisfying  $t_{\text{mix}}/t_{\text{rel}} \rightarrow \infty$ . Connection of hitting times to cutoff is presented in greater generality in [4].

It follows from Theorem 1.1 and [5, Proposition 18.4] that  $t_{\text{rel}} = o(t_{\text{mix}})$ . In the next theorem we give the exact order of the relaxation time. We prove it in Section 4.

We use the notation  $a_k \asymp b_k$  if there exists a constant  $C$  such that  $C^{-1}b_k \leq a_k \leq Cb_k$  for all  $k$  and we write  $a_k \lesssim b_k$  if there exists a constant  $C'$  such that  $a_k \leq C'b_k$  for all  $k$ .

THEOREM 1.3. — *The relaxation time for the lazy simple random walk on the tree  $\mathcal{T}$  satisfies*

$$t_{\text{rel}} \asymp N.$$

To the best of our knowledge the tree  $\mathcal{T}$  is the first example of a tree for which  $t_{\text{mix}}$  is not equivalent to  $t_{\text{rel}}$ . A related problem was studied in [7] and we recall it here.

Suppose that we assign conductances to the edges of a tree in such a way that  $c \leq c(e) \leq c'$  for all edges  $e$ , where  $c$  and  $c'$  are two positive

constants. It is proved in [7, Theorem 9.1] that the mixing time of the weighted lazy random walk is up to constants the same as the mixing time of the lazy simple random walk on the original tree. Since the relaxation time is given by a variational formula, it is immediate that after assigning bounded conductances  $t_{\text{rel}}$  is only changed up to multiplicative constants. Hence if in the original tree the quantities  $t_{\text{rel}}$  and  $t_{\text{mix}}$  are of the same order, then there is no way of assigning weights to the edges in order to make the weighted random walk exhibit cutoff.

## 2. Concentration of the hitting time

Let  $X$  denote a lazy simple random walk on the tree  $\mathcal{T}$ . Define for all  $x \in \mathcal{T}$

$$\tau_x = \inf\{s \geq 0 : X_s = x\}.$$

LEMMA 2.1. — *We have as  $k \rightarrow \infty$*

$$\mathbb{E}_{n_k}[\tau_0] = 2Nk + o\left(N\sqrt{k}\right) \quad \text{and} \quad \text{Var}_{n_k}(\tau_0) \asymp N^2k.$$

We will prove the above concentration lemma in this section. We start by stating standard results about hitting times and excursions that will be used in the proof of Lemma 2.1. We include their proofs for the sake of completeness.

CLAIM 2.2. — *Let  $\tau$  and  $\tilde{\tau}$  denote hitting times of the same state for a discrete time non-lazy and lazy walk respectively. Then*

$$\mathbb{E}[\tilde{\tau}] = 2\mathbb{E}[\tau] \quad \text{and} \quad \text{Var}(\tilde{\tau}) = 4\text{Var}(\tau) + 2\mathbb{E}[\tau],$$

*assuming that both the lazy and non-lazy walks start from the same vertex.*

*Proof.* — It is easy to see that we can write  $\tilde{\tau} = \sum_{i=0}^{\tau-1} \xi_i$ , where  $(\xi_i)_i$  is an i.i.d. sequence of geometric variables with success probability  $1/2$ . By Wald's identity we get

$$\mathbb{E}[\tilde{\tau}] = \mathbb{E}[\tau] \mathbb{E}[\xi_1] = 2\mathbb{E}[\tau].$$

Using the independence between  $\tau$  and the sequence  $(\xi_i)_i$  gives the identity for the variance of  $\tilde{\tau}$ .  $\square$

CLAIM 2.3. — *Let  $T$  be the time spent in an excursion from the root  $o$  by a simple random walk in a binary tree of size  $n$ . Let  $L$  be the length of an excursion from the root conditioned not to hit the leaves. Then for  $C$  a positive constant*

$$\mathbb{E}[T] = n - 1, \quad \mathbb{E}[T^2] \asymp n^2 \quad \text{and} \quad \mathbb{E}[L] \leq C.$$

*Proof.* — It is standard that  $\mathbb{E}[T] = \pi(o)^{-1} = n - 1$ . With probability at least  $1/2$  an excursion from the root hits the leaves. Therefore, the expected hitting time of the root starting from the leaves is upper bounded by  $cn$  for some positive constant  $c$ . Starting from any point  $x$  on the tree, the expected hitting time of the root is thus upper bounded by  $cn$ . Hence by performing independent experiments and using the Markov property we get for a positive constant  $c'$

$$\mathbb{P}(T > 2kn) \leq e^{-c'k}.$$

Therefore, we deduce that  $\mathbb{E}[T^2] \leq c'n^2$ .

For the last part of the claim, first note that the distance from the root behaves as a biased random walk  $X$  on the interval  $[0, h]$ , where  $h$  is the depth of the tree, with

$$\mathbb{P}(X_{k+1} = x + 1 \mid X_k = x) = \frac{2}{3} = 1 - \mathbb{P}(X_{k+1} = x - 1 \mid X_k = x).$$

Conditioning  $X$  to return to 0 before hitting  $h$  we obtain a new process up until it hits 0, which can be dominated by a biased random walk  $Y$  on  $[0, h - 1]$  with

$$\mathbb{P}(Y_{k+1} = x - 1 \mid Y_k = x) = \frac{2}{3} = 1 - \mathbb{P}(Y_{k+1} = x + 1 \mid Y_k = x).$$

We thus conclude that  $\mathbb{E}[L] \leq \mathbb{E}_0[T_0] \leq C$ , where  $T_0 = \inf\{k \geq 1 : Y_k = 0\}$ .  $\square$

CLAIM 2.4. — *Let  $X$  be a simple random walk on the interval  $[0, n]$  started from  $n$  and  $L_i$  be the number of visits to  $i$  before the first time  $X$  hits 0. Then  $L_i$  is a geometric random variable taking values in  $\{1, 2, \dots\}$  and success probability  $(2i)^{-1}$  for  $i \neq n$  and  $1/n$  for  $i = n$ .*

We are now ready to give the proof of the concentration result.

*Proof of Lemma 2.1.* — By Claim 2.2 it suffices to consider a non-lazy random walk. We write  $\tau_0$  for the first hitting time of 0 for a simple random walk on the tree  $\mathcal{T}$ .

Every time we visit a vertex  $n_j$  for some  $j$  with probability  $1/2$  we make an excursion in the binary tree attached to this vertex. Since we are interested in the time it takes to hit 0 we can think of the problem in the following way: we replace a binary tree by a self-loop representing a delay which is the time spent inside the tree in an excursion from the root. It will be helpful to have Figure 2 in mind.

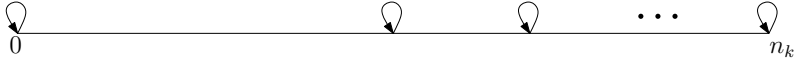


Figure 2. — Delays represented by self loops

Let  $Y$  be a simple random walk on the line  $[0, n_k]$  starting from  $n_k$ . Let  $S$  be the time it takes  $Y$  to reach 0. For  $i = [k/2], \dots, k$  we let  $L_i$  be the local time at  $n_i$  before the first time  $Y$  hits 0, i.e.

$$L_i = \sum_{\ell=0}^S \mathbf{1}(Y_\ell = n_i).$$

For every vertex  $n_i$  we let  $(T_\ell^{(i)})_{\ell \leq L_i}$  be the delays incurred during the  $L_i$  visits to  $n_i$ , i.e.

$$T_\ell^{(i)} = \sum_{m=1}^{G_{i,\ell}} \xi_m^{(i)},$$

where  $(\xi_m^{(i)})_m$  is an i.i.d. sequence of excursions from  $n_i$  in the binary tree rooted at  $n_i$  and  $G_{i,\ell}$  is an independent geometric random variable of success probability  $1/2$ . Note that the random variables  $T_\ell^{(i)}$  are independent over different  $i$  and  $\ell$ . Having defined these times we can now write

$$\tau_0 = S + \sum_{i=[k/2]}^k \sum_{\ell=1}^{L_i} T_\ell^{(i)} = S + D, \tag{2.1}$$

where  $D = \sum_{i=[k/2]}^k \sum_{\ell=1}^{L_i} T_\ell^{(i)}$ . From Claims 2.3 and 2.4 and the independence between  $L_i$  and  $T_\ell^{(i)}$  using the above representation of  $\tau_0$  we immediately get

$$\mathbb{E}_{n_k}[\tau_0] = n_k^2 + \sum_{i=[k/2]}^{k-1} 2n_i \left( \frac{N}{n_i} - 1 \right) + n_k \left( \frac{N}{n_k} - 1 \right) = Nk + o(N\sqrt{k}) \text{ as } k \rightarrow \infty,$$

and hence multiplying by 2 gives the required expression. We now turn to estimate the variance. Using (2.1) we have

$$\begin{aligned} \text{Var}_{n_k}(\tau_0) &= \mathbb{E}_{n_k} \left[ ((S - \mathbb{E}_{n_k}[S]) + (D - \mathbb{E}_{n_k}[D]))^2 \right] \\ &= \text{Var}_{n_k}(S) + \text{Var}_{n_k}(D) + 2\mathbb{E}_{n_k}[(S - \mathbb{E}_{n_k}[S])(D - \mathbb{E}_{n_k}[D])]. \end{aligned}$$

Since  $S$  is the first time that a simple random walk on  $[0, n_k] \cap \mathbb{Z}$  hits 0 started from  $n_k$  it follows that

$$\text{Var}_{n_k}(S) \asymp n_k^4 = o(N^2k). \tag{2.2}$$

This follows by using that  $X_n^4 - 6nX_n^2 + 3n^2 + 2n$  is a martingale when  $X$  is a simple random walk on  $\mathbb{Z}$ .

By Cauchy Schwarz we get

$$\mathbb{E}_{n_k}[(S - \mathbb{E}_{n_k}[S])(D - \mathbb{E}_{n_k}[D])] \leq \sqrt{\text{Var}_{n_k}(S) \text{Var}_{n_k}(D)},$$

so if we prove that

$$\text{Var}_{n_k}(D) \asymp N^2 k, \tag{2.3}$$

then using (2.2) we get  $\sqrt{\text{Var}_{n_k}(S) \text{Var}_{n_k}(D)} \asymp N n_k^2 \sqrt{k} = o(N^2 k)$ , and hence  $\text{Var}_{n_k}(\tau_0) \asymp N^2 k$ . Therefore, it suffices to show (2.3).

To simplify notation further we write  $D_i = \sum_{\ell=1}^{L_i} T_\ell^{(i)}$ . We have

$$\begin{aligned} \text{Var}_{n_k}(D) &= \sum_{i,j=[k/2]}^k \mathbb{E}_{n_k}[(D_i - \mathbb{E}_{n_k}[D_i])(D_j - \mathbb{E}_{n_k}[D_j])] \\ &= \sum_{j=[k/2]}^k \text{Var}_{n_k}(D_j) + 2 \sum_{j=[k/2]}^k \sum_{i=j+1}^k \mathbb{E}_{n_k}[(D_i - \mathbb{E}_{n_k}[D_i])(D_j - \mathbb{E}_{n_k}[D_j])]. \end{aligned} \tag{2.4}$$

By Claims 2.3 and 2.4 and the independence between  $L_i$  and  $T_\ell^{(i)}$ , we get that for all  $i$

$$\text{Var}_{n_k}(D_i) = \mathbb{E}_{n_k} \left[ T_1^{(i)} \right]^2 \text{Var}_{n_k}(L_i) + \mathbb{E}_{n_k}[L_i] \text{Var}_{n_k} \left( T_1^{(i)} \right) \asymp N^2,$$

and hence  $\sum_{i=[k/2]}^k \text{Var}_{n_k}(D_i) \asymp N^2 k$ . In view of that, it suffices to show that for  $i > j$

$$|\mathbb{E}_{n_k}[(D_i - \mathbb{E}_{n_k}[D_i])(D_j - \mathbb{E}_{n_k}[D_j])]| \lesssim N^2 \frac{n_j}{n_i}, \tag{2.5}$$

since then using the double exponential decay of  $(n_\ell)$  completes the proof of the lemma.

Since in order to hit 0 starting from  $n_k$  the random walk must first hit  $n_i$  and then  $n_j$ , it makes sense to split the local time  $L_i$  into two terms: the time  $L_{i,1}$  that  $Y$  spends at  $n_i$  before the first hitting time of  $n_j$  and the time  $L_{i,2}$  that  $Y$  spends at  $n_i$  after the first hitting time of  $n_j$ . Writing

$$D_{i,1} = \sum_{\ell=1}^{L_{i,1}} T_\ell^{(i)} \quad \text{and} \quad D_{i,2} = \sum_{\ell=1}^{L_{i,2}} \tilde{T}_\ell^{(i)},$$



where  $\tilde{T}$  is an independent copy of  $T$ , we have that  $D_{i,1}$  is independent of  $D_j$ , and hence

$$\mathbb{E}_{n_k}[(D_i - \mathbb{E}_{n_k}[D_i])(D_j - \mathbb{E}_{n_k}[D_j])] = \mathbb{E}_{n_k}[(D_{i,2} - \mathbb{E}_{n_k}[D_{i,2}]) (D_j - \mathbb{E}_{n_k}[D_j])]. \quad (2.6)$$

Using the independence between the local times and the delays we get

$$\begin{aligned} \mathbb{E}_{n_k}[D_{i,2}D_j] &= \mathbb{E}_{n_k} \left[ \mathbb{E}_{n_k} \left[ \sum_{\ell=1}^{L_{i,2}} \tilde{T}_\ell^{(i)} \sum_{r=1}^{L_j} T_r^{(j)} \mid L_{i,2}, L_j \right] \right] \\ &= \mathbb{E}_{n_k} \left[ L_{i,2} \mathbb{E}_{n_k} \left[ T_1^{(i)} \right] L_j \mathbb{E}_{n_k} \left[ T_1^{(j)} \right] \right] \\ &= \mathbb{E}_{n_k} [L_{i,2} L_j] \mathbb{E}_{n_k} \left[ T_1^{(i)} \right] \mathbb{E}_{n_k} \left[ T_1^{(j)} \right]. \end{aligned} \quad (2.7)$$

We let  $\tau_x = \inf\{t \geq 0 : Y_t = x\}$  and  $\tau_x^+ = \inf\{t \geq 1 : Y_t = x\}$  be the first hitting time of  $x$  by  $Y$  and the first return time to  $x$  respectively. We then get

$$\mathbb{P}_{n_j} \left( \tau_{n_i} < \tau_0 \wedge \tau_{n_j}^+ \right) = \frac{1}{2(n_i - n_j)}.$$

Once the random walk  $Y$  visits  $n_i$ , then the total number of returns to  $n_i$  before hitting  $n_j$  again is a geometric random variable independent of  $L_j$  and of parameter

$$\mathbb{P}_{n_i} \left( \tau_{n_j} < \tau_{n_i}^+ \right) = \frac{1}{2(n_i - n_j)}.$$

Hence we can write

$$L_{i,2} = \sum_{\ell=1}^{L_j-1} \eta_\ell,$$

where  $\eta_\ell = 0$  with probability  $1 - 1/(2(n_i - n_j))$  and  $\theta_\ell$  with probability  $1/(2(n_i - n_j))$ , where  $\theta_\ell$  is a geometric random variable with  $\mathbb{E}[\theta_\ell] = 2(n_i - n_j)$ . Note that  $\eta_\ell$  is independent of  $L_j$ . Therefore we deduce

$$\begin{aligned} \mathbb{E}_{n_k} [L_{i,2} L_j] &= \mathbb{E}_{n_k} \left[ L_j \sum_{\ell=1}^{L_j-1} \eta_\ell \right] = \mathbb{E}_{n_k} \left[ \mathbb{E} \left[ L_j \sum_{\ell=1}^{L_j-1} \eta_\ell \mid L_j \right] \right] \\ &= (\mathbb{E}_{n_k} [L_j^2] - \mathbb{E}_{n_k} [L_j]) \mathbb{E}_{n_k} [\eta_1] \asymp n_j^2, \end{aligned}$$

where in the last step we used Claim 2.4 and the fact that  $\mathbb{E}_{n_k} [\eta_\ell] = 1$  for all  $\ell$ . Hence combining the above with (2.7) and Claim 2.3 we conclude

$$\mathbb{E}_{n_k} [D_{i,2} D_j] \asymp n_j^2 \frac{N}{n_i} \frac{N}{n_j} = N^2 \frac{n_j}{n_i}. \quad (2.8)$$

Using Wald's identity we obtain

$$\mathbb{E}_{n_k}[D_{i,2}] \mathbb{E}_{n_k}[D_j] \asymp N^2 \frac{n_j}{n_i}$$

and combined with (2.6) and (2.8) proves (2.5) and thus finishes the proof of the lemma.  $\square$

*Proof of Theorem 1.1 (lower bound).* — Let  $t = \mathbb{E}_{n_k}[\tau_0] - \gamma\sqrt{\text{Var}_{n_k}(\tau_0)}$ , where the constant  $\gamma$  will be determined later. By the definition of the total variation distance we get

$$d(t) \geq \|\mathbb{P}_{n_k}(X_t \in \cdot) - \pi\| \geq \pi(\mathcal{T}_0) - \mathbb{P}_{n_k}(X_t \in \mathcal{T}_0) \geq 1 - o(1) - \mathbb{P}_{n_k}(\tau_0 < t),$$

since  $\pi(\mathcal{T}_0) = 1 - o(1)$ . Chebyshev's inequality gives

$$\mathbb{P}_{n_k}(\tau_0 < t) \leq \mathbb{P}_{n_k}\left(|\tau_0 - \mathbb{E}_{n_k}[\tau_0]| > \gamma\sqrt{\text{Var}_{n_k}(\tau_0)}\right) \leq \frac{1}{\gamma^2}.$$

Hence by choosing  $\gamma$  big enough we deduce that for all sufficiently large  $k$

$$d(t) \geq 1 - o(1) - \frac{1}{\gamma^2} > \varepsilon,$$

which implies that  $t_{\text{mix}}(\varepsilon) \geq t$ . By Lemma 2.1 we thus get that

$$t_{\text{mix}}(\varepsilon) \geq 2Nk - c_1 N\sqrt{k} \tag{2.9}$$

for a positive constant  $c_1$ .  $\square$

### 3. Coupling

In this section we prove the upper bound on  $t_{\text{mix}}(\varepsilon)$  via coupling.

*Proof of Theorem 1.1 (upper bound).* — Let  $X_0 = x$  and  $Y_0 \sim \pi$ . Consider the following coupling. We let  $X$  and  $Y$  evolve independently until the first time that  $X$  hits 0. After that we let them continue independently until the first time they collide or reach the same level of the tree  $\mathcal{T}_0$  in which case we change the coupling to the following one: we let  $X$  evolve as a lazy simple random walk and couple  $Y$  to  $X$  so that  $Y$  moves closer to (or further from) the root if and only if  $X$  moves closer to (or further from) the root respectively. Hence they coalesce if they both hit 0.

Let  $\tau$  be the coupling time and  $t = \mathbb{E}_{n_k}[\tau_0] + \gamma\sqrt{\text{Var}_{n_k}(\tau_0)}$ , where the constant  $\gamma$  will be determined later in order to make  $\mathbb{P}(\tau > t)$  as small as we like.

Define  $\tau_x^* = \inf\{s \geq \tau_0 : X_s = x\}$  for all  $x$  and

$$L = \sum_{s=\tau_0}^{\tau_{n_{[k/2]}}^*} \mathbf{1}(X_{s-1} \notin \mathcal{T}_0, X_s \in \mathcal{T}_0),$$

i.e.  $L$  is the number of returns to the tree  $\mathcal{T}_0$  in the time interval  $[\tau_0, \tau_{n_{[k/2]}}^*]$ . Then  $L$  has the geometric distribution with parameter  $1/n_{[k/2]}$ . Setting  $A_L = \{L > \sqrt{n_{[k/2]}}\}$  we get by the union bound

$$\mathbb{P}(A_L^c) = \mathbb{P}(L \leq \sqrt{n_{[k/2]}}) \leq \frac{1}{\sqrt{n_{[k/2]}}}. \tag{3.1}$$

We also define the event that after time  $\tau_0$  the random walk hits the leaves of the tree  $\mathcal{T}_0$  before exiting the interval  $[0, n_{[k/2]}]$ , i.e.

$$E = \left\{ \tau_{\partial\mathcal{T}_0}^* < \tau_{n_{[k/2]}}^* \right\}. \tag{3.2}$$

Since at every return to the tree  $\mathcal{T}_0$  with probability at least  $1/3$  the random walk hits the leaves of  $\mathcal{T}_0$  before exiting the tree  $\mathcal{T}_0$ , it follows that

$$\mathbb{P}\left(\tau_{\partial\mathcal{T}_0}^* > \tau_{n_{[k/2]}}^* \mid L\right) \leq \left(\frac{2}{3}\right)^L. \tag{3.3}$$

By decomposing into the events  $A_L$  and  $E$  we obtain

$$\begin{aligned} \mathbb{P}(\tau > t) &\leq \mathbb{P}(\tau > t, A_L) + \frac{1}{\sqrt{n_{[k/2]}}} \\ &\leq \mathbb{P}(\tau > t, A_L, E) + \left(\frac{2}{3}\right)^{\sqrt{n_{[k/2]}}} + \frac{1}{\sqrt{n_{[k/2]}}}, \end{aligned} \tag{3.4}$$

where the first inequality follows from (3.1) and the second one from (3.3) and the fact that we are conditioning on the event  $\{L > \sqrt{n_{[k/2]}}\}$ . We now define  $S$  to be the first time after  $\tau_{\partial\mathcal{T}_0}^*$  that  $X$  hits 0, i.e.  $S = \inf\{s \geq \tau_{\partial\mathcal{T}_0}^* : X_s = 0\}$ .

Let  $(\xi_i)_i$  be i.i.d. random variables, where  $\xi_1$  is distributed as the length of a random walk excursion on the interval  $[0, n_{[k/2]}]$  conditioned not to hit  $n_{[k/2]}$ . Let  $(\ell_{i,j})_{i,j}$  be i.i.d. random variables with  $\ell_{1,1}$  distributed as the length of a random walk excursion from the root on the tree  $\mathcal{T}_0$  conditioned not to hit the leaves and  $(G_i)_i$  be i.i.d. geometric random variables of success probability  $1/3$ . Then on the event  $E$  we have

$$S - \tau_0 \prec \sum_{i=1}^L \xi_i + \sum_{i=1}^L \sum_{j=1}^{G_i} \ell_{i,j} + \zeta,$$

where  $\zeta$  is independent of the excursion lengths and is distributed as the commute time between the root and the leaves of the tree  $\mathcal{T}_0$  and  $\prec$  denotes stochastic domination. Hence, by Wald's identity we obtain

$$\begin{aligned} \mathbb{E}[(S - \tau_0)\mathbf{1}(E)] &\leq \mathbb{E}[L] \mathbb{E}[\xi_1] + \mathbb{E}[L] \mathbb{E}[G_1] \mathbb{E}[\ell_{1,1}] + \mathbb{E}[\zeta] \\ &\lesssim n_{\lfloor k/2 \rfloor}^2 + n_{\lfloor k/2 \rfloor} + N \lesssim N. \end{aligned} \quad (3.5)$$

Here we used Claim 2.3 to get that  $\mathbb{E}[\ell_{1,1}] \leq C$  for a positive constant  $C$ .

Let  $A = \{Y_{\tau_0} \in \mathcal{T}_0\}$ . Then  $\mathbb{P}(A^c) = o(1)$  as  $k \rightarrow \infty$ , because at time  $\tau_0$  the random walk  $Y$  is stationary, since until this time it evolves independently of  $X$ , and also the stationary probability of the tree is  $1 - o(1)$ . It then follows

$$\mathbb{P}(\tau > t, A_L, E) \leq \mathbb{P}(\tau > t, A_L, E, A) + o(1). \quad (3.6)$$

Let  $\tau_1$  be the time it takes to hit the line  $[0, n_k]$  starting from  $x$ . Let  $\tau_2$  be the time it takes to hit 0 starting from  $X_{\tau_1}$ . Then clearly  $\tau_2$  is smaller than the time it takes to hit 0 starting from  $n_k$ . Thus setting  $B = \{\tau_0 < \mathbb{E}_{n_k}[\tau_0] + \gamma\sqrt{\text{Var}_{n_k}(\tau_0)}/2\}$  we obtain

$$\begin{aligned} \mathbb{P}_x(B^c) &\leq \mathbb{P}_x\left(\tau_2 > \mathbb{E}_{n_k}[\tau_0] + \frac{\gamma\sqrt{\text{Var}_{n_k}(\tau_0)}}{4}\right) + \mathbb{P}_x\left(\tau_1 \geq \frac{\gamma\sqrt{\text{Var}_{n_k}(\tau_0)}}{4}\right) \\ &\leq \mathbb{P}_{n_k}\left(\tau_0 > \mathbb{E}_{n_k}[\tau_0] + \frac{\gamma\sqrt{\text{Var}_{n_k}(\tau_0)}}{4}\right) + o(1) \leq \frac{16}{\gamma^2} + o(1), \end{aligned} \quad (3.7)$$

where the second inequality follows from Markov's inequality and the fact that  $\mathbb{E}_x[\tau_1] \leq N$  for all  $x$  and the third one follows from Chebyshev's inequality. Ignoring the  $o(1)$  terms we get

$$\mathbb{P}(\tau > t, A_L, E, A) \leq \mathbb{P}(\tau > t, A_L, E, A, B) + \frac{16}{\gamma^2}. \quad (3.8)$$

We finally define the event  $F = \{S - \tau_0 > \gamma\sqrt{\text{Var}_{n_k}(\tau_0)}/2\}$ . We note that on the events  $E$  and  $A$  the two walks  $X$  and  $Y$  must have coalesced by time  $S$ . (Indeed, if  $Y$  stays in  $\mathcal{T}_0$  during the time interval  $[\tau_0, \tau_{\partial\mathcal{T}_0}^*]$ , then they must have coalesced. If  $Y$  leaves  $\mathcal{T}_0$  during this time interval, since  $X$  is always in  $[0, n_{\lfloor k/2 \rfloor}]$  until time  $S$  on the event  $E$ , then coalescence must have happened again.) Therefore

$$A \cap E \cap B \cap F^c \subseteq \{\tau < t\}.$$

This in turn implies that for a positive constant  $c_1$

$$\mathbb{P}(\tau > t, A_L, E, A, B) = \mathbb{P}(\tau > t, A_L, E, A, B, F) \leq \mathbb{P}(E, F) \leq \frac{c_1}{\gamma\sqrt{k}}, \quad (3.9)$$

where the last inequality follows by applying Markov's inequality to  $(S - \tau_0)\mathbf{1}(E)$  and using (3.5) and Lemma 2.1. Plugging (3.6), (3.8) and (3.9) into (3.4) gives as  $k \rightarrow \infty$

$$\mathbb{P}(\tau > t) \leq \frac{16}{\gamma^2} + o(1).$$

Hence choosing  $\gamma$  sufficiently large depending on  $\varepsilon$  we can make  $\mathbb{P}(\tau > t) < \varepsilon$  and this shows that for a positive constant  $c_\varepsilon$

$$t_{\text{mix}}(\varepsilon) \leq \mathbb{E}_{n_k}[\tau_0] + \gamma_\varepsilon \sqrt{\text{Var}_{n_k}(\tau_0)} \leq 2Nk + c_2 N\sqrt{k},$$

where the last inequality follows by Lemma 2.1. Combining this with the lower bound on  $t_{\text{mix}}(\varepsilon)$  from (2.9) shows that there exists  $c_\varepsilon > 0$  such that for all  $0 < \varepsilon < 1$

$$t_{\text{mix}}(\varepsilon) - t_{\text{mix}}(1 - \varepsilon) < c_\varepsilon N\sqrt{k}$$

and this completes the proof of the theorem.

#### 4. Relaxation time

In this section we give the proof of Theorem 1.3. We start by stating standard results for random walks on the interval  $[0, n]$  and the binary tree. We include their proofs here for the sake of completeness. A detailed analysis of relaxation time for birth and death chains can be found in [1].

CLAIM 4.1. — *Let  $f$  be a function defined on  $[0, n]$  satisfying  $f(0) = 0$ . Then*

$$\sum_{k=1}^n f(k)^2 \leq n^2 \sum_{\ell=1}^n (f(\ell) - f(\ell - 1))^2.$$

*Proof.* — We set  $\beta_\ell^{-2} = (n - \ell)$  for all  $\ell \in [0, n]$ . Then by Cauchy Schwarz we get

$$f(k)^2 = \left( \sum_{\ell=1}^k (f(\ell) - f(\ell - 1)) \right)^2 \leq \sum_{\ell=1}^k \beta_\ell^2 (f(\ell) - f(\ell - 1))^2 \sum_{\ell=1}^k \beta_\ell^{-2}.$$

Since  $\sum_{\ell=1}^k \beta_\ell^{-2} \leq n^2$  for all  $k \in [0, n]$  we get summing over all  $k$  and interchanging sums

$$\sum_{k=1}^n f(k)^2 \leq n^2 \sum_{\ell=1}^n (f(\ell) - f(\ell - 1))^2$$

and this completes the proof of the claim.  $\square$

CLAIM 4.2. — Let  $\mathcal{T}$  be a binary tree on  $m$  vertices with root  $o$ . Then there exists a universal constant  $c$  such that for all functions  $g$  defined on  $\mathcal{T}$  with  $g(o) = 0$  we have

$$\|g\|^2 \leq cm\mathcal{E}(g, g),$$

where  $\|g\|^2 = \sum_x \pi(x)g(x)^2$  and  $\mathcal{E}$  is the Dirichlet form  $\mathcal{E}(f, g) = \langle f, (I - P)g \rangle_\pi$ . Here  $\pi$  and  $P$  are the stationary distribution and transition matrix of a simple random walk on  $\mathcal{T}$  respectively.

*Proof.* — Since the stationary measure of a simple random walk on the tree satisfies  $\pi(x) \asymp m^{-1}$  for all  $x$  and  $P(x, y) \asymp c$  for all  $x \sim y$ , we will omit them from the expressions.

Let the depth of the tree  $\mathcal{T}$  be  $n = \lceil \log_2 m \rceil$ . Let  $x_k$  be a vertex in  $\mathcal{T}$  of level  $k$ . Then there exists a unique path  $x_0 = o, x_1, \dots, x_k$  going from the root to  $x_k$ . We can now write

$$\begin{aligned} g(x_k)^2 &= \left( \sum_{j=1}^k (g(x_{j-1}) - g(x_j)) \right)^2 = \left( \sum_{j=1}^k (g(x_{j-1}) - g(x_j)) 2^{j/2} \frac{1}{2^{j/2}} \right)^2 \\ &\lesssim \sum_{j=1}^k 2^j (g(x_{j-1}) - g(x_j))^2, \end{aligned}$$

where the last inequality follows by Cauchy Schwarz. Let  $L_k$  denote all the vertices of the tree at distance  $k$  from the root. For any  $x \in \mathcal{T}$  we write

$$G(x) = \sum_{j=1}^{|x|} 2^j (g(y_j) - g(y_{j-1}))^2,$$

where  $|x|$  denotes the level of  $x$  and  $y_0 = o, y_1, \dots, y_{|x|} = x$  is the unique path joining  $x$  to the root. By interchanging sums we obtain

$$\sum_{x \in \mathcal{T}_0} g(x)^2 = \sum_{k=1}^n \sum_{x \in L_k} g(x)^2 \lesssim \sum_k \sum_{x \in L_k} G(x). \quad (4.1)$$

Let  $e$  be an edge of  $\mathcal{T}$ . We write  $e = \langle e^-, e^+ \rangle$  where  $|e^-| < |e^+|$  and  $|\cdot|$  stands for the distance from the root. For every edge  $e$  we let  $N(e)$  be the number of times the term  $(g(e^-) - g(e^+))^2$  appears in the sum appearing on the right hand side of (4.1). We then get

$$\sum_{x \in \mathcal{T}} g(x)^2 \lesssim \sum_{e \in \mathcal{T}} 2^{|e^-|} N(e) (g(e^-) - g(e^+))^2.$$

Notice that  $N(e)$  is the number of paths in  $\mathcal{T}$  joining the root to the leaves and pass through  $e$ . Hence since the tree is of depth  $n$  we get that  $N(e) = 2^{n-|e^-|-1}$ . Therefore we deduce

$$\begin{aligned} \sum_{x \in \mathcal{T}} g(x)^2 &\lesssim \sum_{e \in \mathcal{T}} 2^{|e^-|} 2^{n-|e^-|-1} (g(e^-) - g(e^+))^2 = 2^{n-1} \sum_{e \in \mathcal{T}} (g(e^-) - g(e^+))^2 \\ &\lesssim m\mathcal{E}(g, g) \end{aligned}$$

and this completes the proof of the claim.  $\square$

*Proof of Theorem 1.3.* — To prove the lower bound on  $t_{\text{rel}}$  we use the bottleneck ratio  $\Phi_*$  as in [5, Theorem 13.14]. By setting  $S = \mathcal{T}_0$ , we see that

$$\Phi_* \lesssim \frac{1}{N},$$

and hence  $t_{\text{rel}} \gtrsim N$ . It remains to prove a matching upper bound. We do that by using the variational formula for the spectral gap, which gives

$$t_{\text{rel}} = \sup_{\text{Var}_\pi(f) \neq 0} \frac{\text{Var}_\pi(f)}{\mathcal{E}(f, f)}.$$

Notice that by subtracting from  $f$  its value at 0 the ratio above remains unchanged. So we restrict to functions  $f$  with  $f(0) = 0$ . It suffices to show that for any such  $f$

$$\text{Var}_\pi(f) \lesssim N\mathcal{E}(f, f). \tag{4.2}$$

Let  $f$  be defined on the tree  $\mathcal{T}$  with  $f(0) = 0$ . Then we can write  $f = g + h$ , where  $g$  is zero on  $\mathcal{T}_0^c$  and  $h$  is zero on  $\mathcal{T}_0$  and  $g(0) = h(0) = 0$ . We then have

$$\text{Var}_\pi(f) \leq \|g + h\|^2 = \|g\|^2 + \|h\|^2, \tag{4.3}$$

since by the definition of the functions  $g$  and  $h$  it follows that  $\langle g, h \rangle_\pi = 0$ . Similarly we also get

$$\mathcal{E}(f, f) = \mathcal{E}(g, g) + \mathcal{E}(h, h).$$

CLAIM 4.3. — *There exists a positive constant  $c$  such that*

$$\|h\|^2 \leq cN\mathcal{E}(h, h).$$

*Proof.* — Using Claim 4.2 for the function  $(h(x) - h(n_j))$  restricted to  $x \in \mathcal{T}_j$  we obtain

$$\sum_{v \in \mathcal{T}_j} (h(v) - h(n_j))^2 \lesssim \frac{N}{n_j} \sum_{\substack{u, v \in \mathcal{T}_j \\ u \sim v}} (h(u) - h(v))^2. \tag{4.4}$$

Using that  $(a + b)^2 \leq 2a^2 + 2b^2$  and (4.4) we get

$$\sum_{v \in \mathcal{T}_j} h(v)^2 \leq 2 \sum_{v \in \mathcal{T}_j} (h(v) - h(n_j))^2 + 2 \frac{N}{n_j} h(n_j)^2 \lesssim \frac{N}{n_j} (\mathcal{E}(h, h) + h(n_j)^2).$$

From the above inequality it immediately follows

$$\begin{aligned} \sum_{v \notin \mathcal{T}_0} h(v)^2 &\leq \sum_{v \in [0, n_k]} h(v)^2 + \sum_{j=[k/2]}^k \sum_{v \in \mathcal{T}_j} h(v)^2 \\ &\lesssim \sum_{v \in [0, n_k]} h(v)^2 + \sum_{j=[k/2]}^k \frac{N}{n_j} h(n_j)^2 + N \mathcal{E}(h, h) \end{aligned}$$

and hence it suffices to show

$$\sum_{v \in [0, n_k]} h(v)^2 + \sum_{j=[k/2]}^k \frac{N}{n_j} h(n_j)^2 \lesssim N \sum_{\ell=1}^{n_k} (h(\ell) - h(\ell - 1))^2. \quad (4.5)$$

Claim 4.1 gives

$$\sum_{v \in [0, n_k]} h(v)^2 \leq n_k^2 \sum_{\ell=1}^{n_k} (h(\ell) - h(\ell - 1))^2 \leq N \sum_{\ell=1}^{n_k} (h(\ell) - h(\ell - 1))^2,$$

since  $N = n_k^3$ , and hence it suffices to show

$$\sum_{j=[k/2]}^k \frac{h(n_j)^2}{n_j} \lesssim \sum_{\ell=1}^{n_k} (h(\ell) - h(\ell - 1))^2. \quad (4.6)$$

Setting  $\Delta_\ell = h(\ell) - h(\ell - 1)$  and using Cauchy Schwarz

$$\begin{aligned} h(n_j)^2 &\leq 2h(n_{j-1})^2 + 2(h(n_j) - h(n_{j-1}))^2 \\ &= 2 \left( \sum_{\ell=1}^{n_{j-1}} \Delta_\ell \right)^2 + 2 \left( \sum_{\ell=n_{j-1}+1}^{n_j} \Delta_\ell \right)^2 \\ &\leq 2n_{j-1} \sum_{\ell=1}^{n_{j-1}} \Delta_\ell^2 + (n_j - n_{j-1}) \sum_{\ell=n_{j-1}+1}^{n_j} \Delta_\ell^2, \end{aligned}$$

and hence dividing by  $n_j$  we get

$$\sum_{j=[k/2]}^k \frac{h(n_j)^2}{n_j} \leq 2 \sum_{j=[k/2]}^k \frac{n_{j-1}}{n_j} \sum_{\ell=1}^{n_{j-1}} \Delta_\ell^2 + 2 \sum_{\ell=1}^{n_k} \Delta_\ell^2.$$



If we fix  $\ell \in [0, n_k]$ , then the coefficient of  $\Delta_\ell^2$  in the first sum appearing on the right hand side of the above inequality is bounded from above by  $\sum_{j=1}^k n_{j-1}/n_j < \infty$ , and hence we conclude

$$\sum_{j=\lceil k/2 \rceil}^k \frac{h(n_j)^2}{n_j} \lesssim \sum_{\ell=1}^{n_k} \Delta_\ell^2$$

and this finishes the proof of the claim.  $\square$

Since  $g$  satisfies the assumptions of Claim 4.2 it follows that

$$\|g\|^2 \leq cN\mathcal{E}(g, g).$$

This together with Claim 4.3 and (4.3) proves (4.2) and completes the proof of the theorem.

## 5. Concentration of hitting times

By changing the parameters of our construction, for some  $C < \infty$  and infinitely many  $n$ , an  $n$ -vertex tree  $T_n$  is presented in [6] with the following property: There exist vertices  $x, y$  in  $T_n$  such that lazy simple random walk on  $T_n$  satisfies

$$\mathrm{Var}_x(\tau_y) \leq C \frac{(\mathbb{E}_x[\tau_y])^2}{\log n} \tag{5.1}$$

Is there a general matching lower bound, i.e., is there  $c > 0$  such that for every  $n$ , the inequality

$$\mathrm{Var}_x(\tau_y) \geq c \frac{(\mathbb{E}_x[\tau_y])^2}{\log n} \tag{5.2}$$

holds for lazy simple random walk on all  $n$ -vertex simple graphs?

**Acknowledgments.** — We thank the referee for useful comments.

## Bibliography

- [1] CHEN (G.-Y.) and SALOFF-COSTE (L.). — On the mixing time and spectral gap for birth and death chains, 2013. arXiv:1304.4346.
- [2] CHEN (G.-Y.) and SALOFF-COSTE (L.). — Comparison of cutoffs between lazy walks and Markovian semigroups. *J. Appl. Probab.*, 50(4), p. 943-959 (2013).
- [3] DING (J.), LUBETZKY (E.), and PERES (Y.). — Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields*, 146(1-2), p. 61-85 (2010).
- [4] LANCIA (C.), NARDI (F. A.), and SCOPPOLA (B.). — Entropy-driven cutoff phenomena. *J. Stat. Phys.*, 149(1), p. 108-141 (2012).

- [5] LEVIN (D. A.), PERES (Y.), and WILMER (E. L.). — Markov chains and mixing times. American Mathematical Society, Providence, RI (2009). With a chapter by James G. Propp and David B. Wilson.
- [6] NORRIS (J.), PERES (Y.), and ZHAI (A.). — Surprise probabilities in Markov chains. In Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015, p. 1759-1773 (2015).
- [7] PERES (Y.) and SOUSI (P.). — Mixing times are hitting times of large sets. Journal of Theoretical Probability, p. 1-32 (2013).