
Surprise Probabilities in Markov Chains

JAMES NORRIS^{1†}, YUVAL PERES² and ALEX ZHAI³

¹ Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK
(e-mail: james@statslab.cam.ac.uk)

² Microsoft Research, Redmond, Washington, WA 98052, USA
(e-mail: peres@microsoft.com)

³ Department of Mathematics, Stanford University, Stanford, CA 94305, USA
(e-mail: azhai@stanford.edu)

Received 14 January 2015; revised 22 December 2016; first published online 16 March 2017

In a Markov chain started at a state x , the *hitting time* $\tau(y)$ is the first time that the chain reaches another state y . We study the probability $\mathbb{P}_x(\tau(y) = t)$ that the first visit to y occurs precisely at a given time t . Informally speaking, the event that a new state is visited at a large time t may be considered a ‘surprise’. We prove the following three bounds.

- In any Markov chain with n states, $\mathbb{P}_x(\tau(y) = t) \leq n/t$.
- In a reversible chain with n states, $\mathbb{P}_x(\tau(y) = t) \leq \sqrt{2n}/t$ for $t \geq 4n + 4$.
- For random walk on a simple graph with $n \geq 2$ vertices, $\mathbb{P}_x(\tau(y) = t) \leq 4e \log(n)/t$.

We construct examples showing that these bounds are close to optimal. The main feature of our bounds is that they require very little knowledge of the structure of the Markov chain.

To prove the bound for random walk on graphs, we establish the following estimate conjectured by Aldous, Ding and Oveis-Gharan (private communication): for random walk on an n -vertex graph, for every initial vertex x ,

$$\sum_y \left(\sup_{t \geq 0} p^t(x, y) \right) = O(\log n).$$

2010 *Mathematics subject classification*: 60J10

1. Introduction

Suppose that a Markov chain with n states is run for a long time $t \gg n$. It would be surprising if the state visited on the t th step were not visited at any earlier time; if a state is likely, then we

[†] Partially supported by EPSRC grant EP/103372X/1.

expect it to have been visited earlier, and if it is rare, we do not expect that its first visit will occur at precisely time t . How surprised should we be?

Let $X = \{X_t\}$ be a Markov chain with finite state space S of size n , and let \mathbb{P}_x denote the probability for the chain started at x . Let \mathcal{S}_t denote the event that the state visited at time t was not seen at any previous time. In this paper, we quantify the intuition that \mathcal{S}_t is unlikely by proving upper bounds on $\mathbb{P}(\mathcal{S}_t)$. This question was posed by A. Kontorovich (private communication), who asked for bounds that do not require detailed knowledge of the transition probabilities of the chain.

For $y \in S$, the *hitting time* $\tau(y)$ is defined as the first time that the chain reaches y . We can express \mathcal{S}_t in terms of the hitting time as follows:

$$\mathcal{S}_t = \bigcup_{y \in S} \{\tau(y) = t\}.$$

If we can prove a statement of the form $\mathbb{P}_x(\tau(y) = t) \leq M$ for all $y \in S$, then it follows that $\mathbb{P}_x(\mathcal{S}_t) \leq n \cdot M$. Although naive, it turns out that this approach gives bounds on $\mathbb{P}_x(\mathcal{S}_t)$ which are close to optimal, as we will explain in more detail later.

We start with a simple proposition which bounds $\mathbb{P}_x(\tau(y) = t)$ in the most general setting.

Proposition 1.1. *Let X be a Markov chain with finite state space S of size n , and let x and y be any two states. Then, for all $t > n$,*

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{n}{t}, \quad \text{whence } \mathbb{P}_x(\mathcal{S}_t) \leq \frac{n^2}{t}.$$

Our first main theorem improves this bound for reversible chains.

Theorem 1.2. *Let X be a reversible Markov chain with finite state space S of size n , and let π be the stationary distribution of X . Consider any two states x and y . For all $t > 0$,*

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{2e \cdot \max(1, \log(1/\pi(x)))}{t}.$$

In particular, if X is the random walk on a simple graph with $n \geq 2$ vertices, then

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{4e \cdot \log n}{t} \quad \text{and} \quad \mathbb{P}_x(\mathcal{S}_t) \leq \frac{4e \cdot n \log n}{t}.$$

One limitation of Theorem 1.2 is that in general, the stationary probability $\pi(x)$ can be arbitrarily small. Our second main theorem gives an alternative bound for reversible chains that, in the style of Proposition 1.1, depends only on n and t .

Theorem 1.3. *Let X be a finite reversible Markov chain with n states, and let x and y be any two states. Then, for all $t \geq 4n + 4$,*

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{\sqrt{2n}}{t}, \quad \text{whence } \mathbb{P}_x(\mathcal{S}_t) \leq \frac{n\sqrt{2n}}{t}.$$

Remark. For reversible chains with non-negative eigenvalues, we prove the stronger bound

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{2} \sqrt{\frac{n}{t(t-n)}}.$$

The degrading of the estimate as t approaches n cannot be avoided; if $t < n$, then it is possible to have $\mathbb{P}_x(\tau(y) = t)$ arbitrarily close to 1 by considering a birth-and-death chain on states $1, 2, \dots, n$ where state i transitions to state $i + 1$ with probability $1 - \varepsilon$ for $\varepsilon \rightarrow 0$.

To give a sense of how accurate these estimates are, we give several constructions where $\mathbb{P}_x(\tau(y) = t)$ can be relatively large. In particular, we will show that Proposition 1.1 and Theorem 1.3 are tight up to a constant factor. We also construct a family of simple graphs which achieve

$$\mathbb{P}_x(\tau(y) = t) \geq \frac{c\sqrt{\log n}}{t} \tag{1.1}$$

for some constant $c > 0$. This does not quite match the upper bound of $(4e \log n)/t$ appearing in Theorem 1.2, but it demonstrates that the dependence on n cannot be avoided. All of these constructions can be modified slightly to give lower bounds on $\mathbb{P}_x(\mathcal{S}_t)$ which are of the order of n times the corresponding lower bounds on $\mathbb{P}_x(\tau(y) = t)$.

In the proof of Theorem 1.2, we show a certain ‘maximal probability’ bound that may be of independent interest. For any two states $x, y \in S$, define

$$p^*(x, y) = \sup_{t \geq 0} p^t(x, y),$$

where $p^t(x, y)$ is the transition probability from x to y in t steps. It was asked by Aldous, and independently by Ding and Oveis Gharan, whether for random walks on simple graphs with n vertices, $\sum_{y \in S} p^*(x, y) = O(\log n)$ for any starting vertex x (private communication). Using a theorem of Starr [10], we prove the following proposition, which verifies this.

Proposition 1.4. *Let X be a reversible Markov chain with finite state space S and stationary distribution π . Then, for any $x \in S$,*

$$\sum_{y \in S} p^*(x, y) \leq 2e \cdot \max\left(1, \log \frac{1}{\pi(x)}\right).$$

Remark. For simple random walk on an n -vertex graph, the right-hand side is at most $4e \cdot \log n$. When the graph is a cycle, this bound is tight up to a constant factor (see the end of Section 2.1).

Finally, we mention some situations where stronger bounds for $\mathbb{P}_x(\tau(y) = t)$ of the order of $1/t$ are possible.

Proposition 1.5. *Let X be a Markov chain with state space S and stationary distribution π . Then, for any $y \in S$ and $t > 0$,*

$$\mathbb{P}_\pi(\tau(y) = t) \leq \frac{1}{t},$$

where \mathbb{P}_π denotes the probability for the walk started from a random state sampled from π . In particular, for any state x ,

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{t\pi(x)}.$$

A consequence of Proposition 1.5 is that $\mathbb{P}_x(\tau(y) = t) = O(1/t)$ when t is significantly larger than the mixing time. We define

$$d_x(t) = \|p^t(x, \cdot) - \pi\|_{\text{TV}},$$

where $\|\cdot\|_{\text{TV}}$ denotes total variation distance, and recall that the mixing time $t_{\text{mix}}(\varepsilon)$ is defined as the earliest time t for which $d_x(t) \leq \varepsilon$ for every state x .

Proposition 1.6. *For a Markov chain X with state space S of size n , suppose that we have a bound of the form $\mathbb{P}_x(\tau(y) = t) \leq \psi(t)$ for all $x, y \in S$ (e.g. the bounds of Proposition 1.1 or Theorems 1.2 and 1.3). For any $t > s > 0$,*

$$\mathbb{P}_x(\tau(y) = t) \leq d_x(s)\psi(t - s) + \frac{1}{t - s}.$$

In particular, if $t > 2t_{\text{mix}}(1/4)\lceil \log_2 n \rceil$, we have

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{4}{t}.$$

The rest of the paper is organized as follows. In Section 2 we prove Proposition 1.1 and Theorem 1.2, which are proved by the same method. We prove Proposition 1.4 as an intermediate step, and we also describe constructions giving lower bounds on $\mathbb{P}_x(\tau(y) = t)$. In Section 3 we prove Theorem 1.3. Here we also construct an example to show that the bound is of the right order.

Section 4 is devoted to describing the modified constructions which give lower bounds for $\mathbb{P}_x(\mathcal{S}_t)$. We prove Propositions 1.5 and 1.6 in Section 5. The final three sections contain open problems (Section 6), acknowledgements, and proofs of lemmas deferred from previous sections (Appendix).

2. Proofs of Proposition 1.1 and Theorem 1.2

We start with a bound on $\mathbb{P}_x(\tau(y) = t)$ in terms of the maximal probabilities $p^*(x, y)$.

Lemma 2.1. *Let X be a Markov chain with finite state space S , and let x and y be any two states. Then, for all $t > 0$,*

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{t} \sum_{z \in S} p^*(x, z).$$

Proof. Observe that for each time $s < t$,

$$\mathbb{P}_x(\tau(y) = t) \leq \sum_{z \in S} p^s(x, z) \mathbb{P}_z(\tau(y) = t - s).$$

Summing this inequality over all $s = 0, 1, 2, \dots, t - 1$, we obtain

$$\begin{aligned} t \cdot \mathbb{P}_x(\tau(y) = t) &\leq \sum_{s=0}^{t-1} \sum_{z \in S} p^s(x, z) \mathbb{P}_z(\tau(y) = t - s) \\ &= \sum_{z \in S} \sum_{s=0}^{t-1} p^s(x, z) \mathbb{P}_z(\tau(y) = t - s) \leq \sum_{z \in S} p^*(x, z) \sum_{s=0}^{t-1} \mathbb{P}_z(\tau(y) = t - s) \\ &= \sum_{z \in S} p^*(x, z) \mathbb{P}_z(1 \leq \tau(y) \leq t) \leq \sum_{z \in S} p^*(x, z). \end{aligned}$$

Dividing both sides by t completes the proof. □

Proof of Proposition 1.1. By Lemma 2.1, we have

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{t} \sum_{z \in S} p^*(x, z) \leq \frac{1}{t} \sum_{z \in S} 1 = \frac{n}{t}. \quad \square$$

Under the assumption that Proposition 1.4 holds, we can now also prove Theorem 1.2.

Proof of Theorem 1.2. By Lemma 2.1 and Proposition 1.4,

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{t} \sum_{z \in S} p^*(x, z) \leq \frac{2e \cdot \max(1, \log(1/\pi(x)))}{t}.$$

For the random walk on a simple graph having $n \geq 2$ vertices, the stationary probability of a vertex is proportional to its degree, and in particular it is at least $1/n^2$. Thus, in this case

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{4e \cdot \log n}{t}. \quad \square$$

2.1. Proof of Proposition 1.4

It remains to prove Proposition 1.4. For a Markov chain $X = \{X_t\}$, let P be the transition operator of the chain. That is, for a function $f : S \rightarrow \mathbb{R}$, we have

$$(Pf)(x) = \sum_{y \in S} p(x, y)f(y).$$

We deduce Proposition 1.4 from the following theorem of Starr [10].

Theorem 2.2 (special case of [10], Theorem 1). *Let $X = \{X_t\}$ be a reversible Markov chain with finite state space S and stationary measure π . Then, for any $1 < p < \infty$ and $f : S \rightarrow \mathbb{R}$,*

$$\mathbb{E}_\pi \left| \sup_{n \geq 0} P^{2n} f \right|^p \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}_\pi |f|^p,$$

where for a function $g : S \rightarrow \mathbb{R}$, $\mathbb{E}_\pi g := \sum_{s \in S} \pi(s)g(s)$ denotes expectation with respect to π .

We refer the reader to [1, Section 2.1] for a short proof of this special case of Starr’s theorem.

Proof of Proposition 1.4. Define the two quantities

$$p_e^*(x, y) = \sup_{t \geq 0} p^{2t}(x, y) \quad \text{and} \quad p_o^*(x, y) = \sup_{t \geq 0} p^{2t+1}(x, y)$$

for even and odd times. Take $f(y) = \delta_{x,y}$ in Theorem 2.2. Then, we find that for each t ,

$$(P^t f)(y) = \sum_{z \in S} p^t(y, z) f(z) = p^t(y, x) = \frac{p^t(x, y)\pi(x)}{\pi(y)},$$

and so

$$\sup_{t \geq 0} (P^{2t} f)(y) = \frac{p_e^*(x, y)\pi(x)}{\pi(y)}.$$

We now apply Theorem 2.2 with an exponent p to be specified later. We obtain

$$\sum_{y \in S} \pi(y) \left(\frac{p_e^*(x, y)\pi(x)}{\pi(y)} \right)^p \leq \left(\frac{p}{p-1} \right)^p \cdot \pi(x), \tag{2.1}$$

Let $q = p/(p-1)$ be the conjugate exponent of p , so that $1/p + 1/q = 1$. By Hölder’s inequality,

$$\begin{aligned} \sum_{y \in S} p_e^*(x, y) &= \sum_{y \in S} \pi(y)^{1/q} \frac{p_e^*(x, y)}{\pi(y)^{1/q}} \leq \left(\sum_{y \in S} \pi(y) \right)^{1/q} \left(\sum_{y \in S} \left(\frac{p_e^*(x, y)}{\pi(y)^{1/q}} \right)^p \right)^{1/p} \\ &= \frac{1}{\pi(x)} \left(\sum_{y \in S} \pi(y) \left(\frac{p_e^*(x, y)\pi(x)}{\pi(y)} \right)^p \right)^{1/p} \leq \frac{p}{p-1} \cdot \pi(x)^{1/p-1} = q\pi(x)^{-1/q}, \end{aligned}$$

where the second inequality comes from (2.1). This is valid for all $1 < q < \infty$, and by continuity, also for $q = 1$. Setting $q = \max(1, \log 1/(\pi(x)))$, this yields

$$\sum_{y \in S} p_e^*(x, y) \leq e \cdot \max \left(1, \log \frac{1}{\pi(x)} \right).$$

We also have

$$\begin{aligned} \sum_{y \in S} p_o^*(x, y) &\leq \sum_{y \in S} \sum_{z \in S} p_e^*(x, z) p(z, y) = \sum_{z \in S} p_e^*(x, z) \sum_{y \in S} p(z, y) \\ &= \sum_{z \in S} p_e^*(x, z) \leq e \cdot \max \left(1, \log \frac{1}{\pi(x)} \right), \end{aligned}$$

and so

$$\sum_{y \in S} p^*(x, y) \leq \sum_{y \in S} p_e^*(x, y) + \sum_{y \in S} p_o^*(x, y) \leq 2e \cdot \max \left(1, \log \frac{1}{\pi(x)} \right),$$

which proves the lemma. □

We mention an example that shows Proposition 1.4 is tight up to a multiplicative constant. Consider a cycle of even size $n = 2m$, with vertices labelled by elements of \mathbb{Z}/n in the natural way. Let X be the simple random walk started at 0. Then, for each time $t \geq 0$ and each position $-m < k \leq m - 1$, the probability $p^t(0, k)$ that X is at k is at least the probability $\tilde{p}^t(0, k)$ that a

simple random walk on \mathbb{Z} started at 0 is at k after time t . By the local central limit theorem (see e.g. [6], Theorem 1.2.1), we have for some universal constant $C > 0$ that

$$p^t(0, k) \geq \tilde{p}^t(0, k) \geq \frac{2}{\sqrt{2\pi t}} \exp\left(-\frac{k^2}{2t}\right) - \frac{C}{\sqrt{t} \cdot k^2}.$$

Plugging in $t = k^2$ (which approximately maximizes the right-hand side), we obtain

$$p^*(0, k) \geq \frac{2}{\sqrt{2\pi e}} \cdot \frac{1}{|k|} - \frac{C}{k^3}.$$

Thus

$$\begin{aligned} \sum_{k \in \mathbb{Z}/n} p^*(0, k) &\geq \sum_{k=1}^{m-1} (p^*(0, k) + p^*(0, -k)) \\ &\geq \frac{4}{\sqrt{2\pi e}} \sum_{k=1}^{m-1} \frac{1}{k} - 2C \sum_{k=1}^{m-1} \frac{1}{k^3} \geq \frac{4}{\sqrt{2\pi e}} \log n - C' \end{aligned}$$

for a universal constant $C' > 0$. On the other hand, the stationary probabilities are all $1/n$, so Proposition 1.4 gives the bound

$$\sum_{k \in \mathbb{Z}/n} p^*(0, k) \leq 2e \cdot \log n.$$

Thus, Proposition 1.4 is tight up to a constant factor for random walk on the cycle.

Remark. It would be interesting to determine whether the above cycle example is the ‘worst case’ for Proposition 1.4 applied to regular graphs. More precisely, is it true that for any $\varepsilon > 0$ and sufficiently large n , for a random walk on an n -vertex graph we have

$$\sum_{y \in S} p^*(x, y) \leq \left(\frac{4}{\sqrt{2\pi e}} + \varepsilon \right) \log n ?$$

2.2. Lower bound for Proposition 1.1

The following construction shows that Proposition 1.1 is optimal up to constant factor.

Claim 2.1. For any $n \geq 2$ and $t \geq 2n$, there exists a Markov chain with state space S of size n and states $x, y \in S$ such that

$$\mathbb{P}_x(\tau(y) = t) \geq \frac{n}{8t}.$$

Proof. The following construction is due to Kozma and Zeitouni (private communication). Write $t = r(n - 1) + k$, where $1 \leq k \leq n - 1$ and $r \geq 2$. Consider the Markov chain with states s_1, s_2, \dots, s_{n-1} , and one additional state u (see Figure 1). We take the transition probabilities to be $p(s_i, s_{i+1}) = 1$ for each $1 \leq i \leq n - 2$, and

$$p(s_{n-1}, s_1) = 1 - q, \quad p(s_{n-1}, u) = q, \quad p(u, s_2) = 1,$$

where $q = 1/r$. Note that this chain is periodic with period $n - 1$.

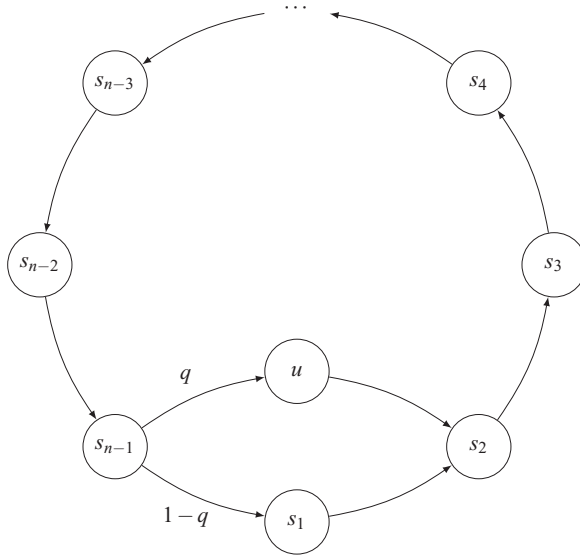


Figure 1. Illustration of Claim 2.1.

We consider the hitting time from s_{n-k} to u . Note that in the first k steps the process deterministically goes to s_{n-1} and then goes to either u or s_1 with probabilities q and $1 - q$, respectively. Thereafter, every $n - 1$ steps it again goes to either u or s_1 with those probabilities. Thus

$$\mathbb{P}_{s_{n-k}}(\tau(u) = t) = \left(1 - \frac{1}{r}\right)^r \cdot \frac{1}{r} \geq \frac{1}{4r} = \frac{n-1}{4t-4k} \geq \frac{n}{8t},$$

completing the construction for $x = s_{n-k}$ and $y = u$. □

2.3. Lower bound for Theorem 1.2

The next construction is a slight modification of a construction of Peres and Sousi [9], who gave it as an example of cutoff in the mixing time of trees. The modified construction yields the lower bound (1.1) mentioned in the Introduction.

Claim 2.2. *There exist simple graphs of n vertices, for arbitrarily large values of n , such that for the random walk started at a vertex x , there is another vertex y and a time t for which*

$$\mathbb{P}_x(\tau(y) = t) \geq \frac{c\sqrt{\log n}}{t},$$

where $c > 0$ is a universal constant.

Remark. In fact, the graphs we construct have maximum degree 4 (corresponding to the vertices w_k defined below). By adding self-loops, we can easily modify these examples to be regular graphs.

For any integer $m > 1$, consider the graph G_m formed by a path of length $2^m - 1$ with vertices denoted by v_1, v_2, \dots, v_{2^m} , together with $m - 1$ attached binary trees. In particular, for each $1 \leq$

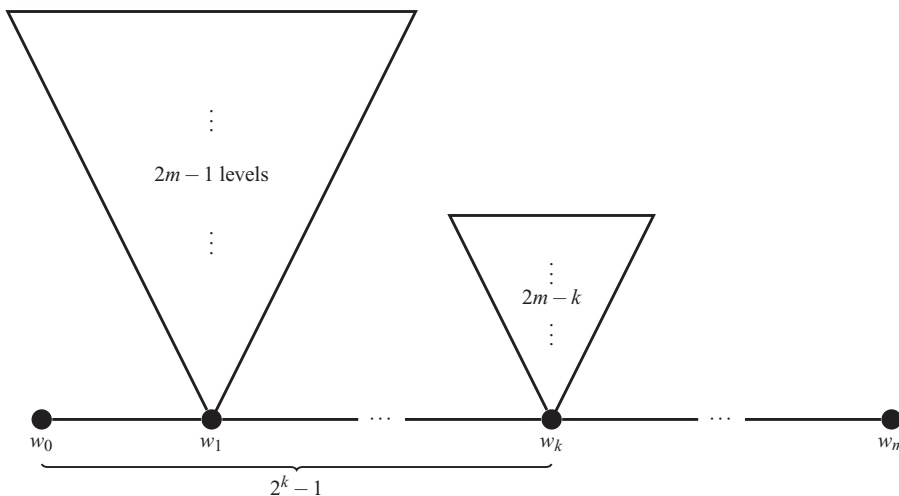


Figure 2. Illustration of G_m .

$k \leq m - 1$, we attach a binary tree of height $2m - k$ rooted at the vertex $w_k := v_{2^k}$ (see Figure 2). The analysis of $\mathbb{P}_x(\tau(y) = t)$ for random walk on G_m hinges on the following concentration of hitting times result similar to Lemma 2.1 of [9].

Lemma 2.3. *For the simple random walk on G_m started at w_m , we have*

$$\mathbb{E}_{w_m} \tau(w_0) = \Theta(m \cdot 2^{2m}) \quad \text{and} \quad \text{Var}_{w_m}(\tau(w_0)) = O(m \cdot 2^{4m}).$$

2.3.1. Proof of Lemma 2.3. First, we recall a standard estimate on trees.

Lemma 2.4 ([9], Claim 2.3). *Let Y_k denote the return time of a random walk on a binary tree of height k started at the root. Then,*

$$\mathbb{E}Y_k = \Theta(2^k) \quad \text{and} \quad \mathbb{E}Y_k^2 = \Theta(2^{2k}).$$

Let \mathcal{T}_k denote the tree (of height $2m - k$) attached to w_k (excluding w_k itself), and let $\mathcal{F} = \bigcup_{k=1}^m \mathcal{T}_k$. For $1 \leq k \leq m$, consider the walk started at w_k and stopped upon hitting w_{k-1} . Define Z_k to be the number of steps either starting or ending in \mathcal{F} . Note that Z_m is deterministically zero. We have the following moment bounds on Z_k .

Lemma 2.5. *For $1 \leq k \leq m - 1$, we have*

$$\mathbb{E}Z_k = \Theta(2^{2m}) \quad \text{and} \quad \mathbb{E}Z_k^2 = \Theta(2^{4m}).$$

Proof. The path of the random walk started at w_k and stopped upon hitting w_{k-1} may be decomposed into components of the following types:

- (i) excursions from w_k hitting w_{k+1} and not hitting \mathcal{T}_k ,
- (ii) excursions from w_k staying entirely within \mathcal{T}_k ,
- (iii) excursions from w_k not hitting $\{w_{k-1}, w_{k+1}\} \cup \mathcal{T}_k$,
- (iv) a path from w_k to w_{k-1} .

Note that components of types (iii) and (iv) do not contribute to the time spent in \mathcal{F} . Components of type (i) can only spend time in \mathcal{F} after first hitting w_{k+1} , so the time they spend in \mathcal{F} is distributed as Z_{k+1} . The time spent in \mathcal{F} by a component of type (ii) is like an excursion from the root for \mathcal{T}_k , so it is distributed as $Y_{2^{m-k}}$. Thus, the law of Z_k is given by the sum of (a random number of) independent copies of Z_{k+1} and $Y_{2^{m-k}}$.

To compute the distribution of how many components of types (i) and (ii) there are, let τ_k be the first time that the random walk started at w_k hits $\{w_{k-1}, w_{k+1}\} \cup \mathcal{T}_k$. By a standard calculation,

$$\begin{aligned} \mathbb{P}_{w_k}(X_{\tau_k} = w_{k-1}) &= \frac{2^{-k}}{2^{-k} + 2^{-k-1} + 1}, \\ \mathbb{P}_{w_k}(X_{\tau_k} = w_{k+1}) &= \frac{2^{-k-1}}{2^{-k} + 2^{-k-1} + 1}, \\ \mathbb{P}_{w_k}(X_{\tau_k} \in \mathcal{T}_k) &= \frac{1}{2^{-k} + 2^{-k-1} + 1}. \end{aligned}$$

It follows that the number K of components of type (i) or (ii) is geometrically distributed with mean

$$\left(\frac{2^{-k}}{2^{-k} + 2^{-k-1} + 1} \right)^{-1} = 2^k + 1 + \frac{1}{2}.$$

Let W_k be a random variable drawn according to Z_{k+1} with probability $1/(2^{k+1} + 1)$ and according to $Y_{2^{m-k}}$ with probability $2^{k+1}/(2^{k+1} + 1)$. Then, conditioned on a component being of type (i) or (ii), the time it spends in T is distributed as W_k . Thus, Z_k is a sum of K i.i.d. copies of W_k . We now give recursive estimates of the first and second moments of Z_k . For the first moment, we have

$$\mathbb{E}Z_k = (\mathbb{E}K) \cdot (\mathbb{E}W_k) = (\mathbb{E}K) \cdot \left(\frac{2^{k+1}}{2^{k+1} + 1} \mathbb{E}Y_{2^{m-k}} + \frac{1}{2^{k+1} + 1} \mathbb{E}Z_{k+1} \right).$$

Applying Lemma 2.4 to estimate $\mathbb{E}Y_{2^{m-k}}$ and using our explicit characterization of K , this gives

$$c \cdot 2^{2m} \leq \mathbb{E}Z_k \leq C \cdot 2^{2m} + \frac{7}{10} \cdot \mathbb{E}Z_{k+1}$$

for some universal constants c and C . Recall that $Z_m = 0$, so this recursive bound implies that $\mathbb{E}Z_k = \Theta(2^{2m})$ for all $1 \leq k \leq m - 1$.

We now bound the second moment of Z_k . Note that our first moment bound immediately implies that $\mathbb{E}W_k = \Theta(2^{2m-k})$. We have

$$\begin{aligned} \mathbb{E}Z_k^2 &= (\mathbb{E}K) \cdot (\mathbb{E}W_k^2) + (\mathbb{E}K(K - 1)) \cdot (\mathbb{E}W_k)^2 \\ &= (\mathbb{E}K) \cdot \left(\frac{2^{k+1}}{2^{k+1} + 1} \mathbb{E}Y_{2^{m-k}}^2 + \frac{1}{2^{k+1} + 1} \mathbb{E}Z_{k+1}^2 \right) + O(2^{4m}) \\ &\leq \frac{7}{10} \cdot \mathbb{E}Z_{k+1}^2 + O(2^{4m}), \end{aligned}$$

where we use the upper bound on $\mathbb{E}Y_{2^{m-k}}^2$ from Lemma 2.4. By the same argument as was used for the first moment, this implies $\mathbb{E}Z_k^2 = O(2^{4m})$ for $1 \leq k \leq m - 1$. Note that we also have $\mathbb{E}Z_k^2 \geq (\mathbb{E}Z_k)^2$, so we have

$$\mathbb{E}Z_k = \Theta(2^{2m}) \quad \text{and} \quad \mathbb{E}Z_k^2 = \Theta(2^{4m}).$$

as desired. □

We are now ready to prove Lemma 2.3.

Proof of Lemma 2.3. Note that a random walk started at w_m and stopped upon hitting w_0 can be decomposed into independent walks started at w_k and stopped upon hitting w_{k-1} for $k = m, m - 1, \dots, 1$. Let $Z = Z_1 + Z_2 + \dots + Z_m$ be the total number of steps starting or ending in \mathcal{F} , and let Z' be the total number of steps completely outside of \mathcal{F} , so that $\tau := Z + Z'$ is the hitting time from w_m to w_0 .

Note that Z' has the distribution of the hitting time of a random walk on a path of length $2^m - 1$ from one end to the other. The following estimates are standard (see e.g. [3], III.7):

$$\mathbb{E}Z' = \Theta(2^{2m}) \quad \text{and} \quad \mathbb{E}Z'^2 = O(2^{4m}).$$

It follows from Lemma 2.5 that

$$\mathbb{E}\tau = \mathbb{E}Z + \mathbb{E}Z' = \Theta(m \cdot 2^{2m}),$$

which proves the first part of the lemma. We also have

$$\begin{aligned} \text{Var}(Z) &= \sum_{k=1}^m \text{Var}(Z_k) = O(m \cdot 2^{4m}), \\ \mathbb{E}Z^2 &= (\mathbb{E}Z)^2 + \text{Var}(Z) = O(m^2 \cdot 2^{4m}). \end{aligned}$$

Thus

$$\begin{aligned} \text{Var}(\tau) &= 2(\mathbb{E}ZZ' - \mathbb{E}Z\mathbb{E}Z') + \text{Var}(Z) + \text{Var}(Z') \\ &\leq 2\sqrt{(\mathbb{E}Z^2)(\mathbb{E}Z'^2)} + \text{Var}(Z) + \text{Var}(Z') = O(m \cdot 2^{4m}), \end{aligned}$$

which is the second statement in the lemma, completing the proof. □

2.3.2. Proof of Claim 2.2. We will show using Lemma 2.3 that G_m satisfies the criteria of Claim 2.2.

Proof of Claim 2.2. For any $m \geq 2$, take

$$n = 2^{2m} - 2^m - 2m + 2$$

to be the number of vertices of G_m . For convenience, write $T = \mathbb{E}_{w_m} \tau(w_0)$. By Lemma 2.3 and Chebyshev's inequality, there is a universal constant C such that

$$\mathbb{P}_{w_m}(|\tau(w_0) - T| \geq C\sqrt{m \cdot 2^{4m}}) \leq \frac{1}{2}.$$

It follows by the pigeonhole principle that for some t with

$$T - C\sqrt{m \cdot 2^{4m}} \leq t \leq T + C\sqrt{m \cdot 2^{4m}} \tag{2.2}$$

we have

$$\mathbb{P}_{w_m}(\tau(w_0) = t) \geq \frac{1}{4C\sqrt{m \cdot 2^{2m}}}.$$

Moreover, note that T is of the order of $m \cdot 2^{2m}$ by Lemma 2.3, so for sufficiently large m , (2.2) implies that $t \geq cm \cdot 2^{2m}$ for a sufficiently small constant $c > 0$. Then,

$$\mathbb{P}_{w_m}(\tau(w_0) = t) \geq \frac{1}{4C\sqrt{m \cdot 2^{2m}}} \geq \frac{c\sqrt{m}}{4Ct} \geq \frac{c}{4\sqrt{2} \cdot C} \cdot \frac{\sqrt{\log n}}{t}.$$

This is the desired bound upon renaming of constants, and m (hence n) can be made arbitrarily large. □

3. Proof of Theorem 1.3

We now turn to Theorem 1.3, which is a ‘worst case’ bound for reversible chains that does not depend on the stationary probability. First, we record several facts about sums of geometric random variables. We say that a random variable Z is *geometrically distributed with parameter* p if $\mathbb{P}(Z = t) = p(1 - p)^{t-1}$ for $t = 1, 2, 3, \dots$

Lemma 3.1. *Let $m > n$ be an integer, and let X_1, X_2, \dots, X_n be i.i.d. geometric random variables with parameter $p = n/m$. Then*

$$\frac{1}{3} \sqrt{\frac{n}{m(m-n)}} \leq \mathbb{P}(X_1 + X_2 + \dots + X_n = m) \leq \frac{1}{2} \sqrt{\frac{n}{m(m-n)}}.$$

Lemma 3.2. *Let X_1, X_2, \dots, X_n be independent geometric random variables, not necessarily identically distributed. Then, for any $t > n$,*

$$\mathbb{P}(X_1 + X_2 + \dots + X_n = t) \leq \frac{1}{2} \sqrt{\frac{n}{t(t-n)}}. \tag{3.1}$$

Remark. In fact, $\mathbb{P}(X_1 + X_2 + \dots + X_n = t)$ is maximized when each X_i has parameter n/t . Lemma 3.2 then follows from Lemma 3.1.

Corollary 3.3. *If $(X_i)_{i=1}^n$ are independent mixtures of geometric random variables, then (3.1) holds.*

Proof. For each i , let θ_i be a random variable on $[0, 1]$ so that X_i is distributed as a geometric with parameter θ_i . By Lemma 3.2, we have

$$\mathbb{P}(X_1 + X_2 + \dots + X_n = t \mid \theta_1, \dots, \theta_n) \leq \frac{1}{2} \sqrt{\frac{n}{t(t-n)}}.$$

Taking the expectation over the θ_i gives the result. □

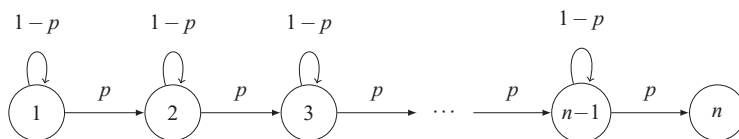


Figure 3. Illustration of Claim 3.1.

Lemma 3.1 is a straightforward calculation using Stirling’s approximation. Lemma 3.2 is a non-trivial bound due to Nazarov (private communication). The proofs of these lemmas are deferred to Sections A.1 and A.2.

Before proving Theorem 1.3, it is instructive to construct the example that attains the lower bound.

Claim 3.1. *For any $n \geq 2$ and $t > n$, there exists a reversible Markov chain X on n states with two states x and y such that*

$$\mathbb{P}_x(\tau(y) = t) \geq \frac{\sqrt{n}}{5t}$$

(see Figure 3).

Proof. Let $p = n/t$. Consider the case where X is a pure-birth chain with states labelled $1, 2, \dots, n$, with transition probabilities

$$p(i, i) = 1 - p, \quad p(i, i + 1) = p \text{ for } 1 \leq i \leq n - 1, \quad p(n, n) = 1.$$

Suppose that X starts at 1, and let T_i be the first hitting time of state i , with $T_1 = 0$. Let $D_i = T_{i+1} - T_i$, so that

$$T_n = \sum_{i=1}^{n-1} D_i.$$

Note that each D_i is a geometric random variable with parameter p , and the D_i are independent. Thus, applying Lemma 3.1 with $m = t$, we have

$$\mathbb{P}(T_n = t) \geq \frac{1}{3} \sqrt{\frac{n-1}{t(t-n+1)}} \geq \frac{\sqrt{n}}{5t},$$

as desired. □

In fact, the above example is in some sense the extremal case. In a pure-birth chain, we saw that the hitting time from the beginning to the end is a sum of geometric random variables. Lemma 3.2 implies that the sum of n geometric random variables cannot be too concentrated, which essentially proves Theorem 1.3 for pure-birth chains.

Then, we will show that the behaviour of the hitting time in a general reversible Markov chain is like a mixture of pure-birth chains. Such a representation was shown by Miclo [8]; in this paper, we give a short elementary proof based on loop erasures of first hitting paths from x to y . By this mixture argument, it follows that the bound for pure-birth chains carries over to the

general case, although we lose approximately a factor of 2 due to the possibility of negative eigenvalues.

The next lemma is a variant of the well-known spectral decomposition of return probabilities in reversible Markov chains. The proof is essentially the same as that of Lemma 2.2 in [4], although our formulation includes an additional non-negativity statement. For the sake of completeness, we include a proof of this lemma in Section A.3.

Lemma 3.4. *Let X be a reversible, irreducible Markov chain with finite state space S . Consider any $x \in S$, and let $U \subset S$ be a subset not containing x . Then, there exist real numbers $a_1, \dots, a_{|S|} \geq 0$ and $\lambda_1, \dots, \lambda_{|S|} \in [-1, 1]$ such that*

$$\mathbb{P}_x(X_t = x, \tau(U) > t) = \sum_{i=1}^{|S|} a_i \lambda_i^t,$$

for each integer $t \geq 0$. Moreover, if the eigenvalues of X (i.e. the eigenvalues of the matrix of transition probabilities) are non-negative, then we may take $\lambda_i \in [0, 1]$ for each i .

Let us now prove Theorem 1.3.

Proof of Theorem 1.3. We will first prove the stronger bound

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{2} \sqrt{\frac{n}{t(t-n)}} \tag{3.2}$$

in the case where all the eigenvalues are non-negative. We then deduce the theorem by considering even times. It is convenient to assume throughout that X is irreducible, so Lemma 3.4 applies. This is valid because we may always restrict to the communicating class of x without increasing the number of states.

Let $Z = (x = z_1, z_2, \dots, z_k = y)$ be a random variable denoting the path taken by the chain started at x and stopped upon hitting y . We let $|Z|$ denote the length of Z , defined as the number of vertices in Z . Note that $|Z| = \tau_x(y) + 1$, and we are interested in bounding $\mathbb{P}(|Z| = t + 1)$.

Define the loop erasure of Z to be the path $(w_1, w_2, \dots, w_\ell)$ determined as follows. We take $w_1 = x$, and inductively for each $i \geq 1$, let k_i be the largest index such that $z_{k_i} \in \{w_1, \dots, w_i\}$. Then, as long as $k_i < k$, define $w_{i+1} = z_{k_i+1}$ and continue the process for $i + 1$. If $k_i = k$, then the path ends. In particular, $k_\ell = k$ and $w_\ell = z_k = y$.

We denote this path, which is a function of Z , by $W(Z)$. A less formal description of the loop erasure is that $W(Z)$ is the path obtained by following Z and, whenever the path forms a loop, removing all vertices from that loop. Loop-erased walks appear in other contexts, including random walks on lattices and uniform sampling of spanning trees (see e.g. [5, 11]).

Fix a loop erasure $W = (x = w_1, w_2, \dots, w_\ell = y)$, and let \mathbb{P}_W denote the conditional probability $\mathbb{P}(\cdot \mid W(Z) = W)$. We now describe a method of sampling from \mathbb{P}_W . For each $1 \leq i < \ell$, let \mathcal{P}_i denote the set of all paths starting and ending at w_i (possibly of length 1) and avoiding $\{y, w_1, w_2, \dots, w_{i-1}\}$. For each i , we independently sample a path $\tilde{Z}_i = (\tilde{z}_{i,1}, \tilde{z}_{i,2}, \dots, \tilde{z}_{i,k})$ from \mathcal{P}_i

with probability

$$c_i \cdot \prod_{j=1}^{k-1} p(\tilde{z}_{i,j}, \tilde{z}_{i,j+1}),$$

where c_i is the normalizing constant which makes all the probabilities sum to 1. We obtain a sample Z from \mathbb{P}_W by taking the concatenation $Z = [\tilde{Z}_1][\tilde{Z}_2] \cdots [\tilde{Z}_{\ell-1}]y$.

Now, observe for each $t \geq 1$ that

$$\begin{aligned} \mathbb{P}(|\tilde{Z}_i| = t) &= \sum_{(v_1, v_2, \dots, v_t) \in \mathcal{P}_i} c_i \cdot \prod_{j=1}^{t-1} p(v_j, v_{j+1}) \\ &= c_i \cdot \mathbb{P}_{w_i}(X_t = w_i, \tau(\{y, w_1, w_2, \dots, w_{i-1}\}) > t - 1), \end{aligned}$$

where the summation includes only those paths in \mathcal{P}_i having length t , and \mathbb{P}_{w_i} denotes probability of the random walk started at w_i . By Lemma 3.4 applied to the last expression, we may write

$$\mathbb{P}(|\tilde{Z}_i| = t) = c_i \sum_{j=1}^n a_j \lambda_j^{t-1},$$

which shows that the probability distribution function of $|\tilde{Z}_i|$ is a non-negative linear combination of the probability distribution functions of geometrics with parameters $(1 - \lambda_1), (1 - \lambda_2), \dots, (1 - \lambda_n)$. Moreover, since probability distribution functions sum to 1, we deduce that this is in fact a convex combination, so $|\tilde{Z}_i|$ is distributed as a mixture of geometric random variables.

We have that

$$|Z| = 1 + |\tilde{Z}_1| + |\tilde{Z}_2| + \cdots + |\tilde{Z}_{\ell-1}|.$$

Hence, by Corollary 3.3, we have

$$\mathbb{P}_W(|Z| - 1 = t) \leq \frac{1}{2} \sqrt{\frac{\ell - 1}{t(t - \ell + 1)}}.$$

Recall that W is a loop erasure and therefore the w_i are all distinct, which implies $\ell \leq n$. Thus, the above inequality implies

$$\mathbb{P}_W(|Z| = t + 1) \leq \frac{1}{2} \sqrt{\frac{\ell - 1}{t(t - \ell + 1)}} \leq \frac{1}{2} \sqrt{\frac{n}{t(t - n)}}.$$

This bound holds for each W , so taking the expectation over all possible loop erasures W , we obtain

$$\mathbb{P}_x(\tau(y) = t) = \mathbb{P}(|Z| = t + 1) \leq \frac{1}{2} \sqrt{\frac{n}{t(t - n)}}.$$

We have thus proved (3.2) when the chain has non-negative eigenvalues. Let us now consider the general case, where the eigenvalues can be negative. Let $Y_t = X_{2t}$, so that Y is also a Markov chain (with transition matrix the square of the transition matrix of X). Note that Y has non-negative eigenvalues. Let $\tau_{x,X}(y)$ denote the hitting time from x to y under the chain X , and similarly let $\tau_{x,Y}(y)$ be the hitting time from x to y under Y . If $\tau_{x,X}(y) = 2k$, then we necessarily

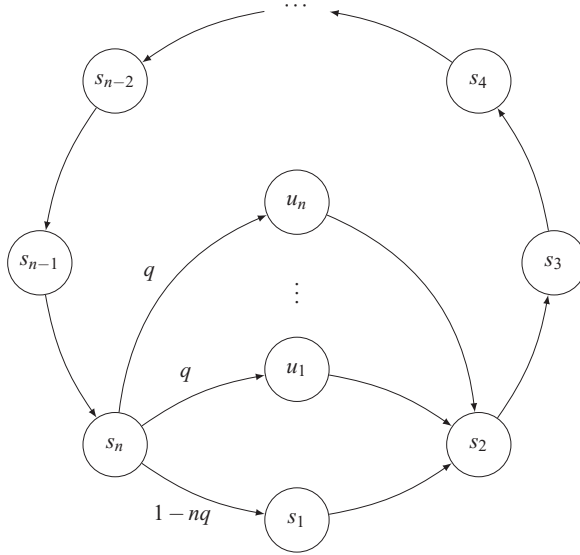


Figure 4. Illustration of Claim 4.1.

have $\tau_{x,Y}(y) = k$. Hence, if $t = 2k$, we immediately have

$$\begin{aligned} \mathbb{P}(\tau_{x,X}(y) = t) &\leq \mathbb{P}(\tau_{x,Y}(y) = k) \\ &\leq \frac{1}{2} \sqrt{\frac{n}{(k-n)k}} < \frac{\sqrt{2n}}{t}, \end{aligned}$$

where we used the assumption $t \geq 4n + 4$. If $t = 2k + 1$, then we have

$$\begin{aligned} \mathbb{P}(\tau_{x,X}(y) = t) &= \sum_{s \in S, s \neq y} p(x, s) \mathbb{P}(\tau_{s,X}(y) = t - 1) \\ &\leq \sum_{s \in S, s \neq y} p(x, s) \cdot \frac{1}{2} \sqrt{\frac{n}{(k-n)k}} \leq \frac{1}{2} \sqrt{\frac{n}{(k-n)k}} \leq \frac{\sqrt{2n}}{t}, \end{aligned}$$

completing the proof. □

4. Lower bound constructions for $\mathbb{P}_x(\mathcal{S}_t)$

We now describe how to modify the constructions in Claims 2.1, 2.2 and 3.1 to give corresponding lower bounds for $\mathbb{P}_x(\mathcal{S}_t)$.

First, recall the cycle construction in the proof of Claim 2.1. Using a similar construction but with n copies of the state u (see Figure 4), we can obtain the following lower bound for $\mathbb{P}_x(\mathcal{S}_t)$.

Claim 4.1. *For any $n \geq 2$ and $t \geq 2n$, there exists a Markov chain with $2n$ states and a starting state x such that*

$$\mathbb{P}_x(\mathcal{S}_t) \geq \frac{n^2}{56t}.$$

Proof. Write $t = rn + k$, where $1 \leq k \leq n$ and $r \geq 2$. Define $q = 1/r$. Consider a Markov chain with states

$$\{s_1, s_2, \dots, s_n\} \cup \{u_1, u_2, \dots, u_n\}$$

whose transition probabilities are

$$\begin{aligned} p(s_i, s_{i+1}) &= 1, && \text{for each } 1 \leq i \leq n-1, \\ p(s_n, s_1) &= 1 - nq, \\ p(s_n, u_i) &= q, \quad p(u_i, s_2) = 1, && \text{for each } 1 \leq i \leq n. \end{aligned}$$

This is very similar to the example in the proof of Claim 2.1, except that u is replaced by many states u_1, u_2, \dots, u_n . Note that the chain is periodic with period n .

We consider the process started at $x := s_{n-k+1}$, so that at time $t - 1$, the process is in state s_n . After $t - 1$ steps, the process has taken r steps starting in s_n , so the probability that a given state u_j is still unvisited is $(1 - q)^r$. Thus, letting Z be the number of states in $\{u_1, \dots, u_n\}$ which are visited by time $t - 1$, we have

$$\mathbb{E}Z = n \cdot (1 - (1 - q)^r) = n \cdot \left(1 - \left(1 - \frac{1}{r}\right)^r\right) \leq \frac{3}{4} \cdot n,$$

and so Markov's inequality implies

$$\mathbb{P}_x\left(Z \geq \frac{7n}{8}\right) \leq \frac{6}{7}.$$

For the event that $Z < (7n)/8$, there are at least $n/8$ states among $\{u_1, \dots, u_n\}$ which have not been visited, so the probability that the chain hits a new state on the t th step is at least $(nq)/8$. Thus

$$\mathbb{P}_x(\mathcal{S}_t) \geq \mathbb{P}_x\left(\mathcal{S}_t \mid Z < \frac{7n}{8}\right) \cdot \mathbb{P}_x\left(Z < \frac{7n}{8}\right) \geq \frac{nq}{56} \geq \frac{n^2}{56t},$$

as desired. □

We can also modify the constructions corresponding to Claims 2.2 and 3.1, to obtain the following.

Claim 4.2. *There exist simple graphs of n vertices, for arbitrarily large values of n , such that for the random walk started at a vertex x , there is a time t for which*

$$\mathbb{P}_x(\mathcal{S}_t) \geq \frac{cn\sqrt{\log n}}{t},$$

where $c > 0$ is a universal constant.

Claim 4.3. *For any n , there exists a reversible Markov chain X on $2n$ states with two states x and y such that*

$$\mathbb{P}_x(\mathcal{S}_t) = \Omega\left(\frac{n\sqrt{n}}{t}\right).$$

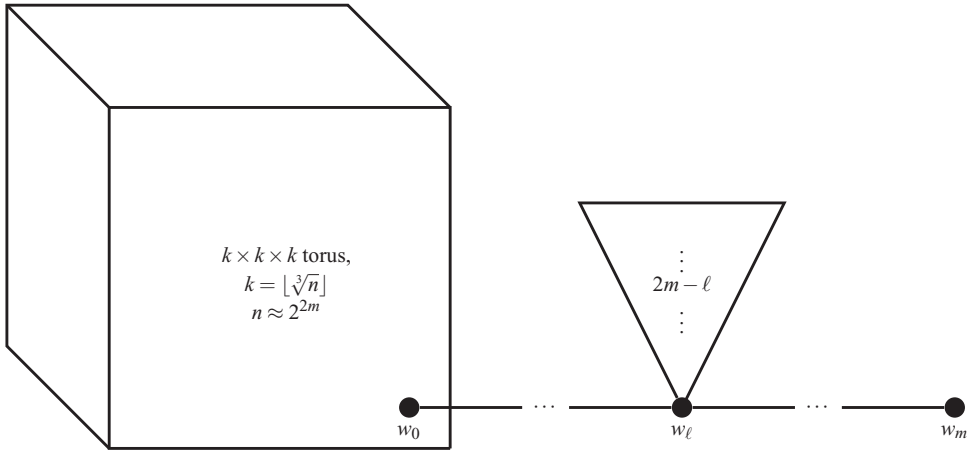


Figure 5. Illustration of \tilde{G}_m .

The constructions we give for the above claims are both based on a general lemma that translates certain lower bounds on $\mathbb{P}_x(\tau(y) = t)$ into lower bounds on $\mathbb{P}_x(\mathcal{S}_t)$.

Lemma 4.1. *Consider a Markov chain with state space S with states $x, y \in S$ and a subset $U \subset S$ such that starting from x , the chain cannot reach U without going through y . Let Z_s denote the number of visited states in U at time s . Then, for any integer $N > 0$, there exists s with $t \leq s < t + 2N$ such that*

$$\mathbb{P}_x(\mathcal{S}_s) \geq \frac{\mathbb{P}_x(t \leq \tau(y) < t + N) \cdot \mathbb{E}_y Z_N}{2N}.$$

Proof. Note that no state in U can be visited before $\tau(y)$. We lower-bound $\mathbb{E}_x(Z_{t+2N-1} - Z_t)$ by only considering the event that $t \leq \tau(y) < t + N$. We have

$$\mathbb{E}_x(Z_{t+2N-1} - Z_t) \geq \sum_{k=0}^{N-1} \mathbb{P}_x(\tau(y) = t + k) \mathbb{E}_y Z_N = \mathbb{P}_x(t \leq \tau(y) < t + N) \cdot \mathbb{E}_y Z_N.$$

This lower-bounds the expected number of new states visited in the time interval $[t, t + 2N)$, so it follows by the pigeonhole principle that for some $t \leq s < t + 2N$ we have

$$\mathbb{P}_x(\mathcal{S}_s) \geq \frac{\mathbb{P}_x(t \leq \tau(y) < t + N) \cdot \mathbb{E}_y Z_N}{2N}. \quad \square$$

We now prove the claims.

Proof of Claim 4.2. Let G_m be as in the proof of Claim 2.2, and let n be the size of G_m . Let $k = \lfloor \sqrt[3]{n} \rfloor$, and let H_k denote the three-dimensional discrete torus of size k^3 , whose vertex set is $(\mathbb{Z}/k\mathbb{Z})^3$, with edges between nearest neighbours.

Define a graph \tilde{G}_m by attaching to G_m a copy of H_k so that $(0, 0, 0)$ in H_k is joined to w_0 in G_m (see Figure 5). We recall the standard fact that the effective resistance between any two vertices

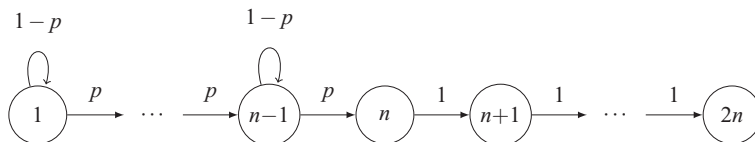


Figure 6. Illustration of Claim 4.3.

y and z in H_k is bounded by a universal constant [7, Exercise 9.1]. This also applies when y and z are considered as vertices in \tilde{G}_m , by Rayleigh’s monotonicity law.

For any $z \in H_k$ (with H_k regarded as a subgraph of \tilde{G}_m), we may apply the commute time identity [7, Proposition 10.6] to w_0 and z , which gives

$$\begin{aligned} \mathbb{E}_{w_0} \tau(z) &\leq \mathbb{E}_{w_0} \tau(z) + \mathbb{E}_z \tau(w_0) \\ &= O(1) \cdot \#\{\text{edges in } \tilde{G}_m\} = O(k^3 + n) \leq Cn, \end{aligned}$$

for a sufficiently large constant C .

Consider a random walk started at w_0 and run for $n' := \lceil 2Cn \rceil$ steps. By Markov’s inequality, each vertex in H_k is visited with probability at least $1/2$, and so the expected number of visited vertices in H_k is at least $\frac{1}{2}k^3$.

Recall also from Lemma 2.3 that $\mathbb{E}_{w_m} \tau(w_0)$ has expectation of the order of $m \cdot 2^{2m} = \Theta(n \log n)$ with fluctuations of the order of $\sqrt{m} \cdot 2^{2m} = \Theta(n\sqrt{\log n})$. It follows that for some t with $t = \Theta(n \log n)$,

$$\mathbb{P}_{w_m}(t \leq \tau(w_0) < t + n') = \Omega\left(\frac{1}{\sqrt{\log n}}\right).$$

Thus, we may apply Lemma 4.1 with U being the vertex set of H_k , $x = w_m$, $y = w_0$, and $N = n'$. We find that for some $s = \Theta(n \log n)$,

$$\begin{aligned} \mathbb{P}_{w_m}(\mathcal{S}_s) &\geq \frac{\mathbb{P}_{w_m}(t \leq \tau(w_0) < t + n') \cdot \mathbb{E}_{w_0} Z_{n'}}{2n} \\ &= \Omega\left(\frac{1}{\sqrt{\log n}}\right) \cdot \left(\frac{1}{2}k^3\right) \cdot \frac{1}{2n} = \Omega\left(\frac{n\sqrt{\log n}}{s}\right), \end{aligned}$$

as desired. □

Proof of Claim 4.3. It suffices to prove the bound for $t \geq n\sqrt{n}$, so we assume this in what follows.

Consider the pure-birth chain from the proof of Claim 3.1 (with states labelled $1, 2, \dots, n$), and introduce n additional states labelled $n + 1, \dots, 2n$. We use the same transition probabilities as in the proof of Claim 3.1, with the modification that $p(i, i + 1) = 1$ for $i = n, n + 1, \dots, 2n - 1$ (see Figure 6).

Recall from the proof of Claim 3.1 that the hitting time τ from 1 to n is distributed as

$$\sum_{i=1}^{n-1} D_i,$$

where the D_i are independent geometrics of parameter $p = n/t$. We can thus calculate

$$\mathbb{E}\tau = \Theta(t) \quad \text{and} \quad \text{Var}(\tau) = O\left(\frac{t^2}{n}\right).$$

It follows that for some $t' = \Theta(t)$,

$$\mathbb{P}(t' \leq \tau < t' + n) = \Omega\left(\frac{n\sqrt{n}}{t}\right).$$

We can apply Lemma 4.1 with $U = \{n, n + 1, \dots, 2n\}$, $x = 1$, $y = n$, and $N = n$. Note that if the chain is started at y , then $Z_n = n$ deterministically in this case. Thus, we obtain for some $s = \Theta(t') = \Theta(t)$,

$$\begin{aligned} \mathbb{P}_x(\mathcal{S}_s) &\geq \frac{\mathbb{P}_x(t' \leq \tau(y) < t' + n) \cdot \mathbb{E}_y Z_n}{2n} \\ &= \Omega\left(\frac{n\sqrt{n}}{t}\right) \cdot n \cdot \frac{1}{2n} = \Omega\left(\frac{n\sqrt{n}}{s}\right), \end{aligned}$$

as desired. □

5. Proofs of Propositions 1.5 and 1.6

Proof of Proposition 1.5. Let S be the state space of X . Note that for $t > 1$,

$$\begin{aligned} \mathbb{P}_\pi(\tau(y) = t) &= \sum_{x \in S \setminus \{y\}} \mathbb{P}_x(\tau(y) = t - 1) \mathbb{P}_\pi(X_1 = x) \\ &\leq \sum_{x \in S} \pi(x) \mathbb{P}_x(\tau(y) = t - 1) = \mathbb{P}_\pi(\tau(y) = t - 1). \end{aligned}$$

Thus, $\mathbb{P}_\pi(\tau(y) = t)$ is non-increasing in t , and so

$$\mathbb{P}_\pi(\tau(y) = t) \leq \frac{1}{t} \sum_{s=0}^{t-1} \mathbb{P}_\pi(\tau(y) = s) \leq \frac{1}{t}.$$

For any particular state $x \in S$, we have

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{\pi(x)} \mathbb{P}_\pi(\tau(y) = t) \leq \frac{1}{t\pi(x)}.$$

□

Proof of Proposition 1.6. We have

$$\begin{aligned} \mathbb{P}_x(\tau(y) = t) &\leq \sum_{z \in S} p^s(x, z) \mathbb{P}_z(\tau(y) = t - s) \\ &\leq \sum_{z \in S, p^s(x, z) \geq \pi(z)} (p^s(x, z) - \pi(z)) \mathbb{P}_z(\tau(y) = t - s) + \sum_{z \in S} \pi(z) \mathbb{P}_z(\tau(y) = t - s) \\ &\leq \sum_{z \in S, p^s(x, z) \geq \pi(z)} (p^s(x, z) - \pi(z)) \psi(t - s) + \sum_{z \in S} \pi(z) \frac{1}{t - s} \\ &\leq d(s) \psi(t - s) + \frac{1}{t - s}, \end{aligned}$$

where between the second and third lines we used Proposition 1.5.

Suppose that $t > 2t_{\text{mix}}(1/4)\lceil \log_2 n \rceil$, and take $s = \lfloor t/2 \rfloor$. By Proposition 1.1, we may take $\psi(u) = n/u$. Note that $s \geq t_{\text{mix}}(1/4)\lceil \log_2 n \rceil$, so

$$d(s) \leq \left(\frac{1}{2}\right)^{\lceil \log_2 n \rceil} \leq \frac{1}{n}.$$

Thus

$$\mathbb{P}_x(\tau(y) = t) \leq \frac{1}{n} \cdot \frac{n}{t-s} + \frac{1}{t-s} = \frac{2}{t-s} \leq \frac{4}{t}.$$

□

6. Open problems

We pose two open problems arising from our work. First, it is natural to try to close the gap between the bound in Theorem 1.2 and the corresponding example given in Claim 2.2. We suspect that Theorem 1.2 is not optimal, and make the following conjecture.

Conjecture 6.1. *Let X be a reversible Markov chain with finite state space, and let π be the stationary distribution of X . Consider any two states x and y . For all $t > 0$,*

$$\mathbb{P}_x(\tau(y) = t) = O\left(\frac{1}{t} \sqrt{\max\left(1, \log \frac{1}{\pi(x)}\right)}\right).$$

The second question relates to hitting times in an interval and is due to Holroyd. The example in Claim 2.1 shows that Proposition 1.1 is optimal up to a constant, but it has the very special property of being periodic with a large period. Consequently, a time t for which $\mathbb{P}_x(\tau(y) = t)$ is large is surrounded by many times t' near t where $\mathbb{P}_x(\tau(y) = t') = 0$. This motivates the following conjecture of Holroyd.

Conjecture 6.2. *Let X be a Markov chain with finite state space S with $|S| = n$, and let x and y be any two states. Then, for all $t > n$,*

$$\mathbb{P}_x(t \leq \tau(y) \leq t + n) = O\left(\frac{n}{t}\right).$$

Acknowledgements

We are grateful to Aryeh Kontorovich for posing the problem that led to this work and to Fedor Nazarov for sharing with us his elegant proof of Lemma 3.2. We thank Jian Ding, Ronen Eldan, Jonathan Hermon, Alexander Holroyd, Gady Kozma, Shayan Oveis Gharan, Perla Sousi, Peter Winkler and Ofer Zeitouni for helpful discussions. We thank the referee for several suggestions to improve the exposition. Finally, we are grateful to Microsoft Research for hosting visits by James Norris and Alex Zhai, which made this work possible.

Appendix: Proofs of lemmas

A.1. Proof of Lemma 3.1

We use the following version of Stirling’s formula, which holds for all integers $N \geq 1$ (see e.g. [3], II.9).

$$\sqrt{2\pi} \leq \frac{N!}{N^{N+1/2}e^{-N}} \leq \sqrt{2\pi} \cdot e^{1/12}. \tag{A.1}$$

Consequently, we have the following estimate on binomial coefficients.

Proposition A.1. *For any positive integers m and n with $m > n$, we have*

$$\frac{1}{3} \sqrt{\frac{m}{(m-n)n}} \cdot \frac{m^m}{(m-n)^{m-n} \cdot n^n} \leq \binom{m}{n} \leq \frac{1}{2} \sqrt{\frac{m}{(m-n)n}} \cdot \frac{m^m}{(m-n)^{m-n} \cdot n^n}.$$

Proof. Apply (A.1) to each factorial term in

$$\binom{m}{n} = \frac{m!}{(m-n)! \cdot n!}.$$

□

Proof of Lemma 3.1. By direct calculation, we have

$$\begin{aligned} \mathbb{P}(X_1 + X_2 + \dots + X_n = m) &= p^n (1-p)^{m-n} \binom{m-1}{n-1} \\ &= \left(\frac{n}{m}\right)^n \left(\frac{m-n}{m}\right)^{m-n} \cdot \binom{m}{n} \cdot \frac{n}{m}. \end{aligned}$$

Applying Proposition A.1 to $\binom{m}{n}$ gives the result. □

A.2. Proof of Lemma 3.2

We will show that $\mathbb{P}(X_1 + X_2 + \dots + X_n = t)$ is maximized when the X_i are identically distributed, and we will then optimize over their common parameter. The key observation due to Nazarov (private communication) is that if X and Y are two independent geometric variables, then $X + Y$ has the same distribution as a mixture of sums of two i.i.d. geometric variables. Indeed, let x and y be the parameters of X and Y , with $x < y$. Define

$$P(x, y, s) = \begin{cases} \frac{xy}{(y-x)s^2} & \text{if } x \leq s \leq y, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $P(x, y, s) \geq 0$, and

$$\int_0^1 P(x, y, s) ds = \int_x^y P(x, y, s) ds = -\frac{xy}{(y-x)s} \Big|_{t=x}^y = 1,$$

and so $P(x, y, \cdot)$ may be regarded as a probability density on $[0, 1]$.

Let S be a random variable drawn from $P(x, y, \cdot)$, and for each $s \in (0, 1)$, let Z_s and Z'_s denote independent geometrics with parameter s . We claim that $Z_s + Z'_s$ has the same distribution as $X + Y$. To verify this, we compute for any integer $m \geq 2$ that

$$\begin{aligned} \mathbb{P}(Z_s + Z'_s = m) &= \int_0^1 P(x, y, s) \sum_{i=1}^{m-1} \mathbb{P}(Z_s = i) \mathbb{P}(Z'_s = m - i) ds \\ &= \int_x^y P(x, y, s) \cdot (m - 1)s^2(1 - s)^{m-2} ds = (m - 1) \int_x^y \frac{xy}{y - x} (1 - s)^{m-2} ds \\ &= \frac{xy}{y - x} ((1 - x)^{m-1} - (1 - y)^{m-1}) = xy \sum_{i=0}^{m-2} (1 - x)^i (1 - y)^{m-2-i} \\ &= \mathbb{P}(X + Y = m). \end{aligned} \tag{A.2}$$

Proof of Lemma 3.2. For each i , let p_i be the parameter of X_i . We can compute by expanding terms that

$$\mathbb{P}(X_1 + X_2 + \dots + X_n = t) = \prod_{i=1}^n p_i \sum_{\substack{j_1 + \dots + j_n = t \\ j_i \in \mathbb{Z}, j_i > 0}} (1 - p_1)^{j_1 - 1} (1 - p_2)^{j_2 - 1} \dots (1 - p_n)^{j_n - 1}. \tag{A.3}$$

The right side of equation (A.3) may be regarded as a continuous function of $p = (p_1, \dots, p_n) \in [0, 1]^n$, which we denote by $F(p)$. Thus, our problem amounts to upper-bounding F .

Now, consider a point $p^* = (p_1^*, \dots, p_n^*) \in [0, 1]^n$ which maximizes F on the compact set $[0, 1]^n$, and suppose for the sake of contradiction that not all of the p_i^* are equal. Then, without loss of generality $p_1^* \neq p_2^*$. However, (A.2) then implies that we can increase F by replacing X_1 and X_2 by two geometrics with the same parameter. More formally, applying equation (A.2) with $X = X_1$ and $Y = X_2$ amounts to saying that

$$F(p_1^*, p_2^*, p_3^*, \dots, p_n^*) = \int_{p_1^*}^{p_2^*} P(p_1^*, p_2^*, s) F(s, s, p_3^*, \dots, p_n^*) ds.$$

Since $P(p_1^*, p_2^*, \cdot)$ is a probability density supported on $[p_1^*, p_2^*]$ and $F(s, s, p_3^*, \dots, p_n^*)$ is non-constant as a function of s for $p_1^* \leq s \leq p_2^*$, it follows that

$$F(p_1^*, p_2^*, p_3^*, \dots, p_n^*) < F(s, s, p_3^*, \dots, p_n^*)$$

for some $s \in [p_1^*, p_2^*]$, a contradiction.

Thus, it follows that the p_i^* all take some common value r . To determine this value r , we may compute

$$F(p^*) = r^n \cdot \binom{t - 1}{n - 1} (1 - r)^{t - n},$$

and optimizing over r , we find that $r = n/t$. Applying Lemma 3.1 finishes the proof. □

A.3. Proof of Lemma 3.4

Proof. Let P be the transition operator of X , and let π be its stationary distribution. Define the inner product $\langle \cdot, \cdot \rangle_\pi$ on \mathbb{R}^S by $\langle f, g \rangle_\pi = \sum_{s \in S} \pi(s) f(s) g(s)$. Note that by reversibility, we have

$$\begin{aligned} \langle Pf, g \rangle_\pi &= \sum_{s \in S} \pi(s) \left(\sum_{s' \in S} P(s, s') f(s') \right) g(s) \\ &= \sum_{s \in S} \sum_{s' \in S} \pi(s') P(s', s) f(s') g(s) \\ &= \sum_{s' \in S} \pi(s') \left(\sum_{s \in S} P(s', s) g(s) \right) f(s') = \langle f, Pg \rangle_\pi. \end{aligned}$$

Hence, P is self-adjoint with respect to the π inner product. Note that because P is stochastic, its eigenvalues lie in the interval $[-1, 1]$.

Now, let $\tilde{S} = S \setminus U$, and let \tilde{P} be the transition matrix of X killed upon hitting U . That is, for $f \in \mathbb{R}^{\tilde{S}}$,

$$(\tilde{P}f)(s) = \sum_{s' \in \tilde{S}} p(s, s') f(s').$$

If we regard P as a symmetric bilinear form, then \tilde{P} is the restriction of P onto the subspace $\mathbb{R}^{\tilde{S}}$. Hence, \tilde{P} is also a symmetric bilinear form, and if P is positive semidefinite, then so is \tilde{P} .

Let \tilde{X}_t be the walk started at x and killed upon hitting U , and let

$$f_t(s) = \frac{1}{\pi(s)} \mathbb{P}(\tilde{X}_t = s)$$

(we are guaranteed that $\pi(s) > 0$ by irreducibility). Note that

$$f_0(s) = \frac{1}{\pi(x)} \delta_{x,s},$$

and

$$\begin{aligned} f_t(s) &= \frac{1}{\pi(s)} \mathbb{P}(\tilde{X}_t = s) = \frac{1}{\pi(s)} \sum_{s' \in \tilde{S}} p(s', s) \mathbb{P}(\tilde{X}_{t-1} = s') \\ &= \sum_{s' \in \tilde{S}} p(s, s') f_{t-1}(s') = (\tilde{P}f_{t-1})(s), \end{aligned}$$

so $f_t = \tilde{P}^t f_0$. We thus have

$$\begin{aligned} \mathbb{P}(X_t = x, \tau(U) > t) &= \mathbb{P}(\tilde{X}_t = x) = \pi(x) f_t(x) \\ &= \pi(x) \cdot \langle f_t, f_0 \rangle_\pi = \pi(x) \cdot \langle \tilde{P}^t f_0, f_0 \rangle_\pi. \end{aligned}$$

Since \tilde{P} is self-adjoint with respect to $\langle \cdot, \cdot \rangle_\pi$, it has an orthonormal basis (in the π inner product) of eigenvectors $\ell_1, \dots, \ell_{|\tilde{S}|} \in \mathbb{R}^{\tilde{S}}$. For each i , let λ_i be the eigenvalue corresponding to ℓ_i . We can write $f_0 = a_1 \ell_1 + \dots + a_{|\tilde{S}|} \ell_{|\tilde{S}|}$, and the last expression becomes

$$\pi(x) \sum_{i=1}^{|\tilde{S}|} a_i^2 \lambda_i^t.$$

This proves the lemma. □

References

- [1] Basu, R., Hermon, J. and Peres, Y. (2015) Characterization of cutoff for reversible Markov chains. *Ann. Probab.*, to appear. An extended abstract appeared in *Proc. Twenty-Sixth Annual ACM–SIAM Symposium on Discrete Algorithms*, SIAM.
- [2] Durrett, R. (2010) *Probability: Theory and Examples*, Cambridge University Press.
- [3] Feller, W. (1957) *An Introduction to Probability Theory and its Applications*, Vol. I, Wiley.
- [4] Gurel-Gurevich, O. and Nachmias, A. (2013) Nonconcentration of return times. *Ann. Probab.* **41** 848–870.
- [5] Lawler, G. (1980) A self-avoiding random walk. *Duke Math. J.* **47** 655–693.
- [6] Lawler, G. (1991) *Intersections of Random Walks*, Birkhäuser.
- [7] Levin, D. A., Peres, Y. and Wilmer, E. L. (2009) *Markov Chains and Mixing Times*, AMS.
- [8] Miclo, L. (2010) On absorption times and Dirichlet eigenvalues. *ESAIM: Probab. Statist.* **14** 117–150.
- [9] Peres, Y. and Sousi, P. (2015) Total variation cutoff in a tree. *Annales de la faculté des sciences de Toulouse* **24** 763–779.
- [10] Starr, N. (1966) Operator limit theorems. *Trans. Amer. Math. Soc.* **121** 90–115.
- [11] Wilson, D. B. (1996) Generating random spanning trees more quickly than the cover time. In *Proc. Twenty-Eighth Annual ACM Symposium on Theory of Computing*, ACM, pp. 296–303.