# A Tight Bound for the Lamplighter Problem

Murali K. Ganapathy* Prasad Tetali†

**Abstract**

We settle an open problem, raised by Y. Peres and D. Revelle, concerning the $L^2$ mixing time of the random walk on the lamplighter graph. We also provide general bounds relating the entropy decay of a Markov chain to the separation distance of the chain, and show that the lamplighter graphs once again provide examples of tightness of our results.

## 1 Introduction

Given a finite connected graph $G$, a vertex of the lamplighter graph $G^\diamondsuit$ consists of a 0-1 labeling of the vertices of $G$, and a marked vertex of $G$. Each vertex has a lamp, the marked vertex indicates the position of a lamplighter and the labeling at any time indicates the off-on status of each lamp (vertex). The lamplighter random walk on $G^\diamondsuit$ corresponds to the lamplighter performing a random walk on $G$, while randomizing the status of each lamp, as he/she visits the corresponding vertex. When $G$ is a cycle or a complete graph, the corresponding lamplighter chains were studied by Häggström and Jonasson [7]. Vertex transitive, other special classes and more general graphs were considered in detail by Peres and Revelle [10], who provided general upper and lower bounds for mixing times of the lamplighter random walk.

By tightening the analysis in [10], we prove an optimal upper bound on the $L^2$ mixing time of the lamplighter Markov chain on a class of graphs considered in [10]. The mixing time and related measures are defined via

**Definition 1.** Let $G$ be a connected $d$-regular undirected graph.

- The relaxation time $T_{rel}(G) = \max_\lambda 1/(1-|\lambda|)$ where the maximum is taken over non-trivial eigenvalues $\lambda$ of the normalized adjacency matrix of $G$.
- $\tau(G)$ ($\tau_2(G)$) is the time for the random walk on $G$ to be within $1/4$ of uniform distribution in total variation distance ($L_2$-distance, respectively).
- $H(G)$ is expected time it takes for the random walk to travel from $x$ to $y$, where the choice of $x$ and $y$ is adversarial.

There is a more popular definition of relaxation time (see e.g. [2, 9]). We choose this definition as it is easier to work with. Standard inequalities imply that the two numbers differ by at most 1.

For a graph $G$, let $\tau_2(G^\diamondsuit)$ denote the $L^2$ mixing time of the lamplighter random walk on $G^\diamondsuit$. Then our main theorem is as follows.

*Google, 1600 Amphitheatre Pkwy, Mountain View, CA - 94043

†School of Mathematics and College of Computing, Georgia Tech., Atlanta, GA 30332-0160. Research supported in part by NSF grant DMS-0401239.

**Theorem 2.** *Suppose $G$ is a regular undirected graph for which, $H(G) \le \kappa |G|$, for some universal constant $\kappa > 0$. Then there exists a constant $c = c(\kappa)$ such that*

$$\tau_2(G^\diamondsuit) \le c\,|G|\big(T_{rel}(G) + \log|G|\big)\,. \tag{1}$$

The above theorem refines and improves upon (by providing a matching upper bound to) the result of Y. Peres and D. Revelle [10], who proved (inter alia) the following theorem.

**Theorem 3** (Peres-Revelle)**.** *Under the hypothesis of Theorem 2, there exist constants, $c_1, c_2$ depending on $\kappa$ such that*

$$c_1\,|G|\big(T_{rel}(G) + \log|G|\big) \;\le \tau_2(G^\diamondsuit) \le c_2\,|G|\big(T_{\mathrm{tv}}(G) + \log|G|\big)\,, \tag{2}$$

*where $T_{\mathrm{tv}}(G)$ is the total variation mixing time of the simple random walk on $G$.*

Note that our theorem shows that for $G = \mathbb{Z}_2^n$, the correct order of magnitude for $\tau_2(G^\diamondsuit)$ to be $n2^n$, since the relaxation time is of order $n$. This settles Problem 4 mentioned at the end of [10], while our theorem itself settles the question raised as Problem 5 in the affirmative.

In [10], the lamplighter random walk on the two-dimensional torus was shown to be an example of a chain for which the relaxation time, the total variation mixing time, and the $L^2$ mixing time were all shown to be distinct orders of magnitude. In this paper, we use the one and two- dimensional tori as examples which further separate the mixing time in entropy (relative to stationarity) from the rest of the above mixing times. These examples also illustrate tightness of the following other result of this paper. We show that in general the *entropy mixing time* is at worst a factor of $\log\log(1/\pi_*)$ larger than the total variation mixing time for reversible Markov chains (see Corollary 12) below.) This is accomplished by relating the relative entropy to the so-called separation distance of a Markov chain.

## 2 The Lamplighter Result

In this section we derive some preliminary technical lemmas and a key theorem from which the main theorem follows. Since random walks on regular undirected graphs are equivalent to reversible Markov chain with uniform stationary distribution, we will use the latter from now on.

Assume that $\mathbb{P}$ is a Markov chain with uniform stationary distribution $\pi$ on a finite state space $\mathcal{X}$. Let $H = \max_{x,y} \mathbb{E}_x T_y$ denote the maximal hitting time (also called the maximum expected first passage time) of the chain. Let $T_{rel}$ denote the relaxation time of the chain, where $\lambda$ is the spectral gap of the chain.

As observed by Peres and Revelle, the $L^2$ mixing time of the lamplighter graph $G^\diamondsuit$ depends upon the moment generating function of the cover time of the underlying graph $G$. More precisely, if $S_t$ denotes the set of unvisited vertices (by the lamplighter) by time $t$, then to get convergence in the $L^2$ (or equivalently, in the uniform metric), one needs $\mathbb{E}2^{|S_t|} \le 1 + \epsilon$, for $\epsilon > 0$. Our main technical contribution is as follows.

Let $\mathbb{P}$ be a reversible Markov Chain on the state space $\mathcal{X}$ with $\pi$ as the stationary distribution.

**Theorem 4.** *Let the chain given by $\mathbb{P}$ start in an initial distribution $\mu$ so that $\mu \ge \pi/2$. Let the maximal hitting time $H = H(\mathbb{P})$ satisfy $H \le c_1|\mathcal{X}|$ for a constant $c_1 \ge 1$. Let $\theta \ge 2$ be arbitrary,* *and let $\mathbf{S}_t$ denote the set of vertices which have not been visited by time $t$. Then there exists a universal constant $c$ such that for all $a, b > 0$, and for $t \geq t' = C_1 |\mathcal{X}| T_{rel} \log\theta + C_2 |\mathcal{X}| \log|\mathcal{X}|$, we have*

$$\mathbb{E}[\theta^{|\mathbf{S}_t|}] \leq 1 + \delta + \delta^2 + \delta^9 ,$$

*where $\delta = \theta^{-(1+2a)T_{rel}} |\mathcal{X}|^{-b}$, $C_1 = 2c\, c_1^2(1+a)$ and $C_2 = c\, c_1^2(1+b)$. In particular, when $a = b = 1$, we have $\mathbb{E}[\theta^{|\mathbf{S}_t|}] < 1.21$.*

Once we have Theorem 4, the main theorem (Theorem 2) follows in a straightforward way, as in [10]. We begin with a few simple lemmas.

Let $\mathbb{P}$ be a Markov chain with uniform stationary distribution $\pi$. Let $\sigma_1$ denote the second largest singular value of $\mathbb{P}$. In particular, if $\mathbb{P}$ is reversible, then $\sigma_1 = \max(\lambda_1, |\lambda_{N-1}|)$, where $1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{N-1}$ denote all the eigenvalues of $\mathbb{P}$. Recall the following basic fact, whose proof we include for completeness.

**Lemma 5.** *Let $f : \mathcal{X} \to \mathbb{R}$. Let $\mathbf{X}_i$ denote the state of the chain $\mathbb{P}$ at time $i$. Then*

$$\mathrm{Cov}(f(\mathbf{X}_1), f(\mathbf{X}_t)) \leq \sigma_1^t \,\mathrm{Var}(f(X_1)) ,$$

*where $\sigma_1$ is the second largest singular value of $\mathbb{P}$.*

*Proof.* Let $g = \mathbb{P}^t f$, so that $\mathrm{Cov}(f(\mathbf{X}_1), f(\mathbf{X}_t)) = \mathrm{Cov}(f, g)$. Hence we have

$$\begin{aligned}
\mathrm{Cov}(f,g) &= \mathbb{E}_\pi[(f - Ef)(g - Eg)] \\
&= \mathbb{E}_\pi[(f - Ef)(\mathbb{P}^t f - E\mathbb{P}^t f)] \\
&= \mathbb{E}_\pi[(f - Ef)((\mathbb{P}^t - E)f)] && (E\mathbb{P} = E) \\
&= \mathbb{E}_\pi[(f - Ef)((\mathbb{P} - E)^t f)] && \mathbb{P}^t - E = (\mathbb{P} - E)^t \\
&\leq \sqrt{\mathbb{E}_\pi((I - E)f)^2}\, \sigma_1^t \sqrt{\mathrm{Var}_\pi(f)} \\
&= \lambda_*^t \,\mathrm{Var}_\pi(f)
\end{aligned}$$

where we used the fact that the operator norm of $\mathbb{P} - E$ is $\sigma_1$. □

**Lemma 6.** *Let $\{X_t\}$ be a reversible Markov chain on $\mathcal{X}$, with uniform stationary distribution $\pi$. Assume that $\Pr\{X_0 = x\} \geq \pi(x)/2$. Let $T_{rel}$ be the relaxation time of the chain. Let $T_x^+$ denote the return time to $x$ and assume there are $\epsilon, \delta > 0$ for which*

$$\Pr_x(T_x^+ \geq \epsilon|\mathcal{X}|) \geq \delta > 0$$

*for all $x \in \mathcal{X}$. Let $\mathcal{Y} \subset \mathcal{X}$ be such that $|\mathcal{Y}| \geq T_{rel}$. Then the probability of hitting at least $\delta\epsilon|\mathcal{Y}|/4$ elements of $\mathcal{Y}$ by time $C\delta^{-2} T_{rel}/\pi(\mathcal{Y})$ is at least $1/2$, where $C \geq 16$ is an absolute constant.*

*Proof.* Let $r = \epsilon T_{rel}/\pi(\mathcal{Y})$. For $1 \leq i \leq r$, let $I_i$ be an indicator random variable for the event $\{X_i \in \mathcal{Y}\}$ and $J_i$ for the event $\{X_i \in \mathcal{Y}\}$ and $\{X_j \neq X_i\}$ for $i < j \leq r$. Finally let $J = \sum_i J_i$ and $I = \sum_i I_i$.

Note that $J$ is the number of distinct elements of $\mathcal{Y}$ which have been visited in the time interval $[1, r]$. Also we have $\Pr\{J_i = 1 | I_i = 1\} \geq \delta$ since $r \leq \epsilon|\mathcal{X}|$. This together with the fact that $\mathbb{E}[I_i] \geq \pi(\mathcal{Y})/2$ (due to our assumption on the initial distribution), gives

$$\mathbb{E}[J] \geq \delta\mathbb{E}[I] \geq \delta r \pi(\mathcal{Y})/2 = \frac{\delta\epsilon T_{rel}}{2} . \tag{3}$$

To conclude $\Pr\{J \geq \epsilon\delta T_{rel}/4\}$ is bounded away from 0, we bound $\mathbb{E}[I^2]$.

Fix $1 \leq i \leq j \leq r$. From Lemma 5 we have

$$\sum_{j \geq i} \mathrm{Cov}(I_i, I_j) \leq \sum_{j \geq i} \lambda_*^{j-i} \mathrm{Var}(I_i) \leq \frac{\mathrm{Var}(I_i)}{1-\lambda_*} = T_{rel} \mathrm{Var}(I_i) \leq T_{rel}\mathbb{E}[I_i] \,.$$

Now we have

$$\mathbb{E}[I^2] \leq 2 \sum_i \sum_{j \geq i} \mathrm{Cov}(I_i, I_j) \leq 2 \sum_i T_{rel}\mathbb{E}[I_i] = 2T_{rel}\mathbb{E}[I] \,.$$

Since $J_i \leq I_i$ we have $\mathbb{E}[J^2] \leq \mathbb{E}[I^2] \leq 2T_{rel}\mathbb{E}[I] \leq 2T_{rel}\mathbb{E}[J]/\delta$, since $\mathbb{E}[J_i] \geq \delta\mathbb{E}[I_i]$. Hence using Equation 3 we have

$$\mathbb{E}[J^2] \leq \frac{4}{\delta^2\epsilon}\mathbb{E}[J]^2 \,. \tag{4}$$

Now let $\alpha$ be the indicator for the event $J \geq \mathbb{E}[J]/2$. Then $\mathbb{E}[J(1-\alpha)] \leq \mathbb{E}[J]/2$ and hence $\mathbb{E}[J\alpha] \geq \mathbb{E}[J]/2$. Now by Cauchy-Schwartz, we have $\mathbb{E}[J\alpha]^2 \leq \mathbb{E}[J^2]\mathbb{E}[\alpha^2]$. Hence

$$\Pr\{J \geq \mathbb{E}[J]/2\} = \mathbb{E}[\alpha^2] \geq \frac{\mathbb{E}[J\alpha]^2}{\mathbb{E}[J^2]} \geq \frac{(\mathbb{E}[J]/2)^2}{\mathbb{E}[J^2]} \geq \frac{\delta^2\epsilon}{16} \,.$$

Thus in a trial of length $r = \epsilon|\mathcal{X}|T_{rel}/|\mathcal{Y}|$, the probability that we do not pick up $\delta\epsilon|\mathcal{Y}|/4$ elements of $\mathcal{Y}$ is less than $1 - \delta^2\epsilon/16$. Hence if we repeat this for $C\delta^{-2}/\epsilon$ intervals of length $r$ each, we can reduce the probability of failure to less than $1/2$. Note that $C \geq 16$ here. □

Suppose that the initial distribution $\mu$ is such that $\mu \geq \pi/2$. Then if the current set of unvisited states is large ($\geq T_{rel}$) then Lemma 6 shows that we visit $\Omega(T_{rel})$ new states within time $O(T_{rel}/\pi(\mathcal{Y}))$ with probability $\geq 1/2$. Once the set of unvisited states gets smaller than $T_{rel}$ things are in better shape. The next lemma establishes the assumption of Lemma 6 and handles the case when the set of unvisited vertices is small.

**Lemma 7.** *Let $\mathbb{P}$ be a Markov chain with uniform stationary distribution and maximal hitting time $H$. For $x \in \mathcal{X}$, let $T_x^+$ denote the expected length of the return time to $x$. Then*

$$\min_x \Pr_x\left\{T_x^+ \geq \frac{\mathcal{X}|}{2}\right\} \geq \frac{|\mathcal{X}|}{2H} \,.$$

*Also, for any $\mathcal{Y} \subseteq \mathcal{X}$, with probability $\geq 1/2$, we visit at least $|\mathcal{Y}|/2$ elements of $\mathcal{Y}$ by time $4H$.*

*Proof.* Since the stationary distribution is uniform, $\mathbb{E}_x[T_x^+] = |\mathcal{X}|$. If after $|\mathcal{X}|/2$ steps we have not yet returned to $x$, and are currently at state $y$, then we expect to visit $x$ within another $H$ steps. Hence

$$|\mathcal{X}| = \mathbb{E}_x[T_x^+] \leq \Pr\{T_x^+ \leq |\mathcal{X}|/2\}|\mathcal{X}|/2 + \Pr\{T^\gamma \geq |\mathcal{X}|/2\}H \,.$$

Rearranging terms, we get the result.

For the second result: Fix $x \in \mathcal{Y}$ and let $\mathbf{H}_x$ denote the time when $x$ is visited. $\mathbb{E}[\mathbf{H}_x] \leq H$. Thus by time $4H$ we visit $x$ with probability $\geq 3/4$.

Let $Y$ denote the number of elements of $\mathcal{Y}$ which have been visited by time $4H$. Then $\mathbb{E}[Y] \geq 3|\mathcal{Y}|/4$. If $q = \Pr\{Y \geq |\mathcal{Y}|/2\}$, we have

$$3|\mathcal{Y}|/4 \leq \mathbb{E}[Y] \leq (1-q)\frac{|\mathcal{Y}|}{2} + q|\mathcal{Y}| \,.$$

Solving for $q$ gives $q \geq 1/2$. □

**Proof of Theorem 4.** Let $r = \lfloor |\mathcal{X}|/T_{rel} \rfloor$ and for $i = 0, 1, \dots, r-1$, let $k_i = |\mathcal{X}| - iT_{rel}$ and for $i = r$, $k_r = 0$. Define stopping times $\mathbf{T}_i$ as the time when $|\mathbf{S}_t| = k_i$ for the first time.

From Lemma 7 and Lemma 6, it then follows that $\mathbf{T}_i - \mathbf{T}_{i-1}$ is stochastically dominated by $\alpha_i = \gamma/k_i \mathbf{Z}_i$ where $\mathbf{Z}_i$ is geometric with mean 2 and $\gamma = C \cdot K^2 |\mathcal{X}| T_{rel}$ and $C \geq 16$ is a universal constant.

Fix $t > 0$, $i < r$ and $\beta > 0$ be arbitrary, Then

$$\begin{aligned}\Pr\{\mathbf{T}_i \geq t\} &= \Pr\Big\{\sum_{j=1}^{i} \alpha_i \geq t\Big\} \leq \Pr\Big\{\sum_{j=1}^{i} \frac{\gamma}{k_i}\mathbf{Z}_i \geq t\Big\} \\ &\leq \exp(-t\beta)\mathbb{E}\Big[\sum_{j=1}^{i} \frac{\gamma\beta}{k_i}\mathbf{Z}_i\Big] \leq \exp(-t\beta)\prod_{j=1}^{i}\mathbb{E}\Big[\frac{\gamma\beta}{k_i}\mathbf{Z}_i\Big]\,.\end{aligned}$$

Choose $\beta$ so that $\beta = k_i/3\gamma$ so that $\gamma\beta \leq k_j/3$ for all $j \leq i$. For $\alpha \leq 1/3$, $\mathbb{E}[\alpha \mathbf{Z}_j] \leq \exp(3\alpha)$. This gives

$$\mathbb{E}\Big[\frac{\gamma\beta}{k_j}\mathbf{Z}_j\Big] \leq \exp(k_i/k_j)\,.$$

Hence we have

$$\Pr\{\mathbf{T}_i \geq t\} \leq \exp\Big(-t\frac{k_i}{3\gamma} + \sum_{j=1}^{i}\frac{k_i}{k_j}\Big) \leq \exp\Big(-t\frac{k_i}{3\gamma} + \frac{k_i}{T_{rel}}\log|\mathcal{X}|\Big)\,.$$

For $i = r$, Lemma 7 implies $(\mathbf{T}_r - \mathbf{T}_{r-1})$ is stochastically dominated by the sum of $\ell = \log_2(2T_{rel})$ independent geometric random variables with mean $4K|\mathcal{X}|$. Applying a Chernoff bound we get

$$\Pr\{\mathbf{T}_r - \mathbf{T}_{r-1} \geq t\} \leq \exp\Big(\ell - \frac{t}{4K|\mathcal{X}|} + \ell\log\Big(\frac{t}{4K|\mathcal{X}|}\Big)\Big)\,. \tag{5}$$

Breaking the values of $|S_t|$ into intervals of size $T_{rel}$ we have

$$\mathbb{E}[\theta^{|S_t|}] \leq 1 + \sum_{i=0}^{r}\theta^{k_i + T_{rel}}\Pr\{|S_t| \geq k_i\} = 1 + \sum_{i=0}^{r}\theta^{k_i+T_{rel}}\Pr\{\mathbf{T}_i \geq t\}\,.$$

For $i < r$, $k_i \geq T_{rel}$ and hence

$$\begin{aligned}\theta^{k_i+T_{rel}}\Pr\{\mathbf{T}_i \geq t\} &\leq \exp\Big((k_i + T_{rel})\log\theta - t\frac{k_i}{3\gamma} + \frac{k_i}{T_{rel}}\log|\mathcal{X}|\Big) \\ &\leq \exp\Big(2k_i\log\theta + \frac{k_i}{T_{rel}}\log|\mathcal{X}| - t\frac{k_i}{3\gamma}\Big)\,.\end{aligned} \tag{6}$$

When $i = r$, $0 = k_r \leq T_{rel} \leq k_{r-1}$ and hence

$$\begin{aligned}\theta^{k_r+T_{rel}}\Pr\{\mathbf{T}_r \geq t\} &\leq \theta^{T_{rel}}\big(\Pr\{\mathbf{T}_{r-1} \geq t/2\} + \Pr\{\mathbf{T}_r - \mathbf{T}_{r-1} \geq t/2\}\big) \\ &\leq \theta^{T_{rel}}\exp\Big(-\frac{t}{2}\frac{k_{r-1}}{3\gamma} + \frac{k_{r-1}}{T_{rel}}\log|\mathcal{X}|\Big) \\ &\quad + \theta^{T_{rel}}\exp\Big(\ell - \frac{t}{8K|\mathcal{X}|} + \ell\log\Big(\frac{t}{8K|\mathcal{X}|}\Big)\Big)\,,\end{aligned} \tag{7}$$

where $\ell = \log_2(2T_{rel})$.

Let $t' = 6CK^2|\mathcal{X}|\big(2(1+a)T_{rel}\log\theta + (1+b)\log|\mathcal{X}|\big)$ for any $a, b > 0$ and hence take $c = 6C$. We now show that for $t \geq t'$, $\mathbb{E}[\theta^{|\mathbf{S}_{t'}|}] - 1$ is small. Recall that $\gamma = CK^2|\mathcal{X}|T_{rel}$, hence we have

$$\begin{aligned} \frac{t'}{3\gamma} &= 4(1+a)\log\theta + 2(1+b)\frac{\log|\mathcal{X}|}{T_{rel}} \\ \frac{t'}{8K|\mathcal{X}|} &= \frac{3CK}{4}\big(2(1+a)T_{rel}\log\theta + (1+b)\log|\mathcal{X}|\big) \\ \ell' := \log\Big(\frac{t'}{8K|\mathcal{X}|}\Big) &\geq \log\Big(\frac{3(1+a)CK\log\theta\cdot 2T_{rel}}{4}\Big) \geq \log(6\cdot(2T_{rel}))\,, \end{aligned}$$

since $C \geq 16, K \geq 1, \theta \geq 2$.

Now for $t \geq t'$ and $i < r$, Equation 6 reduces to

$$\begin{aligned} \theta^{k_i+T_{rel}}\Pr\{\mathbf{T}_i \geq t'\} &\leq \exp\Big(2k_i\log\theta + \frac{k_i}{T_{rel}}\log|\mathcal{X}| - 4(1+a)k_i\log\theta - 2(1+b)\frac{k_i}{T_{rel}}\log|\mathcal{X}|\Big) \\ &= \theta^{-(2+4a)k_i}|\mathcal{X}|^{-(1+2b)k_i/T_{rel}}\,. \end{aligned} \tag{8}$$

And Equation 7 reduces to

$$\begin{aligned} \theta^{T_{rel}}\Pr\{\mathbf{T}_i \geq t'\} \leq \theta^{T_{rel}}&\exp\Big(-2(1+a)k_{r-1}\log\theta - (1+b)\frac{k_{r-1}}{T_{rel}}\log\mathcal{X} + \frac{k_{r-1}}{T_{rel}}\log\mathcal{X}\Big) \\ &+ \theta^{T_{rel}}\exp\Big(\ell - \exp(\ell') + \ell\ell'\Big) \\ &\leq \exp\Big(-(1+2a)k_{r-1}\log\theta - b\frac{k_{r-1}}{T_{rel}}\log|\mathcal{X}|\Big) \\ &+ \exp\Big(T_{rel}\log\theta + (1+\ell')(\ell' - \log(6)) - \exp(\ell')\Big)\,, \end{aligned} \tag{9}$$

using $\ell = \log_2(2T_{rel})$ and $\ell' \geq \ell + \log(6)$. Here we also use the fact that $f(x) = \exp(x)/4 - (1+x)(x - \log(6)) \geq 0$ for all $x \geq 0$, and $k_{r-1} \geq T_{rel}$.

We now have

$$\begin{aligned} \theta^{T_{rel}}\Pr\{\mathbf{T}_i \geq t'\} \leq \theta^{-(1+2a)T_{rel}}&|\mathcal{X}|^{-b} \\ &+ \exp\Big(T_{rel}\log\theta - \frac{9CK}{16}\Big(2(1+a)T_{rel}\log\theta + (1+b)\log|\mathcal{X}|\Big)\Big) \\ &\leq \theta^{-(1+2a)T_{rel}}|\mathcal{X}|^{-b} + \exp\Big(-9(1+2a)T_{rel}\log\theta - 9(1+b)\log|\mathcal{X}|\Big) \\ &\leq \theta^{-(1+2a)T_{rel}}|\mathcal{X}|^{-b} + \theta^{-9(1+2a)T_{rel}}|\mathcal{X}|^{-9b}\,, \end{aligned} \tag{10}$$

using $C \geq 16$.

Combining Equation 10 and Equation 8 we have

$$\mathbb{E}[\theta^{|\mathbf{S}_{t'}|}] \leq 1 + \sum_{i=0}^{r-1}\theta^{-(2+4a)k_i}|\mathcal{X}|^{-(1+2b)k_i/T_{rel}} + \theta^{-(1+2a)T_{rel}}|\mathcal{X}|^{-b} + \theta^{-9(1+2a)T_{rel}}|\mathcal{X}|^{-9b}\,. \tag{11}$$

Now let $\eta = \theta^{-(1+2a)T_{rel}}$. Using $k_{r-i} \geq iT_{rel}$, we get

$$\mathbb{E}[\theta^{|\mathbf{S}_{t'}|}] \leq 1 + \sum_{i=0}^{r-1} \eta^{2(r-i)} |\mathcal{X}|^{-(1+2b)(r-i)} + \eta|\mathcal{X}|^{-b} + \eta^9|\mathcal{X}|^{-9b} .$$

Summing up the geometric progression and simplifying we get

$$\mathbb{E}[\theta^{|\mathbf{S}_{t'}|}] \leq 1 + \eta^2|\mathcal{X}|^{-2b} + \eta|\mathcal{X}|^{-b} + \eta^9|\mathcal{X}|^{-9b} . \tag{12}$$

Observe that by setting $a = b = 1$ and using $\theta \geq 2, T_{rel} \geq 1/2, |\mathcal{X}| \geq 2$, we get $\eta \leq 2^{-3/2}$ and $|\mathcal{X}|^{-b} \leq 1/2$. Finally, substituting in Equation 12 we have $\mathbb{E}[\theta^{|\mathbf{S}_{t'}|}] < 1.21$ . □

**Proof of Theorem 2.** Implicit in the proof of Theorem 1.4 of [10] is that for a suitable constant $C > 0$,

$$\tau_2(G^{\diamondsuit}) \leq C\left(\tau_2(G) + \min_t \left\{\mathbb{E}\left[2^{|S_t|}\right] < 2\right\}\right) . \tag{13}$$

Since $G$ is undirected, the random walk on $G$ is reversible. Hence after time $4\tau(G)$, the distribution $\mu$ of the lamplighter's position satisfies $\mu \geq \pi/2$. Since the random walk on $G$ is reversible, we have (e.g., by [3]) that the mixing time is bounded above by maximal hitting time $H(G)$. Thus by running the Lamplighter chain for an initial $O(H(G))$ steps, we can ensure that the assumption of Theorem 4 holds.

Hence Theorem 4 implies $\mathbb{E}\left[2^{|S_t|}\right] < 2$ for $t = O(H(G)) + O(|G| \cdot (T_{rel} + \log|G|))$. Thus (13) gives

$$\tau_2(G^{\diamondsuit}) = O\big(\tau_2(G) + |G| \cdot (T_{rel} + \log|G|)\big), \tag{14}$$

since $H(G) \leq \kappa|G|$. Finally, the regularity of $G$ implies that the stationary distribution of the random walk on $G$ is uniform. Thus $\tau_2(G) = O(\tau(G)\log|G|)$ implying that $\tau_2(G) = O(|G|\log|G|)$.□

## 3 Separation Distance and Entropy Decay

Recall that Pinsker's inequality lets one bound the total variation mixing time of a Markov chain *from above* by the entropy decay time, up to an absolute constant. In this section, we show that for reversible Markov chains, the time for the relative entropy to decay to within $1/e$ is no larger than $\log\log(1/\pi_*)$ times that of the total variation mixing time. We actually prove a more general result for all Markov chains from which the above will follow under the additional assumption of reversibility.

Once again let $\mathbb{P}$ be a Markov kernel with stationary distribution $\pi$ on a (finite) state space $\mathcal{X}$. First recall the definition of *separation* between a chain at time $n \geq 0$ and $\pi$.

**Definition 8.** For $n \in \mathbb{N}$ and $x \in \mathcal{X}$, set

$$\mathbf{sep}_{\mathbb{P}}(\mathrm{x},\mathrm{n}) = \max_{\mathrm{y}\in\mathcal{X}}\left(1 - \frac{\mathbb{P}^{\mathrm{n}}(\mathrm{x},\mathrm{y})}{\pi(\mathrm{y})}\right), \quad \mathbf{sep}_{\mathbb{P}}(\mathrm{n}) = \max_{\mathrm{x}\in\mathcal{X}} \mathbf{sep}_{\mathbb{P}}(\mathrm{x},\mathrm{n}) .$$

Also set

$$d_{\mathrm{tv}}(\mathbb{P}^n(x,\cdot) - \pi) = \sum_{y \in \pi(y) > \mathbb{P}^n(x,y)} (\pi(y) - \mathbb{P}^n(x,y)) ,$$

$$\|\mathbb{P}^n(x,\cdot)/\pi(\cdot)-1\|_2 = \Big(\sum_{y\in\mathcal{X}} \big(\mathbb{P}^n(x,y)/\pi(y)-1\big)^2\pi(y)\Big)^{1/2}.$$

When understood from the context, we drop the subscript in $\mathbf{sep}_{\mathbb{P}}$. Recall that the function $n \mapsto \mathbf{sep}(\mathrm{n})$ is non-increasing and sub-multiplicative (see [1, 2] for more details.) It is well-known and is easily seen that $d_{\mathrm{tv}}(\mathbb{P}^n(x,\cdot)-\pi) \le \mathbf{sep}(\mathrm{x},\mathrm{n})$:

$$\begin{aligned} d_{\mathrm{tv}}(\mathbb{P}^n(x,\cdot)-\pi) &= \sum_{y\in\pi(y)>\mathbb{P}^n(x,y)} (\pi(y)-\mathbb{P}^n(x,y)) \\ &= \sum_y \pi(y)\Big(1-\frac{\mathbb{P}^n(x,y)}{\pi(y)}\Big) \\ &\le \mathbf{sep}(\mathrm{x},\mathrm{n})\,. \end{aligned}$$

Thus separation bounds total variation. Now we observe that it also controls entropy decay up to a factor of $\log(1/\pi_*)$. Recall that the relative entropy, denoted by $\mathbf{D}(\mu\|\nu)$, of a distribution $\mu$ with respect to $\nu$ is defined as

$$\mathbf{D}(\mu\|\nu) = \sum_x \mu(x)\log(\mu(x)/\nu(x)),$$

where as usual $0\log 0 = 0$, and $\mu$ is assumed to be absolutely continuous with respect to $\nu$ (meaning, $\nu(x)=0$ implies that $\mu(x)=0$.) It is well-known (see e.g. [4]) that $D(\mu\|\nu) \le \log(1/\nu_*)$, where $\nu_* = \min_x \nu(x)$, and that $D(\cdot\|\nu)$ is convex in the sense that, for $0 \le \alpha \le 1$, and for $\mu_1$, $\mu_2$ probability distributions (absolutely continuous with respect to $\nu$),

$$\mathbf{D}(\alpha\mu_1 + (1-\alpha)\mu_2\|\nu) \le \alpha\mathbf{D}(\mu_1\|\nu) + (1-\alpha)\mathbf{D}(\mu_2\|\nu)\,.$$

**Proposition 9.**

$$\mathbf{D}(\mathbb{P}^n(x,y)\|\pi) \le \mathbf{sep}(\mathrm{x},\mathrm{n})\log(1/\pi_*)\,.$$

*Proof.* Let $\mathbf{sep}(\mathrm{x},\mathrm{n}) = \epsilon > 0$. Then $\mathbb{P}^n(x,y) \ge (1-\epsilon)\pi(y)$, for all $x,y\in\mathcal{X}$. Let

$$\mu(y) := (1/\epsilon)[\mathbb{P}^n(x,y) - (1-\epsilon)\pi(y)]\,, \quad \text{for } y\in\mathcal{X}\,.$$

Then $\mu$ is a probability distribution on $\mathcal{X}$ and $\mathbb{P}^n(x,\cdot) = (1-\epsilon)\pi + \epsilon\mu$. (Note here that $\mu$ implicitly depends on $x$.) By the convexity mentioned above,

$$\mathbf{D}(\mathbb{P}^n(x,y)\|\pi) \le \epsilon\,\mathbf{D}(\mu\|\pi) \le \epsilon\,\log(1/\pi_*),$$

hence the proposition. □

**Definition 10.** For $0<\epsilon<1/2$, let the entropy decay (mixing) time be

$$\tau_{\mathrm{ent}}(\epsilon) = \min\{n' :\ n \ge n' \ \Rightarrow\ \max_{x\in\mathcal{X}} \mathbf{D}(\mathbb{P}^n(x,\cdot)\|\pi) \le \epsilon\}\,.$$

and similarly define the other mixing times $\tau_s$, $\tau_{\mathrm{tv}}$, and $\tau_2$ with respect to $\mathbf{sep}(\mathrm{x},\mathrm{n})$, $\mathrm{d}_{\mathrm{tv}}(\mathrm{P}^{\mathrm{n}}(\mathrm{x},\cdot)-\pi)$, and $\|\mathbb{P}^n(x,\cdot)/\pi(\cdot)-1\|_2$, respectively.

It then follows immediately from the above proposition, that:

**Corollary 11.**

$$\tau_{\text{ent}}(\epsilon) \le \tau_s(1/e)[\log\log(1/\pi_*) + \log(1/\epsilon)]\,.$$

It is known that $\tau_s = O\big(\tau_{\text{tv}}(\mathbb{P}) + \tau_{\text{tv}}(\mathbb{P}^*)\big)$, where $\mathbb{P}^*$ denotes the time-reversal of $\mathbb{P}$. Hence the assertion claimed at the beginning of this section follows. Note that the lower bound below does not need reversibility, and uses the general inequality (known as Pinsker's) for two distributions, $\mu$ and $\nu$, one has:

$$2d_{\text{tv}}^2(\mu - \nu) \le D(\mu\|\nu)\,.$$

**Corollary 12.** *If $\mathbb{P}$ is reversible, then*

$$\tau_{\text{tv}}(\epsilon/2) \le \tau_{\text{ent}}(\epsilon) \le C\,\tau_{\text{tv}}(1/2e)\,[\log\log(1/\pi_*) + \log(1/\epsilon)]\,,$$

*for $C > 0$ an absolute constant.*

**Remark 13.** Note that the above result shows that the entropy decay time is in general closer to $\tau_{\text{tv}}$ than to $\tau_2$, since there can be a factor of $\log(1/\pi_*)$ between $\tau_{\text{tv}}$ and $\tau_2$. Similarly this indicates that in general $\rho_0$ of a reversible Markov chain is closer to the spectral gap than it is to the logarithmic Sobolev constant $\rho$.

We now show that the $\log\log 1/\pi_*$ gap in the corollary cannot be improved. While a random walk on the complete graph (with self-loops) can be shown to establish this, we proceed with the following more robust example, which also separates various other mixing times.

**Example 1.** Consider the following lamplighter chain on a discrete circle of size $n$. Unlike the usual lamplighter walk, in this chain each vertex of the circle has an $m$-state lamp for some parameter $m$. However, every time a vertex is visited, the lamplighter completely randomizes the lamp. The (discrete) mixing time of the chain is still related to the time it takes to visit all vertices of the base graph. In particular, it is easy to see that the total variation mixing time is the expected cover time, i.e. $\Theta(n^2)$ and is independent of $m$. From our result above, it then follows that the entropy mixing time of this chain is $O(n^2 \log\log N)$ where $N = m^n$. Thus we have

$$\tau_{\text{ent}} = O(n^2 \cdot (\log n + \log\log m)).$$

Suppose we start the chain at vertex $x$ of the circle and all lamps are in state 0. Let $A$ denote the semi-circle consisting of vertices which are at distance $n/4$ or larger from $x$.

There exists an absolute constant $c > 0$ such that for all $a > 0$, the probability that a random walk on the circle has not touched $A$ after $an^2$ steps is $\ge \exp(-ca)$. Thus after $an^2$ steps, the entropy is at least $\exp(-ca)|A|\log m$ since the entropy of the product distribution is the sum of the component entropies and the entropy of each non-random lamp is $\log m$. Thus in order for this chain to mix in entropy we must have $a = \Omega(\log(n \log m))$. Hence $\tau_{\text{ent}} = \Omega(n^2 \log(n \log m))$ matching the upper bound given by Corollary 12.

Now let us look at the $L^2$ mixing time. In this case, we need to bound the $L^2$ distance of a product space in terms of the independent component $L^2$ distances. Since

$$\|\mu_t/\pi - 1\|_{2,\pi}^2 = \mathrm{E}_\pi(\mu_t/\pi)^2 - 1,$$

it follows that for $t = an^2$, we have

$$1 + \mathrm{Var}_\pi(\mu_t/\pi) \geq \exp(-ca)\Big(\prod_{i \in A} m\Big) = \exp(-ca) m^{n/2}\,.$$

This gives $\tau_2 = \Omega(n^2 \log N)$.

We observe the following from the above example:

- If the number of states is $N = m^n$, then we have a $\Theta(\log N)$ gap between the variation and $L^2$ mixing times.
- [10] shows that the relaxation time of Lamplighter chains (with 2-state lamps) equals the maximal hitting time of the base chain. The $\Omega(\log N)$ gap between $\tau_{tv}$ and $\tau_2$ also shows that in this $m$-state lamp case, the relaxation time of the chain is $\Theta(n^2)$.
- $\tau_{\text{ent}} = \Theta(\tau_{tv} \log\log N)$ here as well.
- Finally that we have a chain where the variation, entropy and $L^2$-mixing times are all different orders of magnitude.

Note that in the above case the variation mixing time and the relaxation time are of the same order of magnitude. To separate these, we need to separate the maximal hitting time and the expected cover time of the underlying chain. So it is natural to consider, the $m$-state lamplighter chain on the $\sqrt{n} \times \sqrt{n}$ *two-dimensional* torus on $n$ vertices. Recall that for this case of the torus, the maximal hitting time is $\Theta(n \log n)$ and the expected cover time is $\Theta(n \log^2 n)$. To show that the entropy time is still separated from the other times, first recall that the the cover time $C(T_n)$ of the torus on $n$ vertices satisfies, $C(T_n) \sim \frac{2}{\pi} n \log^2 n$, and moreover by the precise estimate given in Corollary 9.1 of [5], we also know that for $\beta < 1$, by time $\beta\, C(T_n)$, the set of uncovered vertices has size $n^{1-\beta+o(1)}$. Now using an argument as in the one-dimensional case above, we can get the bound, $\tau_{\text{ent}} = \Omega(n \log(n \log m))$.

## 4 Questions

Suppose $\mathbb{P}, \mathbb{Q}$ are Markov Chains on state space $\mathcal{X}, \mathcal{Y}$ respectively. The Lamplighter chain $\mathbb{Q} \wr \mathbb{P}$ has state space $\mathcal{Y}^{\mathcal{X}} \times \mathcal{X}$, i.e. a configuration of lamps $f$ together with the position of the lamplighter. [10] considered $\mathcal{Y} = \mathbb{Z}_2$ and $\mathbb{Q}$ on $\mathcal{Y}$ which completely randomizes the lamp in one step.

In [6] it is shown that the $L_2$-mixing time of the lamplighter chain on $\mathbb{Q} \wr \mathbb{P}$ is related to the following generalization of the moment generating function considered in this paper.

**Definition 14.** Let $\mathbb{P}$ be a Markov Chain on state space $\mathcal{X}$ and let $\gamma > 0$. For $S > 0$, let $\mathbf{Z}_S^\gamma$ denote the number of states that have been visited fewer than $\gamma$ times up till time $S$ and let $\zeta_S^\gamma(\theta) = \mathbb{E}[\theta^{\mathbf{Z}_S^\gamma}]$.

Quantities similar to $\mathbf{Z}_S^\gamma$, with first moment computations, were considered in [11] and [8]. The first paper to consider moment generating functions of Markov chain related quantities seems to be [10].

**Question 1.** For $\theta > 0, \delta > 0$, find bounds on

$$F(\mathbb{P}, \theta, \gamma, \delta) = \inf_S \{S : \zeta_S^\gamma(\theta) \le 1 + \delta\}\,.$$

For estimating the $L_2$-mixing time of $\mathbb{Q} \wr \mathbb{P}$, where $\mathbb{Q}$ is a Markov chain on $\mathcal{Y}$, the quantity of interest is $F(\mathbb{P}, |\mathcal{Y}|, \mathcal{T}_2(\mathbb{Q}, \epsilon/|\mathcal{X}|), \epsilon)$.

- If $\gamma \ge |\mathcal{X}| \log \theta$, then it is enough to take $S = O(\gamma|\mathcal{X}|)$.
- If $\mathbb{P}$ mixes in one step then it reduces to the coupon collector problem and one can show that for any $\gamma \ge 0$, it is enough to take $S = O((\gamma + \log|\mathcal{X}| + \log\theta)|\mathcal{X}|)$.
- In general, $O((\gamma + \log|\mathcal{X}| + \log\theta)|\mathcal{X}|T_{\mathrm{s}})$ is enough, where $T_{\mathrm{s}}$ is the separation time of $\mathbb{P}$. When $\mathbb{P}$ is reversible $T_{\mathrm{s}} = O(T_{\mathrm{tv}})$.

**Conjecture 1**. Let $\mathbb{P}$ be reversible with uniform stationary distribution and maximal hitting time $H \le \kappa|\mathcal{X}|$ for some constant $\kappa$. Then

$$F(\mathbb{P}, \theta, \gamma, \delta) \le C_\kappa \cdot |\mathcal{X}|(\gamma + T_{rel} + \log|\mathcal{X}| + \log\theta)\,, \tag{15}$$

for $C_\kappa$ depending only on $\kappa$.

**Acknowledgment**. We thank Yuval Peres for general encouragement and for help with examples in Section 3. We also thank the hospitality of the Theory group at Microsoft Research, where this work was completed.

## References

[1] D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Advances in Appl. Math.*, 8:69–97, 1987.

[2] D. Aldous and J.A. Fill. Preliminary version of a book on finite markov chains. URL http://www.stat.berkeley.edu/~aldous/RWG/book.html. In preparation.

[3] D. Aldous. Some inequalities for reversible Markov chains. *J. London Math. Soc. (2)*, 25(3): 564–576, 1982.

[4] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* John-Wiley & Sons, Inc., 1991.

[5] A. Dembo, Y. Peres, J. Rosen, and O. Zeitouni. Cover times for Brownian motion and random walks in two dimensions. *Ann. of Math.*, 160:433–464, 2004.

[6] M.K. Ganapathy. *Robust Mixing.* PhD thesis, University of Chicago, August 2006.

[7] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[8] J. Kahn, J.H. Kim, L. Lovász, and V.H. Vu. The cover time, the blanket time, and the matthews bound. In *41st Annual Symposium on Foundations of Computer Science*, pages 467–475, 2000.

[9] R. Montenegro and P. Tetali. Mathematical aspects of mixing times of markov chains. *Current and Future Trends in Theoretical Computer Science*, May 2006.

[10] Y. Peres and D. Revelle. Mixing times for random walks on finite lamplighter groups. *Electronic Journal of Probability*, 9:825–845, 2004. URL http://www.math.washington.edu/~ejpecp/EjpVol9/paper26.abs.html.

[11] P. Winkler and D. Zuckerman. Multiple cover time. *Random Structures and Algorithms*, 9(4): 403–411, 1996. URL citeseer.ist.psu.edu/139.html.