ALEA, Lat. Am. J. Probab. Math. Stat. 18, 1739–1758 (2021)
DOI: 10.30757/ALEA.v18-64

A probabilistic proof of Cooper & Frieze's "First Visit Time Lemma"

Francesco Manzo[1], **Matteo Quattropani**[2] **and Elisabetta Scoppola**[3]

[1] Dipartimento di Matematica e Fisica, Università di Roma Tre, Largo San Leonardo Murialdo 1, 00146 Roma, Italy.
E-mail address: manzo.fra@gmail.com

[2] Dipartimento di Economia e Finanza, LUISS, Viale Romania 32, 00197 Roma, Italy.
E-mail address: mquattropani@luiss.it

[3] Dipartimento di Matematica e Fisica, Università di Roma Tre, Largo San Leonardo Murialdo 1, 00146 Roma, Italy.
E-mail address: scoppola@mat.uniroma3.it

Abstract. We present an alternative proof of the so-called *First Visit Time Lemma* (FVTL), originally presented by Cooper and Frieze in its first formulation in Cooper and Frieze (2005), and then used and refined in a list of papers by Cooper, Frieze and coauthors. We work in the original setting, considering a growing sequence of irreducible Markov chains on n states. We assume that the chain is rapidly mixing and with a stationary measure with no entry being either too small nor too large. Under these assumptions, the FVTL shows the exponential decay of the distribution of the hitting time of a given state x—for the chain started at stationarity—up to a small multiplicative correction. While the proof by Cooper and Frieze is based on tools from complex analysis, and it requires an additional assumption on a generating function, we present a completely probabilistic proof, relying on the theory of quasi-stationary distributions and on strong-stationary times arguments. In addition, under the same set of assumptions, we provide some quantitative control on the Doob's transform of the chain on the complement of the state x.

1. Introduction

In the early 00's, Cooper and Frieze started a series of papers on which they compute the first order asymptotics of the cover time of random walks on different random graphs, see Cooper and Frieze (2012), Abdullah et al. (2012), Cooper and Frieze (2008), Cooper and Frieze (2007b), Cooper and Frieze (2007a), Cooper et al. (2013), Cooper et al. (2014). Given an arbitrary (possibly directed) graph structure, the cover time is the expected time needed by a simple random walk to visit every vertex of the graph, maximized over all the possible starting positions. One of the key ingredients

Received by the editors 2021-02-22; accepted 2021-07-16.
1991 *Mathematics Subject Classification.* 60J10, 60J28.
Key words and phrases. Hitting times, quasi-stationarity, Doob's transform, stationary times.
M.Q. was partially supported by the GNAMPA-INdAM Project 2020 "Random walks on random games" and PRIN 2017 project ALGADIMAR.

of Cooper and Freze's analysis is the so called *First Visit Time Lemma (FVTL)*, as named by the authors in Cooper and Frieze (2005). The same lemma has been of use in proving also different kind of results, e.g., to estimate expected meeting time of multiple random walks on random graphs, see Cooper et al. (2009/10). The lemma deals with the tail probability of the stopping time τ_x, i.e., the time of the first visit to the state x. Consider a sequence of Markov chains on a growing state space of size n. We assume that for every sufficiently large n the chain is irreducible, admitting a unique invariant measure $\pi = \pi_n$. The framework of the lemma is based on two additional crucial assumptions relating mixing time and spread of the stationary measure, namely, we assume the existence of a time $T = T_n$ such that

$$\max_{x,y} \left| P^T(x,y) - \pi(y) \right| = O\left(\frac{1}{n^3}\right), \tag{1.1}$$

and

$$T \max_x \pi(x) = o(1), \qquad \min_x \pi(x) = \omega(n^{-2}). \tag{1.2}$$

Under the latter assumptions and adding a technical requirement on the generating function of the recurrences to a fixed state x, the authors show that starting from any state y and for all $t > T$:

$$\mathbb{P}_y\left(X_s \neq x,\ \forall s \in [T,t]\right) \sim \left(1 - \frac{\pi(x)}{R_T(x)}\right)^t, \tag{1.3}$$

where $(X_s)_{s\geq 0}$ is the trajectory of the random walk and $R_T(x) \geq 1$ is the expected number of returns in x within the mixing time T. The proof of the latter results, as well as the underlying technical assumptions, evolved with their uses since the first formulation in Cooper and Frieze (2005) to the last (to the best of our knowledge) formulation and proof in Cooper and Frieze (2008). We remark that the assumptions in (1.1)-(1.2) are typically satisfied by random walks on many models of random graphs, e.g., Erdős-Renyi graphs or configuration models.

The techniques used in the proof by Cooper and Frieze rely on probability arguments but also on tools from complex analysis and an analytical expansion of some probability generating functions. In this paper we aim at finding a probabilistic proof of the FVTL, trying to shed some light on the underlying phenomenology. On the technical side, the arguments in our proof are elementary and do not need the additional assumption on the generating function required in the original Cooper and Frieze's proof. We refer to Section 2.2 for a direct comparison of our result with the original one.

Exponential law of hitting times is a classic and widely studied topic in probability. We just recall here the pioneering book by Keilson (1979), the beautiful book by Aldous (1989) and the papers Aldous (1982), Aldous and Brown (1992, 1993). In Aldous (1982), the author recognizes two regimes in which the latter phenomenon takes place:

(1) A single state m is frequently visited before τ_x. When starting from m, the path to x consists of a geometric number of excursions (with mean $(\mathbb{P}_m(\tau_m > \tau_x)^{-1})$ from m to m without touching x, before the final journey to x. The hitting time is dominated by the sum of many i.i.d. excursion times and therefore it is almost exponential. See Keilson (1979).
(2) When the chain is rapidly mixing, then the distribution at time t is near to the stationary distribution even when conditioned on $\tau_x > t$. This case is analyzed in Aldous (1982), where it is shown that

$$\sup_{t\geq 0} \left| \mathbb{P}_\pi(\tau_x > t) - e^{-\frac{t}{\mathbb{E}_\pi[\tau_x]}} \right| \leq \delta, \tag{1.4}$$

where δ is a function of the mixing time of the chain and of the expectation $\mathbb{E}_\pi[\tau_x]$. Aldous shows that, if the hitting of x is a rare event, i.e., the expectation of τ_x is much larger than the mixing time of the chain, then δ is small. Moreover, the heuristic argument developed in Aldous (1989) suggests that the expected hitting time $\mathbb{E}_\pi[\tau_x]$ is well approximated by the ratio between two quantities: the local time spent in x (starting at x) before the process is *well-mixed*, and the stationary distribution at x. As we will see in what follows, this is exactly what is rigorously stated in the First Visit Time Lemma.

In the early years, these two regimes were considered as complementary. One of the main applications of the scenario in (1) has been the study of metastability, namely the behavior of processes that are trapped for a long time in a part of their state space. Before exiting the trap, the process visits many times a "metastable state", reaching an apparent, local equilibrium. In such systems the exit from the trap triggers the relaxation to equilibrium so that relaxation to equilibrium can be discussed as the first hitting to the complement of the trap. We refer to Olivieri and Vares (2005); Bovier and den Hollander (2015) for a general introduction to metastability and to Bianchi and Gaudillière (2016); Bianchi et al. (2020); Fernandez et al. (2015, 2016); Manzo and Scoppola (2019) for a discussion of the extension of metastability methods to other regimes.

In Aldous (1982), scenario (2) is studied by using the quasi-stationary measure introduced in the pioneering paper by Darroch and Seneta (1965) (see also Collet et al. (2013), and Pollett (2008) for a more recent bibliography on the subject). Aldous shows that rapid mixing implies that the stationary and the quasi-stationary measures are closed to each other in total variation distance if $\pi(x)$ is small. By using the fact that the distribution of τ_x is exactly exponential when starting from the quasi-stationary measure, the result in (1.4) is proved by a smart sequence of elementary steps. We have to note that Aldous' result concerns additive error bounds and therefore it cannot provide first-order asymptotics of the exponential approximation when t is large, in contrast to the FVTL where a multiplicative bound is proved.

More recently, these two regimes begin to be understood in a common framework, by generalizing recurrence ideas to measures instead of recurrence to points. The quasi-stationary measure plays the role of a recurrent measure before the hitting. The hitting to the measure can be studied by extending the theory of strong stationary times Aldous and Diaconis (1986, 1987a); Levin and Peres (2017), to quasi-stationarity, see Diaconis and Miclo (2009); Manzo and Scoppola (2019). In particular, the notion of *conditional strong quasi-stationary time (CSQST)* introduced in Manzo and Scoppola (2019), has shown to be useful in providing exact formulas for the distribution of the first hitting time τ_x starting from an arbitrary distribution. An introduction to these tools is given in the following subsection where a rough estimate on the tail of τ_x is given, while a more detailed discussion is provided in Section 5 where we adopt a CSQST perspective, allowing us to consider general regimes and different starting measures.

Under the hypotheses considered in this paper, in particular (1.2), π and the quasi-stationary measure are close in a multiplicative sense, so that we can follow a simple line of argument involving the quasi-stationary measure but not requiring the use of CSQST.

1.1. *A first discussion.* For any $x \in \mathcal{X}$, let τ_x denote the hitting time of x, namely

$$\tau_x = \inf\{t \geq 0 \mid X_t = x\}. \tag{1.5}$$

We will call $[P]_x$ the sub-Markovian probability kernel obtained by removing the x-th row and column by the matrix P. We will assume that $[P]_x$ is a primitive sub-Markovian kernel, i.e., all entries of $([P]_x)^m$ are positive for some $m \in \mathbb{N}$. By the Perron-Froboenius theory (see, e.g., Collet et al. (2013)) there exists a unique probability distribution $\mu_x^\star$ and a real $\lambda_x < 1$

$$\mu_x^\star [P]_x = \lambda_x \mu_x^\star, \tag{1.6}$$

Moreover, we denote by γ_x the corresponding right eigenvector, i.e.,

$$[P]_x \gamma_x = \lambda_x \gamma_x, \tag{1.7}$$

normalized by $\langle \gamma_x, \mu_x^\star \rangle = 1$.

The probability distribution $\mu_x^\star$ is called *quasi-stationary measure* and it is strictly related to the exponential behavior of the tail probability $\mathbb{P}(\tau_x > t)$. Indeed, when looking at the evolution of the process starting from $\mu_x^\star$, by 1.6 we can compute

$$\mathbb{P}_{\mu_x^\star}(\tau_x > t) = \sum_z \mu_x^\star(z)\mathbb{P}_z(\tau_x > t) = \sum_{z\neq x} \mu_x^\star(z) \sum_{y\neq x} \left([P]_x\right)^t(z,y) = \lambda_x^t \sum_{y\neq x} \mu_x^\star(y) = \lambda_x^t. \tag{1.8}$$

For more details see Diaconis and Miclo (2009, 2015); Miclo (2010, 2020), the applications to the metastability regime are discussed in Fernandez et al. (2015, 2016); Manzo and Scoppola (2019).

The right eigenvector γ_x defined in 1.7 controls the dependence on the initial distribution of the probability of the event $\tau_x > t$. Indeed this eigenvector is related to the asymptotic ratios of the right tail probabilities, see Collet et al. (2013, Eq. (3.5))

$$\lim_{t\to\infty} \frac{\mathbb{P}_y(\tau_x > t)}{\mathbb{P}_z(\tau_x > t)} = \frac{\gamma_x(y)}{\gamma_x(z)} \qquad y, z \neq x. \tag{1.9}$$

With this right eigenvector we can construct a *Local Chain* on $\mathcal{X} \setminus \{x\}$, which is usually referred to as *Doob's transform of* X. For any $y, z \neq x$, define the stochastic matrix

$$\widetilde{P}(z,y) := \frac{\gamma_x(y)}{\gamma_x(z)} \frac{P(z,y)}{\lambda_x}. \tag{1.10}$$

More generally

$$\widetilde{P}^t(z,y) = \frac{\gamma_x(y)}{\gamma_x(z)} \frac{\left([P]_x\right)^t(z,y)}{\lambda_x^t} \qquad \forall t \geq 0. \tag{1.11}$$

It is immediate to show that $\widetilde{P}$ is a primitive matrix and has invariant measure

$$\nu(y) := \gamma_x(y)\mu_x^\star(y).$$

For the chain $\widetilde{X}$ we define

$$\tilde{s}^z(t,y) := 1 - \frac{\widetilde{P}^t(z,y)}{\nu(y)} \tag{1.12}$$

and will call separation distance at time t the quantity $\tilde{s}(t)$ defined as

$$\tilde{s}(t) := \sup_{z\neq x} \tilde{s}^z(t) \qquad \text{where} \qquad \tilde{s}^z(t) := \sup_{y\neq x} \tilde{s}^z(t,y). \tag{1.13}$$

Note that $\tilde{s}^z(t) \in [0,1]$ and recall that $\tilde{s}(t)$ has the sub-multiplicative property

$$\tilde{s}(t+u) \leq \tilde{s}(t)\tilde{s}(u),$$

which in particular implies an exponential decay in time of $\tilde{s}$, see Levin and Peres (2017).

Consider any initial measure α on $\mathcal{X} \setminus \{x\}$ and define the transformation

$$\tilde{\alpha}(y) := \frac{\alpha(y)\gamma_x(y)}{\langle \alpha, \gamma_x \rangle}, \qquad \forall y \neq x. \tag{1.14}$$

Then, as shown in Manzo and Scoppola (2019),

$$\mathbb{P}_\alpha(\tau_x > t) = \sum_{y\neq x}\sum_{z\neq x} \alpha(z)\left([P]_x\right)^t(z,y) \tag{1.15}$$

$$= \sum_{y\neq x}\sum_{z\neq x} \alpha(z)\gamma_x(z)\lambda_x^t \mu_x^\star(y) \frac{\widetilde{P}^t(z,y)}{\nu(y)} \tag{1.16}$$

$$= \lambda_x^t \sum_{z\neq x} \alpha(z)\gamma_x(z) \sum_{y\neq x} \mu_x^\star(y)(1 - \tilde{s}^z(t,y)) \tag{1.17}$$

$$= \lambda_x^t \langle \alpha, \gamma_x \rangle \left(1 - \sum_{y\neq x} \mu_x^\star(y)\tilde{s}^{\tilde{\alpha}}(t,y)\right) \tag{1.18}$$

where we call

$$\tilde{s}^{\tilde{\alpha}}(t,y) := \sum_{z\neq x} \tilde{\alpha}(z)\tilde{s}^{z}(t,y) \qquad \text{and} \qquad \tilde{s}^{\tilde{\alpha}}(t) := \sup_{y\neq x} \tilde{s}^{\tilde{\alpha}}(t,y). \tag{1.19}$$

Moreover, again by Manzo and Scoppola (2019), we know that (1.18) can be estimated from above and below by

$$\lambda_x^t \langle \alpha, \gamma_x\rangle \left(1-\tilde{s}^{\tilde{\alpha}}(t)\right) \le \mathbb{P}_\alpha(\tau_x > t) \le \lambda_x^t \langle \alpha, \gamma_x\rangle \left(1+\tilde{s}^{\tilde{\alpha}}(t)\left(\frac{1}{\min_y \gamma_x(y)}-1\right)\right). \tag{1.20}$$

(1.20) suggests that, in the regime in which $|\mathcal{X}| \to \infty$, the first order geometric approximation of the tail probability $\mathbb{P}_\alpha(\tau_x > t)$ can be obtained. In particular, the exponentiality immediately follows from (1.20) for all those Markov chains P, target states x, initial distributions α and time t for which all of the following assumptions hold:

1. $\tilde{s}^{\tilde{\alpha}}(t) = o(1)$, i.e., t is sufficiently large to have that the Doob transform starting at α is well mixed by time t;
2. $\langle \alpha, \gamma_x\rangle \sim 1$, which occurs in particular if γ_x approximates the constant vector;
3. $\min_y \gamma_x(y) = \Omega(1)$, which can be thought of as an additional uniformity requirement to the one in (2).

Despite the intuitions based on (1.20), we are not going to follow exactly the heuristic recipe explained in the list above. In fact our focus is on the special case in which $\alpha = \pi$, which leaded us through a different path toward proving exponentiality. Nevertheless, as a byproduct of our proof of the FVTL we provide uniform upper and lower bounds on the right eigenvector γ_x. We think those bounds can be of independent interest, since they can be turned into a quantitative information on the structure of the Doob's transform of the process X. In particular, for a given model, our bounds could be useful in verifying the conditions in the list above, and therefore in finding—for every fixed choice of the initial distribution α—the right first order approximation of the decay of $\mathbb{P}_\alpha(\tau_x > t)$.

2. Notation and results

We start by presenting the notation and briefly recalling the basic quantities introduced in Section 1. We consider a sequence of Markov chains on a growing state space. Formally:

- $\mathcal{X}^{(n)}$ is a state space of size n.
- $(X^{(n)})_{t\ge 0}$ is a discrete time Markov chain on $\mathcal{X}^{(n)}$.
- $\mathbb{P}^{(n)}$ is the probability law of the Markov chain $(X^{(n)})_{t\ge 0}$, and $\mathbb{E}^{(n)}$ the corresponding expectation.
- $P^{(n)}$ is the transition matrix of $(X^{(n)})_{t\ge 0}$, which is assumed to be ergodic.
- $\pi^{(n)}$ is the stationary distribution of $P^{(n)}$.
- For any probability distribution α on $\mathcal{X}^{(n)}$ and every integer $t \ge 0$, we note by μ_t^α the probability distribution of the chain $X^{(n)}$ starting at α and evolved for t steps, i.e.,
$$\mu_t^\alpha(y) := \sum_{x\in\mathcal{X}^{(n)}} \alpha(x)\big(P^{(n)}\big)^t(x,y), \qquad \forall y \in \mathcal{X}^{(n)}.$$
- For all $x \in \mathcal{X}^{(n)}$, τ_x represents the hitting time of vertex x, defined as in (1.5).
- For all $t \ge 0$ and $x \in \mathcal{X}^{(n)}$, we let the symbol $\zeta_t(x)$ denote the random time spent by the process $X^{(n)}$ in the state x within time t, i.e.,
$$\zeta_t(x) := \sum_{s=0}^{t-1} \mathbf{1}_{X_s^{(n)}=x}. \tag{2.1}$$

- For all $x \in \mathcal{X}^{(n)}$ we denote by $[P^{(n)}]_x$ the sub-Markovian kernel obtained by removing the x-th row and column of $P^{(n)}$. The kernel $[P^{(n)}]_x$ is assumed to be irreducible.
- For all $x \in \mathcal{X}^{(n)}$, λ_x denotes as the leading eigenvalue of $[P^{(n)}]_x$ and $\mu_x^\star$ as the corresponding left eigenvector, normalized so that $\mu_x^\star$ is a probability distribution over $\mathcal{X}^{(n)} \setminus \{x\}$. See (1.6). We remark that by the definitions follows that

$$\mathbb{P}_{\mu_x^\star}(\tau_x > t) = \lambda_x^t, \qquad \forall t \geq 0, \tag{2.2}$$

see (1.8).
- For all $x \in \mathcal{X}^{(n)}$, γ_x denotes the right eigenvector of $[P^{(n)}]_x$ associated to the eigenvalue λ_x. We consider γ_x to be normalized so that $\langle \mu_x^\star, \gamma_x \rangle = 1$.

Since we are interested in asymptotic results when $n \to \infty$, the asymptotic notation will refer to this limit and the explicit dependence on n will be usually dropped.

We will adopt the usual asymptotic notation $(o, O, \Theta, \omega, \Omega)$ and, given two functions $f, g : \mathbb{N} \to \mathbb{R}_+$, we will use the symbols $\sim$ and $\lesssim$ with the meaning

$$f(n) \sim g(n) \qquad \iff \qquad \lim_{n\to\infty} \frac{f(n)}{g(n)} = 1,$$

and

$$f(n) \lesssim g(n) \qquad \iff \qquad \limsup_{n\to\infty} \frac{f(n)}{g(n)} \leq 1,$$

respectively.

2.1. *Results.* We will work under the following asymptotic assumption for the sequence of Markov chains: There exist

- A real number $c > 2$.
- A diverging sequence $T = T(n)$.

such that

(HP1) Fast mixing:

$$\max_{x,y\in\mathcal{X}} |\mu_T^x(y) - \pi(y)| = o(n^{-c}).$$

(HP2) Small $\pi_{\max}$:

$$T \max_{x\in\mathcal{X}} \pi(x) = o(1).$$

(HP3) Large $\pi_{\min}$:

$$\min_{x\in\mathcal{X}} \pi(x) = \omega(n^{-2}).$$

Remark 2.1. It is worth to stress that, under (HP3), the assumption in (HP1) can be read as a L^∞ mixing assumption, namely

(HP1bis) Fast L^∞ mixing:

$$\max_{x\in\mathcal{X}} \|\mu_T^x\|_{\pi,\infty} := \max_{x,y\in\mathcal{X}} \left| \frac{\mu_T^x(y)}{\pi(y)} - 1 \right| = o\big(n^{-(c-2)}\big).$$

Fixed any $x \in \mathcal{X}$ we let $R_T(x)$ denote the expected number of returns at x for the Markov chain starting at x within T. More precisely,

$$R_T(x) = \sum_{t=0}^{T} \mu_t^x(x) \geq 1. \tag{2.3}$$

The precise statement that we prove is the following

Theorem 2.2 (First Visit Time Lemma)**.** *Under the assumptions* (HP1)*,* (HP2) *and* (HP3) *for all* $x \in \mathcal{X}$*, it holds*

$$\sup_{t \geq 0} \left| \frac{\mathbb{P}_\pi(\tau_x > t)}{\lambda_x^t} - 1 \right| \longrightarrow 0, \tag{2.4}$$

and

$$\left| \frac{\lambda_x}{\left(1 - \frac{\pi(x)}{R_T(x)}\right)} - 1 \right| \longrightarrow 0. \tag{2.5}$$

We will see in 4 that it follows as an easy consequence of Theorem 2.2 that the right-eigenvector γ_x asymptotically has mean 1 with respect to the stationary distribution. In other words, the following corollary holds.

Corollary 2.3. *Under the same assumptions of Theorem 2.2: for all* $x \in \mathcal{X}$

$$\sum_{y \in \mathcal{X} \setminus \{x\}} \pi(y)\gamma_x(y) \to 1. \tag{2.6}$$

Moreover, we provide some entry-wise upper and lower bound for the eigenvector γ_x.

Theorem 2.4. *Under the same set of assumptions, for every* $x \in \mathcal{X}$*:*

(1) *For all* $y \in \mathcal{X} \setminus \{x\}$ *it holds*

$$\gamma_x(y) \lesssim 1.$$

(2) *For all* $y \in \mathcal{X} \setminus \{x\}$ *it holds*

$$\gamma_x(y) \gtrsim \left[1 - \mathbb{E}_y\left[\zeta_T(x)\right]\right]_+.$$

Remark 2.5. We remark that the asymptotic lower bound in Theorem 2.4 is in fact not void for most of the models of random graphs which are known to satisfy the assumptions of the FVTL. As an example, if X is the simple random walk on a random regular directed graph of in/out-degree r, then—with high probability with respect to the construction of the environment—for every $x \in \mathcal{X}$ the quantity $\mathbb{E}_y[\zeta_T(x)]$ is strictly smaller than 1 uniformly in $y \in \mathcal{X} \setminus \{x\}$; moreover, $\mathbb{E}_y[\zeta_T(x)] = 0$ for most $y \in \mathcal{X} \setminus \{x\}$. To see the validity of the latter statement, we refer the reader to Caputo and Quattropani (2020, Propositions 4.3 and 4.4).

2.2. *Comparison with Cooper&Frieze's lemma.* In order to facilitate a direct comparison, we write here—using our notation—the claim proved by Cooper and Frieze, stressing the differences with Theorem 2.2.

Theorem 2.6 (See Lemma 6 and Corollary 7 in Cooper and Frieze, 2008)**.** *Consider a sequence of Markov chains satisfying the assumptions* (HP1)*,* (HP2) *and* (HP3) *with* $c = 3$*. Moreover, let*

$$a = \frac{1}{KT}$$

for a suitably large constant K*. Fix* $x \in \mathcal{X}$ *and assume further that the truncated probability generating function*

$$\mathbf{R}(z) = \sum_{t=0}^{T-1} P^t(x,x) z^t, \qquad \forall z \in \mathbb{C}$$

satisfies

$$\min_{|z| \leq 1+a} |\mathbf{R}(z)| \geq \theta \tag{2.7}$$

for some constant $\theta > 0$*. Then, for all* $y \in \mathcal{X}$ *and* $t \geq 0$

$$\mathbb{P}_{\mu_y^T}(\tau_x > t) = \left(1 + O(T\pi(x))\right)\tilde{\lambda}_x^t + o\left(e^{-at/2}\right), \tag{2.8}$$

where

$$\tilde{\lambda}_x = \left(1 + \frac{\pi(x)}{R_T(x)(1+O\left(T\pi(x)\right))}\right)^{-1}.$$

Even at a first sight, there are three main differences between Theorem 2.6 and Theorem 2.2:

(1) First, our proof neglects the technical assumption in (2.7). Indeed, we remark once again we are not going to use any tool from complex analysis, being our proof elementary and completely probabilistic in nature.
(2) Second, the estimate in (2.8) concerns the tail probability of the hitting time when the initial measure is the T-step evolution starting at any fixed vertex y. The latter is in fact a minor difference. In Lemma 3.3 we will show that our estimate in (2.4) holds even when replacing π by μ_y^T, for any choice y.
(3) Finally, our result does not take into account the precise magnitude of the second order corrections. This is because we would like to put the accent of this paper on the underlying phenomenology, trying to keep the paper as easy and readable as possible. We stress that more precise bounds could be obtained through the same set of arguments.

2.3. *First visit to a set.* We now briefly discuss how to extend our result to the more general setting in which one is interested in the hitting time of a set of states. Fix some target set of states $G \subset \mathcal{X}$, and consider the hitting time τ_G. It is natural to ask under which conditions on the chain $(\mathcal{X}, P)$ and on the set G the exponential approximation in (2.3)-(2.4) holds. We stress that the theory of quasi-stationary distributions has been originally developed for such more general setting. The standard approach used by Cooper and Frieze in their series of papers is that of *collapsing* the states in the set G into a new state, g, in a way to preserve the stationary distribution of the other vertices. More precisely, consider the Markov chain on $\tilde{\mathcal{X}} = (\mathcal{X} \setminus G) \cup \{g\}$ with transition matrix $\tilde{P}$ defined as follows

$$\tilde{P}(x,y) := \begin{cases} P(x,y) & \text{if } x, y \neq g \\ \sum_{z\in G} P(x,z) & \text{if } x \neq g,\ y = g \\ \sum_{z\in G} \frac{\pi(z)}{\pi(G)} P(z,y) & \text{if } x = g,\ y \neq g \\ \sum_{v\in G}\sum_{z\in G} \frac{\pi(z)}{\pi(G)} P(z,v) & \text{if } x = y = g \end{cases} \tag{2.9}$$

Called

$$\tilde{\pi}(x) := \begin{cases} \pi(x) & \text{if } x \neq g \\ \pi(G) & \text{if } x = g \end{cases} \tag{2.10}$$

it is immediate to check that

$$\tilde{\pi} = \tilde{\pi}\tilde{P}. \tag{2.11}$$

In other words, the chain $(\tilde{\mathcal{X}}, \tilde{P})$ has the same behavior of $(\mathcal{X}, P)$ out of the set G and the stationary distribution $\tilde{\pi}$ is simply the projection of π on $\tilde{\mathcal{X}}$. Therefore the first hitting time to the set G under the original chain $(\mathcal{X}, P)$, starting at π, has the same law of the first hitting time to state g under the chain $(\tilde{\mathcal{X}}, \tilde{P})$ starting at $\tilde{\pi}$. Notice further that $[P]_G = [\tilde{P}]_g$. Hence, it is enough to check that the modified Markov chain satisfies assumptions (HP1)-(HP2)-(HP3) and consider $x = g$ in Theorem 2.2.

Recall that in most of the cases considered by Cooper and Frieze the underlying Markov chain is the simple random walk on an undirected graph. In order to prove the lower bound on the cover time, the strategy of the authors requires to control the hitting time of a pair of vertices, which are assumed to be "far" in the graph. Take $G = \{x, y\}$ for a pair of vertices $x, y \in \mathcal{X}$ having distance at least 2. Notice that, being the underlying graph undirected, the stationary distribution π is proportional to the degree sequence. Therefore, the associated chain $(\tilde{\mathcal{X}}, \tilde{P})$ coincides with the simple random walk on the graph in which the vertices x and y have been collapsed in a single vertex

g and all of their edges are preserved by g. If the original graph is a sparse connected expander, then it is not hard to see that (HP1)-(HP2)-(HP3) are satisfied by such modified chain[1], and the required control follows again by Theorem 2.2. More in general, such a "collapsing" strategy can work only if the set G is not too large, both in the sense of cardinality and in the sense of its stationary distribution.

2.4. *Overview of the paper.* Section 3 is devoted to the proof of Theorem 2.2. The proof is divided into several steps. We start by showing a first order approximation for the expected hitting time of x starting at stationarity, i.e. $\mathbb{E}_\pi[\tau_x] \sim R_T(x)/\pi(x)$. See Proposition 3.1. In order to show that the latter expectation coincides at first order with $\mathbb{E}_{\mu_x^\star}[\tau_x]$ we prove that the tail probability $\mathbb{P}_\pi(\tau_x > t)$ is asymptotically larger or equal to the the tail of the same probability starting at any other measure. This is the content of Proposition 3.4. To conclude the validity of

$$\mathbb{E}_\pi[\tau_x] \sim \mathbb{E}_{\mu^\star}[\tau_x] = (1-\lambda_x)^{-1}, \tag{2.12}$$

we then use a bootstrap argument: we first show in Lemma 3.9 that $\lambda_x^T \sim 1$, then—in Proposition 3.8—we show that the latter bound can be translated in the sharper estimate in (2.12). Once established (2.12), the exponential approximation can be obtained by using the properties of quasi-stationary distributions.

In Section 4 we use the understanding developed in Section 3 to show the validity of Corollary 2.3 and Theorem 2.4. Namely, we see how the FVTL reflects on the properties of the first right-eigenvector γ_x.

Finally, in Section 5, we aim at framing the FVTL and its setting in the language of *conditional strong quasi-stationary times* introduced in Manzo and Scoppola (2019).

3. Proof of the FVTL

As mentioned in Section 2.4, our proof of Theorem 2.2 is divided into several small steps. The first proposition is devoted to the computation of the average hitting time of x starting at stationarity. The credits for this result go to Abdullah, who presented it in his PhD thesis, Abdullah (2012, Lemma 58). We repeat here the proof for the reader's convenience.

Proposition 3.1 (see Abdullah, 2012)**.** *For all $x \in \mathcal{X}$*

$$\mathbb{E}_\pi[\tau_x] \sim \frac{R_T(x)}{\pi(x)}. \tag{3.1}$$

Proof: By Aldous and Fill (2002, Lemma 2.1) we have

$$\mathbb{E}_\pi[\tau_x] = \frac{Z(x,x)}{\pi(x)},$$

where Z is the so called *fundamental matrix*, defined by

$$Z(x,x) := \sum_{t=0}^{\infty} \mu_t^x(x) - \pi(x). \tag{3.2}$$

By the submultiplicativity of the sequence

$$D(t) := \max_{x,y} \left|\mu_t^x(y) - \pi(y)\right|, \tag{3.3}$$

i.e.,

$$D(t+s) \le 2D(t)D(s), \qquad \forall t,s>0, \tag{3.4}$$

[1] Using, e.g., $T = \log^3(n)$.

and thanks to (HP1), we have

$$\max_{x,y} |\mu^x_{kT}(y) - \pi(y)| \le \left(\frac{2}{n^c}\right)^k, \qquad \forall k \in \mathbb{N}. \tag{3.5}$$

Hence,

$$\begin{aligned} Z(x,x) =& \sum_{t\le T} \left(\mu^x_T(x) - \pi(x)\right) + T\sum_{k\ge 1}\left(\frac{2}{n^c}\right)^k \\ =& R_T(x) + O(T\pi(x)) + O(Tn^{-c}) \\ =& R_T(x)(1+o(1)), \end{aligned}$$

where in the latter asymptotic equality we used $Tn^{-c} \le T\pi_{\max}$, (HP2), and the fact that $R_T(x) \ge 1$. ☐

Remark 3.2. We remark that, by the *eigentime identity* (see Aldous and Fill, 2002; Pitman and Tang, 2018; Miclo, 2015) the trace of the fundamental matrix of an irreducible chain coincides with the sum of the inverse non-null eigenvalues of the generator, which in turn coincide with the expected hitting time of a state sampled accordingly to the stationary distribution. Namely, for all $y \in \mathcal{X}$,

$$\sum_{x\in\mathcal{X}} \pi(x)\mathbb{E}_y\tau_x = \sum_{x\in\mathcal{X}} Z(x,x) = \sum_{i=2}^n \frac{1}{1-\theta_i} \tag{3.6}$$

where

$$1 = \theta_1 > \Re(\theta_2) \ge \cdots \ge \Re(\theta_n) \ge -1$$

are the eigenvalues of P. By Proposition 3.1 we get that, for all $y \in \mathcal{X}$,

$$\sum_{x\in\mathcal{X}} Z(x,x) \sim \sum_{x\in\mathcal{X}} R_T(x). \tag{3.7}$$

In other words, under the assumptions in (HP1), (HP2) and (HP3), the sum of the inverse eigenvalues of $I-P$ can be well approximated by the sum of the expected returns within the mixing time.

A crucial fact that will be used repeatedly in what follows is that under the assumptions in Section 2.1, the tails of τ_x starting at μ^y_T and starting at π coincide at first order.

Lemma 3.3. *For all $x, y \in \mathcal{X}$ and $t > 0$ it holds*

$$\mathbb{P}_{\mu^y_T}(\tau_x > t) \sim \mathbb{P}_\pi(\tau_x > t). \tag{3.8}$$

Proof: By the assumptions (see Remark 2.1) we have that

$$\begin{aligned} \max_{x,y\in\mathcal{X}} \left|\frac{\mu^x_T(y)}{\pi(y)} - 1\right| =& \max_{x,y\in\mathcal{X}} \frac{1}{\pi(y)} |\mu^x_T(y) - \pi(y)| & (3.9) \\ \le& \frac{1}{\min_{y\in\mathcal{X}}\pi(y)} \max_{x,y\in\mathcal{X}} |\mu^x_T(y) - \pi(y)| & (3.10) \\ \text{By (HP1)} \implies \quad \le& \frac{n^{-c}}{\min_{y\in\mathcal{X}}\pi(y)} & (3.11) \\ \text{By (HP3)} \implies \quad =& o(n^{-c+2}), & (3.12) \end{aligned}$$

from which the claim follows. In fact,

$$\begin{aligned} \mathbb{P}_{\mu^T_y}(\tau_x > t) =& \sum_z \mu^T_y(z)\mathbb{P}_x(\tau_x > t) \\ =& (1+o(1))\sum_z \pi(z)\mathbb{P}_x(\tau_x > t) \end{aligned}$$

$$=(1+o(1))\mathbb{P}_\pi(\tau_x > t). \qquad \square$$

The next proposition shows that under the assumptions in Section 2.1 the tail of the hitting time τ_x starting at π coincides—asymptotically—with the tail of τ_x starting at the "furthest" vertex.

Proposition 3.4. *For all $x \in \mathcal{X}$ and for all $t > T$ it holds*

$$\max_{y\in\mathcal{X}} \mathbb{P}_y(\tau_x > t) \sim \mathbb{P}_\pi(\tau_x > t). \tag{3.13}$$

We start by proving a preliminary version of Proposition 3.4, which is expressed by the following lemma.

Lemma 3.5. *For all $x \in \mathcal{X}$ and for all $t > T$ it holds*

$$\max_{y\in\mathcal{X}} \mathbb{P}_y(\tau_x > t) \lesssim \mathbb{P}_\pi(\tau_x > t - T). \tag{3.14}$$

Proof: For all $x, y \in \mathcal{X}$ it holds

$$\mathbb{P}_y(\tau_x > t) = \sum_{z\in\mathcal{X}} \mathbb{P}_y(X_T = z;\ \tau_x > T)\mathbb{P}_z(\tau_x > t - T) \tag{3.15}$$

$$\leq \sum_{z\in\mathcal{X}} \mathbb{P}_y(X_T = z)\mathbb{P}_z(\tau_x > t - T) \tag{3.16}$$

$$=(1+o(1)) \sum_{z\in\mathcal{X}} \pi(z)\mathbb{P}_z(\tau_x > t - T) \tag{3.17}$$

$$\sim\mathbb{P}_\pi(\tau_x > t - T). \qquad \square$$

Roughly, given Lemma 3.5, the proof of Proposition 3.4 follows by showing that the $-T$ term in the right hand side of (3.14) does not affect the asymptotic relation. This fact is made rigorous by Lemma 3.6 and the forthcoming Corollary 3.7. The proof of Lemma 3.6 is based on strong stationary times techniques (see Aldous and Diaconis, 1987b; Diaconis and Fill, 1990; Levin and Peres, 2017) and it is inspired by the recursion in the proof of Fernandez et al. (2016, Lemma 5.4). Before to proceed with the proof, we need to recall some definitions and properties of strong stationary times.

A randomized stopping time τ_π^α is a *Strong Stationary Time (SST)* for the Markov chain X_t with starting distribution α and stationary measure π, if for any $t \geq 0$ and $y \in \mathcal{X}$

$$\mathbb{P}_\alpha\left(X_t = y, \tau_\pi^\alpha = t\right) = \pi(y)\mathbb{P}_\alpha\left(\tau_\pi^\alpha = t\right),$$

which is equivalent to

$$\mathbb{P}_\alpha\left(X_t = y \middle| \tau_\pi^\alpha \leq t\right) = \pi(y). \tag{3.18}$$

If τ_π^α is a SST then

$$\mathbb{P}_\alpha(\tau_\pi^\alpha > t) \geq \mathrm{sep}(\mu_t^\alpha, \pi) := \max_{y\in\mathcal{X}} \left[1 - \frac{\mu_t^\alpha(y)}{\pi(y)}\right], \qquad \forall t \geq 0, \tag{3.19}$$

and when 3.19 holds with the equal sign for every t, the SST is *minimal*. Moreover, a minimal SST always exists, see Levin and Peres (2017, Prop. 6.14).

Lemma 3.6. *For any $t > 0$ it holds*

$$\frac{\mathbb{P}_\pi(\tau_x > t + T)}{\mathbb{P}_\pi(\tau_x > t)} \geq 1 - o(1). \tag{3.20}$$

Proof: We first prove the following inequality

$$\frac{\mathbb{P}_\pi(\tau_x > t + T)}{\mathbb{P}_\pi(\tau_x > t)} \geq 1 - \varepsilon \cdot \frac{\mathbb{P}_\pi(\tau_x > t - T)}{\mathbb{P}_\pi(\tau_x > t)}, \tag{3.21}$$

with $\varepsilon = o(1)$. We start by rewriting

$$\mathbb{P}_\pi(\tau_x > t + T) = \mathbb{P}_\pi(\tau_x > t) - \mathbb{P}_\pi(\tau_x \in [t, t+T]). \tag{3.22}$$

Consider τ_π^z the minimal SST of the process started at z, so that the last term in (3.22) can be written as

$$\begin{aligned}\mathbb{P}_\pi(\tau_x \in [t,t+T]) &= \sum_{z\in\mathcal{X}} \mathbb{P}_\pi(\tau_x > t-T, X_{t-T}=z)\mathbb{P}_z(\tau_x \in [T,2T]) \\ &\le \sum_{z\in\mathcal{X}} \mathbb{P}_\pi(\tau_x > t-T, X_{t-T}=z)\times \\ &\quad \times \Big[\mathbb{P}_z(\tau_x \in [T,2T], \tau_\pi^z \le T) + \mathbb{P}_z(\tau_x \le 2T, \tau_\pi^z > T)\Big] \\ &\le \mathbb{P}_\pi(\tau_x > t-T)\Big[\mathbb{P}_\pi(\tau_x \le 2T) + \max_{z\in\mathcal{X}} \mathbb{P}_z(\tau_\pi^z > T)\Big].\end{aligned}$$

Moreover,

$$\mathbb{P}_\pi(\tau_x \le 2T) = \mathbb{P}_\pi\left(\exists s \le 2T \text{ s.t. } X_s = x\right) \le (2T+1)\pi(x) =: \varepsilon_1 = o(1), \tag{3.23}$$

where we used the assumption (HP2). On the other hand, thanks to Lemma 3.3, we have

$$\max_{z\in\mathcal{X}} \mathbb{P}_z(\tau_\pi^z > T) = \max_{z\in\mathcal{X}} \mathrm{sep}(\mu_T^z, \pi) \le \max_{z\in\mathcal{X}} \left\| \frac{\mu_T^z}{\pi} - 1 \right\|_\infty =: \varepsilon_2 = o(1). \tag{3.24}$$

By plugging (3.23) and (3.24) into (3.22) we get

$$\mathbb{P}_\pi(\tau_x > t+T) \ge \mathbb{P}_\pi(\tau_x > t) - \mathbb{P}_\pi(\tau_x > t-T)(\varepsilon_1 + \varepsilon_2), \tag{3.25}$$

and so (3.21) follows with $\varepsilon := \varepsilon_1 + \varepsilon_2$.

We are now going to exploit (3.21) to prove (3.20). Consider the sequence $(y_i)_{i\ge 1}$

$$y_i := \frac{\mathbb{P}_\pi(\tau_x > (i+1)T)}{\mathbb{P}_\pi(\tau_x > iT)}. \tag{3.26}$$

Thanks to (3.21) we deduce

$$y_{i+1} \ge 1 - \frac{\varepsilon}{y_i}. \tag{3.27}$$

Being $\varepsilon < 1/4$, we can define

$$\bar{\varepsilon} := \frac{1}{2} - \sqrt{\frac{1}{4} - \varepsilon}$$

and get by induction

$$y_i \ge 1 - \bar{\varepsilon}, \qquad \forall i \ge 1. \tag{3.28}$$

Indeed, note that $\varepsilon = \bar{\varepsilon}(1-\bar{\varepsilon}) < \bar{\varepsilon}$

$$y_1 = \frac{\mathbb{P}_\pi(\tau_x > 2T)}{\mathbb{P}_\pi(\tau_x > T)} = 1 - \frac{\mathbb{P}_\pi(\tau_x \in [T,2T])}{\mathbb{P}_\pi(\tau_x > T)} \ge 1 - \frac{(T+1)\pi(x)}{1-(T+1)\pi(x)} \ge 1 - \frac{\varepsilon}{1-\varepsilon} \ge 1 - \bar{\varepsilon}$$

and

$$y_{i+1} \ge 1 - \frac{\varepsilon}{y_i} \ge 1 - \frac{\varepsilon}{1-\bar{\varepsilon}} \ge 1 - \bar{\varepsilon}.$$

The result of the induction in (3.28) can be immediately extended from times iT to general times $t = iT + t_0$ with $t_0 < T$ by noting that again we get

$$1 - \frac{(T+t_0)\pi(x)}{1-t_0\pi(x)} \ge 1 - \frac{\varepsilon}{1-\varepsilon}. \qquad \square$$

Corollary 3.7. *For all $x \in \mathcal{X}$ and for all $t > T$ it holds*

$$\mathbb{P}_\pi(\tau_x > t-T) \sim \mathbb{P}_\pi(\tau_x > t). \tag{3.29}$$

Proof: Notice that it is sufficient to show that

$$\frac{\mathbb{P}_\pi(\tau_x > t)}{\mathbb{P}_\pi(\tau_x > t-T)} \geq 1 - o(1), \tag{3.30}$$

which follows immediately by Lemma 3.6. □

Proof of Proposition 3.4: It follows immediately by Lemma 3.5 and Corollary 3.7. □

The next proposition relates the expected hitting time of x starting at stationarity, with the same expectation but starting at quasi-stationarity.

Proposition 3.8. *For all $x \in \mathcal{X}$*

$$\mathbb{E}_\pi[\tau_x] \sim \mathbb{E}_{\mu_x^\star}[\tau_x] = \frac{1}{1-\lambda_x}.$$

Hence, by Proposition 3.1,

$$1 - \lambda_x \sim \frac{\pi(x)}{R_T(x)}.$$

In order to prove Proposition 3.8, a key ingredient is the following lemma, which states that $1-\lambda_x$ must be much smaller than T^{-1}. We will later see that such a rough bound is sufficient to recover the precise first order asymptotic of λ_x by comparing $\mathbb{E}_{\mu_x^\star}[\tau_x]$ to $\mathbb{E}_\pi[\tau_x]$.

Lemma 3.9. *For all $x \in \mathcal{X}$, it holds*

$$\lambda_x^T \sim 1. \tag{3.31}$$

Proof: Start by noting that

$$\begin{aligned}
\lambda_x^{2T} =& \mathbb{P}_{\mu_x^\star}(\tau_x > 2T) \\
=& \sum_{z\neq x} \mathbb{P}_{\mu_x^\star}(X_T = z,\ \tau_x > T)\, \mathbb{P}_z(\tau_x > T) \\
=& \sum_{z\neq x} \left[\mathbb{P}_{\mu_x^\star}(X_T = z) - \mathbb{P}_{\mu_x^\star}(X_T = z,\ \tau_x \leq T)\right] \mathbb{P}_z(\tau_x > T) \\
\text{Lemma 3.3} \implies \quad \sim& \mathbb{P}_\pi(\tau_x > T) - \sum_{z\neq x} \mathbb{P}_{\mu_x^\star}(X_T = z,\ \tau_x \leq T)\, \mathbb{P}_z(\tau_x > T) \\
\geq& \mathbb{P}_\pi(\tau_x > T) - \max_z \mathbb{P}_z(\tau_x > T)\, \mathbb{P}_{\mu_x^\star}(\tau_x \leq T) \\
\text{Proposition 3.4} \implies \quad \sim& \mathbb{P}_\pi(\tau_x > T)\left(1 - \mathbb{P}_{\mu_x^\star}(\tau_x \leq T)\right) \\
=& \mathbb{P}_\pi(\tau_x > T)\left(1 - (1 - \lambda_x^T)\right).
\end{aligned}$$

Hence

$$\lambda_x^T \gtrsim \mathbb{P}_\pi(\tau_x > T) \geq 1 - (T+1)\pi(x), \tag{3.32}$$

so, by (HP2) we can conclude that $\lambda_x^T \sim 1$. □

Proof of Proposition 3.8: We start with the trivial bounds

$$\sum_{t=T}^{\infty} \mathbb{P}_{\mu_x^\star}(\tau_x > t) \leq \mathbb{E}_{\mu_x^\star}[\tau_x] \leq T + \sum_{t=T}^{\infty} \mathbb{P}_{\mu_x^\star}(\tau_x > t). \tag{3.33}$$

We further notice that

$$\sum_{t=T}^{\infty} \mathbb{P}_{\mu_x^\star}(\tau_x > t) = \sum_{z} \mathbb{P}_{\mu_x^\star}(X_T = z, \tau_x > T) \sum_{t=0}^{\infty} \mathbb{P}_z(\tau_x > t)$$

$$
\begin{aligned}
&= \sum_z \left[\mathbb{P}_{\mu_x^\star}(X_T = z) - \mathbb{P}_{\mu_x^\star}(X_T = z, \tau_x \le T)\right] \sum_{t=0}^{\infty} \mathbb{P}_z(\tau_x > t) \\
&= \sum_z \left[\pi(z)(1 + o(1)) - \mathbb{P}_{\mu_x^\star}(X_T = z, \tau_x \le T)\right] \sum_{t=0}^{\infty} \mathbb{P}_z(\tau_x > t) \\
&=(1 + o(1))\mathbb{E}_\pi[\tau_x] - \sum_z \mathbb{P}_{\mu_x^\star}(X_T = z, \tau_x \le T) \sum_{t=0}^{\infty} \mathbb{P}_z(\tau_x > t). \qquad (3.34)
\end{aligned}
$$

It follows immediately by (3.34) that

$$\sum_{t=T}^{\infty} \mathbb{P}_{\mu_x^\star}(\tau_x > t) \le (1 + o(1))\mathbb{E}_\pi[\tau_x]. \tag{3.35}$$

On the other hand,

$$\sum_z \mathbb{P}_{\mu_x^\star}(X_T = z, \tau_x \le T) \sum_{t=0}^{\infty} \mathbb{P}_z(\tau_x > t) \le \mathbb{P}_{\mu_x^\star}(\tau_x \le T) \cdot \sum_{t=0}^{\infty} \max_z \mathbb{P}_z(\tau_x > t) \tag{3.36}$$

and thanks to Proposition 3.4 we get

$$\sum_{t=0}^{\infty} \max_z \mathbb{P}_z(\tau_x > t) \le T + \sum_{t=T}^{\infty} \mathbb{P}_\pi(\tau_x > t) = (1 + o(1))\mathbb{E}_\pi[\tau_x]. \tag{3.37}$$

At this point, the proof is complete since

$$\mathbb{P}_{\mu_x^\star}(\tau_x \le T) = 1 - \lambda_x^T = o(1), \tag{3.38}$$

where the latter asymptotics follows from Lemma 3.9. □

We are now in shape to prove the main result.

Proof of Theorem 2.2: We start by bounding each entry of the T-step evolution of the quasi-stationary measure. From above we have the trivial bound: for all $x, y \in \mathcal{X}$

$$\mu_T^{\mu_x^\star}(y) \ge \lambda_x^T \mu_x^\star(y). \tag{3.39}$$

The latter immediately implies that for all $x \in \mathcal{X}$ and $t > 0$ it holds

$$\mathbb{P}_\pi(\tau_x > t) \gtrsim \lambda_x^{t+T} \sim \lambda_x^t. \tag{3.40}$$

In fact, by Lemma 3.3,

$$\mathbb{P}_\pi(\tau_x > t) \sim \mathbb{P}_{\mu_T^{\mu_x^\star}}(\tau_x > t) \ge \lambda_x^T \mathbb{P}_{\mu_x^\star}(\tau_x > t) = \lambda_x^{t+T}. \tag{3.41}$$

To conclude the proof, we show a matching upper bound. Component-wise, we can upper bound

$$
\begin{aligned}
\mu_T^{\mu_x^\star}(y) &= \lambda_x^T \mu_x^\star(y) + (1 - \lambda_x) \sum_{s=1}^{T} \lambda_x^s \mu_{T-s}^x(y) \qquad (3.42)\\
&\le \lambda_x^T \mu_x^\star(y) + (1 - \lambda_x)\mathbb{E}_x[\zeta_T(y)], \qquad (3.43)
\end{aligned}
$$

where $\zeta_T(y)$ denotes the local time spent by the chain in the state y within time T, i.e.

$$\zeta_T(y) := \sum_{s=1}^{T} \mathbf{1}_{X_t = y}. \tag{3.44}$$

Notice that for all $x, y \in \mathcal{X}$, holds

$$\sum_{y \in \mathcal{X}} \mathbb{E}_x[\zeta_T(y)] = T. \tag{3.45}$$

Hence,

$$\mathbb{P}_\pi(\tau_x > t) \sim \mathbb{P}_{\mu_T^{\mu_x^\star}}(\tau_x > t) \tag{3.46}$$

$$\leq \sum_{y\in\mathcal{X}} \lambda_x^T \mu_x^\star(y)\mathbb{P}_y(\tau_x > t) + (1-\lambda_x)\sum_{y\in\mathcal{X}} \mathbb{E}_x[\zeta_T(y)]\mathbb{P}_y(\tau_x > t) \tag{3.47}$$

$$\leq \lambda_x^{t+T} + (1-\lambda_x)T \max_y \mathbb{P}_y(\tau_x > t) \tag{3.48}$$

$$= \lambda_x^{t+T} + o\left(\mathbb{P}_\pi(\tau_x > t)\right) \tag{3.49}$$

where in the latter asymptotic equality we used in Lemma 3.9 and Proposition 3.4. We then conclude that for all $x \in \mathcal{X}$ and $t > T$ it holds

$$\mathbb{P}_\pi(\tau_x > t) \lesssim \lambda_x^t. \qquad \square$$

4. Controlling the Doob's transform

We start the section by showing that the unique vector γ_x defined by the requirements

$$\lambda_x \gamma_x = [P]_x \gamma_x, \qquad \langle \mu_x^\star, \gamma_x \rangle = 1, \tag{4.1}$$

can be equivalently characterized by the limits

$$\gamma_x(y) = \lim_{t\to\infty} \frac{\mathbb{P}_y(\tau_x > t)}{\lambda_x^t}, \qquad \forall y \neq x. \tag{4.2}$$

In fact, it is an immediate consequence of (1.9) and $|\mathcal{X}| < \infty$ that for every measure α, α' on $\mathcal{X}$, defining $\gamma_x(x) = 0$ and assuming $\alpha \neq \delta_x$, holds

$$\frac{\langle \gamma_x, \alpha' \rangle}{\langle \gamma_x, \alpha \rangle} = \lim_{t\to\infty} \frac{\mathbb{P}_{\alpha'}(\tau_x > t)}{\mathbb{P}_\alpha(\tau_x > t)}. \tag{4.3}$$

Hence, choosing $\alpha = \mu_x^\star$ and $\alpha' = \delta_y$ in the latter display we get (4.2). Moreover, choosing $\alpha = \mu_x^\star$ and $\alpha' = \pi$ and making use of Theorem 2.2 we get indeed the claim in Corollary 2.3.

We now aim at proving Theorem 2.4. We discuss the upper and the lower bound separately.

Lemma 4.1. *For all $x \in \mathcal{X}$ it holds*

$$\max_{y\in\mathcal{X}\setminus\{x\}} \gamma_x(y) \lesssim 1. \tag{4.4}$$

Proof: Rewrite

$$\max_{y\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y(\tau_x > t) \leq \max_{y\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y\left(\tau_x > t;\ \tau_\pi^y \leq T\right) + \max_{y\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y\left(\tau_x > t;\ \tau_\pi^y > T\right)$$

$$\leq \mathbb{P}_\pi\left(\tau_x > t - T\right) + \max_{y\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y\left(\tau_x > t;\ \tau_\pi^y > T\right). \tag{4.5}$$

Hence, it suffices to show that

$$\max_{y\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y\left(\tau_x > t;\ \tau_\pi^y > T\right) = o\big(\mathbb{P}_\pi\left(\tau_x > t - T\right)\big). \tag{4.6}$$

We decompose the latter by its position at time T, i.e.,

$$\max_{y\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y\left(\tau_x > t;\ \tau_\pi^y > T\right)$$

$$= \max_{y\in\mathcal{X}\setminus\{x\}} \sum_{z\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y\left(\tau_x > T;\ X_T = z;\ \tau_\pi^y > T\right) \mathbb{P}_z\left(\tau_x > t - T\right) \tag{4.7}$$

$$\leq \left(\max_{z\in\mathcal{X}\setminus\{x\}} \mathbb{P}_z\left(\tau_x > t - T\right)\right) \cdot \left(\max_{y\in\mathcal{X}\setminus\{x\}} \mathbb{P}_y(\tau_\pi^y > t)\right)$$

$$\sim \mathbb{P}_\pi\left(\tau_x > t - T\right) \cdot \max_{y\in\mathcal{X}} \mathbb{P}_y(\tau_\pi^y > t)$$

$$= o\big(\mathbb{P}_\pi\left(\tau_x > t - T\right)\big).$$

By inserting the bounds in (4.5) and (4.6) into (4.2) we deduce that

$$\gamma_x(y) = \lim_{t\to\infty} \frac{\mathbb{P}_y(\tau_x > t)}{\lambda_x^t} \lesssim \lim_{t\to\infty} \frac{\mathbb{P}_\pi(\tau_x > t - T)}{\lambda_x^t} = 1 + o(1). \qquad \square$$

Lemma 4.2. *For all $x, y \in \mathcal{X}$ with $x \neq y$ it holds*

$$\gamma_x(y) \gtrsim 1 - \mathbb{E}_y[\zeta_T(x)]. \tag{4.8}$$

Proof: By the same argument of the proof of Lemma 4.1 it is sufficient to show that for all $\varepsilon > 0$

$$\mathbb{P}_y(\tau_x > t) \geq (1 - \varepsilon - \mathbb{E}_y[\zeta_T(x)])\mathbb{P}_\pi\left(\tau_x > t\right). \tag{4.9}$$

Rewrite

$$\mathbb{P}_y(\tau_x > t) \geq \mathbb{P}_y\left(\tau_x > t;\ \tau_\pi^y \leq T\right) \tag{4.10}$$

$$= \sum_{s\leq T} \mathbb{P}_y\left(\tau_x > s;\ \tau_\pi^y = s\right)\mathbb{P}_\pi(\tau_x > t - s) \tag{4.11}$$

$$\geq \mathbb{P}_\pi(\tau_x > t)\sum_{s\leq T} \mathbb{P}_y\left(\tau_x > s;\ \tau_\pi^y = s\right) \tag{4.12}$$

$$= \mathbb{P}_\pi(\tau_x > t)\mathbb{P}_y(\tau_x > \tau_\pi^y; \tau_\pi^y \leq T) \tag{4.13}$$

we are left with showing that

$$\mathbb{P}_y(\tau_x > \tau_\pi^y; \tau_\pi^y \leq T) \geq \mathbb{P}_y(\tau_x > T) - \mathbb{P}_y(\tau_\pi^y > T) \tag{4.14}$$

$$= 1 - \mathbb{P}_y(\tau_x \leq T) - \varepsilon \tag{4.15}$$

$$= 1 - \varepsilon - \sum_{s\leq T} \mathbb{P}_y(\tau_x = s) \tag{4.16}$$

$$\geq 1 - \varepsilon - \sum_{s\leq T} \mathbb{P}_y(X_s = x) \tag{4.17}$$

$$= 1 - \varepsilon - \mathbb{E}_y[\zeta_T(x)], \tag{4.18}$$

which implies the desired result. $\square$

5. A random time perspective on the FVTL

Besides the rough bounds in (1.20) it is possible to have a probabilistic identity that defines the tail probability of the event $\tau_x > t$ when the Markov chain starts at α. In order to provide such a representation, it has been introduced in Manzo and Scoppola (2019) the notion of *conditional strong quasi-stationary time* as an extension of the idea of *strong stationary time* introduced in Aldous and Diaconis (1987b), see also Diaconis and Fill (1990); Levin and Peres (2017). In this last section, we aim at showing how the assumptions leading to the validity of the FVTL reflect on the theory of CSQST and on the mixing behavior of the Doob's transform.

Consider an irreducible Markovian kernel P and a state $x \in \mathcal{X}$ such that $[P]_x$ is irreducible and sub-Markovian. A randomized stopping time $\tau_\star^\alpha$ is a *Conditional Strong Quasi Stationary Time (CSQST)* if for any $y \in \mathcal{X} \setminus \{x\}$, and $t \geq 0$

$$\mathbb{P}_\alpha(X_t = y,\, \tau_\star^\alpha = t) = \mu_x^\star(y)\mathbb{P}_\alpha(\tau_\star^\alpha = t < \tau_x). \tag{5.1}$$

In other words, $\tau_\star^\alpha$ is a CSQST if for any $y \in \mathcal{X} \setminus \{x\}$, and $t \geq 0$

$$\mathbb{P}_\alpha\left(X_t^\alpha = y, \tau_\star^\alpha = t \mid t < \tau_x\right) = \mu_x^\star(y)\mathbb{P}_\alpha\left(\tau_\star^\alpha = t \mid t < \tau_x\right), \tag{5.2}$$

which is equivalent to

$$\mathbb{P}_\alpha\left(X_{\tau_\star^\alpha} = y \mid \tau_\star^\alpha < \tau_x\right) = \mu_x^\star(y). \tag{5.3}$$

By (1.20) we deduce that for any initial distribution α on $\mathcal{X} \setminus \{x\}$ and for any CSQST $\tau_\star^\alpha$ we have for any $t \geq 0$:

$$\mathbb{P}_\alpha(\tau_\star^\alpha \leq t < \tau_x) = \sum_{u \leq t} \lambda^{t-u} \mathbb{P}_\alpha(\tau_\star^\alpha = u < \tau_x) \leq \lambda_x^t \langle \alpha, \gamma_x \rangle (1 - \tilde{s}^{\tilde{\alpha}}(t)).$$

This suggests a new notion of minimality: a conditional strong quasi stationary time $\tau_\star^\alpha$ is *minimal* if for any $t \geq 0$

$$\mathbb{P}_\alpha(\tau_\star^\alpha \leq t < \tau_x) = \lambda_x^t \langle \alpha, \gamma_x \rangle (1 - \tilde{s}^{\tilde{\alpha}}(t)).$$

The existence of minimal CSQSTs is proved in Manzo and Scoppola (2019) where it is shown the validity of the following representation formula: for any minimal CSQST $\tau_\star^\alpha$ and for any $t \geq 0$:

$$\mathbb{P}_\alpha\Big(\tau_x > t\Big) = \lambda_x^t \langle \alpha, \gamma_x \rangle (1 - \tilde{s}^{\tilde{\alpha}}(t)) + \mathbb{P}_\alpha\Big(\tau_{\star,x}^\alpha > t\Big), \tag{5.4}$$

where

$$\tau_{\star,x}^\alpha := \tau_x \wedge \tau_\star^\alpha.$$

As a byproduct of the FVTL and of (5.4) it is possible to show the following result.

Proposition 5.1. *Under the assumptions of the FVTL there exists a minimal CSQST $\tau_{\star,x}^\pi$ such that*

$$\mathbb{P}_\pi(\tau_{\star,x}^\pi = 0) \to 1. \tag{5.5}$$

In physical terms, Proposition 5.1 confirms once again the idea that, under the assumptions of the FVTL, the stationary and the quasi-stationary distributions coincide in the thermodynamic limit.

Proof: We start by rewriting the representation formula in 5.4 in the case $\alpha = \pi$,

$$\mathbb{P}_\pi\Big(\tau_x > t\Big) = \lambda_x^t \langle \pi, \gamma_x \rangle (1 - \tilde{s}^{\tilde{\pi}}(t)) + \mathbb{P}_\pi\Big(\tau_{\star,x}^\pi > t\Big). \tag{5.6}$$

By Theorem 2.2 we know that (5.6) implies that, uniformly in $t \geq 0$,

$$\lambda_x^t \sim \lambda_x^t \langle \pi, \gamma_x \rangle (1 - \tilde{s}^{\tilde{\pi}}(t)) + \mathbb{P}_\pi\Big(\tau_{\star,x}^\pi > t\Big). \tag{5.7}$$

Thanks to Corollary 2.3 we can simplify (5.7) and get

$$\sup_{t \geq 0} \left| \frac{\mathbb{P}_\pi\Big(\tau_{\star,x}^\pi > t\Big)}{\lambda_x^t} - \tilde{s}^{\tilde{\pi}}(t) \right| = o(1). \tag{5.8}$$

We now show that the second term in the left hand side of (5.8) is $o(1)$ uniformly in $t \geq 0$, which implies that the same holds for the first term. In fact, by the monotonicity of the separation distance, the estimate

$$\sup_{t \geq 0} \tilde{s}^{\tilde{\pi}}(t) = o(1), \tag{5.9}$$

is an immediate consequence of

$$\tilde{s}^{\tilde{\pi}}(0) = o(1). \tag{5.10}$$

In order to prove (5.10), start by noting that the stationary distribution of the Doob's transform is given by

$$\nu_x(y) = \mu_x^\star(y) \gamma_x(y), \tag{5.11}$$

while its starting distribution is, by (1.14),

$$\tilde{\pi}(y) = \frac{\pi(y)\gamma_x(y)}{\langle \pi, \gamma_x \rangle} \sim \pi(y)\gamma_x(y), \tag{5.12}$$

where in the latter approximation we used again Corollary 2.3. Hence,

$$\mathrm{sep}(\tilde{\pi},\nu_x)=\max_{y\in\mathcal{X}\setminus\{x\}}\left[1-\frac{\tilde{\pi}(y)}{\nu_x(y)}\right]\sim\max_{y\in\mathcal{X}\setminus\{x\}}\left[1-\frac{\pi(y)}{\mu_x^\star(y)}\right]. \tag{5.13}$$

Therefore, to prove (5.10), it suffices to show that

$$\max_{y\in\mathcal{X}\setminus\{x\}}\left[1-\frac{\pi(y)}{\mu_x^\star(y)}\right]=o(1). \tag{5.14}$$

Notice that for all $y\in\mathcal{X}\setminus\{x\}$ it holds

$$\begin{aligned}
\mu_x^\star(y) &=\lambda_x^{-T}\sum_{z\neq x}\mu_x^\star(z)\big([P]_x\big)^T(z,y) && (5.15)\\
&\le\lambda_x^{-T}\sum_{z\neq x}\mu_x^\star(z)P^T(z,y) && (5.16)\\
\text{Lemma 3.3}\implies\quad &=\lambda_x^{-T}\sum_{z\neq x}\mu_x^\star(z)\pi(y)(1+o(1)) && (5.17)\\
&\sim\lambda_x^{-T}\pi(y)(1+o(1)) && (5.18)\\
\text{Lemma 3.9}\implies\quad &\sim\pi(y). && (5.19)
\end{aligned}$$

The latter chain of asymptotic equalities shows that (5.14) holds, which in turn implies (5.9). Therefore, thanks to (5.8), we conclude that for every minimal CSQST $\tau_{\star,x}^{\pi}$

$$\sup_{t\ge 0}\mathbb{P}_\pi(\tau_{\star,x}^{\pi}>t)=o(1). \tag{5.20}$$

□

Acknowledgements

The authors would like to thank Pietro Caputo for having introduced us to the topic, and the anonymous referee for several fruitful comments on the first draft of this paper.

References

Abdullah, M. The cover time of random walks on graphs. *ArXiv Mathematics e-prints* (2012). PhD thesis. arXiv: 1202.5569.

Abdullah, M., Cooper, C., and Frieze, A. Cover time of a random graph with given degree sequence. *Discrete Math.*, **312** (21), 3146–3163 (2012). MR2957935.

Aldous, D. *Probability approximations via the Poisson clumping heuristic*, volume 77 of *Applied Mathematical Sciences*. Springer-Verlag, New York (1989). ISBN 0-387-96899-7. MR969362.

Aldous, D. and Diaconis, P. Shuffling cards and stopping times. *Amer. Math. Monthly*, **93** (5), 333–348 (1986). MR841111.

Aldous, D. and Diaconis, P. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, **8** (1), 69–97 (1987a). MR876954.

Aldous, D. and Diaconis, P. Strong uniform times and finite random walks. *Adv. in Appl. Math.*, **8** (1), 69–97 (1987b). MR876954.

Aldous, D. and Fill, J. A. Reversible Markov Chains and Random Walks on Graphs (2002). Unfinished monograph, recompiled 2014, available at `http://www.stat.berkeley.edu/~aldous/RWG/book.html`.

Aldous, D. J. Markov chains with almost exponential hitting times. *Stochastic Process. Appl.*, **13** (3), 305–310 (1982). MR671039.

Aldous, D. J. and Brown, M. Inequalities for rare events in time-reversible Markov chains. I. In *Stochastic inequalities (Seattle, WA, 1991)*, volume 22 of *IMS Lecture Notes Monogr. Ser.*, pp. 1–16. Inst. Math. Statist., Hayward, CA (1992). MR1228050.

Aldous, D. J. and Brown, M. Inequalities for rare events in time-reversible Markov chains. II. *Stochastic Process. Appl.*, **44** (1), 15–25 (1993). MR1198660.

Bianchi, A. and Gaudillière, A. Metastable states, quasi-stationary distributions and soft measures. *Stochastic Process. Appl.*, **126** (6), 1622–1680 (2016). MR3483732.

Bianchi, A., Gaudillière, A., and Milanesi, P. On soft capacities, quasi-stationary distributions and the pathwise approach to metastability. *J. Stat. Phys.*, **181** (3), 1052–1086 (2020). MR4160921.

Bovier, A. and den Hollander, F. *Metastability. A potential-theoretic approach*, volume 351 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham (2015). ISBN 978-3-319-24775-5; 978-3-319-24777-9. MR3445787.

Caputo, P. and Quattropani, M. Stationary distribution and cover time of sparse directed configuration models. *Probab. Theory Related Fields*, **178** (3-4), 1011–1066 (2020). MR4168393.

Collet, P., Martínez, S., and San Martín, J. *Quasi-stationary distributions. Markov chains, diffusions and dynamical systems*. Probability and its Applications (New York). Springer, Heidelberg (2013). ISBN 978-3-642-33130-5; 978-3-642-33131-2. MR2986807.

Cooper, C. and Frieze, A. The cover time of random regular graphs. *SIAM J. Discrete Math.*, **18** (4), 728–740 (2005). MR2157821.

Cooper, C. and Frieze, A. The cover time of sparse random graphs. *Random Structures Algorithms*, **30** (1-2), 1–16 (2007a). MR2283218.

Cooper, C. and Frieze, A. The cover time of the preferential attachment graph. *J. Combin. Theory Ser. B*, **97** (2), 269–290 (2007b). MR2290325.

Cooper, C. and Frieze, A. The cover time of the giant component of a random graph. *Random Structures Algorithms*, **32** (4), 401–439 (2008). MR2422388.

Cooper, C. and Frieze, A. Stationary distribution and cover time of random walks on random digraphs. *J. Combin. Theory Ser. B*, **102** (2), 329–362 (2012). MR2885424.

Cooper, C., Frieze, A., and Lubetzky, E. Cover time of a random graph with a degree sequence II: Allowing vertices of degree two. *Random Structures Algorithms*, **45** (4), 627–674 (2014). MR3275701.

Cooper, C., Frieze, A., and Radzik, T. Multiple random walks in random regular graphs. *SIAM J. Discrete Math.*, **23** (4), 1738–1761 (2009/10). MR2570201.

Cooper, C., Frieze, A., and Radzik, T. The cover times of random walks on random uniform hypergraphs. *Theoret. Comput. Sci.*, **509**, 51–69 (2013). MR3123458.

Darroch, J. N. and Seneta, E. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *J. Appl. Probability*, **2**, 88–100 (1965). MR179842.

Diaconis, P. and Fill, J. A. Strong stationary times via a new form of duality. *Ann. Probab.*, **18** (4), 1483–1522 (1990). MR1071805.

Diaconis, P. and Miclo, L. On times to quasi-stationarity for birth and death processes. *J. Theoret. Probab.*, **22** (3), 558–586 (2009). MR2530103.

Diaconis, P. and Miclo, L. On quantitative convergence to quasi-stationarity. *Ann. Fac. Sci. Toulouse Math. (6)*, **24** (4), 973–1016 (2015). MR3434264.

Fernandez, R., Manzo, F., Nardi, F. R., and Scoppola, E. Asymptotically exponential hitting times and metastability: a pathwise approach without reversibility. *Electron. J. Probab.*, **20**, Paper No. 122, 37 (2015). MR3425542.

Fernandez, R., Manzo, F., Nardi, F. R., Scoppola, E., and Sohier, J. Conditioned, quasi-stationary, restricted measures and escape from metastable states. *Ann. Appl. Probab.*, **26** (2), 760–793 (2016). MR3476624.

Keilson, J. *Markov chain models—rarity and exponentiality*, volume 28 of *Applied Mathematical Sciences*. Springer-Verlag, New York-Berlin (1979). ISBN 0-387-90405-0. MR528293.

Levin, D. A. and Peres, Y. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, second edition (2017). ISBN 978-1-4704-2962-1. MR3726904.

Manzo, F. and Scoppola, E. Exact results on the first hitting via conditional strong quasi-stationary times and applications to metastability. *J. Stat. Phys.*, **174** (6), 1239–1262 (2019). MR3934697.

Miclo, L. On absorption times and Dirichlet eigenvalues. *ESAIM Probab. Stat.*, **14**, 117–150 (2010). MR2654550.

Miclo, L. An absorbing eigentime identity. *Markov Process. Related Fields*, **21** (2), 249–262 (2015). MR3442821.

Miclo, L. On metastability. Technical Report 20-1128, Toulouse School of Economics (TSE) (2020). URL `https://EconPapers.repec.org/RePEc:tse:wpaper:124582`.

Olivieri, E. and Vares, M. E. *Large deviations and metastability*, volume 100 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, Cambridge (2005). ISBN 0-521-59163-5. MR2123364.

Pitman, J. and Tang, W. Tree formulas, mean first passage times and Kemeny's constant of a Markov chain. *Bernoulli*, **24** (3), 1942–1972 (2018). MR3757519.

Pollett, P. K. Quasi-stationary distributions: a bibliography (2008). Available at http://www.maths.uq.edu.au/∼pkp/papers/qsds/qsds.pdf.