

Cutoff for the averaging process on the hypercube and complete bipartite graphs*

Pietro Caputo[†] Matteo Quattropani[‡] Federico Sau[§]

Abstract

We consider the averaging process on a graph, that is the evolution of a mass distribution undergoing repeated averages along the edges of the graph at the arrival times of independent Poisson processes. We establish cutoff phenomena for both the L^1 and L^2 distance from stationarity when the graph is a discrete hypercube and when the graph is complete bipartite. Some general facts about the averaging process on arbitrary graphs are also discussed.

Keywords: mixing of Markov chains; cutoff phenomenon; averaging process.

MSC2020 subject classifications: Primary 60K35, Secondary 82B20; 82C26.

Submitted to EJP on December 30, 2022, final version accepted on July 10, 2023.

1 Introduction

The averaging process on a finite graph $G = (V, E)$, referred to as $\text{Avg}(G)$ below, is the stochastic process defined as follows. Start with some probability mass function η on V and consider independent Poisson clocks with rate 1 on each edge $xy \in E$; when xy rings, update both masses $\eta(x)$ and $\eta(y)$ to their average value $\frac{1}{2}(\eta(x) + \eta(y))$. As soon as the underlying graph is connected, whatever the initial distribution η , the dynamics converges to the uniform distribution $\pi \equiv 1/|V|$, and just as for the simple random walk on G , the convergence rate associated to this process is an important feature of the underlying graph. Unlike the case of simple random walks however, the

*P.C. thanks the Miller Institute for Basic Research in Science for funding his visit to UC Berkeley during the Fall 2022. M.Q. thanks the German Research Foundation (project number 444084038, priority program SPP2265) for financial support. F.S. gratefully acknowledges funding by the Lise Meitner fellowship, Austrian Science Fund (FWF): M3211. During the final stage of this work, F.S. was financially supported by “Microgrants 2022”, funded by Regione FVG, legge LR 2/2011.

[†]Università Roma Tre, Italy.

E-mail: pietro.caputo@uniroma3.it

[‡]Sapienza Università di Roma, Italy.

E-mail: matteo.quattropani@uniroma1.it

[§]Università degli Studi di Trieste, Italy.

E-mail: federico.sau@units.it

information available in the literature on convergence rates for $\text{Avg}(G)$ is rather limited. The mathematical analysis of the averaging process started about a decade ago with the work of Aldous and Lanoue [2]. More recently, for the special case of the complete graph a full account was given in [9], where the authors proved the existence of a cutoff phenomenon. On the contrary, in [24] the process was shown to have no cutoff on “finite-dimensional” graphs such as the d -dimensional grid, that is graphs satisfying a suitable Nash inequality which guarantees the existence of a well-defined limiting heat flow. Very little is known for other families of graphs. We refer to the recent work [23] for some partial results, and for some motivating connections to opinion dynamics, random distributed algorithms, and quantum computing.

In this paper we study the convergence to stationarity of $\text{Avg}(G)$ when $G = \{0, 1\}^d$ is the hypercube graph with $n = 2^d$ vertices, and when $G = K_{m,n-m}$ is the complete bipartite graph with the two parts of size m and $n - m$, respectively. In both cases, we are ultimately interested in quantifying, as $n \rightarrow \infty$, the (mean) L^p -distance to equilibrium for $p = 1, 2$, that is

$$(t, \xi) \mapsto \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}}, \tag{1.1}$$

where η_t denotes the mass distribution of $\text{Avg}(G)$ at time $t \geq 0$, while \mathbb{E}_ξ stands for the expectation over the random updates when starting from the configuration $\eta_0 = \xi$, and we write $\|\cdot\|_p$ for the $L^p(\pi)$ -norm with respect to π , the uniform distribution on the vertex set V . For $p \geq 1$, convexity shows that for any fixed ξ the function above is monotone nonincreasing in $t \geq 0$. When $p = 1$, (1.1) is twice the expected value of the usual total variation distance between η_t and π . For worst-case initial conditions ξ , we analyze the distance to equilibrium at times of the form

$$T(a) = t_\star + aw, \quad a \in \mathbb{R},$$

for some appropriate choices of $t_\star = t_\star(n)$ and $w = w(n)$ such that $t_\star \gg w$, and prove *cutoff* results around the mixing time t_\star , with w controlling the size of the so-called *cutoff window*. We refer e.g. to [19] for background on the cutoff phenomenon for Markov chains, and to Tables 1 and 2 below for a summary of results (including ours) on the mixing of the Averaging process.

For the hypercube we establish that both the L^1 and L^2 distances have a cutoff phenomenon with window $w = O(1)$, and mixing time $t_\star = \frac{1}{2} \log d$, which coincides with the L^p -mixing time of the lazy random walk on $\{0, 1\}^d$ where the d coordinates flip independently at rate $\frac{1}{2}$; see, e.g., [19]. This settles a question that was mentioned as an open problem in [2].

Theorem 1.1 ($G = \{0, 1\}^d$: Cutoff in L^p , $p = 1, 2$). *Let $n = 2^d$, and*

$$T(a) := \frac{1}{2} \log d + a, \quad a \in \mathbb{R}.$$

Then,

$$\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{\xi} \mathbb{E}_\xi \left[\left\| \frac{\eta_{T(a)}}{\pi} - 1 \right\|_p^p \right] = 0, \quad p = 1, 2,$$

and

$$\lim_{a \rightarrow -\infty} \liminf_{n \rightarrow \infty} \sup_{\xi} \mathbb{E}_\xi \left[\left\| \frac{\eta_{T(a)}}{\pi} - 1 \right\|_p^p \right] = \begin{cases} 2 & \text{if } p = 1 \\ +\infty & \text{if } p = 2. \end{cases}$$

In the case of complete bipartite graphs $G = K_{m,n-m}$, we establish two cutoff results in L^1 and L^2 , with distinct locations and different sizes for the cutoff windows. Our

results apply to all sizes $1 \leq m \leq n/2$. The two special cases $m = 1$ (star graph) and $m = n/2$ (regular case) were discussed as open problems in [23].

Theorem 1.2 ($G = K_{m,n-m}$: Cutoff in L^1). *Let, for all $1 \leq m \leq n/2$,*

$$T(a) := \frac{n}{2(n-m)} \frac{\log_2 n}{m} + a \frac{\sqrt{\log n}}{m}, \quad a \in \mathbb{R}. \quad (1.2)$$

Then,

$$\lim_{a \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{\xi} \mathbb{E}_{\xi} \left[\left\| \frac{\eta_{T(a)}}{\pi} - 1 \right\|_1 \right] = 0,$$

and

$$\lim_{a \rightarrow -\infty} \liminf_{n \rightarrow \infty} \sup_{\xi} \mathbb{E}_{\xi} \left[\left\| \frac{\eta_{T(a)}}{\pi} - 1 \right\|_1 \right] = 2.$$

Concerning the L^2 -norm, besides determining the size of the cutoff window, we also identify the so-called *cutoff profile*.

Theorem 1.3 ($G = K_{m,n-m}$: Cutoff in L^2). *Let, for all $1 \leq m \leq n/2$,*

$$T(a) := \frac{\log n}{\theta m} + \frac{a}{\theta m}, \quad a \in \mathbb{R}, \quad (1.3)$$

where

$$\theta = \theta_{m,n-m} := \frac{3n}{8m} \left(1 - \sqrt{1 - \frac{32}{9} \frac{m}{n} \binom{n-m}{n}} \right) \in \left[\frac{1}{2}, \frac{2}{3} \right]. \quad (1.4)$$

Then, assuming that $m/n \rightarrow b \in [0, 1/2]$,

$$\lim_{n \rightarrow \infty} \sup_{\xi} \mathbb{E}_{\xi} \left[\left\| \frac{\eta_{T(a)}}{\pi} - 1 \right\|_2^2 \right] = (1 + D(b)) e^{-a}, \quad a \in \mathbb{R},$$

where

$$D(b) := \frac{3 - 4b - \sqrt{9 - 32b + 32b^2}}{2\sqrt{9 - 32b + 32b^2}} \in [0, \frac{\sqrt{6}-2}{4}], \quad b \in [0, 1/2]. \quad (1.5)$$

We refer the reader to Section 2.6 below for further discussions around these results and for a comparison with previously known facts. The rest of the paper is organized as follows. In Section 2 we give formal definitions, recall and establish a few general facts that hold for $\text{Avg}(G)$ on any graph G . In Section 3 we prove Theorem 1.1, in Section 4 we prove Theorem 1.2, and finally in Section 5 we prove Theorem 1.3.

2 The Averaging process on arbitrary graphs: general facts

The goal of this section is to collect a number of preliminary facts about the Averaging process on an arbitrary (unweighted) graph. These general observations allow a first, not necessarily optimal, control on the mixing times. Some of them can be found in the literature only in a discrete-time setting (e.g., [23]), others were formulated for more general reference measures π (e.g., [24]), or for weighted graphs (e.g., [2, 24]); the entropy inequalities from [4] were not originally related to the Averaging process. While not everything discussed in this section will be needed in the proof of our main results, we believe that a unified presentation is beneficial and could be of use as a reference for future research.

2.1 Averaging process vs. (lazy) random walk

We start with a simple observation relating the expectation of $\text{Avg}(G)$ with the transition probability of a random walk on the same graph.

Given a finite connected graph $G = (V, E)$, with $|V| = n$, the Averaging process $\text{Avg}(G)$ is the continuous-time Markov process $(\eta_t)_{t \geq 0}$ on the set $\Delta(V)$ of probability measures on V with infinitesimal generator $\mathcal{L} = \mathcal{L}^{\text{Avg}}$ given by

$$\mathcal{L}f(\eta) = \sum_{xy \in E} (f(\eta^{xy}) - f(\eta)), \quad \eta \in \Delta(V), f : \Delta(V) \rightarrow \mathbb{R}, \quad (2.1)$$

where η^{xy} is defined as

$$\eta^{xy}(z) := \begin{cases} \eta(z) & \text{if } z \neq x, y \\ \frac{\eta(x) + \eta(y)}{2} & \text{otherwise.} \end{cases} \quad (2.2)$$

When starting from a configuration $\xi \in \Delta(V)$, \mathbb{P}_ξ and \mathbb{E}_ξ denote the corresponding law and expectation over the trajectories of the Averaging processes, respectively. The uniform measure $\pi \equiv 1/n$ is a fixed point of the dynamics, i.e., $\mathcal{L}f(\pi) = 0$ for all $f : \Delta(V) \rightarrow \mathbb{R}$.

The process $\text{RW}(G)$ is the continuous-time Markov chain $(X_t)_{t \geq 0}$, with state space V , with generator

$$L^{\text{RW}}\psi(x) = \frac{1}{2} \sum_{y: xy \in E} (\psi(y) - \psi(x)), \quad x \in V, \psi : V \rightarrow \mathbb{R}. \quad (2.3)$$

That is, X_t is the simple random walk on G with jump rate $1/2$. We will refer to it as *lazy random walk* (or, simply, *random walk*) on G , and call $\mathbf{P}_\xi^{\text{RW}}$ the corresponding law when starting from a distribution $\xi \in \Delta(V)$; when $\xi = \mathbf{1}_x$, we simply write \mathbf{P}_x^{RW} . Clearly, $\pi \equiv 1/n$ is the unique equilibrium distribution for $\text{RW}(G)$. Further, we let λ_{gap} denote its *spectral gap*, and t_{rel} its *relaxation time*: $\lambda_{\text{gap}} > 0$ is the smallest non-zero eigenvalue of $-L^{\text{RW}}$, and

$$t_{\text{rel}} = (\lambda_{\text{gap}})^{-1}. \quad (2.4)$$

The first observation is that, for all $x \in V$, $t \geq 0$ and $\xi \in \Delta(V)$,

$$\mathbb{E}_\xi [\eta_t(x)] = \mathbf{P}_\xi^{\text{RW}}(X_t = x) =: \pi_t^\xi(x), \quad (2.5)$$

which equivalently rewrites, for some collection of random variables $Z_t^\xi = (Z_t^\xi(x))_{x \in V}$, as

$$\eta_t = \pi_t^\xi + Z_t^\xi, \quad \text{with } \eta_0 = \xi, \sum_{x \in V} Z_t^\xi(x) = 0, \mathbb{E}_\xi[Z_t^\xi(x)] = 0. \quad (2.6)$$

In other words, the probability mass function at time t is given by the deterministic random walk kernel π_t^ξ plus a mean-zero noise term Z_t^ξ . The identity in (2.5), as emphasized in [2], is an instance of *duality*; see also [20]. An intuitive explanation of (2.5) is that, under repeated averages on G , any given “infinitesimal chunk” of the initial mass performs a (lazy) random walk on G . As a first immediate consequence of (2.5)–(2.6), one has the following general lower bound, stating that the presence of the noise term Z_t^ξ cannot speed up convergence to equilibrium.

Proposition 2.1 ([24, Prop. 5.1]). *For all $p \in [1, \infty)$, $t \geq 0$, and $\xi \in \Delta(V)$,*

$$\mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p^p \right] \geq \left\| \frac{\pi_t^\xi}{\pi} - 1 \right\|_p^p. \quad (2.7)$$

Proof. By using the dual formulation of the L^p -distance, and calling q the conjugate exponent of p , we get

$$\begin{aligned} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p \right] &= \mathbb{E}_\xi \left[\sup_{\|\psi\|_q=1} \sum_{x \in V} (\eta_t(x) - \pi(x)) \psi(x) \right] \\ &\geq \sup_{\|\psi\|_q=1} \sum_{x \in V} \mathbb{E}_\xi [(\eta_t(x) - \pi(x)) \psi(x)] \\ &= \sup_{\|\psi\|_q=1} \sum_{x \in V} (\pi_t^\xi(x) - \pi(x)) \psi(x) = \left\| \frac{\pi_t^\xi}{\pi} - 1 \right\|_p, \end{aligned}$$

where for the second identity we used (2.5). The result follows by Jensen inequality. \square

Remark 2.2 (Sharpness of (2.7)). As we will comment more in detail in Section 2.6 below, for certain families of graphs the lower bound in (2.7) is asymptotically optimal (e.g., for the hypercube), for others it is sub-optimal (e.g., for the L^1 -distance on the complete graph, see [9]).

The lower bound in Proposition 2.1 is not the only relation that one can derive between the L^p -distance of $\text{Avg}(G)$ and that of $\text{RW}(G)$. As first noticed in [2], the infinitesimal mean contraction of the L^2 -distance for $\text{Avg}(G)$ can be expressed in terms of the Dirichlet form of $\text{RW}(G)$, giving rise to a version of Poincaré inequality for $\text{Avg}(G)$. More precisely, recalling the definitions of infinitesimal generators \mathcal{L} and L^{RW} in (2.1) and (2.3), respectively, and letting \mathcal{E}^{RW} denote the Dirichlet form for $\text{RW}(G)$

$$\mathcal{E}^{\text{RW}}(\psi) = \frac{1}{2n} \sum_{xy \in E} (\psi(x) - \psi(y))^2, \quad \psi : V \rightarrow \mathbb{R},$$

a simple computation yields

$$\frac{d}{dt} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = \mathcal{L} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = -\mathbb{E}_\xi \left[\mathcal{E}^{\text{RW}} \left(\frac{\eta_t}{\pi} \right) \right].$$

As a consequence of this, the variational formulation of the spectral gap, Grönwall inequality, and the Markov property of $\text{Avg}(G)$ yield the following result.

Proposition 2.3 ([2, Prop. 2]). *Letting t_{rel} be defined as in (2.4),*

$$\mathbb{E}_\xi \left[\left\| \frac{\eta_{t+s}}{\pi} - 1 \right\|_2^2 \right] \leq e^{-\frac{s}{t_{\text{rel}}}} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right], \quad \xi \in \Delta(V), \quad s, t \geq 0. \quad (2.8)$$

Remark 2.4 (Sharpness of (2.8)). Compared to the more common Poincaré inequality for $\text{RW}(G)$, the inequality in Proposition 2.3 comes with a missing factor 2 in the exponential contraction rate. In general, such an inequality cannot be improved, with the equality being attained when, for instance, G is the complete graph (see [9]).

2.2 Second moments and coupled random walks

As already pointed out in [2], also the second moments of η_t admit a formulation in the spirit of (2.5), at the cost of replacing $\text{RW}(G)$ by a suitable coupling of two random walks. For a given graph $G = (V, E)$, we call $\text{CRW}(G)$ the Markov process $(X_t, Y_t)_{t \geq 0}$ defined as follows. The state space is $V \times V$, interpreted as the set of positions of two labeled particles; as in $\text{Avg}(G)$ we have independent rate 1 Poisson clocks at the edges of G ; when an edge xy rings, each particle on the endpoints of that edge is placed independently at x or at y with probability $1/2$. We note that as long as X_t and Y_t are at (graph) distance strictly larger than 1, they evolve as two independent $\text{RW}(G)$ -processes,

whereas while X_t and Y_t sit on the endpoints of the same edge (on the same vertex or on opposite vertices), they can also experience synchronous jumps.

An argument analogous to that used for (2.5), involving this time the evolution of two infinitesimal portions of masses of $\text{Avg}(G)$, should convince the reader that the following identity holds: for all $x, y \in V$, $t \geq 0$, and $\xi \in \Delta(V)$,

$$\mathbb{E}_\xi [\eta_t(x)\eta_t(y)] = \mathbf{P}_{\xi \otimes \xi}^{\text{CRW}}(X_t = x, Y_t = y), \tag{2.9}$$

with $\mathbf{P}_{\xi \otimes \xi}^{\text{CRW}}$ denoting the law of $\text{CRW}(G)$ when starting from $\xi \otimes \xi$ (two independent samples from ξ). By a generalization of the above two-particle process that goes under the name of *Binomial splitting process*, relations such as (2.9) extend to higher-order moments, see [24]. As the next proposition shows, (2.9) yields two convenient alternative formulations of the L^2 -distance to equilibrium of $\text{Avg}(G)$ in terms of the transition probabilities of $\text{CRW}(G)$.

Proposition 2.5 ([2, 24]). *For all $t \geq 0$ and $\xi \in \Delta(V)$,*

$$\mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = n \mathbf{P}_{\xi \otimes \xi}^{\text{CRW}}(X_t = Y_t) - 1 \tag{2.10}$$

and

$$\mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = \left\| \frac{\pi_t^\xi}{\pi} - 1 \right\|_2^2 + \mathcal{N}_t^\xi, \tag{2.11}$$

where

$$\mathcal{N}_t^\xi := \frac{n}{2} \int_0^t \sum_{xy \in E} (\pi_s^\xi(x) - \pi_s^\xi(y))^2 \Phi_{t-s}(x, y) \, ds,$$

and

$$\Phi_t(x, y) := \frac{1}{2} (\mathbf{P}_{x,x}^{\text{CRW}}(X_t = Y_t) + \mathbf{P}_{y,y}^{\text{CRW}}(X_t = Y_t) - 2 \mathbf{P}_{x,y}^{\text{CRW}}(X_t = Y_t)) \in [0, 1].$$

The identity in (2.10) follows at once by expanding the L^2 -norm and using (2.9). The identity in (2.11) is slightly more involved, and can be obtained by an explicit comparison of the generator of the $\text{CRW}(G)$ -process and that of two independent $\text{RW}(G)$ -processes; we refer to [24, Prop. 5.5] for the details.

2.3 Worst-case initial condition

In the literature on mixing times of Markov chains, the distance to equilibrium at some time t is most often analyzed by starting the chain from a *worst-case initial condition*, namely one which maximizes the distance at time t . By a straightforward convexity argument, when the state space is finite, worst-case initial conditions $\xi \in \Delta(V)$ are always Dirac distributions on some element of the chain's state space. An analogous argument, based on the *linearity* of $\text{Avg}(G)$, shows that the worst-case L^p distance for $\text{Avg}(G)$ is achieved at Dirac masses $\mathbf{1}_x$, $x \in V$, for all $p \geq 1$. A version of this fact already appeared in [23] for the case of the L^1 -distance in the discrete-time model. The same idea was also used in [9] in the context of the complete graph, and in [24] to estimate the L^1 -distance of the Binomial splitting process on finite-dimensional graphs. The simple proof given here covers every graph and all L^p -distances with $p \geq 1$.

Proposition 2.6. *For any graph G , for all $t \geq 0$ and $p \in [1, \infty)$,*

$$\sup_{\xi \in \Delta(V)} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p^p \right] = \sup_{x \in V} \mathbb{E}_{\mathbf{1}_x} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p^p \right].$$

Proof. The inequality ‘ \geq ’ is obvious. The proof of the inverse inequality follows by observing that, when the initial condition is $\xi \in \Delta(V)$, the probability mass function at time t can be represented as

$$\eta_t = \sum_{y \in V} \xi(y) \eta_t^{(y)}, \quad t \geq 0,$$

where, for each $y \in V$, $\eta_t^{(y)}$ is the averaging process on G with initial condition $\mathbf{1}_y$, and the evolutions $(\eta_t^{(y)}, t \geq 0)_{y \in V}$ are coupled by the use of the same edge update sequence. Therefore, letting \mathbb{E} denote the expectation with respect to this coupling, by Jensen inequality,

$$\begin{aligned} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p^p \right] &= \sum_{x \in V} \pi(x) \mathbb{E} \left[\left| \sum_{y \in V} \xi(y) \left(\frac{\eta_t^{(y)}(x)}{\pi(x)} - 1 \right) \right|^p \right] \\ &\leq \sum_{x \in V} \pi(x) \sum_{y \in V} \xi(y) \mathbb{E} \left[\left| \frac{\eta_t^{(y)}(x)}{\pi(x)} - 1 \right|^p \right] = \sum_{y \in V} \xi(y) \mathbb{E}_{\mathbf{1}_y} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p^p \right]. \end{aligned}$$

This concludes the proof. □

2.4 Relative entropy decay

Relative entropy is commonly used to measure the distance from stationarity, and it is natural to relate the mixing properties of a system to its exponential decay rate. In the usual Markov chain setup, this leads to the so-called Log-Sobolev and modified Log-Sobolev inequalities [13, 6]. These inequalities control the global entropy functional in terms of a local Dirichlet form. In the setting of the averaging process, we are led to a new set of inequalities where the global entropy functional is controlled by local versions of the entropy itself rather than by a Dirichlet form. Interestingly, these new inequalities were recently introduced and studied by [4] as a natural factorization statement for various random walk models, see also [8] for a spin system setting. However, the striking connection to the Averaging process that we point out here appears to be new. The latter can be formalized as follows.

Let

$$D(\eta||\pi) := \sum_{x \in V} \eta(x) \log \frac{\eta(x)}{\pi(x)}, \quad \eta \in \Delta(V),$$

denote the *relative entropy* (or *Kullback–Leibler divergence*) of $\eta \in \Delta(V)$ with respect to $\pi \equiv 1/|V|$, and consider its expected value $\mathbb{E}_\xi [D(\eta_t||\pi)]$ under the Averaging process with initial condition $\xi \in \Delta(V)$. Recalling that $n = |V|$, the *local entropy* of $\eta \in \Delta(V)$ at the edge $xy \in E$ is defined as

$$\text{ent}_{xy}(\eta) := \frac{n}{2} \eta(x) \log \frac{\eta(x)}{\frac{1}{2}(\eta(x) + \eta(y))} + \frac{n}{2} \eta(y) \log \frac{\eta(y)}{\frac{1}{2}(\eta(x) + \eta(y))}.$$

Proposition 2.7. *The largest constant $\kappa \geq 0$ satisfying, for all $\xi \in \Delta(V)$ and $t \geq 0$,*

$$\mathbb{E}_\xi [D(\eta_t||\pi)] \leq e^{-\kappa t} D(\xi||\pi), \tag{2.12}$$

coincides with the largest constant $\kappa \geq 0$ satisfying, for all $\eta \in \Delta(V)$,

$$\kappa D(\eta||\pi) \leq \frac{2}{n} \sum_{xy \in E} \text{ent}_{xy}(\eta). \tag{2.13}$$

Proof. Recalling the notation (2.2), and observing that

$$\eta(x) \log \frac{\eta(x)}{\pi(x)} + \eta(y) \log \frac{\eta(y)}{\pi(y)} - \eta^{xy}(x) \log \frac{\eta^{xy}(x)}{\pi(x)} - \eta^{xy}(y) \log \frac{\eta^{xy}(y)}{\pi(y)} = \frac{2}{n} \text{ent}_{xy}(\eta) ,$$

we see that, as a result of an update at edge xy , the relative entropy of η decreases from $D(\eta||\pi)$ to $D(\eta^{xy}||\pi) = D(\eta||\pi) - \frac{2}{n} \text{ent}_{xy}(\eta)$. Therefore,

$$\frac{d}{dt} \mathbb{E}_\xi [D(\eta_t||\pi)] = \mathcal{L} \mathbb{E}_\xi [D(\eta_t||\pi)] = -\frac{2}{n} \sum_{xy \in E} \mathbb{E}_\xi [\text{ent}_{xy}(\eta_t)] .$$

By integrating over time, we see that (2.13) implies (2.12). On the other hand, if (2.12) holds for all $t > 0$, then subtracting $D(\xi||\pi)$ from both sides and taking the limit $t \rightarrow 0^+$ one obtains (2.13). \square

Following [4], we call $\kappa = \kappa(G)$ the *entropy constant* of the graph G . One can compute the value of $\kappa(G)$ for certain simple graphs. For instance, consider the trivial case where G consists of a single edge xy . In this case $n = 2$ and $D(\eta||\pi) = \text{ent}_{xy}(\eta)$ so that $\kappa = 1$. Therefore, using the product structure one easily checks that the entropy constant of the hypercube $\{0, 1\}^d$ is again $\kappa = 1$ for all $d \geq 1$. We refer to [4] for other examples including random walks on hypergraphs and processes with more than one particle such as the Binomial splitting process.

Proposition 2.7 is the entropic analogue of the L^2 -statement in Proposition 2.3. A linearization argument shows that the constant $\kappa = \kappa(G)$ in (2.12) satisfies $\kappa(G) \leq 1/t_{\text{rel}}(G)$ for any graph G . Moreover, one can show (see [4, Lem. 2.1]) that $\kappa(G) \geq 2 \log(2) \beta(G)$, for any G , where $\beta(G)$ denotes the Log-Sobolev constant of G defined as the largest constant $\beta \geq 0$ satisfying, for all $\eta \in \Delta(V)$,

$$\beta D(\eta||\pi) \leq \mathcal{E}^{\text{RW}}(\sqrt{\eta/\pi}) .$$

In particular, this implies that the entropic bound $\mathbb{E}_\xi [D(\eta_t||\pi)] \leq e^{-\beta t} D(\xi||\pi)$ proposed in [2, Prop. 7] can be improved by a factor $2 \log(2)$ in the exponential decay rate. Further relations of interest include upper bounds on $\kappa(G)$ in terms of both $\beta(G)$ and the Modified Log-Sobolev constant of G , see [4] for more details.

Finally, Proposition 2.7 and Pinsker's inequality

$$\left\| \frac{\eta}{\pi} - 1 \right\|_1 \leq \sqrt{2D(\eta||\pi)} , \quad \eta \in \Delta(V) , \tag{2.14}$$

provide the following upper bound on L^1 -distance in terms of the entropy constant κ .

Corollary 2.8. *For all $\xi \in \Delta(V)$, and for all $t \geq 0$*

$$\mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_1 \right] \leq e^{-\kappa t/2} \sqrt{2 \log n} ,$$

where $\kappa = \kappa(G)$ denotes the entropy constant of G .

Proof. From (2.14) and Cauchy-Schwarz inequality, one has

$$\mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_1 \right] \leq \sqrt{2 \mathbb{E}_\xi D(\eta_t||\pi)} \leq e^{-\kappa t/2} \sqrt{2D(\xi||\pi)} .$$

By convexity, $D(\xi||\pi)$ is maximized when ξ is a Dirac mass at a vertex; therefore, since $\pi \equiv 1/n$, $D(\xi||\pi) \leq \log n$. \square

2.5 A lower bound using relative entropy

Concerning lower bounds, the following statement is the continuous-time version of a useful estimate established in the recent work [23].

Proposition 2.9 ([23, Th. 1]). *Let*

$$T_\varepsilon := (1 - \varepsilon) \frac{\log_2 n}{\langle \text{deg} \rangle}, \quad \varepsilon \in (0, 1), \tag{2.15}$$

where $\text{deg}(x) := \text{degree of } x \in V \text{ in the graph } G$, and $\langle \text{deg} \rangle := \frac{1}{n} \sum_x \text{deg}(x)$. Then,

$$\liminf_{n \rightarrow \infty} \sup_{\xi \in \Delta(V)} \mathbb{E}_\xi \left[\left\| \frac{\eta_{T_\varepsilon}}{\pi} - 1 \right\|_1 \right] \geq \varepsilon. \tag{2.16}$$

Proof. As in the proof of Proposition 2.7, an update at edge xy at time t causes the relative entropy of η_t to decrease from $D(\eta_t|\pi)$ to $D(\eta_t^{xy}|\pi) = D(\eta_t|\pi) - \frac{2}{n} \text{ent}_{xy}(\eta_t)$. Since $\frac{2}{n} \text{ent}_{xy}(\eta_t) \leq (\eta_t(x) + \eta_t(y)) \log 2$, integrating and using (2.5), one has

$$\begin{aligned} \mathbb{E}_\xi [D(\eta_t|\pi)] &\geq D(\xi|\pi) - \log 2 \int_0^t \sum_{xy \in E} \mathbb{E}_\xi [\eta_s(x) + \eta_s(y)] ds \\ &= D(\xi|\pi) - \log 2 \int_0^t \sum_{x \in V} \pi_s^\xi(x) \text{deg}(x) ds. \end{aligned} \tag{2.17}$$

Since $\langle \text{deg} \rangle - \text{deg}$ has mean zero with respect to π , there exists $\psi : V \rightarrow \mathbb{R}$ solving

$$L^{\text{RW}} \psi(x) = \log 2 (\langle \text{deg} \rangle - \text{deg}(x)), \quad x \in V. \tag{2.18}$$

Moreover, ψ is defined up to constants, so that we are allowed to select a translation of ψ satisfying

$$\psi \geq 0, \quad \text{and} \quad \psi(y) = 0 \quad \text{for some } y \in V. \tag{2.19}$$

Note that if G is regular (i.e., $\text{deg} \equiv \text{const.}$), then we can take $\psi \equiv 0$. Introducing the function

$$F(\eta) := D(\eta|\pi) - \sum_{x \in V} \eta(x) \psi(x), \quad \eta \in \Delta(V), \tag{2.20}$$

the two conditions in (2.19) ensure, respectively,

$$F(\eta_t) \leq D(\eta_t|\pi), \quad \mathbb{P}_{\mathbf{1}_y}\text{-a.s.}, \tag{2.21}$$

and

$$F(\mathbf{1}_y) = D(\mathbf{1}_y|\pi) = \log n. \tag{2.22}$$

By combining (2.20), (2.5), (2.22), (2.17), and writing $\pi_t^y = \pi_t^{\mathbf{1}_y}$,

$$\begin{aligned} \mathbb{E}_{\mathbf{1}_y} [F(\eta_t)] &= \mathbb{E}_{\mathbf{1}_y} [D(\eta_t|\pi)] - \sum_{x \in V} \pi_t^y(x) \psi(x) \\ &\geq \log n - \int_0^t \sum_{x \in V} (\log 2 \pi_s^y(x) \text{deg}(x) + (L^{\text{RW}} \pi_s^y)(x) \psi(x)) ds \\ &= \log n - \int_0^t \sum_{x \in V} \pi_s^y(x) (\log 2 \text{deg}(x) + L^{\text{RW}} \psi(x)) ds \end{aligned}$$

where for the last identity we used that L^{RW} is symmetric. By (2.18), the expression inside the time integral above equals $\log 2 \langle \text{deg} \rangle$, thus, using (2.21),

$$\mathbb{E}_{\mathbf{1}_y} [D(\eta_t|\pi)] \geq \mathbb{E}_{\mathbf{1}_y} [F(\eta_t)] \geq \log n - \langle \text{deg} \rangle t \log 2, \quad t \geq 0,$$

and choosing $t = T_\varepsilon$ as in (2.15) one has $\mathbb{E}_{\mathbf{1}_y} D(\eta_t || \pi) \geq \varepsilon \log n = \varepsilon D(\mathbf{1}_y || \pi)$. To obtain the desired lower bound on the L^1 -norm one uses the fact that, for π uniform, Fannes-Audenaert inequality (see, e.g., [3])

$$\left\| \frac{\eta}{\pi} - 1 \right\|_1 \geq \frac{D(\eta || \pi)}{\log n} - \frac{1}{e \log n}, \quad \eta \in \Delta(V),$$

holds true. This concludes the proof. □

Remark 2.10. The above proof shows that $\mathbb{E}_{\mathbf{1}_y} D(\eta_{T_\varepsilon} || \pi) \geq \varepsilon D(\mathbf{1}_y || \pi)$. It follows from Proposition 2.7 that $\kappa(G)T_\varepsilon \leq \log(1/\varepsilon)$, and optimizing over $\varepsilon \in (0, 1)$, one obtains the general bound

$$\kappa(G) \leq \frac{\langle \text{deg} \rangle}{\log_2 n}. \tag{2.23}$$

Remarkably, if $G = K_n$ is the complete graph, the estimate (2.23) captures the exact value $\kappa(K_n) = (n - 1) / \log_2(n)$ (see [4, Remark 1.2] and note that a factor 2 has to be removed there because our random walk has jump rate $1/2$). Moreover, the bound (2.23) is saturated also when G is the hypercube, in which case one has $\kappa(G) = 1$.

Remark 2.11. For certain families of graphs such as the complete graphs, the lower bound in (2.16) predicts very sharply the order of the L^1 -mixing time for $\text{Avg}(G)$, see the discussion in Section 2.6. However, due to the ε -dependence in (2.15), this lower bound is not as sharp as needed to establish a cutoff phenomenon.

2.6 Comparison of mixing times and summary of the main results

Following an established convention, we let the L^p -mixing time of $\text{RW}(G)$ be defined as the first time t such that the L^p -distance $\|\pi_t^\xi / \pi - 1\|_p$ is at most $1/2$, uniformly over the initial condition ξ . The L^p -mixing time of $\text{Avg}(G)$ is defined in the same way by using the averaged quantity (1.1) instead of the L^p -distance, i.e.,

$$\inf \left\{ t \geq 0 : \sup_{\xi \in \Delta(V)} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}} \leq \frac{1}{2} \right\}, \quad p \geq 1.$$

In Table 1 and Table 2 below we summarize all known estimates about such mixing times, including the results in the present paper. We display, in particular:

1. lattice approximations of d -dimensional boxes $G = [0, 1]^d \cap \frac{\mathbb{Z}^d}{n}$ (this extends to all finite-dimensional graphs, see [24]);
2. hypercube $G = \{0, 1\}^d$ (Theorem 1.1);
3. complete graph $G = K_n$ (see [9]);
4. complete bipartite graphs $G = K_{m, n-m}$ (Theorem 1.2 and Theorem 1.3).

The purpose of these tables is that of comparing mixing times; hence, we express them all in terms of the relaxation time t_{rel} of $\text{RW}(G)$ defined in (2.4). As a consequence, such expressions will depend on the specific choice of the time normalization solely through the exact value of t_{rel} .

In Table 2, the function $b \in (0, 1/2] \mapsto C(b) \in (3/4, 1]$ is given by (cf. Rmk. 5.3 below)

$$C(b) := \frac{4}{3} \frac{b}{1 - \sqrt{1 - \frac{32}{9} b(1-b)}}, \quad \text{and} \quad \lim_{b \downarrow 0} C(b) = \frac{3}{4}, \quad C\left(\frac{1}{2}\right) = 1. \tag{2.24}$$

We conclude with a discussion of the main results of this paper and some comments on the strategy of proof together with comments on the applications of the general bounds presented in this section.

Table 1: L^1 -mixing times of $\text{RW}(G)$ and $\text{Avg}(G)$.

	$L^1, \text{RW}(G)$	$L^1, \text{Avg}(G)$
$[0, 1]^d \cap \frac{\mathbb{Z}^d}{n}$	$\Theta(t_{\text{rel}})$	$\Theta(t_{\text{rel}})$
$\{0, 1\}^d$	$\frac{t_{\text{rel}}}{2} (\log d + \Theta(1))$	$\frac{t_{\text{rel}}}{2} (\log d + \Theta(1))$
K_n	$\Theta(t_{\text{rel}})$	$\frac{1}{2 \log 2} t_{\text{rel}} (\log n + \Theta(\sqrt{\log n}))$
$K_{m, n-m}$	$\Theta(t_{\text{rel}})$	$\frac{n}{n-m} \frac{1}{4 \log 2} t_{\text{rel}} (\log n + \Theta(\sqrt{\log n}))$

Table 2: L^2 -mixing times of $\text{RW}(G)$ and $\text{Avg}(G)$; $C(\cdot)$ is defined in (2.24) below.

	$L^2, \text{RW}(G)$	$L^2, \text{Avg}(G)$
$[0, 1]^d \cap \frac{\mathbb{Z}^d}{n}$	$\Theta(t_{\text{rel}})$	$\Theta(t_{\text{rel}})$
$\{0, 1\}^d$	$\frac{t_{\text{rel}}}{2} (\log d + \Theta(1))$	$\frac{t_{\text{rel}}}{2} (\log d + \Theta(1))$
K_n	$\frac{t_{\text{rel}}}{2} (\log n + \Theta(1))$	$t_{\text{rel}} (\log n + \Theta(1))$
$K_{m, n-m}$	$\frac{t_{\text{rel}}}{2} (\log n + \Theta(1))$	$C(\frac{m}{n}) t_{\text{rel}} (\log n + \Theta(1))$

2.6.1 Hypercube

Here $G = \{0, 1\}^d$, $n = 2^d$, and with our time normalization one has $t_{\text{rel}} = 1$. Thus, Theorem 1.1 implies that mixing of $\text{Avg}(G)$ on the hypercube occurs with cutoff at the same times $T = \frac{t_{\text{rel}}}{2} (\log d + \Theta(1))$ of that of $\text{RW}(G)$; we refer to [21, Example, p. 987] for further details on mixing for $\text{RW}(G)$ on the hypercube. Therefore, by combining this with Proposition 2.1, we readily obtain an explicit lower bound for L^1 - and L^2 -mixing of $\text{Avg}(G)$. By Jensen inequality, Theorem 1.1 boils down, then, to establishing a matching upper bound for the L^2 -mixing time of $\text{Avg}(G)$.

For this purpose, let us observe that the general approach via L^2 -contraction from Proposition 2.3 would give only $T \leq t_{\text{rel}} \Theta(d)$. On the other hand, the entropy contraction from Proposition 2.7 yields $T \leq t_{\text{rel}} (\log d + \Theta(1))$ for the L^1 -distance, that is an upper bound off by a factor 2. Indeed, this latter bound follows from Corollary 2.8 and the already mentioned fact that $\kappa(G) = 1$ when G is the hypercube. Instead, to prove the upper bound in Theorem 1.1, we provide a tight estimate for $\text{CRW}(G)$ which controls the L^2 -distance via the first decomposition in Proposition 2.5; see Proposition 3.1 below.

Furthermore, we observe that the lower bound based on relative entropy in Proposition 2.9 does not provide the correct order of magnitude for the L^1 -mixing time in the hypercube: since $\langle \text{deg} \rangle = d$, this yields $T \geq (1 - \varepsilon) t_{\text{rel}}$ for every $\varepsilon \in (0, 1)$. Finally, we remark that the hypercube example seems to disprove Conjecture 1 in [23]. Indeed, mixing for $\text{Avg}(G)$ occurs at times $\Theta(t_{\text{rel}} \log \log n)$, while the authors in [23] conjectured that mixing must occur at times either of order $\Theta(t_{\text{rel}})$ or $\Theta(t_{\text{rel}} \log n)$.

2.6.2 Complete bipartite graphs

For complete bipartite graphs $K_{m, n-m}$, both the L^1 - and L^2 -mixing times for $\text{Avg}(G)$ are strictly larger than their analogues for $\text{RW}(G)$, (cf. Tables 1 and Table 2) and therefore the lower bound of Proposition 2.1 turns out to be too poor. Moreover, the leading-order

constants depend also on $\lim_{n \rightarrow \infty} \frac{m}{n} \in [0, 1/2]$: when $m = \frac{n}{2}$ (regular case), one recovers the complete-graph values of the mixing times for both L^1 - and L^2 -mixing, asymptotically saturating Aldous-Lanoue's inequality in (2.8); when $m = 1$ (star graph), L^1 -mixing occurs strictly in between L^1 - and L^2 -mixing for $\text{RW}(G)$, while L^2 -mixing for $\text{Avg}(G)$ sits strictly in between L^2 -mixing for $\text{RW}(G)$ and the upper bound prescribed by (2.8). Analogous observations hold for all other intermediate regimes of m .

For the L^1 and L^2 mixing we use two different strategies of proof. For what concerns L^1 , we extend to complete bipartite graphs $G = K_{m, n-m}$ the probabilistic arguments developed in [9] for the complete graph K_n . As for L^2 , due to the symmetries of the underlying graph, we exploit a representation of the L^2 -distance in terms of a simpler 5-state Markov chain, and analyze its asymptotic behavior in detail.

We remark that, as for the case of the complete graph considered in [23], the entropic lower bound of Proposition 2.9 predicts the correct order of the L^1 -mixing time also for complete bipartite graphs. On the contrary, the upper bound based on relative entropy contraction of Proposition 2.7 yields an estimates that is off by a $\log \log n$ factor in this case.

3 Mixing on the hypercube

This section is devoted to the proof of Theorem 1.1. We let $G = (V, E)$ be the d -hypercube $V = \{0, 1\}^d$, whose edge set consists of unordered pairs $\{x, x^i\}$, with $x = (x_1, \dots, x_d) \in \{0, 1\}^d$ and

$$x_j^i = (x^i)_j := \begin{cases} x_j & \text{if } j \neq i \\ 1 - x_i & \text{otherwise} . \end{cases}$$

By Proposition 2.6 and transitivity of the underlying graph, *all* Dirac masses are worst-case initial conditions for $\text{Avg}(G)$. Hence, throughout this section, we fix an arbitrary $x_0 \in V$, and set $\xi = \mathbf{1}_{x_0}$ as an initial condition.

As already discussed in Section 2.6.1, the claim in Theorem 1.1 reduces to estimating from above the L^2 -mixing time of $\text{Avg}(G)$. In view of (2.10), we have, for all $t \geq 0$,

$$\mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = 2^d \mathbf{P}_{x_0, x_0}^{\text{CRW}}(X_t = Y_t) - 1 . \quad (3.1)$$

In order to estimate the probability on the right-hand side of (3.1), we analyze the Hamming distance (cf. (3.2) below) between these two coupled walks. It is well-known that, in absence of coupling effects, i.e., for two independent copies of $\text{RW}(G)$ on $G = \{0, 1\}^d$, the process induced by this projection corresponds to the Markov chain on $\{0, 1, \dots, d\}$ known as Ehrenfest urn (see, e.g., [19, Sect. 2.3]). The Hamming distance between the two coordinates of $\text{CRW}(G)$ still yields a birth-and-death chain, but this time with a "defect" due to the interaction. For the purpose of proving Theorem 1.1, we carry out an analysis of the L^2 -mixing time of such a perturbation of the Ehrenfest urn. Mixing with respect to L^1 - and separation distances for perturbations of birth-and-death chains has been studied in [14] and [11], see, e.g., [10, Th. 4.10]. However, for the L^2 -distance, the situation seems to be much less generally understood. For this we combine general results on birth-and-death chains with the recent characterization [16] of L^2 -mixing times in the reversible context.

3.1 A perturbation of the Ehrenfest urn

We start with some definitions. Consider the two birth-and-death chains obtained from both $\text{RW}(G) \otimes \text{RW}(G)$ (i.e., two independent copies of $\text{RW}(G)$) and $\text{CRW}(G)$ via the

projection

$$(x, y) \in \{0, 1\}^d \times \{0, 1\}^d \mapsto |x - y| := \sum_{i=1}^d |x_i - y_i| \in \{0, 1, \dots, d\}. \quad (3.2)$$

Let $P_t(k, \ell)$ and $S_t(k, \ell)$, $k, \ell \in \{0, 1, \dots, d\}$, denote the transition kernels of these two Markov chains, and let \mathcal{L}^P and \mathcal{L}^S denote their infinitesimal generators, respectively. Writing $p_k^{P,S}$, resp. $q_k^{P,S}$, for the associated birth, resp. death, rates when in state $k \in \{0, 1, \dots, d\}$, we have

$$p_k^P = d - k, \quad q_k^P = k, \quad (3.3)$$

and

$$p_k^S = \mathbf{1}_{k \neq 0} p_k^P + \mathbf{1}_{k=0} \frac{d}{2}, \quad q_k^S = \mathbf{1}_{k \neq 1} q_k^P + \mathbf{1}_{k=1} \frac{1}{2}, \quad (3.4)$$

while the Binomial distribution ν given by

$$\nu(k) := \binom{d}{k} 2^{-d}, \quad k \in \{0, 1, \dots, d\}, \quad (3.5)$$

is reversible for both chains. Finally, we let P_t and S_t represent the Markov semigroups of the (continuous-time) Ehrenfest urn and its perturbation, respectively.

With these definitions, the right-hand side of (3.1) further reads as

$$2^d S_t(0, 0) - 1 = \left\| S_t \left(\frac{\mathbf{1}_0}{\nu} \right) - 1 \right\|_{L^\infty(\nu)} = \left\| S_{t/2} \left(\frac{\mathbf{1}_0}{\nu} \right) - 1 \right\|_{L^2(\nu)}^2,$$

where the second identity is due to reversibility; see, e.g., [19, Prop. 4.15]. The desired upper bound on the L^2 -mixing of $\text{Avg}(G)$ at times $\frac{1}{2} \log d + \Theta(1)$ then follows by (3.1) as soon as the S -chain is well-mixed in L^2 at times $\frac{1}{4} \log d + \Theta(1)$. This last claim is established in the next proposition, which is the main technical result of this section.

Proposition 3.1. *For every $\varepsilon > 0$, there exists $c = c_\varepsilon > 0$ such that, for all d sufficiently large,*

$$\left\| S_t \left(\frac{\mathbf{1}_0}{\nu} \right) - 1 \right\|_{L^2(\nu)}^2 \leq \varepsilon$$

holds for all $t > \frac{1}{4} \log d + c$.

The proof of this result is postponed to Section 3.2 below.

Remark 3.2. By using the integration by parts formula and well-known monotonicity properties of birth-and-death chains (see, e.g., [15, Lem. 4.1]), we have the following comparison inequalities:

$$P_t(0, 0) \leq S_t(0, 0) \leq P_{t/2}(0, 0), \quad t \geq 0. \quad (3.6)$$

Since the P -chain mixes abruptly at times $t \approx \frac{1}{4} \log d$, the bounds in (3.6) are insufficient to establish cutoff for the S -chain. In this respect, Proposition 3.1 tells us that the first inequality in (3.6) is sharp for times $t \approx \frac{1}{4} \log d$. Therefore, the second inequality — which, roughly speaking, accounts for the early delay of the S -chain when exiting state 0 — needs to be improved for such times.

3.2 Proof of Proposition 3.1

The idea of the proof Proposition 3.1 is based on the probabilistic approach to L^2 -mixing recently introduced in [16], which links L^2 -mixing times to estimates of suitable hitting times. Building on such a machinery, the following lemma represents the main

technical step to derive the claim in Proposition 3.1. Its proof combines two ingredients: the knowledge of the distribution of hitting times for birth-and-death chains ([5]), and a discrete weighted Hardy's inequality ([22]). As a result, we obtain a comparison between the P -chain and the S -chain that is much finer than the one implied by (3.6). This, together with the results in [16], provides a tight comparison between the mixing behavior of the two chains.

In what follows, $\mathbb{P}_k^{P,S}$ denotes the law of the P - or S -chain defined in Section 3.1 when starting from state $k \in \{0, 1, \dots, d\}$, while $\nu([0, M]) := \sum_{k=0}^{M-1} \nu(k)$ (cf. (3.5)).

Lemma 3.3. *Define τ_M to be the first hitting time of the state $M \in \{0, 1, \dots, d\}$. Then, for every $\sigma > 0$, there exists $b = b(\sigma) > 0$ satisfying, for all d large enough,*

$$\mathbb{P}_0^S(\tau_M > t + b) \leq (\nu([0, M]))^\sigma + \mathbb{P}_0^P(\tau_M > t), \quad t > 0, M \in \{1, \dots, d/2\}.$$

Proof. As shown in [5] (see also, e.g., [10]), for any birth-and-death chain on $\{0, 1, \dots, d\}$ with rate matrix \mathcal{L} started in $k = 0$, the hitting time τ_M of the target $M \in \{1, \dots, d\}$ is distributed as the sum of M independent exponential random variables with rates given by $(\lambda_{M,i})_{i=0}^{M-1}$, these being the distinct eigenvalues of $-\mathcal{L}|_M$, the generator of the sub-Markovian chain \mathcal{L} killed upon exiting $\{0, 1, \dots, M - 1\}$.

Now, while the eigenvalues and eigenfunctions of the Ehrenfest urn \mathcal{L}^P defined in Section 3.1 are well-known (see, e.g., [18]), neither those of $\mathcal{L}^P|_M$ nor of $\mathcal{L}^S|_M$ are explicit. Nevertheless, a comparison of eigenvalues is available. Letting $U := \text{diag}(\sqrt{\nu})$ and $\tilde{\mathcal{L}}^{P,S} := U\mathcal{L}^{P,S}U^{-1}$, we have

$$-\tilde{\mathcal{L}}^S = -\tilde{\mathcal{L}}^P + \mathcal{R},$$

with $\mathcal{R} = (\mathcal{R}_{ij})_{i,j=0}^d$ being the following rank-one matrix: $\mathcal{R}_{ij} = 0$ if $i + j > 1$, and

$$\mathcal{R}_{00} = -\frac{d}{2}, \quad \mathcal{R}_{01} = \mathcal{R}_{10} = \sqrt{\frac{d}{2}}, \quad \mathcal{R}_{11} = -\frac{1}{2}.$$

By removing, for every $M \in \{1, \dots, d\}$, the last $d - M + 1$ rows and columns of such matrices, we recover analogous relations for the sub-Markovian chains killed upon exiting $\{0, 1, \dots, M - 1\}$; in particular, the symmetric matrix $-\tilde{\mathcal{L}}^S|_M$ is a rank-one perturbation of $-\tilde{\mathcal{L}}^P|_M$. Therefore, Weyl theorem (see, e.g., [17, Th. 4.3.1]) yields

$$(\lambda_{M,i-1}^P \vee \lambda_{M,i-1}^S) \leq \lambda_{M,i}^S \leq \lambda_{M,i}^P, \quad i = 0, \dots, M - 1. \tag{3.7}$$

Moreover, the variational characterization of the spectral bounds $\lambda_{M,0}^P$ and $\lambda_{M,0}^S$ and the definitions of rates in (3.3)–(3.4) ensure that

$$\frac{1}{2} \lambda_{M,0}^P \leq \lambda_{M,0}^S. \tag{3.8}$$

Now, let $Z_{M,i}^P, \tilde{Z}_{M,i}^P \sim \text{Exp}(\lambda_{M,i}^P)$ and $Z_{M,i}^S, \tilde{Z}_{M,i}^S \sim \text{Exp}(\lambda_{M,i}^S)$ be all mutually independent exponential r.v.'s, with \mathbb{P} denoting their joint law; then, for all $s > 0$ and $M \in \{1, \dots, d\}$, (3.7)–(3.8) yield

$$\mathbb{P}\left(\sum_{i=0}^{M-1} Z_{M,i}^S > s\right) \leq \mathbb{P}\left(Z_{M,0}^S + \sum_{i=1}^{M-1} Z_{M,i}^P > s\right) \leq \mathbb{P}\left(2\tilde{Z}_{M,0}^P + \sum_{i=0}^{M-1} Z_{M,i}^P > s\right).$$

Setting $s = t + b > 0$ in the above inequality, by independence, we further get

$$\mathbb{P}\left(\sum_{i=0}^{M-1} Z_{M,i}^S > t + b\right) \leq \mathbb{P}(Z_{M,0}^P > b/2) + \mathbb{P}\left(\sum_{i=0}^{M-1} Z_{M,i}^P > t\right).$$

In view of $Z_{M,0}^P \sim \text{Exp}(\lambda_{M,0}^P)$ and the aforementioned explicit distribution of τ_M under $\mathbb{P}_0^{P,S}$, the above inequality reads as

$$\mathbb{P}_0^S(\tau_M > t + b) \leq \exp\left(-\frac{b}{2} \lambda_{M,0}^P\right) + \mathbb{P}_0^P(\tau_M > t). \tag{3.9}$$

We estimate the first term on the right-hand side above in Appendix A below, by proving

$$\limsup_{d \rightarrow \infty} \sup_{1 \leq M \leq d/2} (\lambda_{M,0}^P)^{-1} \log\left(\frac{1}{\nu([0, M])}\right) < \infty,$$

which, combined with (3.9), yields the assertion of the lemma. \square

We are now in a good position to prove Proposition 3.1.

Proof of Proposition 3.1. We start by recalling some of the results in [16] (namely, Prop. 3.8, and Thm.s 1.2 & 5.1), suitably adapted to the context of a generic birth-and-death chain on $\{0, 1, \dots, d\}$ with transition matrix Q_t , reversible measure μ , and starting from $k = 0$. Concisely, the following chain of inequalities hold true in this context:

$$\rho_0^Q \leq T_{2,0}^Q \leq \bar{\rho}_0^Q + 5 t_{\text{rel}}^Q \leq \rho_0^Q + 8 t_{\text{LS}}^Q + 10 t_{\text{rel}}^Q, \tag{3.10}$$

where:

- $T_{2,0}^Q$ is the L^2 -mixing time of the Q -chain when starting from $k = 0$, i.e.,

$$T_{2,0}^Q := \inf \left\{ t \geq 0 : \left\| Q_t \left(\frac{\mathbf{1}_0}{\mu} \right) - 1 \right\|_{L^2(\mu)} \leq \frac{1}{2} \right\};$$

- $t_{\text{rel}}^Q := (\lambda_{\text{gap}}^Q)^{-1}$ is the relaxation time of the chain;
- $t_{\text{LS}}^Q := (\beta_{\text{LS}}^Q)^{-1}$ is the inverse of the Log-Sobolev constant of the chain;
- letting $\mu([0, M]) := \sum_{\ell=0}^{M-1} \mu(\ell)$ and $\mu([M, d]) := \sum_{\ell=M}^d \mu(\ell)$, $M \in \{1, \dots, d\}$,

$$\bar{\rho}_0^Q := \inf \left\{ t \geq 0 : \mathbb{P}_0^Q(\tau_M > t) \leq \mu([0, M])^3 \text{ for all } M \in \{1, \dots, d/2\} \right\}; \tag{3.11}$$

- ρ_0^Q is defined analogously to $\bar{\rho}_0^Q$ with $\mu([0, M])^3$ replaced by

$$\mu([0, M]) + \frac{1}{2} \sqrt{\mu([0, M]) \mu([M, d])}.$$

We now specialize the inequalities in (3.10) to our setting of P - and S -chain. It is immediate to check by tensorization and comparison of Dirichlet forms that $t_{\text{rel}}^P, t_{\text{LS}}^P, t_{\text{rel}}^S, t_{\text{LS}}^S = O(1)$. Further, we know the L^2 -mixing of the Ehrenfest urn, namely $T_{2,0}^P \leq \frac{1}{4} \log d + \Theta(1)$. These two observations yield

$$\rho_0^P \leq \bar{\rho}_0^P \leq \frac{1}{4} \log d + a,$$

for some $a > 0$ large enough. In view of the definition of $\bar{\rho}_0^P$ (see (3.11)), and Lemma 3.3, we get, for all $M \in \{1, \dots, d/2\}$,

$$\begin{aligned} \mathbb{P}_0^S \left(\tau_M > \frac{1}{4} \log d + a + b \right) &\leq \nu([0, M])^3 + \mathbb{P}_0^P \left(\tau_M > \frac{1}{4} \log d + a \right) \\ &\leq 2 \nu([0, M])^3, \end{aligned}$$

where $b = b(\sigma) > 0$ is the constant appearing in the statement of Lemma 3.3 for $\sigma = 3$. Note that $2 \nu([0, M])^3 \leq \nu([0, M])$ for all $M \in \{1, \dots, d/2\}$. Therefore, the definition of ρ_0^S , the last inequality in (3.10) and $t_{\text{rel}}^S, t_{\text{LS}}^S = O(1)$ yield the desired upper bound on $T_{2,0}^S$, thus, concluding the proof of the proposition. \square

4 L^1 -mixing on complete bipartite graphs

In this section, we prove Theorem 1.2. All throughout, we let $m \in \{1, \dots, \lfloor \frac{n}{2} \rfloor\}$, and consider the complete bipartite graph with parts $|\mathcal{C}_1| = m$ and $|\mathcal{C}_2| = n - m$. In this context, \mathcal{L} in (2.1) reads as follows:

$$\mathcal{L}f(\eta) = \sum_{x \in \mathcal{C}_1} \sum_{y \in \mathcal{C}_2} (f(\eta^{xy}) - f(\eta)) , \quad \eta \in \Delta(V) , f : \Delta(V) \rightarrow \mathbb{R} .$$

The proof of Theorem 1.2 is divided into several steps. Following [9], we use the discretized dynamics defined as follows. Define

$$H := \lfloor \log_2 n - \log(n)^{1/3} \rfloor .$$

We divide a unit mass into 2^H chunks. To each chunk we assign a label $u \in \{1, \dots, 2^H\}$, a position in V , and a mass equal to 2^{-H} . We add a cemetery state $\{\dagger\}$ and define a dynamics of these mass chunks as follows. For every chunk $u = 1, \dots, 2^H$, we let $e_t(u) \in V \cup \{\dagger\}$ denote its position at time $t \geq 0$, and introduce two sequences $(\beta_t(u))_{t \geq 0}$ and $(\alpha_t(u))_{t \geq 0}$, with $\beta_0(u) = \alpha_0(u) = 0$ and $\beta_t(u) \in \{0, 1\}$ and $\alpha_t(u) \in \mathbb{N}_0$ for all $t > 0$. We write $(w_t(x))_{x \in V \cup \{\dagger\}}$ for the amount of mass at x at time t , that is

$$w_t(x) := \sum_{u=1}^{2^H} 2^{-H} \mathbf{1}_{e_t(u)=x} , \quad x \in V \cup \{\dagger\} , t \geq 0 .$$

The dynamics is given by the following process:

1. all chunks are initially placed on the same site $x_0 \in V$: $e_0(u) = x_0, \forall u$;
2. the process follows the same edge updates of the averaging process: each edge rings independently with rate 1;
3. when edge xy rings at time t , two types of scenarios can occur:
 - (a) good scenarios:
 - i. there are no chunks at the endpoints of xy , and nothing occurs, or
 - ii. exactly one of the endpoints, say y , of the edges is empty, and x contains 2^i chunks with $i > 0$; in this case half of the chunks on x are selected uniformly at random and moved to y ; the other half remains at x ; concurrently, for all chunks u sitting at either x or y , we set $\alpha_t(u) := \alpha_{t-}(u) + 1$;
 - (b) bad scenarios:
 - i. β -event: if both $w_{t-}(x) > 0$ and $w_{t-}(y) > 0$, then set

$$\beta_{t'}(u) = 1 , \quad e_{t'}(u) = \dagger ,$$

for all $t' \geq t$, and all $u = 1, \dots, 2^H$ such that $e_{t-}(u) \in \{x, y\}$. In particular, $w_t(x) = w_t(y) = 0$, and $w_t(\dagger) = w_{t-}(\dagger) + w_{t-}(x) + w_{t-}(y)$.

- ii. α -event: if exactly one of the extremes, say y , of the edges is empty, and x contains exactly one chunk u , i.e. $\alpha_{t-}(u) = H$, then set, for all $t' \geq t$,

$$\alpha_{t'}(u) = H + 1 , \quad e_{t'}(u) = \dagger .$$

In words, the discretized mass splits exactly along an edge if the proposed edge is not occupied at both ends, while if both ends are occupied all mass chunks from that edge are sent to the cemetery. Moreover, if a single mass chunk sits at one of the endpoints of the proposed edge, then it is sent to the cemetery. By construction, the following properties hold a.s.:

Cutoff for the averaging process

1. the total mass of the chunks is preserved by the dynamics, i.e.,

$$w_t(\dagger) + \sum_{x \in V} w_t(x) = 1, \quad t \geq 0,$$

and the mass at \dagger increases when a bad scenario occurs.

2. if $(\eta_t)_{t \geq 0}$ denotes the Averaging process with all the mass initially placed at $x_0 \in V$, then

$$\eta_t(x) \geq w_t(x), \quad x \in V, t \geq 0;$$

3. the process $(w_t)_{t \geq 0}$ eventually sends all chunks to \dagger , i.e.,

$$\lim_{t \rightarrow \infty} w_t(\dagger) = 1.$$

4. for any chunk not in \dagger at time $t \geq 0$, the variable α satisfies

$$w_t(e_t(u)) = 2^{-\alpha_t(u)}, \quad \forall u : e_t(u) \neq \dagger.$$

We start with a technical lemma, and all throughout adopt the following notation (cf. (1.2)):

$$\begin{aligned} t_* &:= \frac{1}{2\gamma} \log_2 n, & \gamma &:= \frac{m(n-m)}{n}, \\ t_*^\lambda &:= t_* + \frac{\lambda}{\gamma} \sqrt{\log n}, & \lambda &\in \mathbb{R}. \end{aligned} \tag{4.1}$$

Lemma 4.1. *For every $\varepsilon > 0$, there exist constants $\lambda = \lambda(\varepsilon) > 0$ and $b = b(\varepsilon) > 0$ such that for all $u = 1, \dots, 2^H$,*

$$\liminf_{n \rightarrow \infty} \mathbb{P} \left(\log_2(n) - b\sqrt{\log n} \leq \alpha_{t_*^\lambda}(u) \leq H, \beta_{t_*^\lambda}(u) = 0 \right) \geq 1 - \varepsilon.$$

The proof of Lemma 4.1 is postponed to the end of this section. Next, we assume its validity and finish the proof of the theorem through the following two propositions.

Proposition 4.2 (Lower bound). *For every $\varepsilon > 0$, there exists $\lambda = \lambda(\varepsilon) > 0$ such that for every $x_0 \in V$*

$$\liminf_{n \rightarrow \infty} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_{t_*^\lambda}}{\pi} - 1 \right\|_1 \right] \geq 2 - \varepsilon.$$

Proof. In view of the definition and properties of the process $(w_t)_{t \geq 0}$, setting $t = t_*^\lambda$,

$$\begin{aligned} \frac{1}{2} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_1 \right] &= \sum_{x \in V} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left(\eta_t(x) - \frac{1}{n} \right)_+ \right] \\ &\geq \sum_{x \in V} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left(\eta_t(x) - \frac{1}{n} \right)_+ \mathbf{1}_{\{w_t(x) \geq \frac{1}{n}\}} \right] \\ &\geq \sum_{x \in V} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left(w_t(x) - \frac{1}{n} \right) \mathbf{1}_{\{w_t(x) \geq \frac{1}{n}\}} \right] \\ &= 2^{-H} \sum_{u=1}^{2^H} \mathbb{P}(e_t(u) \neq \dagger) - \frac{1}{n} \sum_{x \in V} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\mathbf{1}_{\{w_t(x) \geq \frac{1}{n}\}} \right] \\ &\geq \mathbb{P}(\alpha_t(u_0) \leq H, \beta_t(u_0) = 0) - \frac{2^H}{n}, \end{aligned}$$

where u_0 is any fixed chunk label, and we use symmetry and the fact that at most 2^H sites can be non empty at any given time. Finally, since $2^H = o(n)$, Lemma 4.1 concludes the proof. \square

Proposition 4.3 (Upper bound). *For every $\varepsilon > 0$, there exists $\lambda = \lambda(\varepsilon) > 0$ such that for every $x_0 \in V$*

$$\limsup_{n \rightarrow \infty} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_{t_*^\lambda}}{\pi} - 1 \right\|_1 \right] \leq \varepsilon .$$

Proof. Fix constants $c, b > 0$ to be specified later, and let $T^- := t_*^{-c}$ be defined as in (4.1). Given the processes (η_t) and (w_t) , we define a third process $(\tilde{\eta}_t)$ as follows:

1. for times $t < T^-$, $\tilde{\eta}_t = w_t$;
2. for times $t \geq T^-$, we let $(\tilde{\eta}_t)$ evolve as an Averaging process, following the same edge updates of (η_t) , but starting at time T^- from the following configuration (possibly, with total mass strictly less than one): for every $x \in V$,

$$\tilde{\eta}_{T^-}(x) := \begin{cases} w_{T^-}(x) & \text{if } w_{T^-}(x) \leq \frac{2^{b\sqrt{\log n}}}{n} \\ 0 & \text{otherwise ;} \end{cases}$$

3. define $|\tilde{\eta}| := \sum_{x \in V} \tilde{\eta}_{T^-}(x) \in [0, 1]$.

In view of this construction, for all $t \geq T^-$,

$$\begin{aligned} & \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_1 \right] \\ & \leq \sum_{x \in V} \mathbb{E} \left[\left| \eta_t(x) - \frac{1}{n} - \left(\tilde{\eta}_t(x) - \frac{|\tilde{\eta}|}{n} \right) \right| \right] + \sum_{x \in V} \mathbb{E} \left[\left| \tilde{\eta}_t(x) - \frac{|\tilde{\eta}|}{n} \right| \right] \\ & \leq 1 - \mathbb{E} [|\tilde{\eta}|] + \sum_{x \in V} \mathbb{E} [\eta_t(x) - \tilde{\eta}_t(x)] + \mathbb{E} \left[|\tilde{\eta}| \left\| \frac{\tilde{\eta}_t}{|\tilde{\eta}| \pi} - 1 \right\|_1 \right] , \end{aligned}$$

where we use the triangle inequality twice and the fact that $\eta_t(x) \geq \tilde{\eta}_t(x)$ a.s.. As a consequence,

$$\mathbb{E} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_1 \right] \leq 2(1 - \mathbb{E} [|\tilde{\eta}|]) + \mathbb{E} \left[|\tilde{\eta}| \left\| \frac{\tilde{\eta}_t}{|\tilde{\eta}| \pi} - 1 \right\|_1 \right] . \tag{4.2}$$

The first term on the right-hand side of (4.2) is at most 2ε if the constants $c = c(\varepsilon)$, where c is the constant appearing in $T^- = t_*^{-c}$, and $b = b(\varepsilon)$ are large enough. Indeed, by Lemma 4.1,

$$\mathbb{E} [|\tilde{\eta}|] = 2^{-H} \sum_{u=1}^{2^H} \mathbb{P} \left(\log_2(n) - b\sqrt{\log n} \leq \alpha_{T^-}(u) \leq H, \beta_{T^-}(u) = 0 \right) \geq 1 - \varepsilon .$$

As for the second term on the right-hand side of (4.2), letting \mathcal{F}_s denote the σ -algebra generated by the edge updates up to time s , the Markov property yields

$$\begin{aligned} \mathbb{E} \left[|\tilde{\eta}| \left\| \frac{\tilde{\eta}_t}{|\tilde{\eta}| \pi} - 1 \right\|_1 \right] &= \mathbb{E} \left[|\tilde{\eta}| \mathbb{E} \left[\left\| \frac{\tilde{\eta}_t}{|\tilde{\eta}| \pi} - 1 \right\|_1 \middle| \mathcal{F}_{T^-} \right] \right] \\ &\leq \mathbb{E} \left[|\tilde{\eta}| \sqrt{\mathbb{E} \left[\left\| \frac{\tilde{\eta}_t}{|\tilde{\eta}| \pi} - 1 \right\|_2^2 \middle| \mathcal{F}_{T^-} \right]} \right] \\ &\leq \mathbb{E} \left[|\tilde{\eta}| \sqrt{e^{-(t-T^-)/t_{\text{rel}}} \left\| \frac{\tilde{\eta}_{T^-}}{|\tilde{\eta}| \pi} - 1 \right\|_2^2} \right] . \end{aligned}$$

Cutoff for the averaging process

Above we used Cauchy-Schwarz inequality and Aldous-Lanoue's L^2 -upper bound in Proposition 2.3 from time T^- to $t \geq T^-$. By definition of the process $(\tilde{\eta}_t)$,

$$\left\| \frac{\tilde{\eta}_{T^-}}{|\tilde{\eta}|^\pi} - 1 \right\|_2 \leq \left\| \frac{\tilde{\eta}_{T^-}}{|\tilde{\eta}|^\pi} \right\|_2 \leq \|\tilde{\eta}_{T^-}\|_\infty \frac{n}{|\tilde{\eta}|} \leq \frac{2^{b\sqrt{\log n}}}{|\tilde{\eta}|}.$$

Finally, applying this with $t = t_\star^\lambda$ and recalling that $t_{\text{rel}} = 2/m$,

$$\begin{aligned} \mathbb{E} \left[|\tilde{\eta}| \left\| \frac{\tilde{\eta}_{t_\star^\lambda}}{|\tilde{\eta}|^\pi} - 1 \right\|_1 \right] &\leq \exp\left(\frac{-(t_\star^\lambda - T^-)}{2t_{\text{rel}}}\right) 2^{b\sqrt{\log n}} \mathbb{E} \left[\sqrt{|\tilde{\eta}|} \right] \\ &\leq \exp\left(-\frac{n}{n-m} \frac{c+\lambda}{4} \sqrt{\log n}\right) 2^{b\sqrt{\log n}} \\ &\leq \exp\left(\left(b \log 2 - \frac{c+\lambda}{4}\right) \sqrt{\log n}\right). \end{aligned}$$

For all $\varepsilon > 0$ and $b, c > 0$, by choosing $\lambda > 0$ sufficiently large, the right-hand side is bounded above by ε for all n large enough. This concludes the proof. \square

Proof of Lemma 4.1. Set $t := t_\star^{-c}$ and fix $u \in \{1, \dots, 2^H\}$. We start by proving that, for every $c > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\beta_t(u) = 0) = 1. \quad (4.3)$$

Suppose the chunk u starts in \mathcal{C}_2 and note that the hitting time τ_1 of the set \mathcal{C}_1 for u is $\text{Exp}(m/2)$. Once the chunk is in \mathcal{C}_1 it will reach \mathcal{C}_2 at some later time, and will visit again \mathcal{C}_1 at some time τ_2 . Iterating, we define the successive arrival times to \mathcal{C}_1 as $0 < \tau_1 < \tau_2 < \dots$. Then, by neglecting the time spent in \mathcal{C}_1 each time, we can estimate

$$\mathbb{P}(\tau_j < t) \leq \mathbb{P}\left(\sum_{i=1}^j X_i < t\right), \quad j \in \mathbb{N}, \quad (4.4)$$

where X_i are i.i.d. $\text{Exp}(m/2)$, where the factor $\frac{1}{2}$ is due to the laziness of the chunk. If instead u starts in \mathcal{C}_1 then $\tau_1 = 0$ and the inequality in (4.4) holds with the sum on the right hand side going up to $j - 1$ instead of j . Therefore, recalling that $t = O(m^{-1} \log n)$, taking e.g. $j = \log^2(n)$ in (4.4) we have

$$\mathbb{P}(u \text{ hits } \mathcal{C}_1 \text{ more than } \log^2(n) \text{ times before } t) = o(1). \quad (4.5)$$

Let us consider the event E_i that after the i -th hitting of \mathcal{C}_1 , u experiences a β -event before moving to \mathcal{C}_2 or to \dagger because of an α -event. We are going to show that uniformly in $i \geq 1$,

$$\log^2(n) \mathbb{P}(E_i) = o(1). \quad (4.6)$$

In view of (4.5), by a union bound this implies

$$\mathbb{P}(\exists s \leq t \text{ s.t. } e_{s^-}(u) \in \mathcal{C}_1, \beta_s(u) = 1) = o(1).$$

To prove (4.6), observe that, while in \mathcal{C}_1 , u experiences attempts to average with other vertices at a rate $n - m$, and that at any attempt one has the following options:

1. u remains in \mathcal{C}_1 with probability q_0 , which is smaller than $1/2$;
2. u moves to \dagger because of a β -event with probability $q_1 \leq 2^H/(n - m) =: p = O(2^{-\log^{1/3} n})$, since at any given time there are at most 2^H vertices with positive mass;

3. u moves to \mathcal{C}_2 , or to \dagger because of an α -event, with probability $1 - q_0 - q_1 \geq \frac{1}{3}$.

In particular, the overall number of attempts before moving to \mathcal{C}_2 or to \dagger is dominated by a geometric random variable with parameter $1/2$. Thus, a union bound shows that $\mathbb{P}(E_i) \leq 6p = O(2^{-\log^{1/3} n})$, which proves (4.6).

In order to prove (4.3), we are left with proving

$$\mathbb{P}(\exists s \leq t \text{ s.t. } e_{s^-}(u) \in \mathcal{C}_2, \beta_s(u) = 1) = o(1).$$

To this end, observe that, while in \mathcal{C}_2 , u experiences attempts to average with other vertices at a rate m , and thus by the argument in (4.5) we may assume that there are at most $\log^2(n)$ such attempts up to time t . Thus, letting $\Xi = \{0 < \xi_1 < \xi_2 < \dots\}$ denote the set of random times at which these attempts occur, it suffices to show that uniformly in $s \leq t$,

$$\log^2(n) \mathbb{P}(e_{s^-}(u) \in \mathcal{C}_2, \beta_s(u) = 1 \mid s \in \Xi) = o(1).$$

To prove this, suppose $s \in \Xi$, say $s = \xi_i$ for some $i \in \mathbb{N}$. The event $e_{s^-}(u) \in \mathcal{C}_2, \beta_s(u) = 1$ implies that the attempt involves u and a vertex $x \in \mathcal{C}_1$ with $w_s(x) > 0$. Thus

$$\begin{aligned} & \mathbb{P}(e_{s^-}(u) \in \mathcal{C}_2, \beta_s(u) = 1 \mid s \in \Xi) \\ &= \sum_{x \in \mathcal{C}_1} \mathbb{P}(\text{edge}(e_{s^-}(u), x) \text{ rings at time } s, w_{s^-}(x) > 0 \mid s \in \Xi) \\ &= \frac{1}{m} \sum_{x \in \mathcal{C}_1} \mathbb{P}(w_{s^-}(x) > 0 \mid s \in \Xi, \text{edge}(e_{s^-}(u), x) \text{ rings at time } s). \end{aligned}$$

For any $a > 0$, the event $w_{s^-}(x) > 0$ implies that either x has received mass within the time interval $[s - a, s]$, or x has received no mass and experienced less than H updates within the time interval $[s - a, s]$. Therefore,

$$\begin{aligned} & \mathbb{P}(e_{s^-}(u) \in \mathcal{C}_2, \beta_s(u) = 1 \mid s \in \Xi) \\ & \leq \max_{x \in \mathcal{C}_1} \mathbb{P}(w_{s^-}(x) > 0 \mid s \in \Xi, \text{edge}(e_{s^-}(u), x) \text{ rings at time } s) \\ & \leq \max_{x \in \mathcal{C}_1} \mathbb{P}(A_x) + \max_{x \in \mathcal{C}_1} \mathbb{P}(B_x), \end{aligned}$$

where A_x is the event that x received mass within the time interval $[s^- - a, s^-]$, and B_x is the event that x experienced less than H updates within the time interval $[s^- - a, s^-]$. Recalling that there are at most 2^H vertices with positive mass, one has

$$\mathbb{P}(A_x) \leq 2^H a.$$

On the other hand, the probability of less than H updates at x in $[s^- - a, s^-]$ is bounded by the probability that a Poisson random variable with parameter $(n - m)a \geq na/2$ is less or equal than H . Thus,

$$\mathbb{P}(B_x) \leq (na/2)^{H+1} e^{-na/2}.$$

We can take $a = (H2^H)^{-1}$ so that $\mathbb{P}(A_x) = o(1)$ and

$$\mathbb{P}(B_x) \leq 2^{\log^2(n)} e^{-2^{\log^{1/4}(n)}} = o(1).$$

This ends the proof of (4.3).

To conclude the proof of the lemma we are left with estimating $\alpha_t(u)$, that is the number of good-scenario splittings involving the chunk labeled u up to time t . We suppose $e_0(u) \in \mathcal{C}_1$; the proof for the case $e_0(u) \in \mathcal{C}_2$ is analogous. We proceed as follows:

Cutoff for the averaging process

1. let $(Z_i)_{i \in \mathbb{N}}$ be i.i.d. $\text{Geom}(1/2)$; each $Z_i \in \mathbb{N}$ indicates the number of attempts needed for u to succeed in moving to the other block;
2. given $(Z_i)_{i \in \mathbb{N}}$, let $(Y_{i,j})_{j=1}^{Z_i}$ be a family of independent Exponential r.v.'s, with

$$Y_{i,j} \sim \begin{cases} \text{Exp}(n-m) & \text{if } i \text{ odd} \\ \text{Exp}(m) & \text{if } i \text{ even}; \end{cases}$$

the r.v. $\sum_{j=1}^{Z_i} Y_{i,j}$ represents the time u spends during the i -th visit to a block before moving to the other block.

Note that the random variable

$$K_t := \sup \left\{ k \geq 1 : \sum_{i=1}^k \sum_{j=1}^{Z_i} Y_{i,j} \leq t \right\}$$

represents the number of switchings from a block to the other. Therefore, if $\beta_t(u) = 0$,

$$(H+1) \wedge \sum_{i=1}^{K_t} Z_i \leq \alpha_t(u) \leq \sum_{i=1}^{K_t+1} Z_i,$$

where the $+1$ term in the second summation bounds the number of splittings that u experiences in the time intercurring from its last change of block.

Recall the definition of γ from (4.1). On the one hand, since for every $i \geq 1$

$$\mathbb{E} \left[\sum_{j=1}^{Z_i} Y_{i,j} + \sum_{j=1}^{Z_{i+1}} Y_{i,j} \right] = \frac{2}{\gamma},$$

by the CLT, for every $\varepsilon > 0$, there exists $\bar{\zeta} = \bar{\zeta}(\varepsilon) > 0$ such that, for all $\zeta > \bar{\zeta}$ and for all n large enough

$$\mathbb{P} \left(\sum_{i=1}^{\gamma t + \zeta \sqrt{\log n}} \sum_{j=1}^{Z_i} Y_{i,j} > t \right) \geq 1 - \varepsilon. \quad (4.7)$$

Note that $\bar{\zeta}$ can be chosen independently of the constant c in the definition of $t = t_\star^{-c}$. From (4.7) it follows that

$$\mathbb{P} \left(K_t + 1 < \gamma t + \zeta \sqrt{\log n} \right) \geq 1 - \varepsilon.$$

Hence,

$$\begin{aligned} \mathbb{P}(\alpha_t(u) \leq H, \beta_t(u) = 0) &\geq \mathbb{P} \left(\sum_{i=1}^{K_t+1} Z_i \leq H, \beta_t(u) = 0 \right) \\ &\geq \mathbb{P} \left(\sum_{i=1}^{K_t+1} Z_i \leq H, K_t + 1 \leq \gamma t + \zeta \sqrt{\log n} \right) - \varepsilon. \end{aligned} \quad (4.8)$$

To finish the proof, observe that

$$\mathbb{E} \left[\sum_{i=1}^{\gamma t + \zeta \sqrt{\log n}} Z_i \right] = \log_2 n - 2(c - \zeta) \sqrt{\log n}.$$

Thus, by choosing $c = c(\zeta, \varepsilon) > 0$ large enough, another application of the CLT shows that (4.8) is at least $1 - 2\varepsilon$.

For what concerns the lower bound on $\alpha_t(u)$, we proceed in a similar fashion. By the CLT there exists some $\bar{\xi}(\varepsilon) > 0$ such that for all $\xi > \bar{\xi}(\varepsilon)$

$$\mathbb{P} \left(\sum_{i=1}^{\gamma t - \xi \sqrt{\log n}} \sum_{j=1}^{Z_i} Y_{i,j} \leq t \right) \geq 1 - \varepsilon, \quad \mathbb{P} \left(K_t > \gamma t - \xi \sqrt{\log n} \right) \geq 1 - \varepsilon.$$

Therefore, for every $\varepsilon > 0$ and $\xi > \bar{\xi}(\varepsilon) > 0$,

$$\begin{aligned} & \mathbb{P} \left(\alpha_t(u) > \log_2 n - b \sqrt{\log n}, \beta_t(u) = 0 \right) \\ & \geq \mathbb{P} \left(\sum_{i=1}^{K_t} Z_i > \log_2 n - b \sqrt{\log n}, \beta_t(u) = 0 \right) \\ & \geq \mathbb{P} \left(\sum_{i=1}^{K_t} Z_i > \log_2 n - b \sqrt{\log n}, K_t > \gamma t - \xi \sqrt{\log n} \right) - \varepsilon \geq 1 - 2\varepsilon, \end{aligned}$$

where the last step follows by

$$\mathbb{E} \left[\sum_{i=1}^{\gamma t - \xi \sqrt{\log n}} Z_i \right] = \log_2 n - 2(\xi + c) \sqrt{\log n},$$

and the CLT by choosing $b = b(\varepsilon, c, \xi) > 0$ large enough. This completes the proof of Lemma 4.1. \square

5 L^2 -mixing on complete bipartite graphs

This section is devoted to the proof of Theorem 1.3. We keep the same notation as at the beginning of Section 4; in particular, $1 \leq m \leq n/2$, and $|\mathcal{C}_1| = m$, $|\mathcal{C}_2| = n - m$. Furthermore, we remark that, in this context, L^{RW} of $\text{RW}(G)$ reads as

$$L^{\text{RW}}\psi(x) = \frac{1}{2} \sum_{y \in V} (\psi(y) - \psi(x)) (\mathbf{1}_{x \in \mathcal{C}_1} \mathbf{1}_{y \in \mathcal{C}_2} + \mathbf{1}_{x \in \mathcal{C}_2} \mathbf{1}_{y \in \mathcal{C}_1}), \quad x \in V, \psi : V \rightarrow \mathbb{R},$$

from which it is straightforward to check that

$$t_{\text{rel}} = \frac{2}{m}. \tag{5.1}$$

Moreover, we remark that

$$\text{the worst-case initial condition is } \xi = \mathbf{1}_{x_0} \text{ with } x_0 \in \mathcal{C}_2. \tag{5.2}$$

Indeed, observe that:

- starting with all the mass at a single site x_0 is worst-case by Proposition 2.6, and the dependence on x_0 is only through the subset \mathcal{C}_i , $i = 1, 2$, such that $x_0 \in \mathcal{C}_i$;
- starting from $\eta = \mathbf{1}_{x_0}$, $x \in V$, after the first effective update for $\text{Avg}(G)$, the next configuration is $\frac{1}{2}\mathbf{1}_{x_0} + \frac{1}{2}\mathbf{1}_y$, for some $y \in V$ in $\mathcal{C}_i \not\ni x_0$; further, from that time on, mixing will not depend on neither x_0 nor y ;
- letting τ_i denote the random time of the first effective update when $x_0 \in \mathcal{C}_i$,

$$\tau_i \sim \begin{cases} \text{Exp}(n - m) & \text{if } i = 1 \\ \text{Exp}(m) & \text{if } i = 2; \end{cases}$$

Hence, since $n - m \geq m$, such times can be coupled such that, a.s., $\tau_1 \leq \tau_2$.

Before entering the details of the proof of the theorem (Sect. 5.2 below), let us first discuss the particularly relevant case of the star graph, i.e., $G = K_{m, n-m}$ with $m = 1$.

5.1 A motivating example: the star graph

Set $m = 1$, and let x_* denote the unique element in \mathcal{C}_1 (i.e., the center of the star). As pointed out in Section 2.6.2, the L^2 -mixing must occur between times $\frac{t_{\text{rel}}}{2} \log n$ and $t_{\text{rel}} \log n$. As shown in the following corollary, which specializes Theorem 1.3 to the setting $m = 1$, the L^2 -cutoff occurs strictly in between these two extremes, namely around times $\frac{3}{4}t_{\text{rel}} \log n$.

Corollary 5.1 (*L^2 -cutoff on the star graph*). *Let $T(a) := \frac{3}{4}t_{\text{rel}}(\log n + a)$, $a \in \mathbb{R}$. Then,*

$$\lim_{n \rightarrow \infty} \sup_{\xi \in \Delta(V)} \mathbb{E}_\xi \left[\left\| \frac{\eta^{T(a)}}{\pi} - 1 \right\|_2^2 \right] = e^{-a}, \quad a \in \mathbb{R}.$$

Let us briefly comment over this result, with the aim of illustrating that the arguments used to derive Proposition 2.3, while they provide a sharp L^2 -estimate for the case of the complete graph K_n , they do not prove to be useful in the context of the star graph.

A simple computation shows that

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] &= -\mathbb{E}_\xi \left[\mathcal{E}^{\text{RW}} \left(\frac{\eta_t}{\pi} \right) \right] \\ &= -\frac{1}{t_{\text{rel}}} \left(\mathbb{E}_\xi \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] + \mathbb{E}_\xi \left[\left(\frac{\eta_t}{\pi}(x_*) - 1 \right)^2 \right] \right), \end{aligned}$$

for all $\xi \in \Delta(V)$ and $t \geq 0$. By integrating over time and setting $\xi = \mathbf{1}_x$ for some $x \neq x_*$,

$$\mathbb{E}_{\mathbf{1}_x} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = e^{-\frac{t}{t_{\text{rel}}}} \left\| \frac{\mathbf{1}_x}{\pi} - 1 \right\|_2^2 - e^{-\frac{t}{t_{\text{rel}}}} \int_0^t \frac{e^{-\frac{s}{t_{\text{rel}}}}}{t_{\text{rel}}} \mathbb{E}_{\mathbf{1}_x} \left[\left(\frac{\eta_s}{\pi}(x_*) - 1 \right)^2 \right] ds. \quad (5.3)$$

Since $\left\| \frac{\mathbf{1}_x}{\pi} - 1 \right\|_2^2 \approx n$, the first term on the right-hand side above is divergent for $t = (1 - \delta)t_{\text{rel}} \log n$, $\delta \in (0, 1)$, and drops down to order $\Theta(1)$ only if $t = t_{\text{rel}}(\log n + a)$. Therefore, L^2 -mixing of $\text{Avg}(G)$ at the strictly smaller time $T(a) = \frac{3}{4}t_{\text{rel}}(\log n + a)$ must be due to cancellations of diverging terms between first and second terms on the right-hand side of (5.3). In particular, in view of the duality relation in (2.9), controlling this latter term boils down to estimating the transition probabilities of $\text{CRW}(G)$ in the time interval $[0, T(a)]$. This approach proves to be difficult and, instead of dealing with $\text{CRW}(G)$, one might be tempted to bound the last term in (5.3) by relying on the duality with the simpler process $\text{RW}(G)$. In order to do so, one must use Jensen inequality, obtaining

$$e^{-\frac{t}{t_{\text{rel}}}} \int_0^t \frac{e^{-\frac{s}{t_{\text{rel}}}}}{t_{\text{rel}}} \mathbb{E}_{\mathbf{1}_x} \left[\left(\frac{\eta_s}{\pi}(x_*) - 1 \right)^2 \right] ds \geq e^{-\frac{t}{t_{\text{rel}}}} \int_0^t \frac{e^{-\frac{s}{t_{\text{rel}}}}}{t_{\text{rel}}} \left(\mathbb{E}_{\mathbf{1}_x} \left[\frac{\eta_s}{\pi} - 1 \right] \right)^2 ds. \quad (5.4)$$

Nonetheless, it turns out that the right-hand side of (5.4) is of order $O(1)$ for all $t \geq 0$, implying that (5.3) combined with the lower bound in (5.4) is too weak to deduce the desired control.

In the proof of Theorem 1.3 we show how to handle the transition probabilities of $\text{CRW}(G)$ around times $T(a)$ to derive first order estimates on the L^2 -distance without relying directly on (5.3).

5.2 Proof of Theorem 1.3

We now turn to the proof of Theorem 1.3 in the general case $m \in \{1, \dots, \lfloor \frac{n}{2} \rfloor\}$. In preparation of this, we introduce

$$b \in [0, \frac{1}{2}] \mapsto B(b) := \sqrt{9 - 32b + 32b^2} \in [1, 3], \quad (5.5)$$

and note that the function $B(b)$ is non-increasing, as well as $B(0) = 3$ and $B(\frac{1}{2}) = 1$. Finally, we recall the definitions of $\theta = \theta_{m,n-m}$, $T(a)$, and $D(b)$ from the statement of Theorem 1.3, and present three remarks.

Remark 5.2 (Asymptotics of θ). By a first-order Taylor expansion,

$$\lim_{n \rightarrow \infty} \theta_{m,n-m} = \begin{cases} \lim_{n \rightarrow \infty} \frac{3n}{8m} \left(1 - \left(1 - \frac{16}{9} \frac{m}{n} (1 + o(1)) \right) \right) = \frac{2}{3} & \text{if } m/n \rightarrow 0, \\ \frac{3}{8b} \left(1 - \sqrt{1 - \frac{32}{9} b(1-b)} \right) \in [\frac{1}{2}, \frac{2}{3}] & \text{if } m/n \rightarrow b \in (0, \frac{1}{2}]. \end{cases}$$

Remark 5.3 ($T(a)$ in terms of t_{rel}). Recall from (5.1) that $t_{\text{rel}} = \frac{2}{m}$, and (5.5). Define

$$b \in (0, \frac{1}{2}] \mapsto C(b) := \frac{4b}{3-B} \in (\frac{3}{4}, 1],$$

and note that $C(b)$ is increasing on $(0, \frac{1}{2})$, as well as $C(0) := \lim_{b \downarrow 0} C(b) = \frac{3}{4}$, and $C(\frac{1}{2}) = 1$. Then, assuming that $\frac{m}{n} \rightarrow b \in [0, \frac{1}{2}]$, $T(a)$ appearing in Theorem 1.3 can be rewritten in terms of the relaxation time using the following asymptotic equality: for all $a \in \mathbb{R}$,

$$T(a) := \frac{1}{\theta m} (\log n + a) \sim C(b) t_{\text{rel}} (\log n + a).$$

Remark 5.4 (Shape of the cutoff profile). The cutoff profile at time $T(a)$ is exponential in $a \in \mathbb{R}$, and depends on $b = \lim_{n \rightarrow \infty} m/n \in [0, 1/2]$ through $D(b)$ in (1.5), which exhibits the following behavior:

$$D(b) \in [0, \frac{\sqrt{6}-2}{4}], \quad \lim_{b \downarrow 0} D(b) = D(0) = D(\frac{1}{2}) = 0, \quad D(\frac{3}{8}) = \frac{\sqrt{6}-2}{4}.$$

This term $D(b)$ does not appear when considering L^2 -mixing for RW(G) on $G = K_{m,n-m}$.

Proof of Theorem 1.3. Fix $x_0 \in V$ and $t > 0$; then, by duality in (2.9),

$$\begin{aligned} \mathbb{E}_{1_{x_0}} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] &= n \sum_{x \in V} \mathbb{E}_{1_{x_0}} \left[(\eta_t(x))^2 \right] - 1 \\ &= n \sum_{x \in \mathcal{C}_1} \mathbf{P}_{x_0, x_0}^{\text{CRW}} (X_t = x, Y_t = x) + n \sum_{y \in \mathcal{C}_2} \mathbf{P}_{x_0, x_0}^{\text{CRW}} (X_t = y, Y_t = y) - 1. \end{aligned} \tag{5.6}$$

In what follows, we analyze the above transition probabilities for CRW(G) by lumping some states and considering the following auxiliary Markov chain with five states

$$\begin{aligned} 0_1 &:= \{(x, x) : x \in \mathcal{C}_1\} \\ 0_2 &:= \{(y, y) : y \in \mathcal{C}_2\} \\ 1 &:= \{(x, y) \cup (y, x) : x \in \mathcal{C}_1, y \in \mathcal{C}_2\} \\ 2_1 &:= \{(x, x') : x, x' \in \mathcal{C}_1, x \neq x'\} \\ 2_2 &:= \{(y, y') : y, y' \in \mathcal{C}_2, y \neq y'\}, \end{aligned}$$

and corresponding 5×5 transition rate matrix

$$Q = \begin{pmatrix} -\frac{3(n-m)}{4} & \frac{n-m}{4} & \frac{n-m}{2} & 0 & 0 \\ \frac{m}{4} & -\frac{3m}{4} & \frac{m}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & -\frac{n-1}{2} & \frac{m-1}{2} & \frac{n-m-1}{2} \\ 0 & 0 & n-m & -(n-m) & 0 \\ 0 & 0 & m & 0 & -m \end{pmatrix}. \tag{5.7}$$

and reversible measure $\mu : \{0_1, 0_2, 1, 2_1, 2_2\} \rightarrow [0, 1]$ given by

$$\mu = n^{-2} \begin{pmatrix} m & n-m & 2m(n-m) & m(m-1) & (n-m)(n-m-1) \end{pmatrix}. \quad (5.8)$$

This lumping allows us to rewrite (5.6) as follows:

$$\mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = n e^{tQ}(0_i, 0_1) + n e^{tQ}(0_i, 0_2) - 1, \quad x_0 \in \mathcal{C}_i, i \in \{1, 2\}.$$

Let $\{(\rho_j, \phi_j)\}_{j=0}^4$ denote the eigenvalue-eigenfunction pairs of $-Q$, with $(\rho_0 = 0, \phi_0 \equiv 1)$, ρ_1 the second smallest eigenvalue, and $\{\phi_j\}_{j=0}^4$ orthonormal basis in $L^2(\mu)$; hence, we have

$$\begin{aligned} \mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] &= n e^{tQ}(0_i, 0_1) + n e^{tQ}(0_i, 0_2) - 1 \\ &= n \mu(0_1) \sum_{j=0}^4 e^{-\rho_j t} \phi_j(0_i) \phi_j(0_1) + n \mu(0_2) \sum_{j=0}^4 e^{-\rho_j t} \phi_j(0_i) \phi_j(0_2) - 1 \\ &= \sum_{j=1}^4 e^{-\rho_j t} \phi_j(0_i) \left(\frac{m}{n} \phi_j(0_1) + \frac{n-m}{n} \phi_j(0_2) \right), \quad x_0 \in \mathcal{C}_i, i \in \{1, 2\}. \end{aligned} \quad (5.9)$$

Furthermore, the following eigenpair (ρ_2, ϕ_2) for $-Q$ is explicit:

$$\rho_2 = \frac{n}{2}, \quad \phi_2 = \sqrt{\frac{1}{n-m}} \begin{pmatrix} -\frac{n-m}{m} & 1 & \frac{n-2m}{2m} & -\frac{n-m}{m} & 1 \end{pmatrix}.$$

Although the remaining eigenvalues ρ_1, ρ_3, ρ_4 and eigenfunctions ϕ_1, ϕ_3, ϕ_4 of $-Q$ are, in general, not explicit, we now provide sharp asymptotic estimates which will yield the desired claim.

Recall from (5.2) that, in order to analyze the worst-case mixing, it suffices to consider $x_0 \in \mathcal{C}_2$. Hence, with such a choice, (5.9) reads as

$$\mathbb{E}_{\mathbf{1}_{x_0}} \left[\left\| \frac{\eta_t}{\pi} - 1 \right\|_2^2 \right] = \sum_{j=1}^4 e^{-\rho_j t} \Psi_j, \quad (5.10)$$

with

$$\Psi_j := \phi_j(0_2) \left(\frac{m}{n} \phi_j(0_1) + \frac{n-m}{n} \phi_j(0_2) \right), \quad j = 1, \dots, 4. \quad (5.11)$$

Recall that $\lim_{n \rightarrow \infty} \frac{m}{n} = b \in [0, \frac{1}{2}]$. In view of (5.10), we conclude the proof of the theorem as soon as we show that, for $t = T(a)$, $a \in \mathbb{R}$, as given in (1.3), the following three claims hold, as $n \rightarrow \infty$:

$$\Psi_1 = (1 + D(b) + o(1)) n, \quad e^{-\rho_1 t} n = (1 + o(1)) e^{-a}, \quad (5.12)$$

and

$$\sum_{j=2}^4 e^{-\rho_j t} \Psi_j = o(1). \quad (5.13)$$

Actually, in order to have the latter claim in (5.13), it will suffice to prove

$$\min \{\rho_2, \rho_3, \rho_4\} \geq \frac{2}{5} n. \quad (5.14)$$

Cutoff for the averaging process

Indeed, since $\mu(0_2) = \frac{n-m}{n^2}$ and $\|\phi_j\|_{L^2(\mu)} = 1$, we get, for all $j = 0, \dots, 4$,

$$|\phi_j(0_2)| \leq \frac{n}{\sqrt{n-m}},$$

as well as, recalling that $\mu(0_1) = \frac{m}{n^2}$,

$$\Psi_j := \frac{m}{n}\phi_j(0_1) + \frac{n-m}{n}\phi_j(0_2) \leq \sqrt{\frac{m}{n}(\phi_j(0_1))^2 + \frac{n-m}{n}(\phi_j(0_2))^2} \leq \sqrt{n}.$$

By combining these two estimates with that in (5.14), we get

$$\left| \sum_{j=2}^4 e^{-\rho_j t} \Psi_j \right| \leq 3\sqrt{2}n e^{-\frac{2}{5}nt} = o(1),$$

where for the last step we used the expression of $t = T(a) = \frac{1}{\theta m}(\log n + a)$ and $\frac{2}{5}\frac{n}{\theta m} > 1$.

We are now left with proving (5.14) and the two claims in (5.12) about approximating the quantity Ψ_1 and the eigenvalue ρ_1 . Recall the definition of Q in (5.7), of its reversible measure μ in (5.8) and $\{(\rho_j, \phi_j)\}_{j=0}^4$ of its eigenpairs; then, letting $U := \text{diag}(\sqrt{\mu})$, we introduce

$$W := \frac{1}{n}U(-Q)U^{-1} - (U\phi_0)(U\phi_0)^\top.$$

It is immediate to check that W is symmetric and that its eigenpairs $\{(\rho_j^W, \varphi_j^W)\}_{j=0}^4$ are related to those of $-Q$ as follows: $\varphi_j^W = U\phi_j$ for all $j = 0, \dots, 4$, while

$$\rho_0^W = -1 \neq 0 = \rho_0, \quad \rho_j^W = \frac{1}{n}\rho_j, \quad j = 1, \dots, 4. \quad (5.15)$$

Furthermore, it is simple to check that W decomposes as a sum $W = S + R$, with S being a block-diagonal matrix and $R = O(n^{-1/2})$, i.e., all entries of R are of order $O(n^{-1/2})$. More precisely, letting $b = b_n := \frac{m}{n} \in (0, \frac{1}{2}]$, S reads as

$$S = \left(\begin{array}{c|c} S_0 & 0 \\ \hline 0 & S_1 \end{array} \right), \quad \text{with } S_0 := \begin{pmatrix} \frac{3}{4}(1-b) & -\frac{1}{4}\sqrt{b(1-b)} \\ -\frac{1}{4}\sqrt{b(1-b)} & \frac{3}{4}b \end{pmatrix},$$

and

$$S_1 := \begin{pmatrix} \frac{(1-2b)^2}{2} & -\frac{\sqrt{b(1-b)}(1+2b)}{\sqrt{2}} & -\frac{\sqrt{b(1-b)}(3-2b)}{\sqrt{2}} \\ -\frac{\sqrt{b(1-b)}(1+2b)}{\sqrt{2}} & 1-b(1+b) & -b(1-b) \\ -\frac{\sqrt{b(1-b)}(3-2b)}{\sqrt{2}} & -b(1-b) & b(3-b)-1 \end{pmatrix}.$$

The eigenvalues $\{\rho_j^S\}_{j=0}^4$ of S are then the union of those of S_0 and S_1 , which are explicit:

$$\text{spec}(S_0) = \left\{ \frac{3-B}{8}, \frac{3+B}{8} \right\}, \quad \text{spec}(S_1) = \left\{ -1, \frac{1}{2}, 1 \right\},$$

where $B = B(b)$ was defined in (5.5). In particular, the smallest eigenvalue ρ_0^S equals -1 , the second smallest ρ_1^S is $\frac{3-B}{8}$, and, since $B \in [1, 3]$, $\rho_j^S \geq \frac{1}{2}$ for all $j = 2, \dots, 4$. This simple observation implies the following two facts:

1. on the one side, since $W = S + R$, $R = O(n^{-1/2})$ and (5.15), Weyl theorem for eigenvalues of sums of Hermitian matrices (see, e.g., [17, Th. 4.3.1]) yields

$$\left| \frac{1}{n}\rho_j - \rho_j^S \right| = \left| \rho_j^W - \rho_j^S \right| = O(n^{-1/2}), \quad j = 1, \dots, 4, \quad (5.16)$$

and, in turn, (5.14);

2. on the other side, we obtain $\delta_1^S := \min_{j \neq 1} |\rho_1^S - \rho_j^S| \geq \frac{1}{4}$.

By employing the latter claim and $R = O(n^{-1/2})$, we estimate the eigenfunction ϕ_1 of $-Q$ as follows. By applying Davis-Kahan theorem [12] to the $L^2(1)$ -normalized eigenfunctions $\varphi_1^W = U\phi_1$ and φ_1^S associated to the second-smallest eigenvalues of W and S , respectively, we obtain

$$\|U\phi_1 - \varphi_1^S\|_{L^2(1)} = \|\varphi_1^W - \varphi_1^S\|_{L^2(1)} \leq \frac{\|R\|_{L^2(1)}}{\delta_1^S} = O(n^{-\frac{1}{2}}). \tag{5.17}$$

Then, due to the simple form of S and S_0 , the $L^2(1)$ -normalized eigenfunction φ_1^S is explicit and given by $\varphi_1^S = \frac{1}{Z}(A, 1, 0, 0, 0)$, where

$$A = A(b) := \frac{6b - 3 + B}{2\sqrt{b(1-b)}}, \quad Z := \sqrt{A^2 + 1}.$$

Note that if $\frac{m}{n} = b = o(1)$, then $A(b) = (1 + o(1))\frac{1}{3}\sqrt{b}$, while $A(b) \in (0, \infty)$ if $b \in (0, \frac{1}{2}]$. By combining (5.17) with the definitions of φ_1^S , $U := \text{diag}(\sqrt{n})$ and Ψ_1 (see (5.11)), we get

$$\Psi_1 = (1 + o(1))n \left(\frac{A}{A^2 + 1} \frac{\sqrt{b}}{\sqrt{1-b}} + \frac{1}{A^2 + 1} \right) = n(1 + D + o(1)),$$

where $D = D(b)$ was given in (1.5); this corresponds to the first claim in (5.12).

In order to prove the second claim in (5.12) for all values of $b = \frac{m}{n}$, we need the following estimate for ρ_1 :

$$\rho_1 = m \left(\theta + O\left(\frac{1}{n}\right) \right), \tag{5.18}$$

which is, in general, sharper than the one we obtained in (5.16) above. To this aim, we analyze the characteristic polynomial $\lambda \mapsto p(\lambda)$ of $-Q$ divided by $(\lambda - \rho_0)(\lambda - \rho_4) = \lambda(\lambda - \frac{n}{2})$:

$$q(\lambda) := \frac{p(\lambda)}{\lambda(\lambda - \frac{n}{2})} = -\lambda^3 + \frac{7n - 2}{4}\lambda^2 + \frac{2m^2 + (1 - 2m)n - 3n^2}{4}\lambda + \frac{mn(n - m)}{2},$$

and show that, within a window of order $O(\frac{m}{n})$ around θm , $q(\lambda)$ changes sign; since $\frac{m}{n} \ll \theta m < \frac{1}{3}n$, $\min\{\rho_2, \rho_3, \rho_4\} > \frac{2}{5}n$ (see (5.14)) and $\rho_0 = 0$, this would ensure the validity of (5.18). Recall the definition of θ in (1.4); then, writing $\frac{m}{n} = b \in (0, \frac{1}{2}]$, for any $c \in \mathbb{R}$, we have

$$q\left(bn \left(\theta + \frac{c}{n}\right)\right) = -\frac{1 + o(1)}{64} (3 - B + b(18c + 10B - 16) + b^2(16 - 64c) + 64cb^3) n^2.$$

Recall that $B \in [1, 3]$ if $b \in (0, \frac{1}{2}]$, while $B(b) = 3 - \frac{16}{3}b + O(b^2)$ if $b = o(1)$. In either case, by choosing $c = c(b) \in \mathbb{R}$ large (resp. small) enough, the expression in parenthesis above becomes positive (resp. negative). This concludes the proof of the theorem. \square

A Proof of (3.2)

Recall that $\nu(j) := 2^{-d} \binom{d}{j}$, $j \in \{0, 1, \dots, d\}$, and that we want to prove

$$\limsup_{d \rightarrow \infty} \sup_{1 \leq M \leq d/2} (\lambda_{M,0}^P)^{-1} \log \left(\frac{1}{\nu([0, M])} \right) < \infty. \tag{A.1}$$

We divide the proof into two steps.

A.1 Discrete weighted Hardy’s inequality

We start by estimating $\lambda_{M,0}^P$ by means of a discrete weighted Hardy’s inequality (see, e.g., [22]).

Lemma A.1. For every $M \in \{1, \dots, d/2\}$, define

$$C_M := \max_{0 \leq k \leq M-1} \nu([0, k]) \sum_{j=k}^{M-1} \frac{1}{\nu(j)(d-j)}. \tag{A.2}$$

Then,

$$\frac{1}{4C_M} \leq \lambda_{M,0}^P \leq \frac{1}{C_M}.$$

Proof. Recall that $\lambda_{M,0}^P > 0$ is the largest value $\lambda > 0$ satisfying

$$\lambda \nu(f^2) \leq \mathcal{E}_P(f), \quad f \in \mathcal{F}_{M-1} := \{f : \{0, 1, \dots, d\} \rightarrow \mathbb{R} : f = 0 \text{ on } [M, d]\}.$$

We now apply the following version of weighted Hardy’s inequality (e.g., [22] or [10, Prop. A.3]): for all positive measures π, μ on $\{0, 1, \dots, M-1\}$, define

$$\Lambda = \Lambda_{\pi, \mu} := \frac{1}{4} \left(\max_{0 \leq k \leq M-1} \pi([0, k]) \sum_{j=k}^{M-1} \frac{1}{\mu(j)} \right)^{-1};$$

then

$$\Lambda \sum_{k=0}^{M-1} \left(\sum_{j=k}^{M-1} g(j) \right)^2 \pi(k) \leq \sum_{k=0}^{M-1} (g(k))^2 \mu(k), \quad g : \{0, 1, \dots, M-1\} \rightarrow \mathbb{R}.$$

Applying this to $\pi(k) := \nu(k)$, $\mu(k) := \nu(k)(d-k)$, we get, setting, for every $f \in \mathcal{F}_{M-1}$, $g(k) := f(k) - f(k+1)$,

$$4\Lambda \geq \lambda_{M,0}^P \geq \Lambda.$$

This concludes the proof of the lemma. □

In view of Lemma A.1, (A.1) is actually equivalent to

$$\limsup_{d \rightarrow \infty} \sup_{1 \leq M \leq d/2} C_M \log \left(\frac{1}{\nu([0, M])} \right) < \infty, \tag{A.3}$$

where $C_M > 0$ is given in (A.2). The rest of this section is devoted to the proof of (A.3). As a first step, by straightforward monotonicity arguments, we obtain the following upper bounds:

$$\begin{aligned} & \sup_{1 \leq M \leq d/2} C_M \log \left(\frac{1}{\nu([0, M])} \right) \\ &= \sup_{1 \leq M \leq d/2} \sup_{0 \leq k \leq M-1} \sum_{j=k}^{M-1} \frac{\nu([0, k])}{\nu(j)(d-j)} (-\log \nu([0, M-1])) \\ &\leq \sup_{1 \leq M \leq d/2} \sup_{0 \leq k \leq M-1} \frac{1}{d-M+1} \sum_{j=k}^{M-1} \frac{\nu([0, k])}{\nu(j)} (-\log \nu([0, M-1])) \\ &\leq \frac{3}{d} \sup_{0 \leq k < d/2} \nu([0, k]) \sup_{k+1 \leq M \leq d/2} \left(\sum_{j=k}^{M-1} \frac{1}{\nu(j)} \right) (-\log \nu([0, M-1])) \\ &\leq 3 \sup_{0 \leq k < d/2} \frac{1}{d} \nu([0, k]) \log \left(\frac{1}{\nu([0, k])} \right) \left(\sum_{j=k}^{d/2-1} \frac{1}{\nu(j)} \right). \end{aligned}$$

We have just showed that, by defining

$$\Gamma(k) = \Gamma_d(k) := \frac{1}{d} \nu([0, k]) \log \left(\frac{1}{\nu([0, k])} \right) \left(\sum_{j=k}^{d/2-1} \frac{1}{\nu(j)} \right), \quad 0 \leq k < d/2,$$

we have

$$\sup_{1 \leq M \leq d/2} C_M \log \left(\frac{1}{\nu([0, M])} \right) \leq 3 \sup_{0 \leq k < d/2} \Gamma(k). \quad (\text{A.4})$$

A.2 Estimating Γ

Recall from, e.g., [1] the following representation for the binomial distribution:

$$\nu(k) := 2^{-d} \binom{d}{k} = \frac{1}{\sqrt{2\pi d}} \frac{\exp(-dH(\frac{k}{d}) + \mathcal{S}^{(d)}(\frac{k}{d}))}{\sqrt{\frac{k}{d}(1-\frac{k}{d})}}, \quad 1 \leq k \leq d/2 - 1, \quad (\text{A.5})$$

where

$$H(p) = H(p||\frac{1}{2}) := p \log \frac{p}{1/2} + (1-p) \log \frac{1-p}{1/2}, \quad p \in [0, 1],$$

and $\mathcal{S}^{(d)} : (0, 1) \rightarrow \mathbb{R}$ is a function whose explicit form is unimportant for our purposes, yet satisfies the following bounds:

$$\mathcal{S}^{(d)}(p) := \mathcal{S}(d) - \mathcal{S}(pd) - \mathcal{S}((1-p)d), \quad \mathcal{S}(m) \in \left(-\frac{m^{-3}}{360}, \frac{m^{-1}}{12} \right).$$

In particular, there exist constants $s_1 < 0 < s_2$ such that

$$\mathcal{S}^{(d)}(\frac{k}{d}) \in (s_1, s_2), \quad 1 \leq k < d/2. \quad (\text{A.6})$$

It is now straightforward to check that ν satisfies the *convexity hypothesis*, shortly $\text{CONV}(c, \bar{d})$, in [7, App. A], with $\gamma = \nu$, $n_{\min} = 0$, $n_{\max} = d$, $\bar{n} = \lceil d/2 \rceil$, and some $c = c(s_1, s_2) \geq 2$, therein. More precisely,

1. $c^{-1} d/2 \leq d - d/2 \leq c d/2$ is trivial;
2. by monotonicity of ν on $\{0, 1, \dots, d/2\}$,

$$\frac{\nu(k-1)}{\nu(k)} \leq 1 \leq c \exp \left(-\frac{d/2-k}{c(d/2)} \right), \quad 1 \leq k \leq d/2;$$

3. by (A.5), (A.6), and

$$2 \left(\frac{1}{2} - p \right)^2 \leq H(p) \leq 4 \log 2 \left(\frac{1}{2} - p \right)^2, \quad p \in [0, 1],$$

we have, for all $d \in \mathbb{N}$ large enough and $0 \leq k < d/2$,

$$\frac{1}{c\sqrt{d/2}} \exp \left(-2cd \left(\frac{1}{2} - \frac{k}{d} \right)^2 \right) \leq \nu(k) \leq \frac{1}{c\sqrt{d/2}} \exp \left(-\frac{2d}{c} \left(\frac{1}{2} - \frac{k}{d} \right)^2 \right).$$

As a consequence of these three conditions and [7, Lem. A.2–4], we have, for some $C > 0$ independent of d , and all d large enough,

$$\nu([0, k]) \leq C \frac{d/2}{d/2-k} \nu(k), \quad 0 \leq k < d/2,$$

$$\sum_{j=k}^{d/2-1} \frac{1}{\nu(j)} \leq C \frac{d/2}{d/2-k} \frac{1}{\nu(k)}, \quad 0 \leq k < d/2,$$

and

$$\log \left(\frac{1}{\nu([0, k])} \right) \leq C \left(1 + d \left(\frac{1}{2} - \frac{k}{d} \right)^2 \right), \quad 0 \leq k \leq d/2 - \sqrt{d/2}.$$

Finally, combining these estimates as done in the proof of [7, Prop. A.5] yields

$$\limsup_{d \rightarrow \infty} \sup_{0 \leq k < d/2} \Gamma(k) < \infty,$$

thus, by (A.4), (A.3). This concludes the proof.

References

- [1] D. Alfers and H. Dinges. A normal approximation for beta and gamma tail probabilities. *Z. Wahrsch. Verw. Gebiete*, 65(3):399–420, 1984. MR0731229
- [2] David Aldous and Daniel Lanoue. A lecture on the averaging process. *Probab. Surv.*, 9:90–102, 2012. MR2908618
- [3] Koenraad M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *J. Phys. A*, 40(28):8127–8136, 2007. MR2344161
- [4] Alexandre Bristiel and Pietro Caputo. Entropy inequalities for random walks and permutations. *Ann. Inst. Henri Poincaré Probab. Stat.* (to appear), *arXiv:2109.06009*, 2021.
- [5] Mark Brown and Yi-Shi Shao. Identifying Coefficients in the Spectral Representation for First Passage Time Distributions. *Probab. Eng. Inf. Sci.*, 1(1):69–74, 1987.
- [6] Sergey G. Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006. MR2283379
- [7] N. Cancrini, F. Martinelli, and C. Roberto. The logarithmic Sobolev constant of Kawasaki dynamics under a mixing condition revisited. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(4):385–436, 2002. MR1914934
- [8] Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *Comm. Math. Phys.*, 388(2):793–818, 2021. MR4334247
- [9] Sourav Chatterjee, Persi Diaconis, Allan Sly, and Lingfu Zhang. A phase transition for repeated averages. *Ann. Probab.*, 50(1):1–17, 2022. MR4385355
- [10] Guan-Yu Chen and Laurent Saloff-Coste. On the mixing time and spectral gap for birth and death chains. *ALEA Lat. Am. J. Probab. Math. Stat.*, 10(1):293–321, 2013. MR3083928
- [11] Guan-Yu Chen and Laurent Saloff-Coste. Computing cutoff times of birth and death chains. *Electron. J. Probab.*, 20:no. 76, 47, 2015. MR3371435
- [12] Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM J. Numer. Anal.*, 7:1–46, 1970. MR0264450
- [13] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996. MR1410112
- [14] Persi Diaconis and Laurent Saloff-Coste. Separation cut-offs for birth and death chains. *Ann. Appl. Probab.*, 16(4):2098–2122, 2006. MR2288715
- [15] Jian Ding, Eyal Lubetzky, and Yuval Peres. Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields*, 146(1-2):61–85, 2010. MR2550359
- [16] Jonathan Hermon and Yuval Peres. A characterization of L_2 mixing and hypercontractivity via hitting times and maximal inequalities. *Probab. Theory Related Fields*, 170(3-4):769–800, 2018. MR3773799
- [17] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press; New York, second edition, 2012. MR2978290
- [18] Samuel Karlin and James McGregor. Ehrenfest urn models. *J. Appl. Probability*, 2:352–376, 1965. MR0184284

- [19] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. Second edition of [MR2466937], with contributions by Elizabeth L. Wilmer, with a chapter on “Coupling from the past” by James G. Propp and David B. Wilson.
- [20] Thomas M. Liggett. *Interacting particle systems*. Classics in Mathematics. Springer-Verlag, Berlin, 2005. Reprint of the 1985 original. MR2108619
- [21] Eyal Lubetzky and Allan Sly. Cutoff for general spin systems with arbitrary boundary conditions. *Comm. Pure Appl. Math.*, 67(6):982–1027, 2014. MR3193965
- [22] L. Miclo. An example of application of discrete Hardy’s inequalities. *Markov Process. Related Fields*, 5(3):319–330, 1999. MR1710983
- [23] Ramis Movassagh, Mario Szegedy, and Guanyang Wang. Repeated Averages on Graphs. *arXiv:2205.04535*, 2022.
- [24] Matteo Quattropiani and Federico Sau. Mixing of the Averaging process and its discrete dual on finite-dimensional geometries. *Ann. Appl. Probab.*, 33(2):936–971, 2023. MR4564423

Acknowledgments. F.S. wishes to thank Università Roma Tre, Dipartimento di Matematica e Fisica, for the very kind hospitality during an early stage of this work.

Electronic Journal of Probability

Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS¹)
- Easy interface (EJMS²)

Economical model of EJP-ECP

- Non profit, sponsored by IMS³, BS⁴, ProjectEuclid⁵
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund⁶ (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

¹LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

²EJMS: Electronic Journal Management System: <https://vtex.lt/services/ejms-peer-review/>

³IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

⁴BS: Bernoulli Society <http://www.bernoulli-society.org/>

⁵Project Euclid: <https://projecteuclid.org/>

⁶IMS Open Access Fund: <https://imstat.org/shop/donation/>