

Concentration and local smoothness of the averaging process^{*†}

Federico Sau[‡]

Abstract

We consider the averaging process on the discrete d -dimensional torus. On this graph, the process is known to converge to equilibrium on diffusive timescales, not exhibiting cutoff. In this work, we refine this picture in two ways. Firstly, we prove a concentration phenomenon of the averaging process around its mean, occurring on a shorter timescale than the one of its relaxation to equilibrium. Secondly, we establish sharp gradient estimates, which capture its fast local smoothness property. This is the first setting in which these two features of the averaging process — concentration and local smoothness — can be quantified. These features carry useful information on a number of large scale properties of the averaging process. As an illustration of this fact, we determine the limit profile of its distance to equilibrium and derive a quantitative hydrodynamic limit for it. Finally, we discuss their implications on cutoff for the binomial splitting process, the particle analogue of the averaging process.

Keywords: interacting particle systems; mixing of Markov chains; averaging process; hydrodynamic limits.

MSC2020 subject classifications: Primary 60K35, Secondary 60J27; 82B20; 82C26; 91D30.
Submitted to EJP on November 29, 2023, final version accepted on May 31, 2024.

1 Introduction

The *averaging process* on a graph is a random evolution of a probability mass function over its vertices. Its dynamics consists in repeatedly iterating the following two operations:

- (i) first, select an unordered pair of nearest neighbor vertices at the arrival times of i.i.d. Poisson clocks;

^{*}The author acknowledges financial support by “Microgrants 2022”, funded by Regione FVG, legge LR 2/2011.

[†]The author is grateful to Matteo Quattropani and Pietro Caputo for stimulating discussions. The author wishes to thank the two anonymous referees for their careful reading of the manuscript and useful comments.

[‡]Dipartimento di Matematica, Informatica e Geoscienze, Università degli Studi di Trieste, Trieste, Italy.
E-mail: federico.sau@units.it

(ii) then, level out the masses associated to them.

Hence, step (i) is random, while step (ii) is deterministic. In particular, each update is performed at some i.i.d. space-time locations, and conserves the total mass. Therefore, as time runs, for any mass initialization and provided the underlying graph is connected, the averaging dynamics will converge to a deterministic flat distribution. In view of the degeneracy of the equilibrium and of the intrinsic irreversibility of the averaging dynamics, a comprehensive quantitative picture of such a convergence in relation to the underlying graph structure is, despite the recent advances, still a largely open problem of both applied and theoretical interest.

The averaging process appears in the computer science literature as a basic model of opinion dynamics (or randomized distributed/gossip algorithms, see, e.g., [BGPS06, Sha09, MSW22] and references therein), together with a long list of variants, e.g., with updates involving only one vertex at a time (see, e.g., [BCN20] and references therein). In this realm, vertices are typically interpreted as agents, their masses as corresponding to their opinions, and the long-run limiting distribution arising from the local random interactions is then referred to as the consensus.

The mathematical interest on the averaging process as an interacting particle system was revived about a decade ago by a series of lectures and expository articles by Aldous [Ald11] and Aldous and Lanoue [AL12]. Since then, the list of works on the subject rapidly grew (see, e.g., [CDSZ22, Spi22, QS23, MSW22, Cao23, CQS23]). In particular, the seminal paper [AL12] was the first one to link this model to the theory of Markov chains mixing times [AF02, LP17, MT06], and established the first general bounds on the mixing of the averaging process. Only recently, [CDSZ22, QS23, CQS23] provided the first sharp asymptotic mixing results on specific geometries, either proving or excluding the occurrence of the so-called cutoff phenomenon [LP17, Chapter 18].

In this article, we consider the averaging process on large discrete d -dimensional tori. Mixing in this setting is covered by the analysis carried out in [QS23], which shows that the averaging process mixes gradually (i.e., without exhibiting an abrupt cutoff phenomenon) on diffusive timescales. Our aim is to refine this picture, by analyzing two new features of the averaging process — an early concentration phenomenon, and a fast local smoothness property — and extract from them quantitative information on its mixing and scaling behaviors. After a quick recap on the averaging process and its main properties, we present our findings first informally in Section 1.4, and then in detail in Section 2.

1.1 Setting and model definition

For every integer $d \geq 1$, let $\mathbb{T}_N^d := (\mathbb{Z}/N\mathbb{Z})^d$ be the discrete d -dimensional torus of size $N \in \mathbb{N}$. Moreover, let $\mathcal{P}(\mathbb{T}_N^d)$ denote the space of probability mass functions on \mathbb{T}_N^d , namely,

$$\mathcal{P}(\mathbb{T}_N^d) := \left\{ \eta \in [0, 1]^{\mathbb{T}_N^d} \mid \sum_{x \in \mathbb{T}_N^d} \eta(x) = 1 \right\}.$$

For such a compact subset $\mathcal{P}(\mathbb{T}_N^d)$ of $\mathbb{R}^{\mathbb{T}_N^d}$, we further let $\mathcal{C}(\mathcal{P}(\mathbb{T}_N^d))$ indicate the Banach space of continuous functions on $\mathcal{P}(\mathbb{T}_N^d)$ endowed with the uniform norm.

The *averaging process* on \mathbb{T}_N^d (shortly, $\text{Avg}(\mathbb{T}_N^d)$) is the Markov jump process evolving on $\mathcal{P}(\mathbb{T}_N^d)$ and with infinitesimal generator $\mathcal{L} = \mathcal{L}_N$ given, for all $f \in \mathcal{C}(\mathcal{P}(\mathbb{T}_N^d))$, as

$$\mathcal{L}f(\eta) := \frac{1}{2} \sum_{x \in \mathbb{T}_N^d} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} (f(\eta^{xy}) - f(\eta)), \quad \eta \in \mathcal{P}(\mathbb{T}_N^d).$$

Here, $|x - y|$ denotes the usual graph distance between vertices $x, y \in \mathbb{T}_N^d$, while $\eta^{xy} \in \mathcal{P}(\mathbb{T}_N^d)$ is obtained from $\eta \in \mathcal{P}(\mathbb{T}_N^d)$ by updating the masses $\eta(x)$ and $\eta(y)$ with their average value: for all x, y and $z \in \mathbb{T}_N^d$,

$$\eta^{xy}(z) := \begin{cases} \frac{1}{2}(\eta(x) + \eta(y)) & \text{if } z = x \text{ or } z = y \\ \eta(z) & \text{else .} \end{cases}$$

Henceforth, although the updates occur at both random times and locations, as time runs, the deterministic averages and the conservation of mass lead the (random) distribution to become flat, for any mass initialization. In formulas, this reads as

$$\eta_t^\xi \xrightarrow{t \rightarrow \infty} \pi \equiv 1/N^d, \quad \xi \in \mathcal{P}(\mathbb{T}_N^d), \quad (1.1)$$

where $(\eta_t^\xi)_{t \geq 0} = (\eta_t^{\xi, N})_{t \geq 0}$ denotes a trajectory of $\text{Avg}(\mathbb{T}_N^d)$ when starting at time $t = 0$ from $\xi \in \mathcal{P}(\mathbb{T}_N^d)$, $\pi = \pi_N$ is the uniform distribution on \mathbb{T}_N^d , and the above convergence holds a.s. with respect to the random sequence of updates. For the probability law corresponding to such random updates, we write $\mathbb{P} = \mathbb{P}^N$, while $\mathbb{E} = \mathbb{E}^N$ stands for the corresponding expectation.

1.2 Mixing for the averaging process

The statement in (1.1) concerns the qualitative long-run behavior of the averaging process on a fixed-size torus. A mixing analysis aims at quantifying such a convergence in the limit as $N \rightarrow \infty$, with the scope of better capturing the relevant timescales governing the relaxation of η_t^ξ , when starting from a worst-case initial condition $\xi \in \mathcal{P}(\mathbb{T}_N^d)$.

As a sensible notion of distance between the law of η_t^ξ and that of π , we consider the (mean) L^p -distance to equilibrium, given, for all $N \in \mathbb{N}$ and $p \in [1, 2]$, by

$$\mathbb{E} \left[\left\| \frac{\eta_t^\xi}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}} := \left(\sum_{x \in \mathbb{T}_N^d} \pi(x) \mathbb{E} \left[\left| \frac{\eta_t^\xi(x)}{\pi(x)} - 1 \right|^p \right] \right)^{\frac{1}{p}}, \quad \text{with } \xi \in \mathcal{P}(\mathbb{T}_N^d), \quad (1.2)$$

and define the corresponding (worst-case) L^p -mixing time as the first time $t \geq 0$ for which the above quantity falls below $1/2$ for all initial conditions $\xi \in \mathcal{P}(\mathbb{T}_N^d)$. Note that $\|\cdot\|_p = \|\cdot\|_{p, N}$ in (1.2) stands for the usual L^p -norm on (\mathbb{T}_N^d, π) . Further, we remark that (1.2) may also be interpreted as the L^p -Wasserstein metrics between the random $\eta_t^\xi \in \mathcal{P}(\mathbb{T}_N^d)$ and the deterministic $\pi \in \mathcal{P}(\mathbb{T}_N^d)$, when endowing $\mathcal{P}(\mathbb{T}_N^d)$ with the distance induced by $\|\cdot\|_p$. Note that stronger distances, e.g., total variation, are less relevant in the context of the averaging process; indeed, the equilibrium π is deterministic, while η_t^ξ has a non-degenerate random distribution at any time $t > 0$, provided that $\xi \neq \pi$.

Within this setting, a worst-case mixing analysis for $\text{Avg}(\mathbb{T}_N^d)$ has been carried out in [QS23]. There, the authors show that mixing occurs gradually at times $t = \Theta(N^2)$. This result was actually derived for a larger class of graph sequences, referred to as “finite-dimensional” [QS23, Assumption 1], establishing in all these examples that mixing of the averaging process was dictated by that of the corresponding simple random walk on the same graph. Other settings for which sharp asymptotics for the L^p -mixing times, with $p = 1, 2$, have been established are the hypercube [CQS23], and the complete and complete bipartite graphs [CDSZ22, CQS23]. In these last three examples, cutoff occurs. For a quick overview on these results and on some general tools recently developed for the study of mixing times of the averaging process, we refer the interested reader to [CQS23, Section 2].

1.3 Averaging process as a noisy heat flow

The averaging process may also be viewed as the probability distribution of a random walk $(U_t)_{t \geq 0}$ in a highly degenerate dynamic random environment, in which the time-dependent conductances between nearest neighbors are either zero, or become “instantaneously infinite” as soon as the Poisson clock associated to the pair rings. At the occurrence of such an event, the walk originally sitting on one of the vertices will find itself with equal probability on either one of the two nearest neighbors.

With this interpretation, while a realization of $\eta_t^\xi \in \mathcal{P}(\mathbb{T}_N^d)$ coincides with the distribution of U_t for a quenched realization of the random environment, the mean $\mathbb{E}[\eta_t^\xi] \in \mathcal{P}(\mathbb{T}_N^d)$ shall be considered as the corresponding annealed law. As already observed in [AL12, Lemma 1] and exploited in [QS23, CQS23] (see also [Tra23, Section 6.3] for a recent use of this fact), $t \mapsto \mathbb{E}[\eta_t^\xi]$ describes the heat flow of RW(\mathbb{T}_N^d), the “lazy” continuous-time simple random walk $(X_t)_{t \geq 0}$ on \mathbb{T}_N^d , with $X_0 \sim \xi$ and infinitesimal generator $L^{\text{RW}} = L_N^{\text{RW}}$ given, for all $\psi \in \mathbb{R}^{\mathbb{T}_N^d}$, by

$$L^{\text{RW}}\psi(x) = \frac{1}{2} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} (\psi(y) - \psi(x)), \quad x \in \mathbb{T}_N^d. \tag{1.3}$$

Hence, letting $\mathbf{P}_\xi^{\text{RW}}$ denote the law associated to this walk and π_t^ξ its distribution at time $t \geq 0$, we obtain

$$\mathbb{E}[\eta_t^\xi(x)] = \pi_t^\xi(x) := \mathbf{P}_\xi^{\text{RW}}(X_t = x), \quad x \in \mathbb{T}_N^d, t \geq 0, \xi \in \mathcal{P}(\mathbb{T}_N^d). \tag{1.4}$$

In other words, (1.4) states that the “noisy” jump process η_t^ξ has the heat flow π_t^ξ as its mean. Although the heat flow does not necessarily identify the correct mixing behavior of the averaging process on *all* graphs, we will prove that a strong form of concentration of η_t^ξ around π_t^ξ holds on \mathbb{T}_N^d .

1.4 Summary of main results

Let us provide a preliminary informal description of the two main features of Avg that we investigate in this article, together with some of their applications. In what follows, the standard asymptotic notation “ $O(\cdot)$ ”, “ $o(\cdot)$ ”, “ $\Theta(\cdot)$ ”, “ \ll ”, etc., refers to the limit $N \rightarrow \infty$.

In what follows, we focus on worst-case initial conditions $\xi \in \mathcal{P}(\mathbb{T}_N^d)$, which, by simple convexity arguments (see, e.g., [CQS23, Section 2.3]), are all Dirac measures: for all $t \geq 0$,

$$\sup_{\xi \in \mathcal{P}(\mathbb{T}_N^d)} \mathbb{E} \left[\left\| \frac{\eta_t^\xi}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}} = \sup_{x \in \mathbb{T}_N^d} \mathbb{E} \left[\left\| \frac{\eta_t(x, \cdot)}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}}.$$

(The analogous identity for RW(\mathbb{T}_N^d) is well-known.) Here and all throughout, we write $\eta_t(x, \cdot) = \eta_t^\xi$ and, similarly, $\pi_t(x, \cdot) = \pi_t^\xi$, whenever $\xi = \mathbf{1}_x$, for some $x \in \mathbb{T}_N^d$. Finally, in the translation-invariant context of the torus \mathbb{T}_N^d , $\mathbb{E}[\|\frac{\eta_t(x, \cdot)}{\pi} - 1\|_p^p]$ and $\|\frac{\pi_t(x, \cdot)}{\pi} - 1\|_p^p$ do not depend on $x \in \mathbb{T}_N^d$; hence, for notational convenience, we consider $\eta_t(0, \cdot)$ and $\pi_t(0, \cdot)$, i.e., we fix the location of the initial mass as $x = 0 \in \mathbb{T}_N^d$ all throughout.

1.4.1 Early concentration phenomenon

The idea adopted in [QS23] in order to identify the correct timescale at which the averaging process mixes is, in a nutshell, the following. First, identify the timescale at which RW mixes; there, compare the distance-to-equilibrium of Avg with the one of RW, and prove that the resulting error becomes arbitrarily small on that same timescale.

More in detail, this strategy starts from two estimates. The first one — the lower bound in [QS23, Lemma 5.1], which is an immediate consequence of Jensen inequality and (1.4) — states that mixing for $\eta_t(0, \cdot)$ does not occur before that of $\pi_t(0, \cdot)$: for all $p \in [1, 2]$ and $t \geq 0$,

$$\left\| \frac{\pi_t(0, \cdot)}{\pi} - 1 \right\|_p \leq \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}}. \tag{1.5}$$

The second one — a slight sharpening of the upper bound in [QS23, §5.2] — follows by the triangle inequality for the L^p -Wasserstein metrics in (1.2) and Jensen inequality: for all $p \in [1, 2]$ and $t \geq 0$,

$$\mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}} \leq \left\| \frac{\pi_t(0, \cdot)}{\pi} - 1 \right\|_p + \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi} \right\|_2^2 \right]^{\frac{1}{2}}. \tag{1.6}$$

In view of these two bounds, if one expects the mixing of $\eta_t(0, \cdot)$ to be dictated by that of $\pi_t(0, \cdot)$, the key to get matching lower and upper bounds in (1.5)–(1.6) consists in efficiently estimating the second term on the right-hand side of (1.6).

By implementing this strategy for the specific example of \mathbb{T}_N^d , since $\pi_t(0, \cdot)$ mixes gradually at times $t = \Theta(N^2)$, showing

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left\| \frac{\eta_{t_N}(0, \cdot)}{\pi} - \frac{\pi_{t_N}(0, \cdot)}{\pi} \right\|_2^2 \right] = 0, \quad \text{for all } t_N \gg N^2, \tag{1.7}$$

readily implies that also $\eta_t(0, \cdot)$ mixes at times $t = \Theta(N^2)$. Proving (1.7) is the main step in the proofs of [QS23, Proposition 2.8 & Theorem 2.9].

Our first main goal is the following improvement of (1.7):

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left\| \frac{\eta_{t_N}(0, \cdot)}{\pi} - \frac{\pi_{t_N}(0, \cdot)}{\pi} \right\|_2^2 \right] = 0, \quad \text{for all } t_N \gg N^{\frac{2d}{d+2}}. \tag{1.8}$$

Since $N^{2d/(d+2)} \ll N^2$ for any $d \geq 1$, (1.8) captures a new timescale — dimension-dependent, but always shorter than the diffusive one — after which $\eta_t(0, \cdot)$ and its expectation $\pi_t(0, \cdot)$ stay close to each other, although both of them are still far from stationarity. For this reason, we refer to such a property of the averaging process as an *early concentration phenomenon*. For the precise quantitative version of this result, see Theorem 2.4 below.

In view of the bounds (1.5) and (1.6) — which hold true on any graph — early concentration of Avg may be considered as a stronger form of mixing, whenever Avg and RW mix on the same timescale. On the one hand, this suggests that, just like the mixing behaviors of Avg and RW are comparable on a large class of geometries (see, e.g., [QS23, Proposition 2.8] and [CQS23, Theorem 1.1]), early concentration should not be an exclusive instance of the example of $\text{Avg}(\mathbb{T}_N^d)$ considered here, and should be investigated in settings in which scaling limits of discrete gradients of the underlying heat flow are available. On the other hand, this automatically excludes early concentration from settings like the complete and complete bipartite graphs [CDSZ22, CQS23], in which mixing of Avg occurs on a strictly longer timescale than that of RW. For what concerns L^2 -mixing, this is further explained by the following identity, which resembles the inequality in (1.6) with $p = 2$: for all $t \geq 0$,

$$\mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - 1 \right\|_2^2 \right] = \left\| \frac{\pi_t(0, \cdot)}{\pi} - 1 \right\|_2^2 + \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi} \right\|_2^2 \right]. \tag{1.9}$$

The proof of (1.9) combines the definition of L^2 -norm in (1.2) and the identity in (1.4).

A quantitative version of (1.8) bears several useful consequences related to both mixing and scaling limits of $\text{Avg}(\mathbb{T}_N^d)$. For instance, as a refinement of mixing, our estimate combined with a quantitative local CLT for $\text{RW}(\mathbb{T}_N^d)$ identifies the *limit profile* for $\text{Avg}(\mathbb{T}_N^d)$: for all $p \in [0, 1]$ and $t > 0$,

$$\mathbb{E} \left[\left\| \frac{\eta_{tN^2}(0, \cdot)}{\pi} - 1 \right\|_p^p \right]^{\frac{1}{p}} = \|h_t(0, \cdot) - 1\|_{L^p(\mathbb{T}^d)} + O\left(\frac{1}{N}\right),$$

where $h_t(0, \cdot)$ is the heat kernel of the diffusion on \mathbb{T}^d with generator $\frac{1}{2}\Delta$. We detail this result and its simple proof in Proposition 2.5 below. For related recent results about limit profiles also in absence of cutoff, see, e.g., [BDCJ22].

Along the same lines, (1.8) with $t_N = tN^2$ may be also interpreted as a *quantitative hydrodynamic limit* for $\text{Avg}(\mathbb{T}_N^d)$ (Corollary 2.6). Here, the heat equation $\partial_t \rho = \frac{1}{2}\Delta \rho$ on \mathbb{T}^d arises as a hydrodynamic equation, and convergence is established in the stronger L^p -Wasserstein metrics rather than in probability, as most typically done (see, e.g., the classical monograph [KL99]).

1.4.2 Fast local smoothness

Mixing of the averaging process typically deals with measuring how far the averaging process $\eta_t(0, \cdot)$ is from its global equilibrium π . Next to this measure of distance-to-equilibrium, Aldous and Lanoue [AL12, Section 2.3] proposed a natural notion of *local smoothness*, aiming at quantifying more accurately the local flatness of $\eta_t(0, \cdot)$. As intuitively clear, local smoothness should occur at least as fast as global mixing and, as we will show in Theorem 2.4, is intimately related in our context to the phenomenon of early concentration discussed in the previous paragraph. Nevertheless, analyzing local smoothness in general settings turns out to be a delicate issue, and, compared to the range of tools recently developed for the study of global mixing for Avg, techniques and results for local mixing are rather limited (see, e.g., [AL12, Proposition 4] and [BBG20, Corollary 1]).

Following [AL12] and the analogy with the corresponding heat flow, we employ the (mean) Dirichlet form to quantify local roughness of the averaging process. Indeed, while L^p -distances are natural quantities to measure the distance to equilibrium of $\pi_t(0, \cdot)$, the Dirichlet form associated to $\text{RW}(\mathbb{T}_N^d)$ with generator L^{RW} (1.3), given by

$$\mathcal{E}(\psi) := \frac{1}{4N^d} \sum_{x \in \mathbb{T}_N^d} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} (\psi(x) - \psi(y))^2, \quad \psi \in \mathbb{R}^{\mathbb{T}_N^d},$$

relates to an infinitesimal, thus, local, stage of the convergence of $\pi_t(0, \cdot)$, as encoded in

$$\frac{d}{dt} \left\| \frac{\pi_t(0, \cdot)}{\pi} - 1 \right\|_2^2 = -2 \mathcal{E} \left(\frac{\pi_t(0, \cdot)}{\pi} \right), \quad t \geq 0. \tag{1.10}$$

In [AL12, Proposition 2], the authors prove an analogous identity for $\eta_t(0, \cdot)$:

$$\frac{d}{dt} \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - 1 \right\|_2^2 \right] = -\mathbb{E} \left[\mathcal{E} \left(\frac{\eta_t(0, \cdot)}{\pi} \right) \right], \quad t \geq 0. \tag{1.11}$$

These two relations are, indeed, similar, although differing by a factor 2; this dissimilarity is reflected by different exponential decay rates for the (mean) L^2 -distances. Nevertheless, both (1.10) and (1.11) show that quantifying the decay rates of the terms on the right-hand side implies refined mixing results for $\pi_t(0, \cdot)$ and $\eta_t(0, \cdot)$, respectively.

A first simple estimate for the (mean) Dirichlet form along the averaging process may be obtained via a convexity argument as done to derive (1.5). Indeed, Jensen inequality and the identity (1.4) yield the following lower bound:

$$\mathcal{E}\left(\frac{\pi_t(0, \cdot)}{\pi}\right) \leq \mathbb{E}\left[\mathcal{E}\left(\frac{\eta_t(0, \cdot)}{\pi}\right)\right], \quad t \geq 0. \tag{1.12}$$

Our main contribution is to show that, for $\text{Avg}(\mathbb{T}_N^d)$, the inequality in (1.12) can actually be reversed as follows

$$\mathbb{E}\left[\mathcal{E}\left(\frac{\eta_t(0, \cdot)}{\pi}\right)\right] \leq C \phi(t/N^d) \mathcal{E}\left(\frac{\pi_t(0, \cdot)}{\pi}\right), \quad t \geq 0, \tag{1.13}$$

at the negligible cost of including some dimension-dependent constant $C = C(d) > 0$ and a term $\phi(t/N^d)$ which diverges only for very large times $t \gg N^d$. Ultimately, since the Dirichlet form along the heat flow $\pi_t(0, \cdot)$ may be efficiently estimated on \mathbb{T}_N^d , this yields a quantitative control for the local smoothness of $\text{Avg}(\mathbb{T}_N^d)$ for all times $t \geq 0$. This allows, in particular, to capture both the fast local smoothing (on the same timescale at which early concentration in (1.8) occurs) of $\text{Avg}(\mathbb{T}_N^d)$, as well as the correct exponential contraction rate for large times. As an immediate consequence of this, we obtain, by integrating over time this bound, the corresponding control for the global distance-to-equilibrium (Corollary 2.3). The latter result and (1.13) are both employed in Section 7.

The bound in (1.13) and an improved version of (1.12) are the content of Theorem 2.1, and together represent the first sharp results concerning local smoothness for the averaging process. It is worth to emphasize that an analogous result, although for a different model, has been recently proven by Banerjee and Burdzy in [BB21, Theorem 2.6]. There, the authors study (among other things) the *smoothing process* on \mathbb{T}_N^d , a model discussed by Liggett within the class of linear systems [Lig05, Chapter IX]. Like the averaging process, also the smoothing process falls into the class of Markovian mass redistribution models. However, since averages are performed over one vertex at the time, the smoothing dynamics does not conserve mass. This seemingly marginal variation leads to some qualitative differences between the two processes. For instance, equilibrium for the smoothing process is a truly random mass profile, while this is not the case for the averaging. Moreover, as proved in [BB21, Proposition 2.5], averaging over a single vertex deterministically does not increase the value of $\mathcal{E}(\frac{\eta_t}{\pi})$; as simple examples show, this monotonicity property does not hold when performing our edge-averaging dynamics. Nevertheless, despite these dissimilarities, we import some of their arguments into our analysis, yielding comparable local smoothness behaviors of the two processes when considered on \mathbb{T}_N^d .

As already anticipated, fast local smoothness comes with some relevant applications. One of them, and probably the most important, is the early concentration phenomenon (Theorem 2.4). As a second application, we show how to transfer the integral information encoded in the (mean) Dirichlet form into *pointwise gradient estimates* for $\text{Avg}(\mathbb{T}_N^d)$. This result is presented in Proposition 2.2, and may be viewed as an analogue of the annealed gradient estimates for random walks in dynamic random environment recently obtained in [DKS23].

2 Main results

We are now ready to present our main results and some of their consequences in full detail. Before that, recalling the definition of the random walk $\text{RW}(\mathbb{T}_N^d)$ (Section 1.3), we define $\lambda = \lambda^N > 0$ as its *spectral gap*, namely the smallest non-zero eigenvalue of

the (negative) generator $-L^{\text{RW}}$ given in (1.3). It is well-known that $\lambda = 1 - \cos(2\pi/N) = \frac{2\pi^2}{N^2} (1 + O(\frac{1}{N^2}))$. Finally, we write $t_{\text{rel}} = t_{\text{rel}}^N = 1/\lambda = \frac{N^2}{2\pi^2} (1 + O(\frac{1}{N^2}))$ for the corresponding relaxation time.

In what follows, $B, C, C_1, C_2, \dots, C'_1, C'_2, \dots > 0$ denote constants whose exact value is unimportant and may change from line to line. Moreover, unless stated otherwise, such constants may depend on $d \geq 1$, but not on other variables, e.g., $N \in \mathbb{N}$ and $t \geq 0$.

2.1 Local smoothness and gradients estimates

The following is the first of our main results.

Theorem 2.1 (Local smoothness). *For all $d \geq 1$, for all $N \in \mathbb{N}$ large enough, and for all $t \geq 0$, we have*

$$(1 + C_1(t \wedge 1)) \mathcal{E}\left(\frac{\pi_t(0, \cdot)}{\pi}\right) \leq \mathbb{E}\left[\mathcal{E}\left(\frac{\eta_t(0, \cdot)}{\pi}\right)\right] \leq C_2 \exp\left(\frac{Bt}{N^{d+2}}\right) \mathcal{E}\left(\frac{\pi_t(0, \cdot)}{\pi}\right), \quad (2.1)$$

for some constants $B, C_1, C_2 > 0$ (depending only on $d \geq 1$).

Although we formulated the above result as a comparison between the (mean) Dirichlet form of $\text{Avg}(\mathbb{T}_N^d)$ and that of $\text{RW}(\mathbb{T}_N^d)$, the latter quantity is rather explicit thanks to the full knowledge of eigenvalues and eigenfunctions for $\text{RW}(\mathbb{T}_N^d)$ (see also Section 4.1 below for more details). Indeed, one obtains, for all $N \in \mathbb{N}$ large enough and for all $t \geq 0$,

$$C'_1 \Xi(t) \leq \mathcal{E}\left(\frac{\pi_t(0, \cdot)}{\pi}\right) \leq C'_2 \Xi(t), \quad \text{with } \Xi(t) := \frac{N^d \exp(-2t/t_{\text{rel}})}{(N^{d+2} \wedge t^{d/2+1}) \vee 1}, \quad (2.2)$$

for some constants $C'_1, C'_2 > 0$ (depending only on $d \geq 1$).

Plugging the bounds in (2.2) into (2.1) allows to quantify the fast local smoothness of $\text{Avg}(\mathbb{T}_N^d)$. Not only there follows that

$$\mathbb{E}\left[\mathcal{E}\left(\frac{\eta_{t_N}(0, \cdot)}{\pi}\right)\right] = \Theta\left(\frac{1}{N^2}\right), \quad \text{for all } t_N = \Theta(N^2),$$

but we also extract quantitative information for $t \ll N^2$ and $t \gg N^2$. In particular, as long as $t = O(N^2)$, the resulting upper bound states that $\eta_t(0, \cdot)$ gets locally smoother at the polynomial scale $t^{-(d/2+1)}$ (just like $\pi_t(0, \cdot)$). For larger times, instead, local smoothness becomes exponential, with rate $2/t_{\text{rel}} (1 + o(1))$, thus, crucially recovering the factor 2 in the exponential decay rate (cf. (1.10) and (1.11)). Let us further remark that, since $t_{\text{rel}} = \frac{N^2}{2\pi^2} (1 + o(1)) = o(N^{d+2})$, the exponential factor in the upper bound of (2.1) plays no significant role as long as $t = O(N^{d+2})$, a timescale much longer than the diffusive one. As for the lower bound in (2.1), the factor $(1 + C_1(t \wedge 1))$ is a seemingly irrelevant improvement upon (1.12). However, this term is strictly larger than one as soon as $t > 0$, suggesting that, even after a very small time, the fluctuations of the random gradients of $\eta_t(0, \cdot)$ become comparable with their mean.

Local smoothness is an integral (mean) quantity regarding the gradients of the averaging process. As a consequence of the estimate in Theorem 2.1, we derive pointwise (rather than integral) estimates for such gradients. We remark that, keeping in mind the interpretation from Section 1.3 of $\eta_t(x, y)$ as the distribution of a random walk in a dynamic random environment, such bounds may be thought of as the analogues (for $p = 2$) of the annealed gradient estimates recently obtained in [DKS23, Theorem 1.6(i)] for the time-dependent random conductance model with space-time ergodic and uniformly elliptic environment on \mathbb{Z}^d . We emphasize that, although our result resembles that in [DKS23], the two proofs are quite different. We postpone the proof of the following result to Section 6.

Proposition 2.2 (Pointwise gradient estimates). *For all $d \geq 1$, for all $N \in \mathbb{N}$ large enough, for all $t \geq 0$, and for all $x, y \in \mathbb{T}_N^d$ with $|x - y| = 1$, we have*

$$\mathbb{E}[(\eta_t(0, x) - \eta_t(0, y))^2] \leq \frac{C}{(N^{2d+2} \wedge t^{d+1}) \vee 1} \exp\left(-2t/t_{\text{rel}} + \frac{Bt}{N^{d+2}}\right),$$

for some constants $B, C > 0$ (depending only on $d \geq 1$).

Let us conclude this discussion on local smoothness with an important result on mixing of $\text{Avg}(\mathbb{T}_N^d)$, which will turn out useful in Section 7, and which is immediately derived by integrating the infinitesimal relation (1.11) over the time interval $[t, +\infty)$, combined with the upper bounds in (2.1) and (2.2).

Corollary 2.3. *For all $d \geq 1$, for all $N \in \mathbb{N}$ large enough, and for all $t \geq 0$, we have*

$$\mathbb{E}\left[\left\|\frac{\eta_t(0, \cdot)}{\pi} - 1\right\|_2^2\right] \leq \frac{CN^d}{(N^d \wedge t^{d/2}) \vee 1} \exp\left(-2t/t_{\text{rel}} + \frac{Bt}{N^{d+2}}\right),$$

for some constants $B, C > 0$ (depending only on $d \geq 1$).

2.2 Concentration and limit profile

We now present our second main result, whose proof is based on the following identity (which we prove in Lemma 5.1)

$$\mathbb{E}\left[\left\|\frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi}\right\|_2^2\right] = \frac{1}{dN^d} \int_0^t \mathbb{E}\left[\mathcal{E}\left(\frac{\eta_{t-s}(0, \cdot)}{\pi}\right)\right] \mathcal{E}\left(\frac{\pi_s(0, \cdot)}{\pi}\right) ds, \quad t \geq 0,$$

and the findings in Theorem 2.1.

Theorem 2.4 (Concentration). *Recall $\Xi(t)$ from (2.2). For all $d \geq 1$, for all $N \in \mathbb{N}$ large enough, and for all $t \geq 0$, we have*

$$C_1(1 \wedge t)\Xi(t) \leq \mathbb{E}\left[\left\|\frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi}\right\|_2^2\right] \leq C_2(1 \wedge t)\Xi(t) \exp\left(\frac{Bt}{N^{d+2}}\right), \quad (2.3)$$

for some constants $B, C_1, C_2 > 0$ (depending only on $d \geq 1$).

The proof of the following consequence of Theorem 2.4 may be found in Section 6.

Proposition 2.5 (Limit profile). *Fix $d \geq 1$, and let $h_t(0, u)$, $u \in \mathbb{T}^d$ and $t \geq 0$, be the heat kernel of the diffusion on \mathbb{T}^d with generator $\frac{1}{2}\Delta$, started from the origin $0 \in \mathbb{T}^d$.*

Then, for all bounded intervals $[a, b] \subset (0, +\infty)$, there exists $C = C(a, b, d) > 0$ satisfying, for all $p \in [1, 2]$ and for all $N \in \mathbb{N}$ large enough,

$$\sup_{t \in [a, b]} \left| \mathbb{E}\left[\left\|\frac{\eta_{tN^2}(0, \cdot)}{\pi} - 1\right\|_p^p\right]^{\frac{1}{p}} - \|h_t(0, \cdot) - 1\|_{L^p(\mathbb{T}^d)} \right| \leq \frac{C}{N}.$$

As already discussed in Section 1.4.1, specializing the upper bound in (2.3) to times $t = \Theta(N^2)$ yields a quantitative hydrodynamic limit for $\text{Avg}(\mathbb{T}_N^d)$. We report the precise statement below. Its simple proof is analogous to that of Proposition 2.5 and, thus, is left to the reader.

Corollary 2.6 (Quantitative hydrodynamic limit). *Fix $d \geq 1$ and, for all $g \in \mathcal{C}(\mathbb{T}^d)$, let $(t, u) \in [0, \infty) \times \mathbb{T}^d \mapsto h_t^g(u)$ be the unique solution to the heat equation $\partial_t \rho = \frac{1}{2}\Delta \rho$ on \mathbb{T}^d with initial condition $\rho_0 = g$.*

Then, for all bounded intervals $[a, b] \subset (0, +\infty)$, there exists $C = C(a, b, d) > 0$ satisfying, for all $N \in \mathbb{N}$ large enough, for all $\xi \in \mathcal{P}(\mathbb{T}_N^d)$, and $\Psi \in \mathcal{C}(\mathbb{T}_N^d)$,

$$\sup_{t \in [a, b]} \mathbb{E}\left[\left|\sum_{x \in \mathbb{T}_N^d} \eta_{tN^2}^\xi(x) \Psi\left(\frac{x}{N}\right) - \int_{\mathbb{T}^d} h_t^g(u) \Psi(u) du\right|\right] \leq \|\Psi\|_\infty \left(\frac{C}{N} + \left\|\frac{\xi}{\pi} - \frac{g(\cdot)}{\pi}\right\|_1\right).$$

2.3 Structure of the paper

The rest of the paper is organized as follows. In Section 3, we discuss some preliminary facts which allow us to conveniently reformulate our problem. In Sections 4 and 5 we present the proofs of our main results, Theorems 2.1 and 2.4, respectively. The proofs of Propositions 2.2 and 2.5 are the content of Section 6. As a further application of our main results, in Section 7, we prove cutoff (in total variation distance) for an interacting particle system dual to the averaging process, known as *binomial splitting process*, strengthening a result in [QS23].

3 Preliminaries

In this section, we introduce two auxiliary Markov processes, whose properties will play a key role in Sections 4 and 5 below, dedicated to the proofs of our main results.

3.1 Coupled random walks

As discussed in (1.4), first-order moments of the averaging process may be expressed in terms of the random walk X_t . This property is a particular instance of *duality* (see, e.g., [Lig05]). This connection was already noted in [AL12], further exploited and generalized in [QS23, CQS23], and holds more generally for all k^{th} -order moments. In particular, in order to prove our main results, we will need to introduce a dual system of $k = 2$ interacting walks, which we now describe in detail.

Place two particles on the vertices of \mathbb{T}_N^d , and endow each unordered pair of nearest neighbors of \mathbb{T}_N^d with exponential clocks of unit rate, independent over the pairs. Assume that the clock associated to the pair $\{x, y\}$, $|x - y| = 1$, rings. Particles sitting at that time on either x or y decide, independently from each other and with probability $1/2$, to change vertex, i.e., moving from x to y if originally at x , or vice versa. Let $((X_t, Y_t))_{t \geq 0} \subset \mathbb{T}_N^d \times \mathbb{T}_N^d$ denote the Markov process of the positions of these two particles. Both marginals X_t and Y_t evolve as two copies of $\text{RW}(\mathbb{T}_N^d)$. Moreover, they move independently as long as they are at graph distance larger than one. However, when sitting at distance smaller than two, the two walks may interact by experiencing synchronous jumps. For this reason, we refer to (X_t, Y_t) as a system of two *coupled random walks*, shortly $\text{CRW}(\mathbb{T}_N^d)$. We let $\mathbf{P}_\nu^{\text{CRW}}$ denote its law when $(X_0, Y_0) \sim \nu$.

The importance of $\text{CRW}(\mathbb{T}_N^d)$ lies in the following second-order duality relation, analogous to that in (1.4):

$$\mathbb{E}[\eta_t^\xi(x)\eta_t^\xi(y)] = \mathbf{P}_{\xi \otimes \xi}^{\text{CRW}}(X_t = x, Y_t = y), \quad t \geq 0, \quad x, y \in \mathbb{T}_N^d, \quad \xi \in \mathcal{P}(\mathbb{T}_N^d). \quad (3.1)$$

In analogy with the notation used for $\text{RW}(\mathbb{T}_N^d)$, when $\xi \otimes \xi' = \mathbb{1}_x \otimes \mathbb{1}_y$ for some $x, y \in \mathbb{T}_N^d$, we simply write $\mathbf{P}_{x,y}^{\text{CRW}} = \mathbf{P}_{\xi \otimes \xi'}^{\text{CRW}}$. The introduction of $\text{CRW}(\mathbb{T}_N^d)$ and the duality relation with $\text{Avg}(\mathbb{T}_N^d)$ allow us to efficiently rewrite the mean Dirichlet form of $\text{Avg}(\mathbb{T}_N^d)$.

Proposition 3.1. *For all $d \geq 1$, $N \in \mathbb{N}$, $t \geq 0$ and $e \in \mathbb{T}_N^d$ with $|e| = 1$, we have*

$$\frac{1}{dN^d} \mathbb{E} \left[\mathcal{E} \left(\frac{\eta_t(0, \cdot)}{\pi} \right) \right] = \mathbf{P}_{0,0}^{\text{CRW}}(X_t - Y_t = 0) - \mathbf{P}_{0,0}^{\text{CRW}}(X_t - Y_t = e). \quad (3.2)$$

Proof. By definition of $\mathcal{E}(\cdot)$ and the duality relation (3.1), we obtain

$$\begin{aligned} \mathbb{E} \left[\mathcal{E} \left(\frac{\eta_t(0, \cdot)}{\pi} \right) \right] &= \frac{N^d}{4} \sum_{\substack{x, y \in \mathbb{T}_N^d \\ |x-y|=1}} \mathbb{E} [(\eta_t(0, x) - \eta_t(0, y))^2] \\ &= \frac{N^d}{2} \sum_{x \in \mathbb{T}_N^d} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} \{ \mathbb{E}[\eta_t(0, x)^2] - \mathbb{E}[\eta_t(0, x)\eta_t(0, y)] \} \end{aligned}$$

$$= \frac{N^d}{2} \sum_{x \in \mathbb{T}_N^d} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} \{ \mathbf{P}_{0,0}^{\text{CRW}}(X_t = x, Y_t = x) - \mathbf{P}_{0,0}^{\text{CRW}}(X_t = x, Y_t = y) \} .$$

We get the desired conclusion by rearranging the terms in the last summation as follows:

$$\begin{aligned} \sum_{x \in \mathbb{T}_N^d} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} \mathbf{P}_{0,0}^{\text{CRW}}(X_t = x, Y_t = y) &= 2d \mathbf{P}_{0,0}^{\text{CRW}}(X_t = Y_t) , \\ \sum_{x \in \mathbb{T}_N^d} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} \mathbf{P}_{0,0}^{\text{CRW}}(X_t = x, Y_t = y) &= \sum_{\substack{e' \in \mathbb{T}_N^d \\ |e'|=1}} \mathbf{P}_{0,0}^{\text{CRW}}(Y_t = X_t + e') \\ &= 2d \mathbf{P}_{0,0}^{\text{CRW}}(X_t - Y_t = e) , \end{aligned}$$

where for the last identity we used the translation and permutation invariance of the marginals laws of (X_t, Y_t) , as well as $(X_t, Y_t) \stackrel{\text{Law}}{=} (Y_t, X_t)$ under $\mathbf{P}_{0,0}^{\text{CRW}}$. \square

3.2 Difference process

Next to $\text{CRW}(\mathbb{T}_N^d)$, motivated by the result in Proposition 3.1, we consider a second auxiliary process. Given $((X_t, Y_t))_{t \geq 0}$ evolving as $\text{CRW}(\mathbb{T}_N^d)$, define $(Z_t)_{t \geq 0}$ as

$$Z_t := X_t - Y_t , \quad t \geq 0 .$$

We refer to Z_t as the *difference process* associated to (X_t, Y_t) . In our context, Z_t turns out to be a Markov process. Indeed, if we were considering the difference process $Z_t^0 := X_t^0 - Y_t^0$ associated to (X_t^0, Y_t^0) , two *independent* copies of $\text{RW}(\mathbb{T}_N^d)$, then it is well-known that Z_t^0 is the simple random walk on \mathbb{T}_N^d with infinitesimal generator $A_0 = 2L^{\text{RW}}$, i.e., Z_t^0 jumps like $\text{RW}(\mathbb{T}_N^d)$, but at a double rate. As for Z_t , due to the interaction between X_t and Y_t when $|X_t - Y_t| \leq 1$, Z_t moves like Z_t^0 with a defect represented by slow bonds attached to the origin. More in detail, the infinitesimal generator A of Z_t reads as

$$A = A_0 + R ,$$

where we recall that $A_0 = 2L^{\text{RW}}$ is the generator of Z_t^0 , while, for all $\psi : \mathbb{T}_N^d \rightarrow \mathbb{R}$,

$$R\psi(z) := \begin{cases} -\frac{1}{2} A_0 \psi(0) & \text{if } z = 0 \\ -\frac{1}{2} (\psi(0) - \psi(z)) + \frac{1}{4} (\psi(-z) - \psi(z)) & \text{if } |z| = 1 \\ 0 & \text{else .} \end{cases}$$

Moreover, for all $t \geq 0$, let $S_t = e^{tA}$ (resp. $S_t^0 = e^{tA_0}$) denote the transition kernels of the random walk Z_t (resp. Z_t^0). Observe that, since both A_0 and R are symmetric kernels, A and S_t are symmetric, too. Symmetry of S_t and Proposition 3.1 imply the following result.

Proposition 3.2. *For all $t \geq 0$, we have*

$$\frac{1}{dN^d} \mathbb{E} \left[\mathcal{E} \left(\frac{\eta_t(0, \cdot)}{\pi} \right) \right] = S_t(0, 0) - S_t(e, 0) . \tag{3.3}$$

We conclude this section by observing that similar computations to those employed to derive (3.2) readily yield the following analogue of (3.3): for all $t \geq 0$,

$$\frac{1}{dN^d} \mathcal{E} \left(\frac{\pi_t(0, \cdot)}{\pi} \right) = S_t^0(0, 0) - S_t^0(e, 0) . \tag{3.4}$$

Hence, by (3.3), (3.4), and Jensen inequality, we obtain, for all $t \geq 0$,

$$\begin{aligned} S_t(0, 0) - S_t(e, 0) &= \frac{1}{dN^d} \mathbb{E} \left[\mathcal{E} \left(\frac{\eta_t(0, \cdot)}{\pi} \right) \right] \\ &\geq \frac{1}{dN^d} \mathcal{E} \left(\frac{\pi_t(0, \cdot)}{\pi} \right) = S_t^0(0, 0) - S_t^0(e, 0). \end{aligned} \tag{3.5}$$

4 Proof of Theorem 2.1

This section is devoted to the proof of Theorem 2.1, which we split into three parts. First, we exploit the representations in (3.3) and (3.4) to express the mean Dirichlet form of $\text{Avg}(\mathbb{T}_N^d)$ as an infinite series (Section 4.1). Then, for this infinite series, we provide lower and upper bounds in Sections 4.2 and 4.3, respectively. We gather all these facts together in Section 4.4.

4.1 A renewal-type equation

Our main goal in this section is to find a closed-form expression for the remainder in the inequality (3.5). For this purpose, we employ the identities in (3.3) and (3.4) involving Dirichlet forms and the transition kernels S_t and S_t^0 , respectively, both introduced in Section 3.2. Let us further recall the definition of the associated generators $A = A_0 + R$ and A_0 from Section 3.2.

We start by applying the integration by parts formula: for all $x \in \mathbb{T}_N^d$ and $t \geq 0$,

$$\begin{aligned} S_t(x, 0) - S_t^0(x, 0) &= - \int_0^t \sum_{y \in \mathbb{T}_N^d} S_{t-s}^0(x, y) (A_0 - A) S_s \mathbf{1}_0(y) \, ds \\ &= \int_0^t S_{t-s}^0(x, 0) R S_s \mathbf{1}_0(0) \, ds + \int_0^t \sum_{\substack{y \in \mathbb{T}_N^d \\ |y|=1}} S_{t-s}^0(x, y) R S_s \mathbf{1}_0(y) \, ds \\ &= -\frac{1}{2} \int_0^t S_{t-s}^0(x, 0) \sum_{\substack{y \in \mathbb{T}_N^d \\ |y|=1}} (S_s(y, 0) - S_s(0, 0)) \, ds \\ &\quad - \frac{1}{2} \int_0^t \sum_{\substack{y \in \mathbb{T}_N^d \\ |y|=1}} S_{t-s}^0(x, y) (S_s(0, 0) - S_s(y, 0)) \, ds \\ &= \frac{1}{2} \int_0^t \sum_{\substack{y \in \mathbb{T}_N^d \\ |y|=1}} (S_s(0, 0) - S_s(y, 0)) (S_{t-s}^0(x, 0) - S_{t-s}^0(x, y)) \, ds. \end{aligned}$$

Noting that $S_s(e, 0) = S_s(e', 0)$ for all $e, e' \in \mathbb{T}_N^d$ with $|e| = |e'| = 1$, we further obtain

$$S_t(x, 0) - S_t^0(x, 0) = \int_0^t (S_s(0, 0) - S_s(e, 0)) \sum_{\substack{y \in \mathbb{T}_N^d \\ |y|=1}} \frac{1}{2} (S_{t-s}^0(x, 0) - S_{t-s}^0(x, y)) \, ds.$$

By evaluating the above expression at $x = 0$ and $x = e \in \mathbb{T}_N^d$, and subtracting, we get

$$\begin{aligned} S_t(0, 0) - S_t(e, 0) &= S_t^0(0, 0) - S_t^0(e, 0) \\ &+ \int_0^t (S_s(0, 0) - S_s(e, 0)) \sum_{\substack{y \in \mathbb{T}_N^d \\ |y|=1}} \frac{1}{2} (S_{t-s}^0(0, 0) - S_{t-s}^0(0, y) - S_{t-s}^0(e, 0) + S_{t-s}^0(e, y)) \, ds. \end{aligned} \tag{4.1}$$

By exploiting the symmetry and translation invariance of the random walk Z_t^0 , we may simplify the summation above as follows: for $e, e' \in \mathbb{T}_N^d$ with $|e| = |e'| = 1$ and $e' \neq \pm e$,

$$\begin{aligned} & \sum_{\substack{y \in \mathbb{T}_N^d \\ |y|=1}} \frac{1}{2} (S_t^0(0, 0) - S_t^0(0, y) - S_t^0(e, 0) + S_t^0(e, y)) \\ &= (d + 1/2) S_t^0(0, 0) - 2d S_t^0(e, 0) + (d - 1) S_t^0(e + e', 0) + 1/2 S_t^0(2e, 0) . \end{aligned}$$

Hence, adopting the shorthand notation

$$f(t) := (d + 1/2) S_t^0(0, 0) - 2d S_t^0(e, 0) + (d - 1) S_t^0(e + e', 0) + 1/2 S_t^0(2e, 0) , \quad (4.2)$$

$$g(t) := S_t^0(0, 0) - S_t^0(e, 0) , \quad u(t) := S_t(0, 0) - S_t(e, 0) , \quad (4.3)$$

the identity in (4.1) reads as the following renewal equation

$$u(t) = g(t) + \int_0^t u(s) f(t - s) ds , \quad t \geq 0 , \quad (4.4)$$

for which, by iterating this integral relation, a solution may be expressed in terms of an infinite series expansion involving only the functions f and g :

$$u = g + \sum_{k=1}^{\infty} (g * f^{*k}) = \sum_{k=0}^{\infty} (g * f^{*k}) , \quad \text{with } f^{*0} := \delta_0 , \quad f^{*k} := f * f^{*(k-1)} . \quad (4.5)$$

Here, the symbol $*$ denotes the usual convolution for functions defined on \mathbb{R} , although we will always apply it to functions which vanish on $(-\infty, 0)$. Let us further remark that, with our notation, $f^{*1} = f$. Moreover, note that both f and g are bounded and continuous. Therefore, in order to ensure that $u = \sum_k (g * f^{*k})$ is the unique solution to (4.4), it suffices to show that the functions f and g (and, thus, their iterated convolutions) are non-negative. We prove this property in Lemma 4.1 below. In the same lemma, we gather other simple properties of these functions to be employed later.

Instrumentally to the proof of this lemma, we remark that the functions f and g may be further simplified, since they only depend on transition probabilities of the simple random walk Z_t^0 (without defects) on \mathbb{T}_N^d and Z_t^0 moves independently among each of the d components. Indeed, letting $p_t(i) = p_t(i, 0)$, for $i \in \mathbb{T}_N$ and $t \geq 0$, denote the transition probabilities to the origin of the one-dimensional simple random walk on \mathbb{T}_N (i.e., Z_t^0 for $d = 1$), we obtain, for all $d \geq 1$, $x = (i_1, \dots, i_d) \in \mathbb{T}_N^d$ and $t \geq 0$,

$$S_t^0(x, 0) = \prod_{\ell=1}^d p_t(i_\ell) .$$

Henceforth, we have, for all $d \geq 1$,

$$g(t) = p_t(0)^{d-1} (p_t(0) - p_t(1)) , \quad (4.6)$$

while, for $d = 1$,

$$f(t) = 3/2 p_t(0) - 2 p_t(1) + 1/2 p_t(2) , \quad (4.7)$$

and, for $d \geq 2$,

$$f(t) = p_t(0)^{d-2} \left[d (p_t(0) - p_t(1))^2 + 1/2 (p_t(0)^2 - 2 p_t(1)^2 + p_t(0) p_t(2)) \right] . \quad (4.8)$$

In what follows, we will repeatedly employ this convenient rewriting in combination with the explicit eigendecomposition (or, equivalently, the Laplace inversion formula) of $p_t(i)$ (see, e.g., [LP17, Section 12.3.1]): for all $i \in \mathbb{T}_N$ and $t \geq 0$,

$$p_t(i) = \frac{1}{N} \sum_{j=0}^{N-1} \cos(2\pi ij/N) \exp(-2\lambda_j^N t), \quad \text{with } \lambda_j^N := 1 - \cos(2\pi j/N) \geq 0. \quad (4.9)$$

Lemma 4.1. *For every $d \geq 1$, we have:*

- (a) *For all $t > 0$, $f(t)$ and $g(t)$ are positive and uniformly (with respect to $N \in \mathbb{N}$) bounded away from zero.*
- (b) *For all $N \in \mathbb{N}$ large enough, f and g are decreasing.*
- (c) *For all $N \in \mathbb{N}$ large enough, $f \leq (d + 1/2)g$.*

Proof. By the inversion formula (4.9), we obtain at once that

$$t \mapsto p_t(0) \quad \text{and} \quad t \mapsto p_t(0) - p_t(1)$$

are decreasing and, for each fixed $t \geq 0$, uniformly (in $N \in \mathbb{N}$) bounded away from zero. A similar conclusion holds for

$$t \mapsto 3/2 p_t(0) - 2 p_t(1) + 1/2 p_t(2) \quad \text{and} \quad t \mapsto p_t(0)^2 - 2 p_t(1)^2 + p_t(0) p_t(2),$$

as a consequence of the following two inequalities, respectively:

$$\begin{aligned} 3/2 - 2 \cos a + 1/2 \cos 2a &= 4 (\sin(a/2))^4 \geq 0, & a \in \mathbb{R}, \\ 1 - 2 \cos a \cos b + 1/2 \cos 2a + 1/2 \cos 2b &= (\cos a - \cos b)^2 \geq 0, & a, b \in \mathbb{R}. \end{aligned}$$

Therefore, the claims in items (a)–(b) hold true because f and g in (4.6)–(4.8) are products and sums of functions satisfying such properties.

As for item (c), we claim that $p_t(2) \leq p_t(1)$ for all $t \geq 0$. For this purpose, let us define $t \in [0, \infty) \mapsto q_t := p_t(1) - p_t(2)$, and observe that, for all $N \geq 4$, $q_0 = 0$, whereas, since $p_t(1), p_t(2) \rightarrow \frac{1}{N}$ as $t \rightarrow \infty$, $\lim_{t \rightarrow \infty} q_t = 0$. In view of (4.9), we have, for all $t \geq 0$,

$$q'_t = \frac{1}{N} \sum_{j=0}^{N-1} \{2 (\cos(2\pi j/N) - \cos(4\pi j/N)) (\cos(2\pi j/N) - 1)\} \exp(-2\lambda_j^N t), \quad (4.10)$$

from which we obtain $q'_0 = 1$. Indeed, $\frac{2}{N} \sum_{j=0}^{N-1} \cos(2\pi ij/N) \cos(2\pi \ell j/N) = 0$ if $(i, \ell) = (0, 1), (0, 2), (1, 2)$, whereas $= 1$ if $(i, \ell) = (1, 1)$. Furthermore, $t \mapsto q'_t$ must become negative because $q_0 = \lim_{t \rightarrow \infty} q_t = 0$, and we now argue that it changes sign only once, thus, ensuring that $q_t \geq 0$ for all $t \geq 0$. The expression within curly brackets in (4.10) reads as

$$\kappa_j^N := -16 (\sin(\pi j/N))^4 (1/2 + \cos(2\pi j/N)),$$

which is non-negative if and only if $j \in [N/3, 2N/3]$. Moreover, since $\kappa_j^N = \kappa_{N-j}^N$ and $\lambda_j^N = \lambda_{N-j}^N$, we get the following decomposition of q'_t in terms of non-negative and non-positive parts (in what follows, we omit to write the integer part of $N/3, N/2, 2N/3$):

$$q'_t = r_t^- + r_t^+ := \frac{2}{N} \sum_{j=1}^{N/3} \kappa_j^N \exp(-2\lambda_j^N t) + \frac{1}{N} \sum_{j=N/3+1}^{2N/3-1} \kappa_j^N \exp(-2\lambda_j^N t).$$

Let $T = T_N > 0$ be the first time $t > 0$ satisfying $q'_t = 0$. Then, since $j \in \{1, \dots, N - 1\} \mapsto \lambda_j^N \in (0, 2]$ is symmetric around $N/2$ and increasing on $\{1, \dots, N/2\}$, we get, for all $s \geq 0$,

$$q'_{T+s} \leq r_T^- \exp\left(-2\lambda_{\lfloor N/3 \rfloor}^N s\right) + r_T^+ \exp\left(-2\lambda_{\lfloor N/3 \rfloor + 1}^N s\right) \leq q'_T \exp\left(-2\lambda_{\lfloor N/3 \rfloor}^N s\right) \leq 0,$$

where the last step used the definition of T . Altogether, since we showed that $q'_0 = 1 > 0$, this proves that, for all $N \geq 4$ and $t \geq 0$, $q_t \geq 0$, that is,

$$p_t(2) \leq p_t(1), \quad t \geq 0. \tag{4.11}$$

Finally, inserting (4.11) into the expressions for f in (4.7)–(4.8), we obtain, for $d = 1$,

$$f(t) \leq 3/2 (p_t(0) - p_t(1)) = 3/2 g(t), \quad t \geq 0,$$

while, for $d \geq 2$,

$$\begin{aligned} f(t) &\leq p_t(0)^{d-2} \left[d(p_t(0) - p_t(1))^2 + 1/2 (p_t(0) - p_t(1))^2 \right] \\ &= (d + 1/2) p_t(0)^{d-2} (p_t(0) - p_t(1))^2 \\ &\leq (d + 1/2) p_t(0)^{d-1} (p_t(0) - p_t(1)) = (d + 1/2) g(t), \quad t \geq 0, \end{aligned}$$

where for the last inequality we used the crude bound $p_t(0) - p_t(1) \leq p_t(0)$. This concludes the proof of the lemma. \square

4.2 Lower bound in Theorem 2.1

In view of (3.3)–(3.4), (4.2)–(4.3), and the infinite series representation in (4.5), we recast the lower bound in (2.1) in terms of the functions $u(t)$ and $g(t)$. This is the content of the following lemma. Remark that, as in Theorem 2.1, for $t > 0$, we aim at a strict, uniform in $N \in \mathbb{N}$, inequality $u \gtrsim g$, improving upon the obvious estimate $u \geq g$ given in (3.5).

Lemma 4.2. *For all $d \geq 1$, there exists $c = c(d) \in (0, 1)$ satisfying, for all $N \in \mathbb{N}$ large enough and for all $t \geq 0$,*

$$u(t) \geq \exp(c(1 \wedge t)) g(t).$$

Proof. We show that, for all $d \geq 1$ and $N \in \mathbb{N}$,

$$u(t) \geq \left(\sup_{s \in [0, t]} \exp(sf(s)) \right) g(t), \quad t \geq 0, \tag{4.12}$$

holds true. This suffices since f is decreasing, $f(0) = d + 1/2$ and, for each $s > 0$, $f(s)$ is uniformly (in $N \in \mathbb{N}$) bounded away from zero (Lemma 4.1).

In what follows, we derive (4.12) by showing that

$$\int_0^t f^{*k}(s) ds \geq \frac{(tf(t))^k}{k!}, \quad t \geq 0, \tag{4.13}$$

holds for all $k \in \mathbb{N}$. We show (4.13) by induction on $k \in \mathbb{N}$. The claim for $k = 1$, namely $\int_0^t f(s) ds \geq tf(t)$, clearly holds since f is decreasing. Fix $k \geq 2$, and assume (4.13) to be true for $k - 1$; then, for all $t \geq 0$,

$$\begin{aligned} \int_0^t f^{*k}(s) ds &= \int_0^t \int_0^s f(s-r) f^{*(k-1)}(r) dr ds \\ &\geq \int_0^t f(s) \int_0^s f^{*(k-1)}(r) dr ds \end{aligned}$$

$$\begin{aligned} &\geq \int_0^t f(s)^k \frac{s^{k-1}}{(k-1)!} ds \\ &\geq f(t)^k \int_0^t \frac{s^{k-1}}{(k-1)!} ds = \frac{(tf(t))^k}{k!}, \end{aligned}$$

where for the first and third inequalities we used the monotonicity and positivity of f , while for the second one we used the induction hypothesis. This shows the validity of (4.13). The desired conclusion now follows by the infinite-series representation of $u(t)$ in (4.5), as well as the monotonicity of g and the positivity of f (Lemma 4.1): for all $t \geq s \geq 0$,

$$\begin{aligned} u(t) &= g(t) + \sum_{k=1}^{\infty} \int_0^t g(t-r) f^{*k}(r) dr \\ &\geq g(t) + g(t) \sum_{k=1}^{\infty} \int_0^s f^{*k}(r) dr \\ &\geq g(t) \left(\sum_{k=0}^{\infty} \frac{(sf(s))^k}{k!} \right) = g(t) \exp(sf(s)), \end{aligned}$$

where the last inequality is a consequence of (4.13) and positivity of g . Optimizing over $s \leq t$ yields (4.12) and, thus, concludes the proof. \square

4.3 Upper bound in Theorem 2.1

Also for the derivation of the upper bound in Theorem 2.1, we exploit the infinite-series representation (4.5) for u involving convolutions of the functions f and g . However, since we are not interested in optimizing constants in this case, we may employ $f \leq (d + 1/2)g$ from Lemma 4.1(c) and positivity of f and g (Lemma 4.1(a)) to estimate $u(t)$ from above by $\tilde{u}(t)$, defined as

$$\tilde{u} := \sum_{k=1}^{\infty} \tilde{g}^{*k}, \quad \text{with } \tilde{g} := (d + 1/2)g. \tag{4.14}$$

Hence, the following estimate combined with (2.2) (cf. (4.3) and (3.4)) would conclude the proof of the upper bound in Theorem 2.1.

Lemma 4.3. *For all $d \geq 1$ and for all $N \in \mathbb{N}$ large enough, we have*

$$\sum_{k=1}^{\infty} \tilde{g}^{*k}(t) \leq \frac{C}{N^{d+2} \wedge t^{d/2+1}} \exp\left(-2t/t_{\text{rel}} + \frac{Bt}{N^{d+2}}\right), \quad t > 0, \tag{4.15}$$

for some constants $B, C > 0$ (depending only on $d \geq 1$).

Before entering the details of the proof of Lemma 4.3, we remark that, compared to the bound

$$\tilde{g}(t) \leq \frac{C'}{N^{d+2} \wedge t^{d/2+1}} \exp(-2t/t_{\text{rel}}), \quad t \geq 0, \tag{4.16}$$

for some $C' = C'(d) > 0$, estimating the infinite series in (4.15) requires an extra factor $\exp\left(\frac{Bt}{N^{d+2}}\right)$. With this approach, this exponential factor is actually unavoidable, as we now explain. Recall the lower bound for g (thus, for \tilde{g} , too) in (2.2), analogous to (4.16): for some $C'' = C''(d) > 0$,

$$\tilde{g}(t) \geq \frac{C''}{(N^{d+2} \wedge t^{d/2+1}) \vee 1} \exp(-2t/t_{\text{rel}}), \quad t \geq 0.$$

Finally, letting $c = c(t) := \frac{C''}{N^{d+2}} \mathbf{1}_{[0, \infty)}(t)$, we get $\tilde{g}(t) \geq c \exp(-2t/t_{\text{rel}})$ and, thus,

$$\sum_{k=1}^{\infty} \tilde{g}^{*k}(t) \geq \exp(-2t/t_{\text{rel}}) \sum_{k=1}^{\infty} c^{*k}(t) = \exp(-2t/t_{\text{rel}}) \sum_{k=1}^{\infty} \frac{(ct)^k}{k!} = \exp(-2t/t_{\text{rel}} + ct) .$$

Proof of Lemma 4.3. Since we can follow closely the proof in [BB21, Lemma 5.6], our task boils down to establishing the analogues of the two preliminary lemmas therein, namely [BB21, Lemmas 5.4 & 5.5].

We start by showing that, for all $N \in \mathbb{N}$ large enough, we have

$$\int_0^{\infty} \tilde{g}(t) dt \leq \frac{d + 1/2}{2d} , \tag{4.17}$$

from which we then derive, for some $B = B(d) > 0$ and $\theta \in (0, 1)$,

$$\int_0^{\infty} \exp\left(2t/t_{\text{rel}} - \frac{Bt}{N^{d+2}}\right) \tilde{g}(t) dt \leq \theta . \tag{4.18}$$

In order to prove (4.17), recall $\tilde{g} := (d + 1/2)g$, as well as (4.3) and (3.4). Henceforth,

$$\int_0^{\infty} \tilde{g}(t) dt = \frac{d + 1/2}{2dN^d} \int_0^{\infty} 2\mathcal{E}\left(\frac{\pi_t(0, \cdot)}{\pi}\right) dt = \frac{d + 1/2}{2dN^d} \left\| \frac{\mathbf{1}_0}{\pi} - 1 \right\|_2^2 \leq \frac{d + 1/2}{2d} ,$$

where the last step follows from $\left\| \frac{\mathbf{1}_0}{\pi} - 1 \right\|_2^2 = N^d - 1$. This yields (4.17).

For what concerns (4.18), arguing similarly as in the proof of [BB21, Lemma 5.5], we split the integral in (4.18) into three pieces, so to obtain

$$\int_0^{\infty} \exp\left(2t/t_{\text{rel}} - \frac{Bt}{N^{d+2}}\right) \tilde{g}(t) dt = \int_0^{\varepsilon N^2} \dots + \int_{\varepsilon N^2}^{N^2} \dots + \int_{N^2}^{\infty} \dots , \tag{4.19}$$

for some well-chosen $\varepsilon \in (0, 1)$. We first deal with the first integral on the right-hand side above. Since $t_{\text{rel}} \geq \delta N^2$, for some $\delta > 0$, for all $\gamma \in (0, 1/3)$, there exists $\varepsilon = \varepsilon(\delta, \gamma) > 0$ small enough such that, for all $N \in \mathbb{N}$ large enough,

$$\exp(2\varepsilon N^2/t_{\text{rel}}) \leq 1 + \gamma .$$

With such a choice of $\gamma, \varepsilon > 0$, (4.17) ensures that

$$\int_0^{\varepsilon N^2} \exp\left(2t/t_{\text{rel}} - \frac{Bt}{N^{d+2}}\right) \tilde{g}(t) dt \leq (1 + \gamma) \int_0^{\infty} \tilde{g}(t) dt \leq \frac{(1 + \gamma)(d + 1/2)}{2d} < 1 . \tag{4.20}$$

Note the strict inequality above. Now, the second and third integrals on the right-hand side of (4.19) may be turned as small as desired via the estimate in (4.16), by choosing first $N \in \mathbb{N}$ large enough and then $B > 0$ large enough, respectively. Indeed,

$$\int_{\varepsilon N^2}^{N^2} \exp\left(2t/t_{\text{rel}} - \frac{Bt}{N^{d+2}}\right) \tilde{g}(t) dt \leq C' \int_{\varepsilon N^2}^{N^2} t^{-(d/2+1)} dt \leq CN^{-d} , \tag{4.21}$$

for some $C = C(d, \varepsilon) > 0$, and

$$\int_{N^2}^{\infty} \exp\left(2t/t_{\text{rel}} - \frac{Bt}{N^{d+2}}\right) \tilde{g}(t) dt \leq \frac{C'}{N^{d+2}} \int_{N^2}^{\infty} \exp\left(-\frac{Bt}{N^{d+2}}\right) dt = \frac{C'}{B} . \tag{4.22}$$

Hence, (4.20), (4.21) and (4.22) yield (4.18) for some $\theta \in (\frac{(1+\gamma)(d+1/2)}{2d}, 1)$ by choosing $\gamma > 0$ small enough, and, then, $N \in \mathbb{N}$ and $B > 0$ large enough.

From this stage, the rest of the proof follows as in [BB21, Lemma 5.6]; we streamline the main arguments in [BB21, pp. 1150–1153] for the reader’s convenience, while implementing the small modifications required. Recall $\theta \in (0, 1)$ from (4.18), and fix $\sigma \in (0, 1)$ such that $\theta/\sigma \in (0, 1)$. For such a $\sigma \in (0, 1)$, we define the sequence

$$a_m := \frac{1}{2} \frac{\sigma^{\frac{2(m-1)}{d+2}}}{\sum_{j=1}^{\infty} \sigma^{\frac{2(j-1)}{d+2}}}, \quad m \geq 1,$$

which, by definition, satisfies

$$\sum_{m=1}^{\infty} a_m = 1/2 \quad \text{and} \quad a^* := \prod_{m=1}^{\infty} (1 - a_m) > 0. \tag{4.23}$$

Next, by adopting the following shorthand notation for the integrand in (4.18),

$$h(t) := \exp\left(2t/t_{\text{rel}} - \frac{Bt}{Nd+2}\right) \tilde{g}(t) \leq \frac{C'}{t^{d/2+1} \wedge Nd+2} =: \tilde{h}(t), \quad t > 0, \tag{4.24}$$

we claim that the following holds true: for all $k \in \mathbb{N}$ and $t > 0$,

$$h^{*k}(t) \leq 2\theta^k \tilde{h}\left(t - t \sum_{m=1}^k a_m\right) + \theta^{k-1} \sum_{m=1}^k \tilde{h}(ta^* a_m). \tag{4.25}$$

We refer to [BB21, Eq. (5.35) & pp. 1150–1151] for the proof by induction of this inequality. Here, we only remark that $t \mapsto \tilde{h}(t)$ given in (4.24) (and, thus, the right-hand side of (4.25)) is a non-increasing function; this fact, (4.23), and the inequalities in (4.18) and (4.24) are the only inputs required for the proof of (4.25).

By (4.25) and the monotonicity of $t \mapsto \tilde{h}(t)$, we obtain, for all $t > 0$,

$$\begin{aligned} \sum_{k=1}^{\infty} h^{*k}(t) &\leq 2 \sum_{k=1}^{\infty} \theta^k \tilde{h}\left(t - t \sum_{m=1}^k a_m\right) + \sum_{k=1}^{\infty} \theta^{k-1} \sum_{m=1}^k \tilde{h}(ta^* a_m) \\ &\leq \frac{2\theta}{1-\theta} \tilde{h}(t/2) + \frac{1}{1-\theta} \sum_{m=1}^{\infty} \theta^{m-1} \tilde{h}(ta^* a_m). \end{aligned}$$

Now, for $t \leq 2N^2$, since $ta^* a_m \leq ta_m \leq N^2$, we further get, for some $C = C(d) > 0$,

$$\begin{aligned} \sum_{k=1}^{\infty} h^{*k}(t) &\leq \frac{2\theta C}{1-\theta} \frac{1}{t^{d/2+1}} + \frac{C'}{1-\theta} \sum_{m=1}^{\infty} \frac{\theta^{m-1}}{(ta^* a_m)^{d/2+1}} \\ &\leq \frac{2\theta C}{1-\theta} \frac{1}{t^{d/2+1}} + \frac{C''}{1-\theta} \frac{1}{t^{d/2+1}} \sum_{m=1}^{\infty} \left(\frac{\theta}{\sigma}\right)^{m-1} \\ &\leq \frac{C}{t^{d/2+1}}, \end{aligned}$$

where for the last inequality we used that $\theta/\sigma \in (0, 1)$. Analogously, we have

$$\sum_{k=1}^{\infty} \tilde{h}^{*k}(t) \leq \frac{C}{N^{d+2}}, \quad t > 2N^2.$$

Since

$$\sum_{k=1}^{\infty} h^{*k}(t) = \exp\left(2t/t_{\text{rel}} - \frac{Bt}{Nd+2}\right) \sum_{k=1}^{\infty} \tilde{g}^{*k}(t), \quad t \geq 0,$$

this concludes the proof of the lemma. □

4.4 Conclusion of the proof of Theorem 2.1

Proof of Theorem 2.1. Recall (3.3), (3.4), and (4.3). Then, the lower bound in (2.1) is the content of Lemma 4.2. The upper bound for times $t \geq 1$ may be obtained from $u \leq \tilde{u}$ (defined in (4.14)), Lemma 4.3, and the first inequality in (2.2). The same upper bound, but for small times $t \in [0, 1]$, follows by the very definition of u in (4.3) and

$$u(t) = S_t(0, 0) - S_t(e, 0) \leq S_t(0, 0) \leq 1 . \tag{4.26}$$

This concludes the proof of the theorem. □

5 Proof of Theorem 2.4

We divide the proof of Theorem 2.4 on the phenomenon of early concentration into two parts. First, we express the quantity of interest in terms of (mean) Dirichlet forms for $\text{Avg}(\mathbb{T}_N^d)$ and $\text{RW}(\mathbb{T}_N^d)$. This is carried out in Lemma 5.1, by specializing some expressions obtained in [QS23, CQS23] to our setting. Finally, thanks to this rewriting, we exploit the estimates (and the proof arguments) from Theorem 2.1 to conclude.

Lemma 5.1. *For all $d \geq 1$, $N \in \mathbb{N}$ and $t \geq 0$, we have*

$$\mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi} \right\|_2^2 \right] = \frac{1}{dN^d} \int_0^t \mathbb{E} \left[\mathcal{E} \left(\frac{\eta_{t-s}(0, \cdot)}{\pi} \right) \right] \mathcal{E} \left(\frac{\pi_s(0, \cdot)}{\pi} \right) ds \tag{5.1}$$

$$= dN^d \int_0^t (S_{t-s}(0, 0) - S_{t-s}(e, 0)) (S_s^0(0, 0) - S_s^0(e, 0)) ds , \tag{5.2}$$

for some $e \in \mathbb{T}_N^d$ with $|e| = 1$.

Proof. Let \mathcal{N}_t denote the left-hand side of (5.1). Then, by the last two displays in [CQS23, Proposition 2.5], we have

$$\mathcal{N}_t = \frac{N^d}{2} \int_0^t \frac{1}{2} \sum_{x \in \mathbb{T}_N^d} \sum_{\substack{y \in \mathbb{T}_N^d \\ |x-y|=1}} (\pi_s(0, x) - \pi_s(0, y))^2 \Phi_{t-s}(x, y) ds , \tag{5.3}$$

where $\Phi_t(x, y)$ is defined in terms of $\text{CRW}(\mathbb{T}_N^d)$ (Section 3.1) as follows:

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

By translation invariance of the dynamics of $\text{CRW}(\mathbb{T}_N^d)$ and recalling the definition of the difference process Z_t and its kernel S_t (Section 3.2), we obtain, for all nearest neighbor vertices $x, y \in \mathbb{T}_N^d$,

$$\Phi_t(x, y) = \mathbf{P}_{0,0}^{\text{CRW}}(X_t = Y_t) - \mathbf{P}_{0,e}^{\text{CRW}}(X_t = Y_t) = S_t(0, 0) - S_t(e, 0) ,$$

where $e \in \mathbb{T}_N^d$ is any vertex satisfying $|e| = 1$. Plugging this identity into (5.3), we get

$$\mathcal{N}_t = \int_0^t (S_{t-s}(0, 0) - S_{t-s}(e, 0)) \mathcal{E} \left(\frac{\pi_s(0, \cdot)}{\pi} \right) ds .$$

The identities in (5.1) and (5.2) follow by (3.3) and (3.4), respectively. □

Proof of Theorem 2.4. Let us adopt the notation from Section 4. Recalling the definitions of $u(t)$ and $g(t)$ in (4.3), the identity in (5.2) reads as

$$\frac{1}{dN^d} \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi} \right\|_2^2 \right] = (u * g)(t) , \quad t \geq 0 .$$

For the upper bound in (2.3), since $u \leq \tilde{u}$ and $0 \leq g \leq \tilde{g}$ (cf. (4.5) and (4.14)), we obtain

$$u * g \leq \tilde{u} * g = \sum_{k=1}^{\infty} (\tilde{g}^{*k} * g) \leq \sum_{k=2}^{\infty} \tilde{g}^{*k} \leq \sum_{k=1}^{\infty} \tilde{g}^{*k} = \tilde{u} ,$$

and observe that the right-hand side was estimated, for all $t > 0$ and $N \in \mathbb{N}$ large enough, in Lemma 4.3. This provides the desired estimate for $t \geq 1$; for $t \in [0, 1]$, it suffices to recall $g \leq u \leq 1$ (cf. (4.5) and (4.26)), yielding

$$(u * g)(t) \leq t , \quad t \geq 0 .$$

For the lower bound in (2.3), using again $0 \leq g \leq u$ and the monotonicity of $t \mapsto g(t)$ (Lemma 4.1(b)), we get, for all $t \geq 0$,

$$(u * g)(t) \geq (g * g)(t) = 2 \int_0^{t/2} g(t-s)g(s) \, ds \geq 2g(t) \int_0^{t/2 \wedge 1} g(s) \, ds \geq c(t \wedge 1)g(t) ,$$

where the last inequality follows from Lemma 4.1(a)–(b). The first inequality in (2.2) and (3.4), (4.3) yield the desired lower bound, thus, concluding the proof of the theorem. \square

6 Proofs of Propositions 2.2 and 2.5

We start with the proof of Proposition 2.2.

Proof of Proposition 2.2. Recall that, for \mathbb{P} -a.e. realization of the Poisson point process used for the updates, $\eta_t(0, \cdot)$ may also be interpreted as the probability distribution over \mathbb{T}_N^d of the “infinitesimal chunk” of mass $U_{0,t} \in \mathbb{T}_N^d$, which is a time-inhomogeneous random walk, started in $0 \in \mathbb{T}_N^d$, and evolving at later times as follows: nothing happens, until the vertex on which it sits, say $x \in \mathbb{T}_N^d$, experiences an update with a nearest neighbor, say $y \in \mathbb{T}_N^d$; in that case, the chunk moves from x to y with probability $1/2$, while with the remaining probability it stays put. Since we are describing a time-inhomogeneous Markov process, we adopt, only in this proof, the slightly more convenient notation $\eta_{s,t}(\cdot, \cdot)$, $s \leq t$, indicating both starting and terminal times.

In view of this representation, Chapman-Kolmogorov formula holds in our context, and reads as follows: for all $0 \leq s \leq t$,

$$\eta_{0,t}(0, x) - \eta_{0,t}(0, y) = \sum_{z \in \mathbb{T}_N^d} \eta_{0,s}(0, z) (\eta_{s,t}(z, x) - \eta_{s,t}(z, y)) . \tag{6.1}$$

Then, by (6.1) and Cauchy-Schwarz inequality, we get, for all $0 \leq s \leq t$,

$$\begin{aligned} \mathbb{E}[(\eta_t(0, x) - \eta_t(0, y))^2] &= \mathbb{E}\left[\left(\sum_{z \in \mathbb{T}_N^d} \eta_{0,s}(0, z) (\eta_{s,t}(z, x) - \eta_{s,t}(z, y))\right)^2\right] \\ &\leq \sum_{z \in \mathbb{T}_N^d} \mathbb{E}[\eta_{0,s}(0, z) (\eta_{s,t}(z, x) - \eta_{s,t}(z, y))^2] \\ &= \sum_{z \in \mathbb{T}_N^d} \mathbb{E}[\eta_s(0, z)] \mathbb{E}[(\eta_{t-s}(z, x) - \eta_{t-s}(z, y))^2] \\ &= \sum_{z \in \mathbb{T}_N^d} \pi_s(0, z) \mathbb{E}[(\eta_{t-s}(z, x) - \eta_{t-s}(z, y))^2] , \end{aligned}$$

where the third step is a consequence of the fact that the Poisson updates over the time intervals $(0, s)$ and $(s, +\infty)$ are independent, while for the fourth one we used (1.4). By

using the well-known heat kernel estimate $\pi_s(0, z) \leq C((s^{d/2} \wedge N^d) \vee 1)^{-1}$, we further get

$$\begin{aligned} \mathbb{E}[(\eta_t(0, x) - \eta_t(0, y))^2] &\leq \frac{C}{(s^{d/2} \wedge N^d) \vee 1} \sum_{z \in \mathbb{T}_N^d} \mathbb{E}[(\eta_{t-s}(z, x) - \eta_{t-s}(z, y))^2] \\ &= \frac{C}{(s^{d/2} \wedge N^d) \vee 1} \frac{2}{dN^d} \mathbb{E} \left[\mathcal{E} \left(\frac{\eta_{t-s}(0, \cdot)}{\pi} \right) \right], \end{aligned}$$

where for the last step we used the duality relation (3.1), the reversibility of $\text{CRW}(\mathbb{T}_N^d)$ with respect to the counting measure on $\mathbb{T}_N^d \times \mathbb{T}_N^d$, and Proposition 3.1. After inserting $s = t/2 \wedge N^2$ in this last estimate, the upper bounds in (2.1) and (2.2) yield the desired result. \square

We then conclude this section with the proof of Proposition 2.5 about the limit profile.

Proof of Proposition 2.5. Recall that, although we omit $N \in \mathbb{N}$ from the notation, $\|\cdot\|_p$ denotes the L^p -norm on (\mathbb{T}_N^d, π) . By the triangle inequality and (1.5)–(1.6), we get

$$\begin{aligned} &\left| \mathbb{E} \left[\left\| \frac{\eta_{tN^2}(0, \cdot)}{\pi} - 1 \right\|_p^{\frac{1}{p}} \right] - \|h_t - 1\|_{L^p(\mathbb{T}^d)} \right| \\ &\leq \mathbb{E} \left[\left\| \frac{\eta_{tN^2}(0, \cdot)}{\pi} - 1 \right\|_p^{\frac{1}{p}} - \left\| \frac{\pi_{tN^2}(0, \cdot)}{\pi} - 1 \right\|_p \right] \\ &\quad + \left| \left\| \frac{\pi_{tN^2}(0, \cdot)}{\pi} - 1 \right\|_p - \|h_t(\cdot/N) - 1\|_p \right| + \left| \|h_t(\cdot/N) - 1\|_p - \|h_t - 1\|_{L^p(\mathbb{T}^d)} \right| \\ &\leq \mathbb{E} \left[\left\| \frac{\eta_{tN^2}(0, \cdot)}{\pi} - \frac{\pi_{tN^2}(0, \cdot)}{\pi} \right\|_2^2 \right]^{\frac{1}{2}} \\ &\quad + \left\| \frac{\pi_{tN^2}(0, \cdot)}{\pi} - h_t(\cdot/N) \right\|_p + \frac{1}{N} \sum_{i=1}^d \|\partial_i h_t\|_{C(\mathbb{T}^d)}, \end{aligned}$$

where the last term has been estimated via a first-order Taylor expansion (recall that $h_t \in C^\infty(\mathbb{T}^d)$, for all $t > 0$). The first term on the right-hand side above is $O(\frac{1}{N})$ thanks to Theorem 2.4, while the second term is $o(\frac{1}{N})$ by the quantitative local CLT for $\text{RW}(\mathbb{T}_N^d)$ (see, e.g., [LL10] for the analogous result on \mathbb{Z}^d). Since all these estimates are uniform over finite time intervals bounded away from zero and infinity, this concludes the proof. \square

7 An application to cutoff of the dual particle system

As a further consequence of Corollary 2.3 and the quantitative concentration result in Theorem 2.4, we provide sharp estimates for the mixing time of the particle system discussed in [QS23], dual to the averaging process, known as *binomial splitting process*. In a few words, this model is the k -particle generalization of the processes $\text{RW}(\mathbb{T}_N^d)$ and $\text{CRW}(\mathbb{T}_N^d)$ previously introduced. Here, we provide a quick description of this particle system and refer to [QS23, Section 2.1] (see also [PR23, Section 1.2]) for more details, results, and background.

Given the discrete torus \mathbb{T}_N^d , $d \geq 1$, and an integer $k \in \mathbb{N}$, let $\Omega_k = \Omega_{N,k}$ denote the configuration space of k indistinguishable particles on \mathbb{T}_N^d , namely,

$$\Omega_k := \left\{ \zeta \in \mathbb{N}_0^{\mathbb{T}_N^d} \mid \sum_{x \in \mathbb{T}_N^d} \zeta(x) = k \right\}.$$

On such a configuration space, we consider $(\zeta_t^k)_{t \geq 0} = (\zeta_t^{N,k})_{t \geq 0}$ as the Markov process which evolves as follows: start with a particle configuration $\zeta \in \Omega_k$ at time $t = 0$; at the Poisson event times of $\text{Avg}(\mathbb{T}_N^d)$ involving, say, the nearest neighbors $x, y \in \mathbb{T}_N^d$, reassign, independently and uniformly at random, x or y as the new position of each particle which was originally sitting on either x or y . The name “binomial splitting” is now explained: the update corresponds to placing on x a binomially-distributed fraction of the total $\zeta(x) + \zeta(y)$ particles. Further, the dynamics conserves the total number of particles, and, although particles interact when performing simultaneous jumps, at equilibrium, they are distributed as if they were independent. Indeed, a simple detailed balance computation ensures that

$$\mu_{k,\pi} = \text{Multinomial}(k, \pi) ,$$

is the unique reversible measure for the particle system with k particles. Finally, it is immediate to see that the particle systems with $k = 1$ and $k = 2$ particles correspond, respectively, to the processes $\text{RW}(\mathbb{T}_N^d)$ and $\text{CRW}(\mathbb{T}_N^d)$ described in Sections 1.3 and 3.1, once the particles’ labels are removed.

We are interested in proving a total variation (TV) cutoff phenomenon for this process, as $k = k_N \rightarrow \infty$, strengthening the result in [QS23, Theorem 2.3]. For this purpose, let, for all $k \in \mathbb{N}$, $(P_t^k)_{t \geq 0}$ denote the Markov transition kernel associated to $(\zeta_t^k)_{t \geq 0}$, and write μP_t^k for the distribution of ζ_t^k when $\zeta_0^k \sim \mu$, for some $\mu \in \mathcal{P}(\Omega_k)$. (Here, in analogy with $\mathcal{P}(\mathbb{T}_N^d)$, $\mathcal{P}(\Omega_k)$ stands for the space of probability distributions on Ω_k .) Further, we consider

$$d_k(t) = d_{N,k}(t) := \sup_{\mu \in \mathcal{P}(\Omega_k)} \|\mu P_t^k - \mu_{k,\pi}\|_{\text{TV}} , \quad t \geq 0 ,$$

to encode the worst-case TV-distance to equilibrium of the process as a function of time. Then, recalling that $t_{\text{rel}} = t_{\text{rel}}^N > 0$ denotes the spectral gap of $\text{RW}(\mathbb{T}_N^d)$ (which, by [QS23, Theorem 2.1], coincides with the spectral gap of the k -particle binomial splitting process), and letting

$$T(a) = T_{N,k}(a) := \frac{t_{\text{rel}}}{2} (\log k + aw_k) , \quad a \in \mathbb{R} , \tag{7.1}$$

the result in [QS23, Theorem 2.3] specialized to the setting of the torus \mathbb{T}_N^d reads as follows:

$$\lim_{a \rightarrow -\infty} \liminf_{N \rightarrow \infty} d_k(T(a)) = 1 , \tag{7.2}$$

$$\lim_{a \rightarrow \infty} \limsup_{N \rightarrow \infty} d_k(T(a)) = 0 , \tag{7.3}$$

for $w_k \equiv 1$ and for all $k = k_N \rightarrow \infty$ satisfying $k = O(N^{2d})$. Our main improvement consists in extending this result from quadratic to exponentially-many particles, paying the price of enlarging the so-called cutoff window w_k .

Proposition 7.1 (Cutoff for binomial splitting process). *Letting $w_k := 1 \vee \frac{\log k}{N^d}$, we have:*

(a) *the claim in (7.2) holds for all $k \rightarrow \infty$ satisfying $\log k = o(N^d \log N)$;*

(b) *the claim in (7.3) holds for all $k \rightarrow \infty$.*

In particular, TV-cutoff for $(\zeta_t^k)_{t \geq 0}$ holds for $k \rightarrow \infty$ satisfying $\log k = o(N^d \log N)$.

We anticipate that, on the one hand, the strategy in [QS23] and our estimates on the fast local smoothness of $\text{Avg}(\mathbb{T}_N^d)$ suffice to yield the desired upper bound, i.e., Proposition 7.1(b). On the other hand, for the lower bound, our results on early concentration of $\text{Avg}(\mathbb{T}_N^d)$ combined with the proof in [QS23] fail to provide sharp

estimates in the regime $\log k \geq N^d$. With the purpose of covering also this case, here, we follow a different proof strategy.

Finally, the key observation of the proof of Proposition 7.1 is that $(\zeta_t^k)_{t \geq 0}$ and $\text{Avg}(\mathbb{T}_N^d)$ are related to each other via the following *intertwining relation* [QS23, Proposition 3.1]: for all $\xi \in \mathcal{P}(\mathbb{T}_N^d)$, $k \in \mathbb{N}$, and $t \geq 0$,

$$\mu_{k,\xi} P_t^k = \mathbb{E}[\mu_{k,\eta_t^\xi}], \quad \text{with } \mu_{k,\eta} := \text{Multinomial}(k, \eta), \quad \text{for all } \eta \in \mathcal{P}(\mathbb{T}_N^d). \quad (7.4)$$

We crucially exploit this fact for establishing both claims in Proposition 7.1.

Proof of Proposition 7.1. We divide the proof into two parts, starting with item (b).

Upper bound (b). By (a slight generalization of) the intertwining relation in (7.4), we have the following estimate [QS23, Lemma 6.4]: for all $k \in \mathbb{N}$ and $t \geq 0$,

$$d_k(t) := \sup_{\mu \in \mathcal{P}(\Omega_k)} \|\mu P_t^k - \mu_{k,\pi}\|_{\text{TV}} \leq \sqrt{ek} \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - 1 \right\|_2^2 \right]^{\frac{1}{2}}.$$

Thanks to Corollary 2.3, we further get, for some $C = C(d) > 0$ and for all $t \geq N^2$,

$$d_k(t) \leq \frac{C\sqrt{k} N^{d/2}}{(N^{d/2} \wedge t^{d/4}) \vee 1} \exp\left(-t/t_{\text{rel}} + \frac{Bt}{2N^{d+2}}\right) = C\sqrt{k} \exp\left(-t/t_{\text{rel}} + \frac{Bt}{2N^{d+2}}\right),$$

where for the last step we used $t \geq N^2$. Recall $t_{\text{rel}} = \Theta(N^2)$, $k \rightarrow \infty$, and $w_k := 1 \vee \frac{\log k}{N^d}$. Hence, by substituting $t = T(a) \gg N^2$ given in (7.1), the right-hand side above reads as

$$C \exp\left(-\frac{a}{2} w_k + \frac{Bt_{\text{rel}}}{4N^{d+2}} (\log k + aw_k)\right),$$

which vanishes taking first $N \rightarrow \infty$ and then $a \rightarrow \infty$. This concludes the proof of the upper bound.

Lower bound (a). The approach we follow goes via the intertwining relation (7.4), which we employ as follows:

$$\begin{aligned} d_k(t) &\geq \|\mu_{k,\xi} P_t^k - \mu_{k,\pi}\|_{\text{TV}} = \sup_{A \subset \Omega_k} (\mathbb{E}[\mu_{k,\eta_t^\xi}(A)] - \mu_{k,\pi}(A)) \\ &\geq \mathbb{E}[\mu_{k,\eta_t^\xi}(A_*)] - \mu_{k,\pi}(A_*), \end{aligned} \quad (7.5)$$

for some $A_* \subset \Omega_k$ and $\xi \in \mathcal{P}(\mathbb{T}_N^d)$ that we now specify.

First, we set $\xi = \mathbb{1}_0$ and $t = T(a)$, for some $a < 0$. With this choice, we have (see, e.g., [LP17, Lemma 20.11])

$$\|\pi_t(0, \cdot) - \pi\|_{\text{TV}} := \sup_{B \subset \mathbb{T}_N^d} (\pi_t(0, B) - \pi(B)) \geq \frac{e^{-t/t_{\text{rel}}}}{2} = \frac{e^{-aw_k/2}}{2\sqrt{k}}.$$

Let $B_* = B_*(N, a) \subset \mathbb{T}_N^d$ be a proper subset which attains the above supremum; hence,

$$\pi_t(0, B_*) \geq \pi(B_*) + \frac{e^{-aw_k/2}}{2\sqrt{k}}. \quad (7.6)$$

Now, for some $b > 0$ to be fixed later, define, for all $N \in \mathbb{N}$ large enough,

$$A_* = A_*(N, a, b) := \left\{ \zeta \in \Omega_k : \sum_{x \in B_*} \zeta(x) \geq k(\pi(B_*) + \frac{b}{\sqrt{k}}) \right\}.$$

Let us continue from (7.5). Since, under $\mu_{k,\eta}$, $k \in \mathbb{N}$ and $\eta \in \mathcal{P}(\Gamma_N^d)$, we have $Z^{k,\eta(B_*)} := \sum_{x \in B_*} \zeta(x) \sim \text{Bin}(k, \eta(B_*))$, we obtain (recall $t = T(a)$)

$$\begin{aligned} d_k(t) &\geq \mathbb{E}[\mathbb{P}(Z^{k,\eta_t(0,B_*)} \geq k(\pi(B_*) + \frac{b}{\sqrt{k}}))] - \mathbb{P}(Z^{k,\pi(B_*)} \geq k(\pi(B_*) + \frac{b}{\sqrt{k}})) \\ &\geq \mathbb{E}[\mathbb{P}(Z^{k,\eta_t(0,B_*)} \geq k(\pi(B_*) + \frac{b}{\sqrt{k}}))] - \frac{\pi(B_*)(1 - \pi(B_*))}{b^2} \\ &\geq \mathbb{E}[\mathbb{P}(Z^{k,\eta_t(0,B_*)} \geq k(\pi(B_*) + \frac{b}{\sqrt{k}}))] - \frac{1}{b^2}, \end{aligned}$$

where for the second step we applied Cauchy-Schwarz inequality. Let us now provide a lower bound for the first term on the right-hand side above. For this purpose, for every $c > b > 0$,

$$\begin{aligned} &\mathbb{E}[\mathbb{P}(Z^{k,\eta_t(0,B_*)} \geq k\pi(B_*) + b\sqrt{k})] \\ &\geq \mathbb{E}[\mathbb{P}(Z^{k,\eta_t(0,B_*)} \geq k\pi(B_*) + b\sqrt{k}) \mathbb{1}_{\{\eta_t(0,B_*) \geq \pi(B_*) + \frac{c}{\sqrt{k}}\}}] \\ &\geq \mathbb{P}(Z^{k,\pi(B_*) + \frac{c}{\sqrt{k}}} \geq k\pi(B_*) + b\sqrt{k}) \mathbb{P}(\eta_t(0, B_*) \geq \pi(B_*) + \frac{c}{\sqrt{k}}) \\ &= \mathbb{P}(Z^{k,\pi(B_*) + \frac{c}{\sqrt{k}}} \geq k\pi(B_*) + c\sqrt{k} - (c - b)\sqrt{k}) \mathbb{P}(\eta_t(0, B_*) \geq \pi(B_*) + \frac{c}{\sqrt{k}}) \\ &\geq \frac{(c - b)^2}{(c - b)^2 + 1} \mathbb{P}(\eta_t(0, B_*) \geq \pi(B_*) + \frac{c}{\sqrt{k}}), \end{aligned}$$

where for the second inequality we used the stochastic domination $Z^{k,q} \succeq Z^{k,p}$ if $q > p$, while for the last one we used Cantelli inequality. Furthermore, provided that $e^{-aw_k/2} > 4c$, we get, again by Cantelli inequality,

$$\begin{aligned} \mathbb{P}(\eta_t(0, B_*) \geq \pi(B_*) + \frac{c}{\sqrt{k}}) &= \mathbb{P}(\eta_t(0, B_*) \geq \pi_t(0, B_*) - (\pi_t(0, B_*) - \pi(B_*) - \frac{c}{\sqrt{k}})) \\ &\geq \frac{1}{1 + \sigma^2/r^2}, \end{aligned}$$

with

$$\sigma := \mathbb{E}[(\eta_t(0, B_*) - \pi_t(0, B_*))^2]^{\frac{1}{2}}, \quad r := \pi_t(0, B_*) - \pi(B_*) - \frac{c}{\sqrt{k}} \geq \frac{e^{-aw_k/2}/2 - c}{\sqrt{k}} > 0.$$

Hence, we are done as soon as we can show that $\sigma^2/r^2 \rightarrow 0$. Recall (7.6) and $e^{-aw_k} > 4c$; then, we have

$$r^2 \geq \frac{1}{k} \left(\frac{e^{-aw_k/2}}{2} - c \right)^2 \geq \frac{e^{-aw_k}}{16k}.$$

Moreover, recalling that $\|\frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi}\|_1 = 2 \sup_{B \subset \Gamma_N^d} (\eta_t(0, B) - \pi_t(0, B))$, we get

$$\begin{aligned} \sigma^2 &:= \mathbb{E}[(\eta_t(0, B_*) - \pi_t(0, B_*))^2] \leq \frac{1}{4} \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi} \right\|_1^2 \right] \\ &\leq \frac{1}{4} \mathbb{E} \left[\left\| \frac{\eta_t(0, \cdot)}{\pi} - \frac{\pi_t(0, \cdot)}{\pi} \right\|_2^2 \right] \\ &\leq \frac{C}{N^2} \frac{e^{-aw_k}}{k} \exp \left(\frac{Bt_{\text{rel}}}{2N^{d+2}} (\log k + aw_k) \right), \end{aligned}$$

where for the second line we used Jensen inequality, while for the third one we used the second inequality in (2.3) with $t = T(a) \geq N^2$. (Here, $B, C > 0$ depend only on $d \geq 1$.) Hence,

$$\frac{\sigma^2}{r^2} \leq \frac{C}{N^2} \exp \left(\frac{Bt_{\text{rel}}}{2N^{d+2}} (\log k + aw_k) \right) \leq \frac{C}{N^2} \exp \left(\frac{Bt_{\text{rel}}}{2N^{d+2}} \log k \right) \xrightarrow{N \rightarrow \infty} 0,$$

provided that $\log k = o(N^d \log N)$. Finally, setting $c > b > 0$ large enough and $a < 0$ small enough concludes the proof. \square

References

- [AF02] David Aldous and James Allen Fill. Reversible Markov Chains and Random Walks on Graphs, 2002. Unfinished monograph, recompiled 2014. Available at <https://www.stat.berkeley.edu/users/aldous/RWG/book.pdf>.
- [AL12] David Aldous and Daniel Lanoue. A lecture on the averaging process. *Probab. Surv.*, 9:90–102, 2012. MR2908618
- [Ald11] David Aldous. Finite Markov Information-Exchange Processes, 2011. Available at <https://www.stat.berkeley.edu/~aldous/260-FMIE/index.html>.
- [BB21] Sayan Banerjee and Krzysztof Burdzy. Rates of convergence to equilibrium for potlatch and smoothing processes. *Ann. Probab.*, 49(3):1129–1163, 2021. MR4255140
- [BBG20] Raphaël Berthier, Francis Bach, and Pierre Gaillard. Tight Nonparametric Convergence Rates for Stochastic Gradient Descent under the Noiseless Linear Model. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS’20, Red Hook, NY, USA, 2020. Curran Associates Inc.
- [BCN20] Luca Becchetti, Andrea Clementi, and Emanuele Natale. Consensus dynamics: an overview. *ACM SIGACT News*, 51(1):58–104, 2020. MR4086635
- [BDCJ22] Gerardo Barrera, Conrado Da Costa, and Milton Jara. Gradual convergence for Langevin dynamics on a degenerate potential. *arXiv:2209.11026*, 2022.
- [BGPS06] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE Trans. Inform. Theory*, 52(6):2508–2530, 2006. MR2238556
- [Cao23] Fei Cao. Explicit decay rate for the Gini index in the repeated averaging model. *Math. Methods Appl. Sci.*, 46(4):3583–3596, 2023. MR4551331
- [CDSZ22] Sourav Chatterjee, Persi Diaconis, Allan Sly, and Lingfu Zhang. A phase transition for repeated averages. *Ann. Probab.*, 50(1):1–17, 2022. MR4385355
- [CQS23] Pietro Caputo, Matteo Quattropiani, and Federico Sau. Cutoff for the averaging process on the hypercube and complete bipartite graphs. *Electron. J. Probab.*, 28:Paper No. 100, 31, 2023. MR4617937
- [DKS23] Jean-Dominique Deuschel, Takashi Kumagai, and Martin Slowik. Gradient estimates of the heat kernel for random walks among time-dependent random conductances. *arXiv:2309.09675*, 2023.
- [KL99] Claude Kipnis and Claudio Landim. *Scaling limits of interacting particle systems*, volume 320 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1999. MR1707314
- [Lig05] Thomas M. Liggett. *Interacting particle systems*. Classics in Mathematics. Springer-Verlag, Berlin, 2005. Reprint of the 1985 original. MR2108619
- [LL10] Gregory F. Lawler and Vlada Limic. *Random walk: a modern introduction*, volume 123 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. MR2677157
- [LP17] 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.
- [MSW22] Ramis Movassagh, Mario Szegedy, and Guanyang Wang. Repeated Averages on Graphs. *arXiv:2205.04535*, 2022.
- [MT06] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006. MR2341319
- [PR23] Richard Pymar and Nicolás Rivera. Mixing of the symmetric beta-binomial splitting process on arbitrary graphs. *arXiv:2307.02406*, 2023.
- [QS23] 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
- [Sha09] Devavrat Shah. Gossip algorithms. *Found. Trends Netw.*, 3(1):1–125, 2009. MR2969074

Concentration and local smoothness

- [Spi22] Sam Spiro. An averaging process on hypergraphs. *J. Appl. Probab.*, 59(2):495–504, 2022. MR4444030
- [Tra23] Hong-Quan Tran. Cutoff for the Glauber-Exclusion process in the full high-temperature regime: an information percolation approach. *arXiv:2310.15810*, 2023.