The Annals of Probability
2013, Vol. 41, No. 5, 3420–3461
DOI: 10.1214/12-AOP813

MEAN FIELD CONDITIONS FOR COALESCING RANDOM WALKS[1]

BY ROBERTO IMBUZEIRO OLIVEIRA

IMPA

The main results in this paper are about the *full coalescence time* C of a system of coalescing random walks over a finite graph G. Letting $\mathsf{m}(G)$ denote the mean meeting time of two such walkers, we give sufficient conditions under which $\mathbf{E}[\mathsf{C}] \approx 2\mathsf{m}(G)$ and $\mathsf{C}/\mathsf{m}(G)$ has approximately the same law as in the "mean field" setting of a large complete graph. One of our theorems is that mean field behavior occurs over all vertex-transitive graphs whose mixing times are much smaller than $\mathsf{m}(G)$; this nearly solves an open problem of Aldous and Fill and also generalizes results of Cox for discrete tori in $d \geq 2$ dimensions. Other results apply to nonreversible walks and also generalize previous theorems of Durrett and Cooper et al. Slight extensions of these results apply to voter model consensus times, which are related to coalescing random walks via duality.

Our main proof ideas are a strengthening of the usual approximation of hitting times by exponential random variables, which give results for nonstationary initial states; and a new general set of conditions under which we can prove that the hitting time of a union of sets behaves like a minimum of independent exponentials. In particular, this will show that the first meeting time among k random walkers has mean $\approx \mathsf{m}(G)/\binom{k}{2}$.

1. Introduction. Start a continuous-time random walk from each vertex of a finite, connected graph G. The walkers evolve independently, except that when two walkers *meet*—that is, lie on the same vertex at the same time—they coalesce into one. One may easily show that there will almost surely be a finite time at which only one walk will remain in this system. The first such time is called the *full coalescence time* for G and is denoted by C.

The main goal of this paper is to show that one can estimate the law of C for a large family of graphs G, and that this law only depends on G through a single rescaling parameter. More precisely, we will prove results of the following form: if the *mixing time* t^G_{mix} of G (defined in Section 2) is "small," then there exists a parameter $\mathsf{m}(G) > 0$ such that the law $\mathsf{C}/\mathsf{m}(G)$ takes a universal shape. Slight extensions of these results will be used to study the so-called *voter model consensus time* on G.

Received September 2011; revised June 2012.

[1]Supported by a *Universal* grant and a *Bolsa de Produtividade em Pesquisa* from CNPq, Brazil.

MSC2010 subject classifications. Primary 60K35; secondary 60J27.

Key words and phrases. Coalescing random walks, voter model, hitting times, exponential approximation.

The universal shape of $\mathsf{C}/\mathsf{m}(G)$ comes from a *mean field* computation over a large complete graph K_n. In this case the distribution of C can be computed exactly (cf. [2], Chapter 14),

$$\frac{\mathsf{C}}{(n-1)/2} =_d \sum_{i=2}^{n} \mathsf{Z}_i,$$

where:

$$\text{(1.1)} \quad \text{The } \mathsf{Z}_i\text{'s are independent and } \forall i \geq 2, t \geq 0 \qquad \mathbf{P}(\mathsf{Z}_i \geq t) = e^{-t\binom{i}{2}}.$$

In words, C is a rescaled sum of independent exponential random variables with means $1/\binom{i}{2}$, $2 \leq i \leq n$.

The scaling factor $(n-1)/2$ is the expected meeting time of two independent random walks over K_n, and we see that

$$\frac{\mathsf{C}}{(n-1)/2} \to_w \sum_{i\geq 2} \mathsf{Z}_i \quad \text{and} \quad \frac{\mathbf{E}[\mathsf{C}]}{(n-1)/2} \to 2 \qquad \text{when } n \text{ grows.}$$

This suggests the general problem we address in this paper:

PROBLEM 1.1. *Given a graph G, let $\mathsf{m}(G)$ denote the expected meeting time of two independent random walks over G, both started from stationarity. Give sufficient conditons on G under which C has mean-field behavior, that is,*

$$\text{(1.2)} \qquad \mathrm{Law}\big(\mathsf{C}/\mathsf{m}(G)\big) \approx \mathrm{Law}\left(\sum_{i\geq 2} \mathsf{Z}_i\right)$$

and

$$\text{(1.3)} \qquad \mathbf{E}[\mathsf{C}] \approx \mathsf{m}(G)\mathbf{E}\left[\sum_{i\geq 2} \mathsf{Z}_i\right] = 2\mathsf{m}(G).$$

A version of this problem was posed in Aldous and Fill's 1994 draft [2], Chapter 14, and much more recently by Aldous [1]. However, as far as we know there are only two families of examples where the problem has been fully solved. Discrete tori $G = (\mathbf{Z}/m\mathbf{Z})^d$ with with $d \geq 2$ fixed and $m \gg 1$ were considered in Cox's 1989 paper [7]. More recently, Cooper, Frieze and Radzik [6] proved mean field behavior in large random d-regular graphs (d bounded). Partial results were also obtained by Durrett [8, 9] for certain models of large networks.

We note that mean-field behavior is not universal over all large graphs. One counterexample comes from a sequence of growing cycles, where the limiting law of C was also computed by Cox [7]. Stars with n vertices are also not mean field: C is lower bounded by the time the last edge of the star is crossed by some walker, which is about $\log n$, whereas $\mathsf{m}(G)$ is uniformly bounded.

1.1. *Results for transitive, reversible chains.* Our results in this paper address (1.2) and (1.3) simultaneously by proving approximation bounds in L_1 Wasserstein distance, which implies closeness of first moments; cf. Section 2.2.

The first theorem implies that mean field behavior occurs whenever G is vertex-transitive, and its *mixing time* (defined in Section 2) is much smaller than $\mathsf{m}(G)$. This nearly solves a problem posed by Aldous and Fill in [2], Chapter 14. In their open Problem 12, they ask for an analogous result with the relaxation time replacing the mixing time (more on this below).

The natural setting for this first theorem is that of walkers evolving according to the same reversible, transitive Markov chain (the definition of C easily generalizes to this case), where *transitive* means that for any two states x and y one can find a permutation of the state space mapping x to y and leaving the transition rates invariant. Clearly, the standard continuous-time random walk on a vertex-transitive graph is transitive in this sense.

NOTATIONAL CONVENTION 1.1. In this paper we will use "$b = O(a)$" in the following sense: there exist universal constants $C, \xi > 0$ such that $|a| \le \xi \Rightarrow |b| \le C|a|$.

THEOREM 1.1 (Mean field for transitive, reversible chains). *Let Q be the* (*generator of a*) *transitive*, *reversible*, *irreducible Markov chain over a finite state space* $\mathbf{V}$, *with mixing time t^Q_{mix}. Define $\mathsf{m}(Q)$ to be the expected meeting time of two independent continuous-time random walks over $\mathbf{V}$ that evolve according to Q, when both are started from stationarity. Denote by C the full coalescence time for walks evolving according to Q. Finally, define $\{\mathsf{Z}_i\}_{i=2}^{+\infty}$ as in* (1.1). *Then*

$$d_W\left(\mathrm{Law}\left(\frac{\mathsf{C}}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i\ge 2} \mathsf{Z}_i\right)\right) = O\left(\left[\rho(Q)\ln\left(\frac{1}{\rho(Q)}\right)\right]^{1/6}\right),$$

where

$$\rho(Q) \equiv \frac{t^Q_{\mathrm{mix}}}{\mathsf{m}(Q)},$$

and d_W denotes L_1 Wasserstein distance. In particular,

$$\mathbf{E}[\mathsf{C}] = \left\{2 + O\left(\left[\rho(Q)\ln\left(\frac{1}{\rho(Q)}\right)\right]^{1/6}\right)\right\}\mathsf{m}(Q).$$

This result generalizes Cox's theorem [7] for $(\mathbf{Z}/m\mathbf{Z})^d$ with $d \ge 2$ and growing m. In this case, for any fixed d, the mixing time grows as m^2 whereas $\mathsf{m}(G) \approx m^2 \ln m$ for $d = 2$ and $\mathsf{m}(G) \approx m^d$ for larger d. The original problem posed by Aldous and Fill remains open, but we note that:

- For transitive, reversible chains, the mixing time is at most a $C \ln |\mathbf{V}|$ factor away from the relaxation time, with $C > 0$ universal (this is true whenever the stationary distribution is uniform). This means we are not too far off from a full solution;
- Any counterexample to their problem would have to come from a vertex-transitive graph with mixing time of the order of $\mathsf{m}(G)$ *and* relaxation time asymptotically smaller than the mixing time. To the best of our knowledge, such an object is not known to exist.

1.2. *Results for other chains.* We also have results on coalescing random walks evolving according to arbitrary generators Q on finite state spaces $\mathbf{V}$. Again, we only require that the mixing time t_{mix}^{Q} of Q be sufficiently small relative to other parameters of the chain.

THEOREM 1.2 (Mean field for general Markov chains). *Let Q denote (the generator of) a mixing Markov chain over a finite set $\mathbf{V}$, with unique stationary distribution π. Denote by $q_{\max}$ the maximum transition rate from any $x \in \mathbf{V}$ and by $\pi_{\max}$ the maximum stationary probability of an element of $\mathbf{V}$. Let $\mathsf{m}(Q)$ denote the expected meeting time of two random walks evolving according to Q, both started from π. Finally, let C denote the full coalescence time of random walks evolving in $\mathbf{V}$ according to Q. Then*

$$d_W\left(\mathrm{Law}\left(\frac{\mathsf{C}}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i\geq 2} \mathsf{Z}_i\right)\right) = O\left(\left(\alpha(Q)\ln\left(\frac{1}{\alpha(Q)}\right)\ln^4|\mathbf{V}|\right)^{1/6}\right),$$

where

$$\alpha(Q) = \left(1 + q_{\max} t_{\mathrm{mix}}^{Q}\right)\pi_{\max},$$

and d_W again denotes L^1 Wasserstein distance. In particular,

$$\mathbf{E}[\mathsf{C}] = \left\{2 + O\left(\left[\alpha(Q)\ln\left(\frac{1}{\alpha(Q)}\right)\ln^4|\mathbf{V}|\right]^{1/6}\right)\right\}\mathsf{m}(Q).$$

We note that this theorem does *not* imply Theorem 1.1: for instance, it does *not* work for two-dimensional discrete tori. However, the well-known formula for π over graphs gives the following corollary:

COROLLARY 1.1 (Proof omitted). *Assume G is a connected graph with vertex set $\mathbf{V}$, where each vertex $x \in \mathbf{V}$ has degree $\deg_G(x)$. Assume that $\varepsilon \in (|\mathbf{V}|^{-1}, 1)$ is such that*

$$\left(\frac{\max_{x\in\mathbf{V}} \deg_G(x)}{|\mathbf{V}|^{-1}\sum_{x\in\mathbf{V}} \deg_G(x)}\right) t_{\mathrm{mix}}^{G} \leq \frac{\varepsilon|\mathbf{V}|}{\ln^4|\mathbf{V}|\ln\ln|\mathbf{V}|}.$$

Then

$$d_W\left(\mathrm{Law}\left(\frac{\mathsf{C}}{\mathsf{m}(G)}\right), \mathrm{Law}\left(\sum_{i\geq 2} \mathsf{Z}_i\right)\right) = O\left(\left[\varepsilon\left(1+\frac{\ln(1/\varepsilon)}{\ln\ln|\mathbf{V}|}\right)\right]^{1/6}\right).$$

This corollary suffices to prove mean field behavior over a variety of examples, such as:

- *all graphs with bounded ratio of maximal to average degree and mixing time at most of the order* $|\mathbf{V}|/\ln^5|\mathbf{V}|$: this includes expanders [6] and supercritical percolation clusters in $(\mathbf{Z}/m\mathbf{Z})^d$ with $d \geq 3$ fixed [5, 15];
- *all graphs with maximal degree* $\leq |\mathbf{V}|^{1-\eta}$ *(*$\eta > 0$ *fixed) and mixing time that is polylogarithmic in* $|\mathbf{V}|$: this includes the giant component of a typical Erdös–Rényi graph $G_{n,d/n}$ with $d > 1$ [10] and the models of large networks considered by Durrett [8, 9].

Let us briefly comment on the case of large networks. Durrett has estimated $\mathsf{m}(G)$ in these models, and has proven results similar to ours for a bounded number of walkers. We do not attempt to compute $\mathsf{m}(G)$ here, which in general is a model-specific parameter. However, we do show that mean field behavior for C follows from "generic" assumptions about networks that hold for many different models. This is important because recent measurements of real-life social networks [11] suggest that known models of large networks are very inaccurate with respect to most network characteristics outside of degree distributions and conductance. In fairness, coalescing random walks and voter models over large networks are not particularly realistic either, but at the very least we know that mean field behavior is not an artifact of a particular class of models. We also observe that our Theorem 1.2 also works for nonreversible chains, for example, random walks on directed graphs.

1.3. *Results for the voter model.* The voter model is a very well-known process in the interacting particle systems literature [13]. The configuration space for the voter model is the power set $\mathcal{O}^{\mathbf{V}}$ of functions $\eta : \mathbf{V} \to \mathcal{O}$, where $\mathbf{V}$ is some nonempty set, and $\mathcal{O}$ is a nonempty set of possible opinions. The evolution of the process is determined by numbers $q(x, y)$ $(x, y \in V, x \neq y)$ and is informally described as follows: at rate $q(x, y)$, node x copies y's opinion. That is, there is a transition at rate $q(x, y)$ from any state $\eta : \mathbf{V} \to \mathcal{O}$ to the corresponding state $\eta^{x\leftarrow y}$, where

$$\eta^{x\leftarrow y}(z) = \begin{cases} \eta(y), & \text{if } z = x; \\ \eta(z), & \text{for all other } z \in V \setminus \{x\}. \end{cases}$$

A classical duality result relates this voter model to a system of coalescing random walks with transition rates $q(\cdot, \cdot\cdot)$ and corresponding generator Q. More precisely, suppose that $\mathbf{V} = \{x(1), \dots, x(n)\}$ and that $(\overline{X}_t(i))_{t\geq 0, 1\leq i\leq n}$ is a system of coalescing random walks evolving according to Q with $X_0(i) = x(i)$ for each $1 \le i \le n$.

PROPOSITION 1.1 (Duality [2]). *Choose $\eta_0 \in \mathcal{O}^{\mathbf{V}}$. Then the configuration*

$$\hat{\eta}_t : x(i) \in \mathbf{V} \mapsto \eta_0(\overline{X}_t(i)) \in \mathcal{O} \qquad (1 \le i \le n)$$

has the same distribution as the state η_t of the voter model at time t, when the initial state is η_0. In particular, the consensus time *for the voter model*

$$\tau \equiv \inf\{t \ge 0 : \forall i, j \in \mathbf{V}, \eta_t(i) = \eta_t(j)\}$$

satisfies $\mathbf{E}[\tau] \le \mathbf{E}[\mathsf{C}] < +\infty$.

Now assume that the initial state $\eta_0 \in \mathcal{O}^{\mathbf{V}}$ is random and that the random variables $\{\eta_0(x)\}_{x \in V}$ are i.i.d. and have common law μ which is not a point mass. In this case one can show via duality that the law of the consensus time τ is that of $\mathsf{C}_{K \wedge n}$, where K is a $\mathbf{N}$-valued random variable independent of the coalescing random walks, defined by

$$K = \min\{i \in \mathbf{N} : U_{i+1} \ne U_1\} \qquad \text{where } U_1, U_2, U_3, \dots, \text{ are i.i.d. draws from } \mu,$$

and for each $1 \le k \le n$,

$$\mathsf{C}_k \equiv \min\{t \ge 0 : |\{\overline{X}_t(i) : 1 \le i \le n\}| = k\}.$$

Thus the key step in analyzing the voter model via our techniques is to prove approximations for the distribution of C_k. Theorems 1.1 and 1.2 imply mean-field behavior for $\mathsf{C} = \mathsf{C}_1$. A quick inspection of the proofs reveals that the same bounds for Wasserstein distance can be obtained for C_k for any $1 \le k \le n$. It follows that:

THEOREM 1.3 (Proof omitted). *Let $\mathbf{V}$, $\mathcal{O}$ and μ be as above, and consider the voter model defined by $\mathbf{V}$, $\mathcal{O}$ and by the generator Q corresponding to transition rates $q(x, y)$. Assume that the sequence $\{\mathsf{Z}_i\}_{i \ge 2}$ is defined as in* (1.1), *and also that K has the law described above and is independent from the Z_i. Define $\rho(Q)$ and $\alpha(Q)$ as in Theorems 1.1 and 1.2. Then the consensus time τ for this voter model satisfies*

$$d_W\left(\mathrm{Law}\left(\frac{\tau}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i > K} Z_i\right)\right) = O\big((\rho(Q)\ln(1/\rho(Q)))^{1/6}\big)$$

if Q is reversible and transitive, and

$$d_W\left(\mathrm{Law}\left(\frac{\tau}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i > K} Z_i\right)\right) = O\big((\alpha(Q)\ln(1/\alpha(Q))\ln^4|\mathbf{V}|)^{1/6}\big),$$

otherwise.

1.4. *Main proof ideas*. Our proofs of Theorems 1.1 and 1.2 both start from the formula (1.1) for the terms in the distribution of C over K_n. Crucially, each term Z_i has a specific meaning: Z_i is the time it takes for a system with i particles to evolve to a system with $i-1$ particles, rescaled by the expected meeting time of two walkers. For $i=2$, this is just the (rescaled) meeting time of a pair of particles, which is an exponential random variable with mean 1. For $i>2$, we are looking at the first meeting time among $\binom{i}{2}$ pairs of particles. It turns out that these pairwise meeting times are independent; since the minimum of k independent exponential random variables with mean μ is an exponential r.v. with mean μ/k, we deduce that Z_i is exponential with mean $1/\binom{i}{2}$.

The bulk of our proof consists of proving something similar for more general chains Q. Fix some such Q, with state space $\mathbf{V}$, and let C_i denote the time it takes for a system of coalescing random walks evolving according to Q to have i uncoalesced particles. Clearly, $M \equiv \mathsf{C}_1 - \mathsf{C}_2$ is the meeting time of a pair of particles, which is the hitting time of the diagonal set

$$\Delta \equiv \{(x,x) : x \in \mathbf{V}\}$$

by the Markov chain $Q^{(2)}$ given by a pair of independent realizations of Q. More generally, $M^{(i+1)} = \mathsf{C}_i - \mathsf{C}_{i+1}$ is the hitting time of

$$\Delta^{(i+1)} = \{(x(1),\dots,x(i+1)) : \exists 1 \le i_1 < i_2 \le i+1, x(i_1) = x(i_2)\}.$$

The mean-field picture suggests that each $M^{(i+1)}$ should be close in distribution to Z_i. Indeed, it is known that:

General principle: Let H_A be the hitting time of a subset A of states. If the mixing time t_{mix}^{Q} is small relative to $\mathbf{E}[H_A]$, then H_A is approximately exponentially distributed.

This is a general meta-result for small subsets of the state space of a Markov chain; precise versions (with different quantitative bounds) are proven in [3, 4] when the chain starts from the stationary distribution. However, we face a few difficulties when trying to use these off-the-shelf results:

(1) For each i, $M^{(i+1)}$ is the first hitting time of $\Delta^{(i+1)}$ after time C_{i+1}. The random walkers are *not* stationary at this random time, so we need to "do" exponential approximation from nonstationary starting points.

(2) In order to get Wasserstein approximations, we need better control of the tail of $M^{(i+1)}$.

(3) To prove that Z_i and $M^{(i+1)}/\mathsf{m}(Q)$ are close, we must show something like that $\mathbf{E}[M^{(i+1)}] \approx \mathbf{E}[M]/\binom{i+1}{2}$, that is, that $M^{(i+1)}$ behaves like the minimum of $\binom{i+1}{2}$ independent exponentials.

(4) Finally, we should not expect the exponential approximation to hold when $\Delta^{(i+1)}$ is too large. That means that the "big bang" phase (to use Durrett's phrase) at the beginning of the process has to be controlled by other means.

It turns out that we can deal with points 1 and 2 via a different kind exponential approximation result, stated as Theorem 3.1. This result will give bounds of the following form:

$$\mathbf{P}_x(H_A > t) = \big(1 + o(1)\big) \exp\left(-\frac{t}{(1 + o(1))\mathbf{E}[H_A]}\right) \tag{1.4}$$

as long as

$$t^Q_{\text{mix}} = o\big(\mathbf{E}[H_A]\big) \quad \text{and} \quad \mathbf{P}_x\big(H_A \le t^Q_{\text{mix}}\big) = o(1).$$

Notice that this holds even for *nonstationary* starting points x if the chain started from x is unlikely to hit A before the mixing time. This is discussed in Section 3 below. We also take some time in that section to develop a specific notion of "near exponential random variable." Although this takes up some space, we believe it provides a useful framework for tackling other problems. We note that a version of Theorem 3.1 for stationary initial states result is implicit in [3].

We now turn to point 3. The key difficulty in our setting is that, unlike Cox [7] or Cooper et al. [6], we do not have a good "local" description of the graphs under consideration which we could use to compute $\mathbf{E}[M^{(i+1)}]$ directly. We use instead a simple general idea, which we believe to be new, to address this point. Clearly, $M^{(i+1)}$ is a minimum of $\binom{i+1}{2}$ hitting times. Let us consider the *general* problem of understanding the law of

$$H_B = \min_{1\le i\le \ell} H_{B_i} \qquad \text{where } B = \bigcup_{i=1}^{\ell} B_i,$$

under the assumption that $\mathbf{E}[H_{B_i}] = \mu$ does not depend on i when the initial distribution is stationary (this covers the case of $M^{(i+1)}$). Assume also that (1.4) holds for all $A \in \{B, B_1, B_2, \dots, B_\ell\}$. Then the following holds for ε in a suitable range:

$$\forall A \in \{B, B_1, B_2, \dots, B_\ell\} \qquad \mathbf{P}\big(H_A \le \varepsilon \mathbf{E}[H_A]\big) \approx \varepsilon.$$

Morally speaking, this means that $\varepsilon\mathbf{E}[H_A]$ is the ε-quantile of H_A for all A as above; this is implicit in [3] and is made explicit in our own Theorem 3.1. Now apply this to $A = B$, with ε replaced by $\varepsilon\mu/\mathbf{E}[H_B]$, and obtain

$$\frac{\varepsilon\mu}{\mathbf{E}[H_B]} \approx \mathbf{P}(H_B \le \varepsilon\mu) = \mathbf{P}\left(\bigcup_{i=1}^{\ell}\{H_{B_i} \le \varepsilon\mu\}\right).$$

If we can show that the pairwise correlations between the events $\{H_{B_i} \le \varepsilon\mu\}$ are sufficiently small, then we may obtain

$$\frac{\varepsilon\mu}{\mathbf{E}[H_B]} \approx \mathbf{P}\left(\bigcup_{i=1}^{\ell}\{H_{B_i} \le \varepsilon\mu\}\right) \approx \sum_{i=1}^{\ell} \mathbf{P}(H_{B_i} \le \varepsilon\mu) = \ell\varepsilon.$$

This gives

$$\mathbf{E}[H_B] \approx \frac{\mu}{\ell}$$

as if the times $H_{B_1}, \dots, H_{B_\ell}$ were independent exponentials. The reasoning presented here is made rigorous and quantitative in Theorem 3.2 below.

Finally, we need to take care of point 4, that is, the "big bang" phase. In the setting of Theorem 1.2, we simply use our results on the coalescence times for smaller number of particles, which seems wasteful but is enough to prove our results. For the reversible/transitive case, we use a bound from [14] which is of the optimal order. Incidentally, the differences in the bounds of the two theorems come from this better bound for the big bang phase and from a more precise control of the correlations between meeting times of different pairs of walkers.

1.5. *Outline*. The remainder of the paper is organized as follows. Section 2 contains several preliminaries. Section 3 contains a general discussion of random variables with nearly exponential distribution and our general approximation results for hitting times. In Section 4 we apply these results to the first meeting time among k particles, after proving some technical estimates. Section 5 contains the formal definition of the coalescing random walks process and proves mean field behavior for a moderate initial number of walkers. Finally, Section 6 contains the proofs of Theorems 1.1 and 1.2. Related results and open problems are discussed in the final sections.

2. Preliminaries.

2.1. *Basic notation*. We write $\mathbf{N}$ for nonnegative integers and $[k] = \{1, 2, \dots, k\}$ for any $k \in \mathbf{N} \setminus \{0\}$. Given a set S, we let $|S|$ denote its cardinality. Moreover, for $k \in \mathbf{N}$, we let

$$\binom{S}{k} \equiv \{A \subset S : |A| = k\}.$$

Notice that with this notation,

$$|S| \text{ finite} \Rightarrow \left|\binom{S}{k}\right| = \binom{|S|}{k} \equiv \frac{|S|!}{k!(|S|-k)!}.$$

We will often speak of universal constants $C > 0$. These are numbers that do not depend on any of the parameters or mathematical objects under consideration in a given problem. We will also use the notation "$a = O(b)$" in the universal sense prescribed in Notational convention 1.1. In this way we can write down expressions such as

$$e^b = 1 + b + O(b^2) \quad \text{and} \quad \ln\left(\frac{1}{1-b}\right) = b + O(b^2) = O(b).$$

Given a finite set S, we let $M_1(S)$ denote the set of all probability measures over S. Given $p, q \in M_1(S)$, their total variation disance is defined as follows:

$$d_{\mathrm{TV}}(p,q) \equiv \frac{1}{2}\sum_{s\in S}\bigl|p(s)-q(s)\bigr| = \sup_{A\subset S}\bigl[p(A)-q(A)\bigr],$$

where $p(A)=\sum_{a\in A} p(a)$. For S not finite, $M_1(S)$ will denote the set of all probability measures over the "natural" σ-field over S. For instance, for $S=\mathbf{R}$ we consider the Borel σ-field, and for $S=\mathbf{D}([0,+\infty),\mathbf{V})$ (see Section 2.3.1 for a definition) we use the σ-field generated by projections.

If X is a random variable taking values over S, we let $\mathrm{Law}(X)\in M_1(S)$ denote the distribution (or law) of X. Here we again assume that there is a "natural" σ-field to work with.

2.2. *Wasserstein distance.* The L_1 *Wasserstein distance* is a metric over probability measures over $\mathbf{R}$ with finite first moments, given by

$$d_W(\lambda_1,\lambda_2) = \int_{\mathbf{R}}\bigl|\lambda_1(x,+\infty]-\lambda_2(x,+\infty]\bigr|\,dx \qquad \bigl(\lambda_1,\lambda_2\in M_1(\mathbf{R})\bigr).$$

A classical duality result gives

$$d_W(\lambda_1,\lambda_2) = \sup_{f:\mathbf{R}\to\mathbf{R}\,1\text{-Lipschitz}}\biggl(\int_{\mathbf{R}} f(x)\lambda_1(dx) - \int_{\mathbf{R}} f(x)\lambda_2(dx)\biggr).$$

NOTATIONAL CONVENTION 2.1. Whenever we compute Wasserstein distances, we will *assume* that the distributions involved have first moments. This can be checked in each particular case.

REMARK 2.1. If Z_1, Z_2 are random variables, we sometimes write

$$d_W(Z_1,Z_2) \text{ instead of } d_W\bigl(\mathrm{Law}(Z_1),\mathrm{Law}(Z_2)\bigr).$$

Note that

$$d_W(Z_1,Z_2) = \int_{\mathbf{R}}\bigl|\mathbf{P}(Z_1\ge t)-\mathbf{P}(Z_2\ge t)\bigr|\,dt.$$

Also notice that

$$\bigl|\mathbf{E}[Z_1]-\mathbf{E}[Z_2]\bigr| \le d_W(Z_1,Z_2).$$

This is an equality if $Z_1\ge 0$ a.s. and $Z_2 = CZ_1$ for some constant $C>0$,

$$\forall C\in\mathbf{R} \qquad d_W(Z_1,CZ_1) = |C-1|\mathbf{E}[Z_1], \tag{2.1}$$

since $|f(CZ_1)-f(Z_1)|\le |C-1|Z_1$ for every 1-Lipschitz function $f:\mathbf{R}\to\mathbf{R}$.

We note here three useful lemmas on Wasserstein distance. These are probably standard, but we could not find references for them, so we provide proofs for the latter two lemmas in Section 7 of the Appendix. The first lemma is immediate.

LEMMA 2.1 (Sum lemma for Wasserstein distance; Proof omitted). *For any two random variables X, Y with finite first moments and defined on the same probability space*,

$$d_W(X, X+Y) \le \mathbf{E}\big[|Y|\big].$$

For the next lemma, recall that, given two real-valued random variables X, Y, we say that X is stochastically dominated by Y and write $X \preceq_d Y$ if $\mathbf{P}(X > t) \le \mathbf{P}(Y > t)$ for all $t \in \mathbf{R}$.

LEMMA 2.2 (Sandwich lemma for Wasserstein distance). *Let Z, Z_-, Z_+ and W be real-valued random variables with finite first moments and $Z_- \preceq_d Z \preceq_d Z_+$. Then*

$$d_W(Z, W) \le d_W(Z_-, W) + d_W(Z_+, W).$$

LEMMA 2.3 (Conditional lemma for Wasserstein distance). *Let W_1, W_2, Z_1, Z_2 be real-valued random variables with finite first moments. Assume that Z_1 and Z_2 independent and that W_1 is $\mathcal{G}$-measurable for some sub-σ-field $\mathcal{G}$. Then*

$$\begin{aligned} &d_W\big(\mathrm{Law}(W_1 + W_2), \mathrm{Law}(Z_1 + Z_2)\big) \\ &\quad \le d_W\big(\mathrm{Law}(W_1), \mathrm{Law}(Z_1)\big) + \mathbf{E}\big[d_W\big(\mathrm{Law}(W_2 \mid \mathcal{G}), \mathrm{Law}(Z_2)\big)\big]. \end{aligned}$$

REMARK 2.2. Here we are implicitly assuming that $\mathrm{Law}(W_2 \mid \mathcal{G})$ is given by some regular conditional probability distribution.

2.3. *Continuous-time Markov chains*.

2.3.1. *State space and trajectories*. Let $\mathbf{V}$ be some nonempty finite set, called the *state space*. We write $\mathbf{D} \equiv \mathbf{D}([0, +\infty), \mathbf{V})$ for the set of all paths

$$\omega : t \ge 0 \mapsto \omega_t \in \mathbf{V}$$

for which there exist $0 = t_0 < t_1 < t_2 < \cdots < t_n < \cdots$ with $t_n \nearrow +\infty$ and ω constant over each interval $[t_n, t_{n+1})$ ($n \in \mathbf{N}$). Such paths will sometimes be called *càdlàg*.

For each $t \ge 0$, we let $X_t : \mathbf{D} \to \mathbf{V}$ be the projection map sending ω to ω_t. We also define $X = (X_t)_{t \ge 0}$ as the identity map over $\mathbf{D}$. Whenever we speak about probability measures and events over $\mathbf{D}$, we will implicitly use the σ-field $\sigma(\mathbf{D})$ generated by the maps X_t, $t \ge 0$. We define an associated filtration as follows:

$$\mathcal{F}_t \equiv \sigma\{X_s : 0 \le s \le t\} \qquad (t \ge 0).$$

We also define the time-shift operators

$$\Theta_T : \omega(\cdot) \in \mathbf{D} \mapsto \omega(\cdot + T) \in \mathbf{D} \qquad (T \ge 0).$$

2.3.2. *Markov chains and their generators*. Let $q(x,y)$ be nonnegative real numbers for each pair $(x,y)\in\mathbf{V}^2$ with $x\neq y$. Define a linear operator $Q:\mathbf{R}^{\mathbf{V}}\to\mathbf{R}^{\mathbf{V}}$, which maps $f\in\mathbf{R}^{\mathbf{V}}$ to $Qf\in\mathbf{R}^{\mathbf{V}}$ satisfying

$$(Qf)(x)\equiv\sum_{y\in\mathbf{V}\backslash\{x\}}q(x,y)\big(f(x)-f(y)\big)\qquad(x\in\mathbf{V}).$$

It is a well-known result that there exists a unique family of probability measures $\{\mathbf{P}_x\}_{x\in\mathbf{V}}$ with the properties listed below:

(1) for all $x\in\mathbf{V}$, $\mathbf{P}_x(X_0=x)=1$;
(2) for all distinct $x,y\in\mathbf{V}$, $\lim_{\varepsilon\searrow 0}\frac{\mathbf{P}_x(X_\varepsilon=y)}{\varepsilon}=q(x,y)$;
(3) *Markov property*: for any $x\in\mathbf{V}$ and $T\geq 0$, the conditional law of $X\circ\Theta_T$ given $\mathcal{F}_T$ under measure $\mathbf{P}_x$ is given by $\mathbf{P}_{X_T}$.

The family $\{\mathbf{P}_x\}_{x\in\mathbf{V}}$ satisfying these properties is the *Markov chain with generator Q*. We will often abuse notation and omit any distinction between a Markov chain and its generator in our notation.

For $\lambda\in M_1(\mathbf{V})$, $\mathbf{P}_\lambda$ denotes the mixture

$$\mathbf{P}_\lambda\equiv\sum_{x\in\mathbf{V}}\lambda(x)\mathbf{P}_x.$$

This corresponds to starting the process from a random state distributed according to λ. For $x\in\mathbf{V}$ or $\lambda\in M_1(\mathbf{V})$ and $Y:\mathbf{D}\to S$ a random variable, we let $\mathrm{Law}_x(Y)$ or $\mathrm{Law}_\lambda(Y)$ denote the law of Y under $\mathbf{P}_x$ or $\mathbf{P}_\lambda$ (resp.).

2.3.3. *Stationary measures and mixing*. Any Markov chain Q as above has at least one stationary measure $\pi\in M_1(\mathbf{V})$; this is a measure such that for any $T\geq 0$,

$$\mathrm{Law}_\pi(X\circ\Theta_T)=\mathrm{Law}_\pi(X).$$

We will be only interested in *mixing Markov chains*, which are those Q with a unique stationary measure that satisfy the following condition:

$$\forall\alpha\in(0,1),\exists T\geq 0,\forall x\in\mathbf{V}\qquad d_{\mathrm{TV}}\big(\mathrm{Law}(x)X_T,\pi\big)\leq\alpha.$$

The smallest such T is called the α-mixing time of Q and is denoted by $t_{\mathrm{mix}}^Q(\alpha)$. By the Markov property and the definition of total-variation distance, we also have that for all $\alpha\in(0,1)$, all $t\geq t_{\mathrm{mix}}^Q(\alpha)$, all $x\in\mathbf{V}$ and all events S,

$$\big|\mathbf{P}_x(X\circ\Theta_t\in S)-\mathbf{P}_\pi(X\in S)\big|\leq\alpha.$$

The specific value $t_{\mathrm{mix}}^Q\equiv t_{\mathrm{mix}}^Q(1/4)$ is called *the* mixing time of Q. We note that for all $\varepsilon\in(0,1/2)$,

$$t_{\mathrm{mix}}^Q(\varepsilon)\leq C\ln(1/\varepsilon)t_{\mathrm{mix}}^Q,\tag{2.2}$$

where $C>0$ is universal; this is proven in [12], Section 4.5, for discrete time chains, but the same argument works here.

2.3.4. *Product chains.* Letting Q be as above, we may consider the joint trajectory of k independent realizations of Q,

$$X_t^{(k)} = (X_t(1), \dots, X_t(k)) \qquad (t \ge 0)$$

where each $(X_t(i))_{t \ge 0}$ has law $\mathbf{P}_{x(i)}$. It turns out that this corresponds to a Markov chain $Q^{(k)}$ on $\mathbf{V}^k$ with transition probabilities

$$q^{(k)}(x^{(k)}, y^{(k)}) = \begin{cases} q(x(i), y(i)), & \text{if } x(i) \neq y(i) \wedge \forall j \in [k] \setminus \{i\}, x(j) = y(j); \\ 0, & \text{otherwise.} \end{cases}$$

REMARK 2.3. In what follows we will always denote elements of $\mathbf{V}^k$ [resp., $M_1(\mathbf{V}^k)$] by symbols like $x^{(k)}, y^{(k)}, \dots$ (resp., $\lambda^{(k)}, \rho^{(k)}, \dots$). We will then denote the distribution of $Q^{(k)}$ started from $x^{(k)}$ or $\lambda^{(k)}$ by $\mathbf{P}_{x^{(k)}}$ or $\mathbf{P}_{\lambda^{(k)}}$. This is a slight abuse of our convention for the Q chain, but the initial state/distribution will always make it clear that we are referring to the product chain.

The following result on $Q^{(k)}$ will often be useful.

LEMMA 2.4. *Assume Q is mixing and has (unique) stationary distribution π. Then $Q^{(k)}$ is also mixing, and the product measure $\pi^{\otimes k}$ is its (unique) stationary distribution. Moreover, the mixing times of $Q^{(k)}$ satisfy*

$$\forall \alpha \in (0, 1/2) \qquad t_{\mathrm{mix}}^{Q^{(k)}}(\alpha) \le t_{\mathrm{mix}}^{Q}(\alpha/k) \le C \ln(k/\alpha) t_{\mathrm{mix}}^{Q}$$

with $C > 0$ universal.

PROOF SKETCH. Notice that the law of $X_T^{(k)}$ has a product form

$$\mathrm{Law}_{x^{(k)}}(X_T^{(k)}) = \mathrm{Law}_{x(1)}(X_T) \otimes \mathrm{Law}_{x(2)}(X_T) \otimes \cdots \otimes \mathrm{Law}_{x(k)}(X_T).$$

It is well known (and not hard to show) that the total-variation distance between product measures is at most the sum of the distances of the factors. This gives

$$d_{\mathrm{TV}}(\mathrm{Law}_{x^{(k)}}(X_T^{(k)}), \pi^{\otimes k}) \le \sum_{i=1}^{k} d_{\mathrm{TV}}(\mathrm{Law}_{x(i)}(X_T), \pi).$$

The RHS is $\le \alpha$ if each term in the sum is less than α/k. This is achieved when $T \ge t_{\mathrm{mix}}^{Q}(\alpha/k)$; (2.2) then finishes the proof. □

3. Nearly exponential hitting times.

3.1. *Basic definitions*. We first recall a standard definition: the *exponential distribution* with mean $m > 0$, denoted by $\mathsf{Exp}(m)$, is the unique probabilty dstribution $\mu \in M_1(\mathbf{R})$ such that, if Z is a random variable with law μ,

$$\mathbf{P}(Z \geq t) = e^{-t/m} \qquad (t \geq 0).$$

We write $Z =_d \mathsf{Exp}(m)$ when Z is a random variable with $\mathrm{Law}(Z) = \mathsf{Exp}(m)$.

Similarly, given $m > 0$ as above and parameters $\alpha > 0$, $\beta \in (0, 1)$, we say that a measure $\mu \in M_1(\mathbf{R})$ has distribution $\mathsf{Exp}(m, \alpha, \beta)$ if it is the law of a random variable $\widetilde{Z}$ with $\widetilde{Z} \geq 0$ almost surely, and for all $t > 0$,

$$(1 - \alpha)e^{-t/((1-\beta)m)} \leq \mathbf{P}(\widetilde{Z} \geq t) \leq (1 + \alpha)e^{-t/((1+\beta)m)}.$$

We will write $\mu = \mathsf{Exp}(m, \alpha, \beta)$ or $\widetilde{Z} =_d \mathsf{Exp}(m, \alpha, \beta)$ as a shorthand for this. Notice that $\mathsf{Exp}(m, \alpha, \beta)$ does not denote a single distribution, but rather a family of distributions that obey the above property, but we will mostly neglect this minor issue.

Random variables with law $\mathsf{Exp}(m, \alpha, \beta)$ will naturally appear in our study of hitting times of Markov chains. We compile here some simple results about them. The first proposition is trivial and we omit its proof.

PROPOSITION 3.1 (Proof omitted). *If* $\mu \in M_1(\mathbf{R})$ *satisfies*

$$\mu = \mathsf{Exp}(m, \alpha, \beta),$$

and $m' > 0$, $\gamma \in (0, 1)$ *are such that* $\beta + \gamma + \beta\gamma < 1$,

$$(1 - \gamma)m' \leq m \leq (1 + \gamma)m',$$

then

$$\mu = \mathsf{Exp}\big(m', \alpha, \beta + \gamma + \beta\gamma\big).$$

We now show that random variables $\mathsf{Exp}(m, \alpha, \beta)$ are close to the corresponding exponentials.

LEMMA 3.1 [Wasserstein distance error for $\mathsf{Exp}(m, \alpha, \beta)$]. *We have the following inequality for all* $\alpha > 0$, $0 < \beta < 1$:

$$d_W\big(\mathsf{Exp}(m), \mathsf{Exp}(m, \alpha, \beta)\big) \leq 2(\alpha + \beta)m.$$

That is, if $\widetilde{Z} =_d \mathsf{Exp}(m, \alpha, \beta)$, *the Wasserstein distance between* $\mathrm{Law}(\widetilde{Z})$ *and* $\mathsf{Exp}(m)$ *is at most* $2\alpha m + 2\beta m$.

PROOF. Assume $\widetilde{Z} =_d \mathsf{Exp}(m, \alpha, \beta)$ and $Z =_d \mathsf{Exp}(m)$ are given. By convexity,

$$\begin{aligned}
d_W(\widetilde{Z}, Z) &= \int_0^{+\infty} \left|\mathbf{P}(\widetilde{Z} \geq t) - e^{-t/m}\right| dt \\
&\leq \int_0^{\infty} \max_{\xi \in \{-1,+1\}} \left|(1+\xi\alpha)_+ e^{-t/((1+\xi\beta)m)} - e^{-t/m}\right| dt \\
&\leq \int_0^{\infty} \left|(1+\alpha) e^{-t/((1+\beta)m)} - e^{-t/m}\right| dt \\
&\quad + \int_0^{\infty} \left|(1-\alpha)_+ e^{-t/((1-\beta)m)} - e^{-t/m}\right| dt \\
&=: (I) + (II).
\end{aligned}$$

For the first term on the RHS, we note that

$$\forall t \geq 0 \qquad (1+\alpha) e^{-t/((1+\beta)m)} - e^{-t/m} \geq 0,$$

hence

$$(I) = \int_0^{\infty} \left\{(1+\alpha) e^{-t/((1+\beta)m)} - e^{-t/m}\right\} dt = [\alpha + \beta + \alpha\beta] m.$$

Similarly, for term (II) we have

$$\forall t \geq 0 \qquad (1-\alpha)_+ e^{-t/((1-\beta)m)} - e^{-t/m} \leq 0$$

hence

$$(II) = \int_0^{\infty} \left\{e^{-t/m} - (1-\alpha)_+ e^{-t/((1-\beta)m)}\right\} dt \leq [\alpha + \beta - \alpha\beta] m.$$

Hence

$$d_W(\widetilde{Z}, Z) \leq (I) + (II) = 2(\alpha + \beta) m. \qquad \square$$

3.2. *Hitting times are nearly exponential.* In this section we consider a mixing continuous-time Markov chain $\{\mathbf{P}_x\}_{x \in \mathbf{V}}$ with generator Q, taking values over a finite state space $\mathbf{V}$, with unique stationary distribution π. Given a nonempty $A \subset \mathbf{V}$ with $\pi(A) > 0$, we define the *hitting time of* A to be

$$H_A(\omega) \equiv \inf\{t \geq 0 : \omega(t) \in A\} \qquad (\omega \in \mathbf{D}([0, +\infty), \mathbf{V})).$$

The condition $\pi(A) > 0$ ensures that $\mathbf{E}_x[H_A] < +\infty$ for all $x \in \mathbf{V}$.

Our first result in this section presents sufficient conditions on A and $\mu \in M_1(\mathbf{V})$ that ensure that H_A is approximately exponentially distributed.

THEOREM 3.1. *In the above Markov chain setting, assume that* $0 < \varepsilon < \delta < 1/5$ *are such that*

$$\mathbf{P}_\pi\left(H_A \leq t_{\mathrm{mix}}^{Q}(\delta\varepsilon)\right) \leq \delta\varepsilon.$$

Let $t_\varepsilon(A)$ *be the* ε*-quantile of* $\mathrm{Law}_\pi(H_A)$, *that is, the unique number* $t_\varepsilon(A) \in [0, +\infty)$ *with* $\mathbf{P}_\pi(H_A \le t_\varepsilon(A)) = \varepsilon$ [*this is well defined since* $\mathbf{P}_\pi(H_A \le t)$ *is a continuous and strictly increasing function of* t *in our setting*]. *Given* $\lambda \in M_1(\mathbf{V})$, *write*

$$r_\lambda \equiv \mathbf{P}_\lambda\big(H_A \le t^Q_{\mathrm{mix}}(\delta\varepsilon)\big).$$

Then

$$\mathrm{Law}_\lambda(H_A) = \mathsf{Exp}\bigg(\frac{t_\varepsilon(A)}{\varepsilon}, O(\varepsilon) + 2r_\lambda, O(\delta)\bigg).$$

Moreover,

$$\left|\frac{\varepsilon \mathbf{E}_\pi[H_A]}{t_\varepsilon(A)} - 1\right| = O(\delta)$$

and

$$\mathrm{Law}_\lambda(H_A) =_d \mathsf{Exp}\big(\mathbf{E}_\pi[H_A], O(\varepsilon) + 2r_\lambda, O(\delta)\big).$$

We emphasize that results similar to this are not new in the literature [3, 4], but the lower-tail part of our result does not seem to be explicit anywhere. The proof is strongly related to that in [3], but we wish to stress the relationship between the quantile $t_\varepsilon(A)$ and the exponential approximation, which we will need below.

The second result considers what happens when we have an union of events

$$A = A_1 \cup A_2 \cup \cdots \cup A_\ell.$$

As described in the Introduction, we give a sufficient condition under which the hitting time H_A behaves like a minimum of independent exponentials.

THEOREM 3.2. *Assume that the set* A *considered above can be written as*

$$A = \bigcup_{i=1}^{\ell} A_i,$$

where the sets $A_1, \dots, A_\ell$ *are nonempty and*

$$m := \mathbf{E}_\pi[H_{A_1}] = \mathbf{E}_\pi[H_{A_2}] = \cdots = \mathbf{E}_\pi[H_{A_\ell}].$$

Assume $0 < \delta < 1/5$, $0 < \varepsilon < \delta/2\ell$ *are such that for all* $1 \le i \le \ell$,

$$\forall i \in [\ell] \qquad \mathbf{P}_\pi\big(H_{A_i} \le t^Q_{\mathrm{mix}}(\delta\varepsilon/2)\big) \le \frac{\delta\varepsilon}{2}.$$

Then for all $\lambda \in M_1(\mathbf{V})$,

$$\mathrm{Law}_\lambda(H_A) = \mathsf{Exp}\bigg(\frac{m}{\ell}, 2r_\lambda + O(\ell\varepsilon), O(\delta + \xi)\bigg),$$

where

$$r_\lambda \equiv \mathbf{P}_\lambda\big(H_A \le t^Q_{\mathrm{mix}}(\delta\varepsilon)\big)$$

and

$$\xi \equiv \frac{1}{\ell\varepsilon} \sum_{1\le i<j\le \ell} \mathbf{P}_\pi(H_{A_i} \le \varepsilon m, H_{A_j} \le \varepsilon m).$$

REMARK 3.1. If the H_{A_i} are in fact independent, then $\xi = O(\varepsilon\ell)$.

The remainder of the section is devoted to the proof of these two results.

3.3. *Hitting time of a single set*: *Proofs*. We first present the proof of Theorem 3.1 modulo two important lemmas, and subsequently prove those lemmas.

PROOF OF THEOREM 3.1. Let $\lambda \in M_1(\mathbf{V})$ be arbitrary. Throughout the proof we will assume implicitly that $\delta + r_\lambda + \varepsilon$ is smaller than some sufficiently small absolute constant; the remaining case is easy to handle by increasing the value of C_0 if necessary.

We begin with an upper bound for $\mathbf{P}_\lambda(H_A \ge t)$ in terms of $t_\varepsilon(A)$.

LEMMA 3.2 (Proven in Section 3.3.1). *Under the assumptions of Theorem* 3.1,

$$\forall t \ge 0 \qquad \mathbf{P}_\lambda(H_A \ge t) \le \big(1 + O(\varepsilon)\big)e^{-\varepsilon(1+O(\delta))t/t_\varepsilon(A)}.$$

In particular, this implies

$$\forall \mu \in M_1(\mathbf{V}) \qquad \mathbf{E}_\mu[H_A] = \int_0^{+\infty} \mathbf{P}_\mu(H_A \ge t)\,dt \le \big(1 + O(\delta)\big)\frac{t_\varepsilon(A)}{\varepsilon}. \tag{3.1}$$

It turns out that the upper bound in the above lemma can be nearly reversed if we start from some distribution that is "far" from A.

LEMMA 3.3 (Proven in Section 3.3.2). *With the assumptions of Theorem* 3.1, *if* $2\varepsilon + r_\lambda < 1/2$,

$$\forall t \ge 0 \qquad \mathbf{P}_\lambda(H_A \ge t) \ge \big(1 - O(\varepsilon) - r_\lambda\big)_+ e^{-\varepsilon(1+O(\delta))t/t_\varepsilon(A)}.$$

Notice that the combination of these two lemmas already implies the first statement in the proof, as it shows that for all $t \ge 0$,

$$\mathbf{P}_\lambda(H_A \ge t) \\ \in \big[\big(1 - O(\varepsilon) - 2r_\lambda\big)e^{-\varepsilon t/((1+O(\delta))t_\varepsilon(A))}, \big(1 + O(\varepsilon)\big)e^{-\varepsilon t/((1+O(\delta))t_\varepsilon(A))}\big].$$

To see this, notice that the upper bound is always valid by Lemma 3.2. For the lower bound, we use Lemma 3.3 if $2\varepsilon + r_\lambda \le 1/2$, and note that the lower bound is 0 if $2\varepsilon + r_\lambda > 1/2$ and the constant in the $O(\varepsilon)$ term is at least 4.

We now prove the assertion about expectations in the theorem. We use Lemma 3.1 and deduce

$$\left|\mathbf{E}_\pi[H_A] - \frac{t_\varepsilon(A)}{\varepsilon}\right| \le d_W\big(\mathrm{Law}_\pi(H_A), \mathsf{Exp}\big(\varepsilon^{-1}t_\varepsilon(A)\big)\big)$$
$$\le O(\delta + r_\pi)\frac{t_\varepsilon(A)}{\varepsilon},$$

and the assertion follows from dividing by $\varepsilon^{-1}t_\varepsilon(A)$ and noting that

$$r_\pi = \mathbf{P}_\pi\big(H_A \le t^Q_{\mathrm{mix}}(\delta\varepsilon)\big) \le \delta\varepsilon$$

by assumption. The final assertion in the theorem then follows from Proposition 3.1. □

3.3.1. *Proof of Lemma* 3.2.

PROOF. Set $T = t^Q_{\mathrm{mix}}(\delta\varepsilon)$. We note for later reference that $T < t_\varepsilon(A)$, since

$$\mathbf{P}_\pi(H_A \le T) \le \delta\varepsilon < \varepsilon = \mathbf{P}_\pi\big(H_A \le t_\varepsilon(A)\big).$$

Our main goal will be to show the following inequality:

$$\forall k \in \mathbf{N} \qquad \mathbf{P}_\lambda\big(H_A > (k+1)t_\varepsilon(A)\big) \le (1 - \varepsilon + 2\delta\varepsilon)\mathbf{P}_\lambda\big(H_A > kt_\varepsilon(A)\big). \tag{3.2}$$

Once established, this goal will imply

$$\forall k \in \mathbf{N} \qquad \mathbf{P}_\lambda\big(H_A \ge kt_\varepsilon(A)\big) \le (1 - \varepsilon + 2\delta\varepsilon)^k$$

and

$$\forall t \ge 0 \qquad \mathbf{P}_\lambda(H_A \ge t) \le e^{-\varepsilon(1+O(\delta))\lfloor t/t_\varepsilon(A)\rfloor} = \big(1 + O(\varepsilon)\big)e^{-\varepsilon t/((1+O(\delta))t_\varepsilon(A))},$$

which is the desired result. To achieve the goal, we fix some $k \in \mathbf{N}$ and use $T \le t_\varepsilon(A)$ to bound

$$\mathbf{P}_\lambda\big(H_A > (k+1)t_\varepsilon(A)\big) \le \mathbf{P}_\lambda\begin{pmatrix} H_A > kt_\varepsilon(A), \\ H_A \circ \Theta_{kt_\varepsilon(A)+T} > t_\varepsilon(A) - T\end{pmatrix},$$
$$(\text{Markov prop.}) = \mathbf{P}_\lambda\big(H_A > kt_\varepsilon(A)\big)\mathbf{P}_\Lambda\big(H_A > t_\varepsilon(A) - T\big),$$

where Λ is the law of $X_{kt_\varepsilon(A)+T}$ conditioned on $\{H_A > kt_\varepsilon(A)\}$. Since this event belongs to $\mathcal{F}_{kt_\varepsilon(A)}$ and $T = t^Q_{\mathrm{mix}}(\delta\varepsilon)$, Λ is $\delta\varepsilon$-close to π in total variation distance. We deduce

$$\frac{\mathbf{P}_\lambda(H_A > (k+1)t_\varepsilon(A))}{\mathbf{P}_\lambda(H_A > kt_\varepsilon(A))} \le \mathbf{P}_\pi\big(H_A > t_\varepsilon(A) - T\big) + \delta\varepsilon. \tag{3.3}$$

Now observe that

$$\begin{aligned}
\mathbf{P}_\pi\big(H_A > t_\varepsilon(A) - T\big) &\le \mathbf{P}_\pi\big(H_A > t_\varepsilon(A)\big) \\
&\quad + \mathbf{P}_\pi\big(H_A \in \big(t_\varepsilon(A) - T, t_\varepsilon(A)\big]\big) \\
&\le \mathbf{P}_\pi\big(H_A > t_\varepsilon(A)\big) \\
&\quad + \mathbf{P}_\pi(H_A \circ \Theta_{t_\varepsilon(A)-T} \le T), \\
\big(\text{defn. of } t_\varepsilon(A)\big) &= 1 - \varepsilon + \mathbf{P}_\pi(H_A \circ \Theta_{t_\varepsilon(A)-T} \le T), \\
(\pi \text{ stationary}) &= 1 - \varepsilon + \mathbf{P}_\pi(H_A \le T), \\
\big(T = t_{\mathrm{mix}}^Q(\delta\varepsilon) + \text{assumption}\big) &\le 1 - \varepsilon + \delta\varepsilon,
\end{aligned}$$

and plugging this into (3.3) gives

$$\frac{\mathbf{P}_\lambda(H_A > (k+1)t_\varepsilon(A))}{\mathbf{P}_\lambda(H_A > kt_\varepsilon(A))} \le \big(1 - \varepsilon(1 - 2\delta)\big)$$

as desired. $\square$

3.3.2. *Proof of Lemma* 3.3.

PROOF. The general scheme of the proof is similar to that of Lemma 3.2, but we will need to be a bit more careful in our estimates. In particular, we will need that $(1 + 5\delta)\varepsilon < 1/2$ and $2\varepsilon + r_\lambda < 1/2$.

Define $T \equiv t_{\mathrm{mix}}^Q(\delta\varepsilon)$ as in the proof of Lemma 3.2 in Section 3.3.1. Again observe that $T < t_\varepsilon(A)$. Define

$$f(k) \equiv \mathbf{P}_\lambda\big(H_A \ge kt_\varepsilon(A)\big) \qquad (k \in \mathbf{N}).$$

Clearly, $f(0) = 1$ and

$$\begin{aligned}
f(1) &\ge \mathbf{P}_\lambda\big(H_A \circ \Theta_T \ge t_\varepsilon(A)\big) - \mathbf{P}_\lambda(H_A \le T) \ge 1 - \varepsilon - \delta\varepsilon - r_\lambda \\
&\ge 1 - 2\varepsilon - r_\lambda
\end{aligned} \tag{3.4}$$

since $T = t_{\mathrm{mix}}^Q(\delta\varepsilon)$, and by the properties of mixing times,

$$\mathbf{P}_\lambda\big(H_A \circ \Theta_T \ge t_\varepsilon(A)\big) \ge \mathbf{P}_\pi\big(H_A \ge t_\varepsilon(A)\big) - \delta\varepsilon.$$

We now claim the following:

CLAIM 3.1. *For all* $k \in \mathbf{N} \setminus \{0\}$,

$$\frac{f(k+1)}{f(k)} \ge (1 - \varepsilon - 5\delta\varepsilon).$$

Notice that the claim and (3.4) imply

$$\begin{aligned}\forall t \geq 0 \qquad \mathbf{P}_\lambda(H_A \geq t) &\geq f\big(\lceil t/t_\varepsilon(A)\rceil\big)\\ &\geq (1-2\varepsilon - r_\lambda)(1-\varepsilon-5\delta\varepsilon)^{\lceil t/t_\varepsilon(A)\rceil - 1}\\ &= \big(1-O(\varepsilon)-r_\lambda\big)(1-\varepsilon-5\delta\varepsilon)^{t/t_\varepsilon(A)}\\ &\geq \big(1-O(\varepsilon)-r_\lambda\big)e^{-(1+O(\delta))\varepsilon t/t_\varepsilon(A)},\end{aligned}$$

which is precisely the bound we wish to prove. We spend the rest of this proof proving the claim.

Fix some $k \geq 1$, and notice that

$$\begin{aligned} f(k+1) &\geq \mathbf{P}_\lambda\big(H_A \geq kt_\varepsilon(A), H_A \circ \Theta_{kt_\varepsilon(A)+T} \geq t_\varepsilon(A) - T\big)\\ &\qquad - \mathbf{P}_\lambda\big(H_A \geq kt_\varepsilon(A), H_A \circ \Theta_{kt_\varepsilon(A)} < T\big) \qquad (3.5)\\ &=: (I) - (II).\end{aligned}$$

We bound the two terms (I), (II) separately. By the Markov property,

$$(I) = \mathbf{P}_\lambda\big(H_A \geq kt_\varepsilon(A)\big)\mathbf{P}_\Lambda\big(H_A \geq t_\varepsilon(A)\big),$$

where Λ is the conditional law of $X_{kt_\varepsilon(A)+T}$ given $H_A \geq kt_\varepsilon(A)$. Since $T = t_{\text{mix}}^Q(\varepsilon\delta)$, Λ is within distance $\delta\varepsilon$ from π. We deduce

$$(I) \geq \mathbf{P}_\lambda\big(H_A \geq kt_\varepsilon(A)\big)\big(\mathbf{P}_\pi\big(H_A \geq t_\varepsilon(A)\big) - \delta\varepsilon\big) = f(k)(1-\varepsilon-\delta\varepsilon). \qquad (3.6)$$

We now upper bound term (II) in (3.5). Notice that (again because of the Markov property)

$$(II) \leq \mathbf{P}_\pi\big(H_A \geq (k-1)t_\varepsilon(A), H_A \circ \Theta_{kt_\varepsilon(A)} < T\big) = f(k-1)\mathbf{P}_{\Lambda'}(H_A < T),$$

where Λ' is the law of $X_{kt_\varepsilon(A)}$ conditioned on $\{H_A \geq (k-1)t_\varepsilon(A)\}$. Recalling that $t_\varepsilon(A) \geq T = t_{\text{mix}}^Q(\delta\varepsilon)$, we see that Λ' is $\delta\varepsilon$-close to π. Since we have also assumed that $\mathbf{P}_\pi(H_A \leq T) \leq \delta\varepsilon$, we deduce

$$(II) \leq f(k-1)\big(\mathbf{P}_\pi(H_A < T) + \delta\varepsilon\big) \leq 2\delta\varepsilon f(k-1).$$

We combine this with (3.6) and (3.5) to obtain

$$\forall k \in \mathbf{N} \setminus \{0,1\} f(k+1) \geq f(k)(1-\varepsilon-\delta\varepsilon) - f(k-1)(2\delta\varepsilon).$$

One can argue inductively that $f(k)/f(k-1) \geq 1/2$ for all $k \geq 1$. Indeed, this holds for $k \geq 2$ by the claim applied to $k-1$. For $k = 1$ we may use (3.4) and the assumption on $2\varepsilon + r_\lambda$ to deduce the same result. Applying this to the previous inequality, we obtain

$$\forall k \in \mathbf{N} \setminus \{0\} f(k+1) \geq f(k)(1-\varepsilon-5\delta\varepsilon),$$

which finishes the proof of the claim and of the lemma. $\square$

3.4. *Hitting times of a union of sets*: *Proofs*. We present the proof of Theorem 3.2 below.

PROOF OF THEOREM 3.2. There are three main steps in the proof, here outlined in a slightly oversimplified way:

(1) We show that Theorem 3.1 is applicable to the hitting times of $A_1, \dots, A_\ell$. In particular, this shows that $\mathbf{P}_\pi(H_{A_i} \le \varepsilon m) \approx \varepsilon$.

(2) We show that

$$\mathbf{P}_\pi(H_A \le \varepsilon m) \approx \sum_{i=1}^{\ell} \mathbf{P}_\pi(H_{A_i} \le \varepsilon m) \approx \ell\varepsilon,$$

so that $t_{\ell\varepsilon}(A) \approx \varepsilon m$.

(3) Finally, we apply Theorem 3.1 to H_A and deduce that this random variable is approximately exponential with mean

$$\mathbf{E}_\pi[H_A] \approx t_{\varepsilon\ell}(A)/\varepsilon\ell \approx m/\ell.$$

The actual proof is only slightly more complicated than this outline. We begin with a claim corresponding to step 1 above.

CLAIM 3.2. *For all* $1 \le i \le \ell$,

$$\varepsilon_i \equiv \mathbf{P}_\pi(H_{A_i} \le \varepsilon m) = \bigl(1 + O(\delta)\bigr)\varepsilon.$$

PROOF. Consider some $\varepsilon' \in [\varepsilon/2, 2\varepsilon]$. Notice that

$$t_{\mathrm{mix}}^{Q}(\delta\varepsilon') \le t_{\mathrm{mix}}^{Q}(\delta\varepsilon/2),$$

and therefore,

$$\mathbf{P}_\pi\bigl(H_{A_i} \le t_{\mathrm{mix}}^{Q}(\delta\varepsilon')\bigr) \le \frac{\delta\varepsilon}{2} \le \delta\varepsilon'.$$

This shows that Theorem 3.1 is applicable with A_i replacing A and ε' replacing ε. We deduce in particular that

$$\forall \frac{\varepsilon}{2} \le \varepsilon' \le 2\varepsilon \qquad \left|\frac{\varepsilon'\mathbf{E}_\pi[H_{A_i}]}{t_{\varepsilon'}(A_i)} - 1\right| \le O\bigl(\delta + \varepsilon'\bigr) = O(\delta).$$

In particular, there exists a universal constant $c > 0$ such that if $\varepsilon' \le (1 - c\delta)\varepsilon$, then $t_{\varepsilon'}(A_i) < \varepsilon\mathbf{E}_\pi[H_{A_i}]$, whereas if $\varepsilon' > (1 + c\delta)\varepsilon$, $t_{\varepsilon'}(A_i) > \varepsilon\mathbf{E}_\pi[H_{A_i}]$. In other words,

$$(1 - c\delta)\varepsilon \le \mathbf{P}_\pi\bigl(H_{A_i} \le \varepsilon\mathbf{E}_\pi[H_{A_i}]\bigr) \le (1 + c\delta)\varepsilon. \qquad \square$$

We now come to the second part of the proof.

CLAIM 3.3. *Let ξ be as in the statement of Theorem 3.2. Then*

$$\mathbf{P}_\pi(H_A \le \varepsilon m) = \big(1 + O(\delta + \xi)\big)\ell\varepsilon.$$

In particular, there exists a number $\eta = (1 + O(\delta + \xi))\ell\varepsilon$ with $\varepsilon m = t_\eta(A)$.

PROOF. To see this, we note that

$$\{H_A \le \varepsilon m\} = \bigcup_{i=1}^{\ell} \{H_{A_i} \le \varepsilon m\}.$$

The union bound gives

$$\mathbf{P}_\pi(H_A \le \varepsilon m) \le \sum_{i=1}^{\ell} \mathbf{P}_\pi(H_{A_i} \le \varepsilon m) \le \big(1 + O(\delta)\big)\ell\varepsilon.$$

A lower bound can be obtained via the Bonferroni inequality,

$$\begin{aligned}\mathbf{P}_\pi(H_A \le \varepsilon m) &\ge \sum_{i=1}^{\ell} \mathbf{P}_\pi(H_{A_i} \le \varepsilon m) \\ &\quad - \sum_{1 \le i < j \le \ell} \mathbf{P}_\pi(H_{A_i} \le \varepsilon m, H_{A_j} \le \varepsilon m) \\ &= \big(1 + O(\delta + \xi)\big)\ell\varepsilon,\end{aligned}$$

using the definition of ξ. □

We now need to show that the assumptions of Theorem 3.1 are applicable to H_A, with the value of η in Claim 3.3 replacing ε. We assume that $\delta + \xi$ is small enough, which we may do because otherwise the theorem is trivial. In particular, we can assume that the $O(\xi + \delta)$ term in the expression for η is between $-1/2$ and 1, so that

$$\frac{\varepsilon\ell}{2} \le \eta \le 2\varepsilon\ell.$$

Since we also assumed $\varepsilon < \delta/2\ell$, we have $\eta < \delta$. Moreover, $t^Q_{\text{mix}}(\delta\eta) \le t^Q_{\text{mix}}(\delta\varepsilon/2)$. This implies

$$\mathbf{P}_\pi\big(H_A \le t^Q_{\text{mix}}(\delta\eta)\big) \le \sum_{i=1}^{\ell} \mathbf{P}_\pi\big(H_{A_i} \le t^Q_{\text{mix}}(\delta\varepsilon/2)\big) \le \ell\delta\varepsilon/2 \le \delta\eta.$$

We may now apply Theorem 3.1 (with η replacing ε) to deduce that for any $\lambda \in M_1(\mathbf{V})$,

$$\text{Law}_\lambda(H_A) = \mathsf{Exp}\big(t_\eta(A)/\eta, O(\eta) + 2r_\lambda, O(\delta + \xi)\big).$$

To finish the proof, we note that $\eta = O(\ell\varepsilon)$,

$$t_\eta(A)/\eta = \varepsilon m / \big(1 + O(\delta + \xi)\big)\ell\varepsilon = \big(1 + O(\delta + \xi)\big)\frac{m}{\ell}$$

and apply Proposition 3.1. □

4. Meeting times of multiple random walks. We now put our two exponential approximation results to use, showing that the meeting times we are interested in are well approximated by exponential random variables. Much of the work needed for this is contained in technical estimates whose proofs can be safely skipped in a first reading.

4.1. *Basic definitions*. For the remainder of this section, $\mathbf{V}$ is a finite set, and Q is the generator of a mixing Markov chain over $\mathbf{V}$ with mixing times $t_{\mathrm{mix}}^{Q}(\cdot)$ and stationary measure π. For each $k \in \mathbf{N} \setminus \{0, 1\}$ we will also consider the Markov chains $Q^{(k)}$ over $\mathbf{V}^k$ that correspond to k independent realizations of Q from prescribed initial states, as defined in Section 2.3.4. We will also follow the notation from that section.

For $k = 2$, we define the first meeting time

$$M \equiv \inf\{t \geq 0 : X_t(1) = X_t(2)\} \tag{4.1}$$

and the parameters

$$\mathsf{m}(Q) \equiv \mathbf{E}_{\pi^{\otimes 2}}[M], \tag{4.2}$$

$$\rho(Q) \equiv \frac{t_{\mathrm{mix}}^{Q}}{\mathsf{m}(Q)}. \tag{4.3}$$

We also define an extra prameter $\mathsf{err}(Q)$ which will appear as an error term at several different points in the paper. Ths parameter $\mathsf{err}(Q)$ is defined as

$$\mathsf{err}(Q) = c_0\sqrt{\rho(Q)\ln(1/\rho(Q))} \qquad \text{if } Q \text{ is reversible and transitive.} \tag{4.4}$$

For other Q, we define it as

$$\mathsf{err}(Q) = c_1 \sqrt{(1 + q_{\max} t_{\mathrm{mix}}^{Q})\pi_{\max} \ln\left(\frac{1}{(1 + q_{\max} t_{\mathrm{mix}}^{Q})\pi_{\max}}\right)}. \tag{4.5}$$

The numbers $c_0, c_1 > 0$ are universal constants that we do not specify explicitly. We choose them so as to satisfy Propositions 4.1, 4.4 and 4.5 below.

We now take $k > 2$ and consider the process $Q^{(k)}$, with trajectories

$$(X_t^{(k)} = (X_t(1), X_t(2), \dots, X_t(k)))_{t \geq 0}$$

corresponding to k independent realizations of Q; cf. Section 2.3.4. This has stationary distribution $\pi^{\otimes k}$.

We write $M^{(k)}$ for the first meeting time among these random walks,

$$M^{(k)} \equiv \inf\{t \geq 0 : \exists 1 \leq i < j \leq k, X_t(i) = X_t(j)\}. \tag{4.6}$$

One may note that

$$M^{(k)} = \min_{\{i,j\} \in \binom{[k]}{2}} M_{i,j},$$

where $\binom{[k]}{2}$ was defined in Section 2, and for $1 \le i < j \le k$,

$$M_{i,j} = M_{j,i} \equiv \inf\{t \ge 0 : X_t(i) = X_t(j)\} \tag{4.7}$$

is distributed as M for a realization of $Q^{(2)}$ starting from $(X_0(i), X_0(j))$.

4.2. *Technical estimates for reversible and transitive chains*. In this subsection we collect the estimates that we will use in the case of chains that are reversible and transitive.

PROPOSITION 4.1. *Assume Q is reversible and transitive and define $\mathsf{err}(Q)$ accordingly. If $\mathsf{err}(Q) \le 1/4$, then*

$$\mathbf{P}_{\pi^{\otimes 2}}\big(M \le t_{\mathrm{mix}}^{Q}(\mathsf{err}(Q)^2)\big) \le \mathsf{err}(Q)^2.$$

REMARK 4.1. The proof is entirely general, but we will only use this estimate in the transitive/reversible case.

PROOF. We will prove a result in contrapositive form: if $0 < \beta < 1/4$ is such that

$$\mathbf{P}_{\pi^{\otimes 2}}\big(M \le t_{\mathrm{mix}}^{Q}(\beta)\big) > \beta,$$

then $\beta < c_0^2 \rho(Q) \ln(1/\rho(Q))$ for some universal $c_0 > 0$.

Notice that for any $x^{(2)} \in \mathbf{V}^2$,

$$\begin{aligned}
\mathbf{P}_{x^{(2)}}\big(M > t_{\mathrm{mix}}^{Q}(\beta/4) + t_{\mathrm{mix}}^{Q}(\beta)\big) &\le \mathbf{P}_{x^{(2)}}\big(M \circ \Theta_{t_{\mathrm{mix}}^{Q}(\beta/4)} > t_{\mathrm{mix}}^{Q}(\beta)\big) \\
&\le \mathbf{P}_{\pi^{\otimes 2}}\big(M > t_{\mathrm{mix}}^{Q}(\beta)\big) + \beta/2 \\
&< (1 - \beta/2),
\end{aligned}$$

where the middle inequality follows from the fact that $t_{\mathrm{mix}}^{Q}(\beta/4)$ is an upper bound for the $\beta/2$-mixing time of $Q^{(2)}$; cf. Lemma 2.4. A standard argument using the Markov property implies that for any $k \in \mathbf{N}$,

$$\mathbf{P}_{x^{(2)}}\big(M > k\big(t_{\mathrm{mix}}^{Q}(\beta/4) + t_{\mathrm{mix}}^{Q}(\beta)\big)\big) < (1 - \beta/2)^k,$$

so that

$$\mathsf{m}(Q) = \mathbf{E}_{\pi^{\otimes 2}}[M] \le C \frac{t_{\mathrm{mix}}^{Q}(\beta) + t_{\mathrm{mix}}^{Q}(\beta/4)}{\beta}.$$

Since $t_{\mathrm{mix}}^{Q}(\alpha) \le C \ln(1/\alpha) t_{\mathrm{mix}}^{Q}$, we deduce that

$$\frac{\beta}{c \ln(1/\beta)} < \rho(Q),$$

with $c > 0$ universal, which implies the desired result. $\square$

We now prove an estimate on correlations.

PROPOSITION 4.2. *Assume Q is transitive. Then for all $t,s\geq 0$ and $\{i,j\},\{\ell,r\}\subset\mathbf{V}$ with $\{i,j\}\neq\{r,\ell\}$,*

$$\mathbf{P}_{\pi^{\otimes k}}(M_{i,j}\leq t, M_{\ell,r}\leq s)\leq 2\mathbf{P}_{\pi^{\otimes 2}}(M\leq s)\mathbf{P}_{\pi^{\otimes 2}}(M\leq t).$$

PROOF. If $\{i,j\}\cap\{\ell,r\}=\varnothing$, the events $\{M_{i,j}\leq t\}$ and $\{M_{\ell,r}\leq t\}$ are independent. Since the laws of both $M_{i,j}$ and $M_{\ell,r}$ under $\pi^{\otimes k}$ are equal to the law of M under $\pi^{\otimes 2}$, we obtain

$$\mathbf{P}_{\pi^{\otimes k}}(M_{i,j}\leq t, M_{\ell,r}\leq s)\leq \mathbf{P}_{\pi^{\otimes 2}}(M\leq s)\mathbf{P}_{\pi^{\otimes 2}}(M\leq t)$$

in this case. Assume now $\{i,j\}\cap\{\ell,r\}$ has one element. Without loss of generality we may assume $k=3$, $\{i,j\}=\{1,2\}$ and $\{\ell,r\}=\{1,3\}$. We have

$$\begin{aligned}\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2}\leq t, M_{1,3}\leq s)&\leq \mathbf{P}_{\pi^{\otimes 3}}(M_{1,2}\leq t, M_{1,3}\circ\Theta_{M_{1,2}}\leq s)\\ &\quad+\mathbf{P}_{\pi^{\otimes 3}}(M_{1,3}\leq s, M_{1,2}\circ\Theta_{M_{1,3}}\leq t).\end{aligned}\tag{4.8}$$

Consider the first term on the RHS. By the Markov property,

$$\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2}\leq t, M_{1,3}\circ\Theta_{M_{1,2}}\leq s)=\mathbf{P}_{\pi^{\otimes 2}}(M\leq t)\mathbf{P}_{\lambda^{(2)}}(M\leq s),$$

where $\lambda^{(2)}$ is the law of $X_{M_{1,2}}(1), X_{M_{1,2}}(3)$ conditionally on $M_{1,2}\leq t$. Since $(X_t(3))_t$ is stationary and independent from this event, $\lambda^{(2)}=\lambda\otimes\pi$ for some $\lambda\in M_1(\mathbf{V})$ which is the law of $X_{M_{1,2}}(1)$ under $\mathbf{P}_{\pi^{\otimes 2}}$. The transitivity of Q (which implies that π is uniform) implies that $\lambda=\pi$ and therefore

$$\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2}\leq t, M_{1,3}\circ\Theta_{M_{1,2}}\leq s)=\mathbf{P}_{\pi^{\otimes 2}}(M\leq t)\mathbf{P}_{\pi^{\otimes 2}}(M\leq s).$$

The same bound can be shown for the other term in the RHS of (4.8), and this implies the proposition. □

4.3. *Technical estimates for the general case.* We will need the following general result:

PROPOSITION 4.3. *For any $\lambda\in M_1(\mathbf{V})$ and $T\geq 0$,*

$$\mathbf{P}_{\lambda\otimes\pi}(M\leq T)\leq(1+2Tq_{\max})\pi_{\max}.$$

PROOF. Let $(X_t)_t$ be a single realization of Q. One may imagine that the trajectory of $(X_t)_{t\geq 0}$ is sampled as follows. First, let $\mathcal{P}$ be a Poisson process with intensity $q_{\max}$ independent from the initial state X_0. At each time $t\in\mathcal{P}$, one updates the value of X_t as follows: if $X_s=x$ for s immediately before t, one sets

$$X_t=y\text{ with probability}\qquad\frac{q(x,y)}{q_{\max}}\big(y\in\mathbf{V}\setminus\{x\}\big)$$

and $X_t=x$ with the remaining probability. This implies that, at the points of the Poisson process, X_t is updated as in the discrete-time Markov chain with matrix $P=(I+Q/q_{\max})$, and it is easy to see that π is stationary for this chain.

Now let $X_t(1), X_t(2)$ be independent trajectories of Q, with $X_t(1)$ started from λ and $X_t(2)$ started from the stationary distribution π. We will imagine that each $X_t(i)$ has its own Poisson process $\mathcal{P}(i)$ and was generated in the way described above. It then follows that

$$\mathbf{P}_{\lambda\otimes\pi}\big(M \le T \mid \mathcal{P}(1), \mathcal{P}(2)\big) \le \mathbf{P}_{\lambda\otimes\pi}\big(X_0(1) = X_0(2)\big) + \sum_{t\in\mathcal{P}} \mathbf{P}_{\lambda\otimes\pi}\big(X_t(1) = X_t(2) \mid \mathcal{P}(1), \mathcal{P}(2)\big),$$

where $\mathcal{P} = \mathcal{P}(1) \cup \mathcal{P}(2)$, since the processes can only change values at the times of the two Poisson processes. At time 0, we have

$$\mathbf{P}_{\lambda\otimes\pi}\big(X_0(1) = X_0(2)\big) = \sum_{x\in\mathbf{V}} \lambda(x)\pi(x) \le \sum_{x\in\mathbf{V}} \lambda(x)\pi_{\max} \le \pi_{\max}.$$

For $t \in \mathcal{P}$, the law of $X_t(1), X_t(2)$ equals

$$\big(\lambda P^{k_1}\big) \otimes \big(\pi P^{k_2}\big),$$

where $k_i = |\mathcal{P}(i) \cap (0, t]|$ $(i = 1, 2)$. Crucially, π is stationary for P, hence $\pi P^{k_2} = \pi$, and we obtain

$$\mathbf{P}_{\lambda\otimes\pi}\big(X_t(1) = X_t(2) \mid \mathcal{P}(1), \mathcal{P}(2)\big) = \sum_{x\in\mathbf{V}} \big(\lambda P^{k_1}\big)(x)\pi(x) \le \pi_{\max}$$

as for $t = 0$. We deduce

$$\mathbf{P}_{\lambda\otimes\pi}\big(M \le T \mid \mathcal{P}(1), \mathcal{P}(2)\big) \le \big(1 + \big|\big(\mathcal{P}(1) \cup \mathcal{P}(2)\big) \cap (0, T]\big|\big)\pi_{\max}.$$

The proposition follows from taking expectations on both sides and noticing that

$$\mathbf{E}\big[\big|\big(\mathcal{P}(1) \cup \mathcal{P}(2)\big) \cap (0, T]\big|\big] = 2T q_{\max}. \qquad \square$$

We now prove an estimate corresponding to Proposition 4.1 in this general setting.

PROPOSITION 4.4. *Assume* $\mathsf{err}(Q)$ *is as defined in* (4.5). *Then*

$$\mathbf{P}_{\pi^{\otimes 2}}\big(M \le t^{Q}_{\mathrm{mix}}(\mathsf{err}(Q)^2)\big) \le \mathsf{err}(Q)^2.$$

PROOF. The previous proposition implies

$$\begin{aligned}\mathbf{P}_{\pi^{\otimes 2}}\big(M \le t^{Q}_{\mathrm{mix}}(\mathsf{err}(Q)^2)\big) &\le \big(1 + 2t^{Q}_{\mathrm{mix}}(\mathsf{err}(Q))q_{\max}\big)\pi_{\max} \\ &\le C\big(1 + 2t^{Q}_{\mathrm{mix}} q_{\max}\big)\pi_{\max} \ln\big(1/\mathsf{err}(Q)\big).\end{aligned}$$

This is $\le \mathsf{err}(Q)^2$ by definition of this quantity, if we choose c_1 in (4.5) to be large enough. $\square$

We now prove an estimate on correlations that is similar to Proposition 4.2, but with an extra term. Recall that $\binom{[k]}{2}$ was defined in Section 2.1.

PROPOSITION 4.5. *For any mixing Markov chain Q, if one defines* $\mathsf{err}(Q)$ *as in* (4.5), *we have the following inequality for $k \geq 3$ and all distinct pairs* $\{i, j\}, \{\ell, r\} \in \binom{[k]}{2}$:

$$\mathbf{P}_{\pi^{\otimes k}}(M_{i,j} \leq t, M_{\ell,r} \leq s) \leq 2\mathbf{P}_{\pi^{\otimes 2}}(M \leq t)\mathbf{P}_{\pi^{\otimes 2}}(M \leq s) + O\big(\mathsf{err}(Q)^2\big).$$

PROOF. The case $\{i, j\} \cap \{\ell, r\} = \varnothing$ follows as in the proof of Proposition 4.2. In case $\{i, j\} \cap \{\ell, r\}$ has one element, we may again assume that $i = \ell = 1$, $j = 2$ and $k = 3$. Equation (4.8) still applies, so we proceed to bound

$$\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t, M_{1,3} \circ \Theta_{M_{1,2}} \leq s),$$

which is upper bounded by

$$\begin{aligned}
&\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t, M_{1,3} \circ \Theta_{M_{1,2}} \leq s) \\
&\quad \leq \mathbf{P}_{\pi^{\otimes 3}}\big(M_{1,2} \leq t, M_{1,3} \circ \Theta_{M_{1,2}} \leq t^Q_{\mathrm{mix}}(\eta)\big) \\
&\qquad + \mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t, M_{1,3} \circ \Theta_{M_{1,2}+t^Q_{\mathrm{mix}}(\eta)} \leq s) = (I) + (II)
\end{aligned}$$

for some $\eta \in (0, 1/4)$ to be chosen later.

Term (I) is equal to

$$\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t)\mathbf{P}_{\lambda^{(2)}}\big(M \leq t^Q_{\mathrm{mix}}(\eta)\big),$$

where $\lambda^{(2)}$ is the law of $(X_{M_{1,2}}(2), X_{M_{1,2}}(3))$ conditionally on $\{M_{1,2} \leq t\}$. As in the previous proof, $(X_t(3))_t$ is stationary and independent from the conditioning, hence $\lambda^{(2)} = \lambda \otimes \pi$ for some $\lambda \in M_1(\mathbf{V})$. We use Proposition 4.3 to deduce

$$(I) \leq \mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t)O\big(\big(1 + t^Q_{\mathrm{mix}}(\eta)q_{\max}\big)\pi_{\max}\big).$$

The analysis of term (II) is simpler: we have

$$(II) = \mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t)\mathbf{P}_{\lambda_* \otimes \pi}(M \leq s)$$

for some $\lambda_* \in M_1(\mathbf{V})$ which is the law of $X_{M_{1,2}+t^Q_{\mathrm{mix}}(\eta)}$ conditionally on $\{M_{1,2} \leq t\}$. The time shift by $t^Q_{\mathrm{mix}}(\eta)$ implies that λ_* is η-close to stationary, hence

$$(II) \leq \mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t)\big(\eta + \mathbf{P}_{\pi^{\otimes 2}}(M \leq s)\big).$$

We deduce that

$$\begin{aligned}
&\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \leq t, M_{1,3} \circ \Theta_{M_{1,2}} \leq s) \\
&\quad \leq \mathbf{P}_{\pi^{\otimes 2}}(M \leq t)\mathbf{P}_{\pi^{\otimes 2}}(M \leq s) \\
&\qquad + \mathbf{P}_{\pi^{\otimes 2}}(M \leq t)O\big(\eta + \big(1 + t^Q_{\mathrm{mix}}(\eta)q_{\max}\big)\pi_{\max}\big).
\end{aligned}$$

Recall $t^Q_{\mathrm{mix}}(\eta) \leq C t^Q_{\mathrm{mix}} \ln(1/\eta)$ for some universal $C > 0$. If

$$\eta_0 \equiv \big(1 + q_{\max} t^Q_{\mathrm{mix}}\big)\pi_{\max} \leq 1/2,$$

we may take $\eta = \eta_0$ to obtain

$$\mathbf{P}_{\pi^{\otimes 3}}(M_{1,2} \le t, M_{1,3} \circ \Theta_{M_{1,2}} \le s) \le \mathbf{P}_{\pi^{\otimes 2}}(M \le t)\mathbf{P}_{\pi^{\otimes 2}}(M \le s) + \mathbf{P}_{\pi^{\otimes 2}}(M \le t) O\left(\eta \ln\left(\frac{1}{\eta}\right)\right).$$

The case of $\eta_0 \ge 1/2$ is covered "automatically" by the big-oh notation.

An analogous bound can be obtained with the roles of $(t, 2)$ and $(s, 3)$ reversed. Plugging these into (4.8) gives the desired bound. □

4.4. *Exponential approximation for a pair of particles*. We now come back to the setting of Section 4.1 and show M is approximately exponentially distributed.

LEMMA 4.1. *Define* $\mathsf{err}(Q)$ *as in* (4.4) (*if* Q *is reversible and transitive*) *or as in* (4.5) (*if not*). *Then* $\forall \lambda^{(2)} \in M_1(\mathbf{V}^{(2)})$

$$\mathrm{Law}_{\lambda^{(2)}}(M) = \mathsf{Exp}\big(\mathsf{m}(Q), O(\mathsf{err}(Q)) + 2r\lambda^{(2)}, O(\mathsf{err}(Q))\big),$$

where

$$r_{\lambda^{(2)}} = \mathbf{P}_{\lambda^{(2)}}\big(M \le t_{\mathrm{mix}}^{Q}(\mathsf{err}(Q)^2)\big).$$

PROOF. This is a direct application of Theorem 3.1 to the hitting time of the diagonal set

$$\Delta \equiv \{(x, x) : x \in \mathbf{V}\} \subset \mathbf{V}^2$$

by the chain with generator $Q^{(2)}$ defined in Section 2 and with $\varepsilon = \mathsf{err}(Q)$, $\delta = 2\mathsf{err}(Q)$. All we need to show is that

$$\mathbf{P}_{\pi^{\otimes 2}}\big(M \le t_{\mathrm{mix}}^{Q^{(2)}}(\delta\varepsilon)\big) \le \varepsilon\delta,$$

where $t_{\mathrm{mix}}^{Q^{(2)}}(\cdot)$ denotes the mixing times of $Q^{(2)}$. This inequality follows from

$$t_{\mathrm{mix}}^{Q^{(2)}}\big(2\mathsf{err}(Q)^2\big) \le t_{\mathrm{mix}}^{Q}\big(\mathsf{err}(Q)^2\big) \qquad \text{(Lemma 2.4)}$$

and

$$\mathbf{P}_{\pi^{\otimes 2}}\big(M \le t_{\mathrm{mix}}^{Q}(\mathsf{err}(Q)^2)\big) \le \mathsf{err}(Q)^2 < \delta\varepsilon,$$

which follows from Proposition 4.1 in the reversible/transitive case and Proposition 4.4 in the general case. □

4.5. *Exponential approximation for many random walkers.* We now consider the more complex problem of bounding the meeting times among $k \geq 2$ particles. We take the notation in Section 4.1 for granted.

LEMMA 4.2. *Let $\ell = \binom{k}{2} > 0$, and assume that the quantity $\mathsf{err}(Q)$ defined in (4.4) (if Q is reversible and transitive) or as in (4.5) (if not) satisfies $\mathsf{err}(Q) \leq 1/10\ell$. Then for all $\lambda^{(k)} \in M_1(\mathbf{V}^k)$,*

$$\mathrm{Law}_{\lambda^{(k)}}(M^{(k)}) = \mathsf{Exp}\left(\frac{\mathsf{m}(Q)}{\ell}, O(k^2\mathsf{err}(Q)) + 2r_{\lambda^{(k)}}, O(k^2\mathsf{err}(Q))\right),$$

where $r_{\lambda^{(k)}} = \mathbf{P}_{\lambda^{(k)}}(M^{(k)} \leq t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2))$.

PROOF. $M^{(k)}$ is the hitting time of a union of ℓ sets:

$$\Delta^{(k)} \equiv \bigcup_{\{i,j\}\in\binom{k}{2}} \Delta_{\{i,j\}} \qquad \text{where } \Delta_{\{i,j\}} \equiv \{x^{(k)} \in \mathbf{V}^k : x^{(k)}(i) = x^{(k)}(j)\}.$$

We will apply Theorem 3.2, applied to the product chain $Q^{(k)}$, to show that this hitting time is approximately exponential. We set $\delta = 2\ell\mathsf{err}(Q)$, $\varepsilon = \mathsf{err}(Q)$ and verify the conditions of the theorem:

- $0 < \delta < 1/5$, $0 < \varepsilon < \delta/2\ell$: These conditions follow from $\mathsf{err}(Q) < 1/10\ell$.
- $\mathbf{P}_{\pi^{\otimes k}}(M_{i,j} \leq t^{Q^{(k)}}_{\mathrm{mix}}(\delta\varepsilon/2)) \leq \delta\varepsilon/2$. To prove this we simply observe that
$$t^{Q^{(k)}}_{\mathrm{mix}}(\delta\varepsilon/2) \leq t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2) \qquad \text{(Lemma 2.4 and defn. of } \varepsilon, \delta)$$
and that
$$\mathbf{P}_{\pi^{\otimes 2}}(M \leq t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)) \leq \mathsf{err}(Q)^2 = \frac{\delta\varepsilon}{2\ell} \leq \frac{\delta\varepsilon}{2}$$
by Proposition 4.1 (in the reversible/transitive case) or by Proposition 4.4 (in general).
- $\mathbf{E}_{\pi^{\otimes k}}[H_{\Delta_{\{i,j\}}}] = \mathsf{m}(Q)$ is the same for all $\{i,j\} \in \binom{[k]}{2}$: this is obvious.

The lemma will then follow once we show that the ξ quantity in Theorem 3.2, which in this case equals

$$\xi = \sum_{\{i,j\}\neq\{\ell,r\} \text{ in } \binom{[k]}{2}} \frac{\mathbf{P}_{\pi^{\otimes k}}(M_{\{i,j\}} \leq \varepsilon\mathsf{m}(Q), M_{\{\ell,r\}} \leq \varepsilon\mathsf{m}(Q))}{\ell\varepsilon},$$

and satisfies $\xi = O(k^2\mathsf{err}(Q))$. To start, we go back to Claim 3.2 in the proof of Theorem 3.2 and observe that whenever the assumptions of that theorem hold,

$$\mathbf{P}_{\pi^{\otimes k}}(M_{\{i,j\}} \leq \varepsilon\mathsf{m}(Q)) = O(\varepsilon). \tag{4.9}$$

Now note that Propositions 4.2 (in the reversible/transitive case) and 4.5 (in the general case) imply that each term in the sum defining ξ is $O(\mathsf{err}(Q)^2)$. We deduce

$$\xi \leq \frac{O(\varepsilon^2)\binom{\ell}{2}}{\ell\varepsilon} \leq O(\ell\varepsilon) = O(k^2\mathsf{err}(Q)). \qquad \square$$

5. Coalescing random walks: Basics. In this section we formally define the coalesing random walks process. We then show that if the initial number of particles is not large, mean field behavior follows from the exponential approximation of meeting times.

5.1. *Definitions*. Fix a Markov chain Q on a finite state space $\mathbf{V}$. Given a number $k \in [|\mathbf{V}|] \setminus \{1\}$ and an initial state $x^{(k)} \in \mathbf{V}^k$, consider a realization of $Q^{(k)}$

$$(X^{(k)})_{t\geq 0} \equiv (X_t(1), \dots, X_t(k))_{t\geq 0}.$$

We build the coalescing random walks process from $X^{(k)}$ by defining the trajectories of the k walkers one by one. We first set

$$\overline{X}_t(1) = X_t(1), \qquad t \geq 0.$$

Given $j \in [k] \setminus \{1\}$, assume that $\overline{X}_t(i)$ has been defined for all $1 \leq i < j$ and $t \geq 0$. We let T_j be the first time $t \geq 0$ at which $X_t(j) = \overline{X}_t(I_j)$ for some $1 \leq I_j < i$, and then set

$$\overline{X}_t(j) \equiv \begin{cases} X_t(j), & t < T_j; \\ \overline{X}_t(I_j), & t \geq T_j. \end{cases}$$

Intuitively, this says that as soon as j encounters a walker with lower index, it starts moving along with it. The process

$$(\overline{X}_t^{(k)})_{t\geq 0} \equiv (\overline{X}_t(j))_{t\geq 0}$$

is what we call the *coalescing random walks process* based on Q, with initial state $x^{(k)}$.

REMARK 5.1. For any $j \geq 3$, there might be more than one index $i < j$ such that $\overline{X}_{T_j}(i) = X_{T_j}(j)$. However, it is easy to see that all such i will have the same trajectory after time T_j because they must have met by that time. This implies that there is no ambiguity in the definition of $\overline{X}_t(j)$ for any j.

We also define

$$\mathsf{C}_i \equiv \inf\{t \geq 0 : |\{\overline{X}_t(j) : j \in [k]\}| \leq i\}$$

and $\mathsf{C} \equiv \mathsf{C}_1$. The fact that we are working in continuous time implies the following:

PROPOSITION 5.1 (Proof omitted). *Assume that the initial state* $x^{(k)} = (x(1), x(2), \dots, x(k))$ *is such that* $x(i) \neq x(j)$ *for all* $1 \leq i < j \leq k$. *Then* $\mathsf{C}_k = 0 < \mathsf{C}_{k-1} < \mathsf{C}_{k-2} < \cdots < \mathsf{C}_1$ *almost surely.*

It is sometimes useful to view the coalescing random walks process as a process with *killings*. Define a random $2^{[k]}$-valued process $(A_t)_{t\geq 0}$ as follows:

- $1 \in A_t$ for all t;
- proceeding recursively, for each $j \in [k] \setminus \{1\}$, we have $j \in A_t$ if and only if $\tau_j > t$, where τ_j is the first time t at which $X_t(i) = X_t(j)$ for some $i < j$ with $i \in A_t$.

Intuitively, A_t is the set of all walkers that are "alive" at time $t \geq 0$, and a walker dies at the first time it meets an alive walker with smaller index. One may check that coalescing random walks is equivalent to the killed process in the following sense.

PROPOSITION 5.2 (Proof omitted). *We have* $\tau_j = T_j$ *for all* $j \in [k] \setminus \{1\}$. *Moreover, for all* $t \geq 0$, *we have*

$$\{X_t(j) : j \in A_t\} = \{\overline{X}_t(j) : j \in [k]\}.$$

Finally, for all $i \in [k-1]$,

$$\mathsf{C}_i = \inf\{t \geq 0 : |A_t| \leq i\}.$$

Recall that $M_{i,j}$ is the meeting time between walkers i and j; cf. (4.7). We have the following simple proposition:

PROPOSITION 5.3 (Proof omitted). *Assume that the initial state*

$$x^{(k)} = (x(1), x(2), \dots, x(k))$$

is such that $x(i) \neq x(j)$ *for all* $1 \leq i < j \leq k$. *Then for each* $1 \leq p \leq k-1$,

$$\mathsf{C}_p - \mathsf{C}_{p+1} = \min_{\{i,j\} \subset A_{\mathsf{C}_{p+1}}} M_{i,j} \circ \Theta_{\mathsf{C}_{p+1}}.$$

Moreover, each time C_p *equals* $M_{i,j}$ *for some* $\{i, j\} \in \binom{[k]}{2}$.

5.2. *Mean-field behavior for moderately large* k. We now prove a mean-field-like result for an initial number of particles k that is not too large, assuming that meeting times of up to k walkers satisfy our exponential approximation property.

LEMMA 5.1. *Assume that* Q, $\mathsf{err}(Q)$ *and* k *satisfy the assumptions of Lemma* 4.2. *Let* $x^{(k)} \in \mathbf{V}^k$. *Then for all* $p \in [k-1]$,

$$d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i=p+1}^{k} \mathsf{Z}_i\right)\right) = \frac{O(k^2 \mathsf{err}(Q)) + 12\eta(x^{(k)})}{p},$$

where

$$\eta(x^{(k)}) = \mathbf{P}_{x^{(k)}}(M^{(k)} \leq t_{\mathrm{mix}}^{Q}(\mathsf{err}(Q)^2)) + \mathbf{P}_{x^{(k)}}\left(\begin{array}{c} \exists \{i,j\}, \{\ell, r\} \in \binom{[k]}{2} : \\ \{\ell, r\} \neq \{i, j\} \text{ but } M_{i,j} \circ \Theta_{M_{\ell,r}} \leq t_{\mathrm{mix}}^{Q}(\mathsf{err}(Q)^2) \end{array}\right),$$

and the Z_i *are the random variables described in* (1.1).

PROOF. Write $x^{(k)} = (x(1), \dots, x(k))$. We will prove the similar bound

$$\begin{aligned}
&\text{“}\forall 1 \le i < j \le k : x(i) \neq x(j)\text{”} \\
&\Rightarrow d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i=p+1}^{k} \mathsf{Z}_i\right)\right) \\
&= \frac{O(k^2 \mathsf{err}(Q)) + 4\eta(x^{(k)})}{p}.
\end{aligned} \tag{5.1}$$

To see how this implies the general result, consider some $x^{(k)}$ such that some of its coordinates are equal, so that in particular $\eta(x^{(k)}) \ge 1$. One still has the trivial bound

$$d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i=p+1}^{k} \mathsf{Z}_i\right)\right) \le \mathbf{E}_{x^{(k)}}\left[\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right] + \mathbf{E}\left[\sum_{i=p+1}^{k} \mathsf{Z}_i\right].$$

The second term on the RHS is $\le 2/p$. For the first term, let j be the number of distinct coordinates of x and

$$y^{(j)} = \big(y(1), \dots, y(j)\big) \in \mathbf{V}^j$$

have distinct coordinates with

$$\big\{y(1), \dots, y(j)\big\} = \big\{x(1), \dots, x(k)\big\}.$$

Then clearly,

$$\mathbf{E}_{x^{(k)}}\left[\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right] = \mathbf{E}_{y^{(j)}}\left[\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right].$$

If $p \ge j$, the RHS is 0. If not, it can be upper bounded using the bound in (5.1),

$$\begin{aligned}
\mathbf{E}_{y^{(j)}}\left[\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right] &\le \mathbf{E}\left[\sum_{i=p+1}^{k} \mathsf{Z}_i\right] + d_W\left(\mathrm{Law}_{y^{(j)}}\left(\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i=p+1}^{j} \mathsf{Z}_i\right)\right) \\
&\le \frac{2 + 4\eta(y^{(j)}) + O(k^2 \mathsf{err}(Q))}{p}.
\end{aligned}$$

Since $\eta(x^{(k)}) \ge 1 \ge \eta(y^{(j)})/2$ in this case, we obtain

$$d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}_p}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i=p+1}^{k} \mathsf{Z}_i\right)\right) \le \frac{12\eta(x^{(k)}) + O(k^2 \mathsf{err}(Q))}{p}$$

for such $x^{(k)}$ with repetitions, which gives the lemma in general.

We prove (5.1) by reverse induction on p. The case $p = k-1$ is trivial: C_{k-1} is simply $M^{(k)}$, and $\eta(x^{(k)})$ is an upper bound for $r_{\delta_{x^{(k)}}}$, so we may apply Lemma 4.2 to deduce the desired bound.

For the inductive step, consider $p_0 < k-1$, and assume the result is true for all $p_0 < p \le k-1$. We will use the easily proven fact that C_{p_0+1} is a stopping time for the process $(X_t^{(k)})_{t\ge 0}$ process. Consider the corresponding σ-field $\mathcal{F}_{\mathsf{C}_{p_0+1}}$. We will apply Lemma 2.3 with

$$\begin{aligned}
Z_1 &= \sum_{i=p_0+2}^{k} \mathsf{Z}_i,\\
Z_2 &= \mathsf{Z}_{p_0+1},\\
W_1 &= \frac{\mathsf{C}_{p_0+1}}{\mathsf{m}(Q)},\\
W_2 &= \frac{\mathsf{C}_{p_0} - \mathsf{C}_{p_0+1}}{\mathsf{m}(Q)} = \frac{\min_{\{i,j\}\subset A_{\mathsf{C}_{p_0+1}}} M_{i,j} \circ \Theta_{\mathsf{C}_{p_0+1}}}{\mathsf{m}(Q)},\\
\mathcal{G} &= \mathcal{F}_{\mathsf{C}_{p_0+1}}.
\end{aligned}$$

(We used Proposition 5.3 to obtain the second expression for W_2 above.) Applying Lemma 2.3 in conjunction with the induction hypothesis gives

$$\begin{aligned}
&d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}_{p_0}}{\mathsf{m}(Q)}\right), \sum_{i=p_0+1}^{k} \mathsf{Z}_i\right)\\
&\quad\le \frac{O(k^2\mathsf{err}(Q)) + 4\eta(x^{(k)})}{p_0+1}\\
&\qquad + \mathbf{E}_{x^{(k)}}\left[d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\min_{\{i,j\}\subset A_{\mathsf{C}_{p_0+1}}} M_{i,j}\circ\Theta_{\mathsf{C}_{p_0+1}}}{\mathsf{m}(Q)}\Big|\mathcal{F}_{C_{p_0+1}}\right),\right.\right.\\
&\qquad\qquad\left.\left. \mathsf{Z}_{p_0+1}\right)\right].
\end{aligned}\tag{5.2}$$

Note that $A_{\mathsf{C}_{p_0+1}}$ is $\mathcal{F}_{\mathsf{C}_{p_0+1}}$-measurable. The strong Markov property for $Q^{(k)}$ implies that

$$\mathrm{Law}_{x^{(k)}}\left(\frac{\min_{\{i,j\}\subset A_{\mathsf{C}_{p_0+1}}} M_{i,j}\circ\Theta_{\mathsf{C}_{p_0+1}}}{\mathsf{m}(Q)}\Big|\mathcal{F}_{C_{p_0+1}}\right)$$

is the same as

$$\mathrm{Law}_{X^{(k)}_{\mathsf{C}_{p_0+1}}}\left(\frac{\min_{\{i,j\}\subset A_{\mathsf{C}_{p_0+1}}} M_{i,j}}{\mathsf{m}(Q)}\right).$$

Now define $Y^{(p_0+1)}$ as the vectors whose coordinates are the p_0+1 distinct points $X_{\mathsf{C}_{p_0+1}}(i)$ with $i \in A_{p_0+1}$ (the order of the coordinates does not matter). Clearly,

$$\mathrm{Law}_{X^{(k)}_{\mathsf{C}_{p_0+1}}}\left(\frac{\min_{\{i,j\}\subset A_{\mathsf{C}_{p_0+1}}} M_{i,j}}{\mathsf{m}(Q)}\right) = \mathrm{Law}_{Y^{(p_0+1)}}\left(\frac{M^{(p_0+1)}}{\mathsf{m}(Q)}\right).\tag{5.3}$$

By Lemma 4.2, this last law is approximately exponential,

$$\mathsf{Exp}\left(\frac{1}{\binom{p_0+1}{2}}, O\big(k^2\mathsf{err}(Q)\big) + 2r_{\delta_{Y^{(p_0+1)}}}, O\big(k^2\mathsf{err}(Q)\big)\right),$$

and Lemma 3.1 gives

$$d_W\left(\mathrm{Law}_{Y^{(p_0+1)}}\left(\frac{M^{(p_0+1)}}{\mathsf{m}(Q)}\right), \mathsf{Z}_{p_0+1}\right) \le \frac{O(k^2\mathsf{err}(Q)) + 4r_{\delta_{Y^{(p_0+1)}}}}{p_0(p_0+1)}.$$

Using the definition of $r_{\delta_{Y^{(p_0+1)}}}$, we obtain from (5.2) the following inequality:

$$\begin{aligned}
&d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}_{p_0}}{\mathsf{m}(Q)}\right), \sum_{i=p_0+1}^{k} \mathsf{Z}_i\right) \\
&\quad \le \frac{O(k^2\mathsf{err}(Q)) + 4\eta(x^{(k)})}{p_0+1} \\
&\qquad + \frac{O(k^2\mathsf{err}(Q)) + 4\mathbf{E}_{x^{(k)}}[\mathbf{P}_{Y^{(p_0+1)}}(M^{(p_0+1)} \le t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2))]}{p_0(p_0+1)}.
\end{aligned} \tag{5.4}$$

To finish, we need to show that the expected value on the RHS is $\le \eta(x^{(k)})$. For this we recall (5.3) to note that

$$\begin{aligned}
&\mathbf{P}_{Y^{(p_0+1)}}\big(M^{(p_0+1)} \le t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\big) \\
&\quad = \mathbf{P}_{X^{(k)}_{\mathsf{C}_{p_0+1}}}\Big(\min_{\{i,j\}\subset A_{\mathsf{C}_{p_0+1}}} M_{i,j} \le t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\Big) \\
&\quad = \mathbf{P}_{x^{(k)}}\big(\mathsf{C}_{p_0} - \mathsf{C}_{p_0+1} \le t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2) \mid \mathcal{F}_{\mathsf{C}_{p_0+1}}\big),
\end{aligned}$$

where the last line uses Proposition 5.3 and the strong Markov property. Averaging shows that the expectation on the RHS of (5.4) is

$$\mathbf{P}_{x^{(k)}}\big(\mathsf{C}_{p_0} - \mathsf{C}_{p_0+1} \le t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\big),$$

and Proposition 5.3 implies that this is at most

$$\mathbf{P}_{x^{(k)}}\left(\bigcup_{\{i,j\}\neq\{\ell,r\}} \{M_{i,j} \circ \Theta_{M_{\ell,r}} \le t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\}\right).$$

Since the RHS is $\le \eta(x^{(k)})$, we are done. $\square$

6. Proofs of the main theorems.

6.1. *The full coalescence time in the transitive case.* In this section we prove Theorem 1.1.

PROOF OF THEOREM 1.1. Recall that $\mathsf{C} = \mathsf{C}_1$ by definition. Lemma 5.1 gives the following bound for any $k \leq \sqrt{1/4\mathsf{err}(Q)} \wedge |\mathbf{V}|$ and $x^{(k)} \in \mathbf{V}^k$:

$$d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}}{\mathsf{m}(Q)}\right), \sum_{i=2}^{k} \mathsf{Z}_i\right) \leq 12\eta(x^{(k)}) + O(k^2\mathsf{err}(Q)).$$

Notice that

$$d_W\left(\sum_{i=2}^{k} \mathsf{Z}_i, \sum_{i=2}^{+\infty} \mathsf{Z}_i\right) \leq \mathbf{E}\left[\sum_{j\geq k+1} \mathsf{Z}_j\right] = \frac{2}{k+1},$$

hence

$$d_W\left(\mathrm{Law}_{x^{(k)}}\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}\right), \sum_{i=2}^{+\infty} \mathsf{Z}_i\right) = 12\eta(x^{(k)}) + O\left(k^2\mathsf{err}(Q) + \frac{1}{k}\right).$$

Convexity of d_W implies

PROPOSITION 6.1. *Under the assumptions of Theorem 1.1, the following holds for* $\mathsf{err}(Q) \leq 1/4$, $1 \leq k \leq \sqrt{1/4\mathsf{err}(Q)} \wedge |\mathbf{V}|$ *and* $\lambda^{(k)} \in M_1(\mathbf{V}^k)$:

$$\begin{aligned} d_W\left(\mathrm{Law}_{\lambda^{(k)}}\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}\right), \sum_{i=2}^{+\infty} \mathsf{Z}_i\right) &\leq 12 \int \eta(x^{(k)})\, d\lambda^{(k)}(x^{(k)}) \\ &\quad + O\left(k^2\mathsf{err}(Q) + \frac{1}{k}\right). \end{aligned} \tag{6.1}$$

Notice that our control of C_1 gets worse as k increases, and we cannot use the above bound to approximate the law of C_1 started with one particle at each vertex of $\mathbf{V}$. What we use instead is a truncation argument combined with the Sandwich lemma for d_W (Lemma 2.2 above). For this we need to find two random variables

$$\mathsf{C}_- \preceq_d \mathsf{C}_1 \preceq_d \mathsf{C}_+$$

such that both $\mathsf{C}_-/\mathsf{m}(Q)$ and $\mathsf{C}_+/\mathsf{m}(Q)$ are close to $\sum_{i=2}^{+\infty} \mathsf{Z}_i$. More specifically, we will show that

$$d_W\left(\frac{\mathsf{C}_\pm}{\mathsf{m}(Q)}, \sum_{i\geq 2} \mathsf{Z}_i\right) = O\left(k^2\mathsf{err}(Q) + k^4\mathsf{err}(Q)^2 + \frac{1}{k+1} + \rho(Q)\ln(1/\rho(Q))\right).$$

Before we continue, let us show how this last bound implies our result. The Sandwich Lemma 2.2 gives

$$d_W\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}, \sum_{i\geq 2}\mathsf{Z}_i\right) = O\left(k^2\mathsf{err}(Q) + k^4\mathsf{err}(Q)^2 + \frac{1}{k} + \rho(Q)\ln\frac{1}{\rho(Q)}\right).$$

Since $\rho(Q)\ln(1/\rho(Q)) = O(\mathsf{err}(Q))$, we may choose $k = (\mathsf{err}(Q))^{-1/3}$ [which works for $\mathsf{err}(Q)$ sufficiently small] to obtain

$$d_W\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}, \sum_{i\geq 2}\mathsf{Z}_i\right) = O\bigl(\mathsf{err}(Q)^{1/3}\bigr),$$

and this is precisely the bound we seek because

$$\mathsf{err}(Q) = O\Bigl(\sqrt{\rho(Q)\ln\bigl(1/\rho(Q)\bigr)}\Bigr).$$

We now construct C_-, C_+ and prove that they have the required properties.

Construction of C_-: pick $x(1), \dots, x(k) \in \mathbf{V}$ from distribution π, independently and with replacement. Let C_- denote the full coalescence time for k walkers started from these positions. This might be degenerate: there might be more than one walker starting from some element of $\mathbf{V}$, but this only means those particles will coalesce instantly.

Clearly, $\mathsf{C}_- \preceq_d \mathsf{C}_1$. Moreover,

$$\mathrm{Law}\left(\frac{\mathsf{C}_-}{\mathsf{m}(Q)}\right) = \mathrm{Law}_{\pi^{\otimes k}}\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}\right).$$

Therefore by Proposition 6.1,

$$\begin{aligned} d_W\left(\mathrm{Law}\left(\frac{\mathsf{C}_-}{\mathsf{m}(Q)}\right), \sum_{i=2}^{+\infty}\mathsf{Z}_i\right) &= d_W\left(\mathrm{Law}_{\pi^{\otimes k}}\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}\right), \sum_{i=2}^{+\infty}\mathsf{Z}_i\right) \\ &= O\left(\int \eta\bigl(x^{(k)}\bigr)\,d\pi^{\otimes k} + k^2\mathsf{err}(Q) + \frac{1}{k}\right). \end{aligned}$$

Notice that the integral on the RHS is at most

$$\begin{aligned} \int \eta\bigl(x^{(k)}\bigr)\,d\pi^{\otimes k} &\leq \sum_{\{i,j\}\in\binom{[k]}{2}} \mathbf{P}_{\pi^{\otimes k}}\bigl(M_{i,j} \leq t_{\mathrm{mix}}^{Q}\bigl(\mathsf{err}(Q)^2\bigr)\bigr) \\ &\quad + \sum_{\substack{\{i,j\},\{\ell,r\}\in\binom{[k]}{2}: \\ \{i,j\}\neq\{\ell,r\}}} \mathbf{P}_{\pi^{\otimes k}}\left(\begin{array}{c} M_{i,j}\circ\Theta_{M_{\ell,r}} \\ \leq t_{\mathrm{mix}}^{Q}\bigl(\mathsf{err}(Q)^2\bigr)\end{array}\right) \\ &= O\bigl(k^4\mathsf{err}(Q)^2\bigr) \end{aligned} \tag{6.2}$$

as can be deduced from the proofs of Propositions 4.2 and 4.1. We conclude that

$$
\begin{aligned}
(6.3)\qquad & d_W\left(\mathrm{Law}\left(\frac{\mathsf{C}_-}{\mathsf{m}(Q)}\right), \mathrm{Law}\left(\sum_{i\geq 2}\mathsf{Z}_i\right)\right) \\
& = O\left(k^4\mathsf{err}(Q)^2 + k^2\mathsf{err}(Q) + \frac{1}{k}\right).
\end{aligned}
$$

Construction of C_+: we will use the following simple stochastic domination result, which we describe in the language of the process with killings. Let $\tau \leq \sigma$ be stopping times for the $X^{(k)}$ process. If all killings are suppressed between time τ and σ, the resulting full coalescence time C_+ stochastically dominates C_1. We will use this result, whose proof we omit, with the following choice of τ and σ:

$$\tau = \mathsf{C}_k \quad \text{and} \quad \sigma = \mathsf{C}_k + t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2).$$

Lemma 2.1 implies

$$d_W\left(\frac{\mathsf{C}_+}{\mathsf{m}(Q)}, \frac{\mathsf{C}_1\circ\Theta_\sigma}{\mathsf{m}(Q)}\right) \leq \frac{\mathbf{E}[\sigma]}{\mathsf{m}(Q)} = \frac{\mathbf{E}[\mathsf{C}_k]}{\mathsf{m}(Q)} + \frac{t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)}{\mathsf{m}(Q)}.$$

Since Q is transitive, $\mathsf{m}(Q)$ can be bounded from below in terms of the maximal hitting time in Q [2], Chapter 14. Theorem 1.2 in [14] implies

$$\mathbf{E}[\mathsf{C}_k] \leq \frac{C\mathsf{m}(Q)}{k} + Ct^Q_{\mathrm{mix}}$$

for some universal $C > 0$. Recalling the definition of $\rho(Q)$ in (4.3), we obtain

$$\frac{\mathbf{E}[\mathsf{C}_k]}{\mathsf{m}(Q)} = O\left(\frac{1}{k} + \rho(Q)\right).$$

Moreover, we also have

$$t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2) = O\big(\ln(1/\mathsf{err}(Q))t^Q_{\mathrm{mix}}\big) = O\big(t^Q_{\mathrm{mix}}\ln(1/\rho(Q))\big),$$

hence

$$d_W\left(\frac{\mathsf{C}_+}{\mathsf{m}(Q)}, \frac{\mathsf{C}_1\circ\Theta_\sigma}{\mathsf{m}(Q)}\right) = O\left(\frac{1}{k} + \rho(Q)\ln\big(1/\rho(Q)\big)\right).$$

This shows

$$d_W\left(\frac{\mathsf{C}_+}{\mathsf{m}(Q)}, \sum_{i=2}^k \mathsf{Z}_i\right) = O\left(\frac{1}{k} + \rho(Q)\ln\big(1/\rho(Q)\big)\right) + d_W\left(\frac{\mathsf{C}_1\circ\Theta_\sigma}{\mathsf{m}(Q)}, \sum_{i=2}^k \mathsf{Z}_i\right).$$

Now consider the time $\mathsf{C}_1\circ\Theta_\sigma$. Since all killings were suppressed between times $\tau = \mathsf{C}_k$ and $\sigma = \mathsf{C}_k + t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)$, there are k alive particles at time σ_-. Letting $\lambda^{(k)}$ denote their law, we have

$$\mathrm{Law}\left(\frac{\mathsf{C}_1\circ\Theta_\sigma}{\mathsf{m}(Q)}\right) = \mathrm{Law}_{\lambda^{(k)}}\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}\right),$$

and Proposition 6.1 implies

$$d_W\left(\frac{\mathsf{C}_1 \circ \Theta_\sigma}{\mathsf{m}(Q)}, \sum_{i=2}^{k} \mathsf{Z}_i\right) = O\left(\int \eta(x^{(k)})\, d\lambda^{(k)}(x^{(k)}) + k^2 \mathsf{err}(Q) + \frac{1}{k}\right).$$

Now observe that

$$t_{\mathrm{mix}}^{Q}(\mathsf{err}(Q)^2) \geq t_{\mathrm{mix}}^{Q^{(k)}}(k\mathsf{err}(Q)^2) \qquad \text{(cf. Lemma 2.4)},$$

hence the law of the k particles at time $\mathsf{C}_k + t_{\mathrm{mix}}^{Q}(\mathsf{err}(Q)^2)$ is $k\mathsf{err}(Q)^2$-close to stationary, irrespective of their states at time C_k. We deduce that $\lambda^{(k)}$ is $k\mathsf{err}(Q)^2$-close to stationary, and

$$d_W\left(\frac{\mathsf{C}_1 \circ \Theta_\sigma}{\mathsf{m}(Q)}, \sum_{i=2}^{k} \mathsf{Z}_i\right) = O\left(\int \eta(x^{(k)})\, d\pi^{\otimes k} + k^2 \mathsf{err}(Q) + k\mathsf{err}(Q)^2 + \frac{1}{k}\right).$$

The integral on the RHS was estimated in (6.2), and we deduce

$$d_W\left(\frac{\mathsf{C}_1 \circ \Theta_\sigma}{\mathsf{m}(Q)}, \sum_{i=2}^{k} \mathsf{Z}_i\right) = O\left(k^2 \mathsf{err}(Q) + k^4 \mathsf{err}(Q)^2 + \frac{1}{k}\right),$$

and we deduce

$$d_W\left(\frac{\mathsf{C}_+}{\mathsf{m}(Q)}, \sum_{i=2}^{k} \mathsf{Z}_i\right) = O\left(k^2 \mathsf{err}(Q) + k^4 \mathsf{err}(Q)^2 + \frac{1}{k} + \rho(Q)\ln(1/\rho(Q))\right). \qquad \square$$

6.2. *The general setting.* We now come to the proof of Theorem 1.2.

PROOF OF THEOREM 1.2. The proof is essentially the same as in the reversible/transitive case, but with the definition of $\mathsf{err}(Q)$ given in (4.5). In particular, we can still use the same definition of C_- used in that proof to obtain

$$d_W\left(\frac{\mathsf{C}_-}{\mathsf{m}(Q)}, \sum_{i=2}^{+\infty} \mathsf{Z}_i\right) = O\left(k^2 \mathsf{err}(Q) + k^4 \mathsf{err}(Q)^2 + \frac{1}{k}\right). \tag{6.4}$$

We will need a different strategy in the analysis of C_+, where we need to bound $\mathbf{E}[\mathsf{C}_k]$ by different means. Note that $\mathsf{C}_k \geq t$ if and only if there exist distinct $y(1), \dots, y(k) \in \mathbf{V}$ such that there is *no coalescence* among the walkers started from these vertices. The probability of this "no coalescence event" for a given choice of $y(i)$'s is $\mathbf{P}_{y^{(k)}}(M^{(k)} \geq t)$ for $y^{(k)} = (y(1), \dots, y(k))$. Therefore,

$$\mathbf{P}(\mathsf{C}_k \geq t) \leq \left(\sum_{y^{(k)} \in \mathbf{V}^k} \mathbf{P}_{y^{(k)}}\left(M^{(k)} \geq t\right)\right) \wedge 1.$$

By Lemma 4.2, each term in the RHS satisfies

$$\mathbf{P}_{y^{(k)}}\left(M^{(k)} \geq t\right) \leq C e^{-t\binom{k}{2}/((1+O(k^2\mathsf{err}(Q)))\mathsf{m}(Q))}$$

for some universal $C > 0$. Since there are $\leq |\mathbf{V}|^k$ terms in the sum, we have

$$\mathbf{P}(\mathsf{C}_k \geq t) \leq \left(C|\mathbf{V}|^k e^{-t\binom{k}{2}/((1+O(k^2\mathsf{err}(Q)))\mathsf{m}(Q))}\right) \wedge 1.$$

Integrating the RHS gives

$$\frac{\mathbf{E}[\mathsf{C}_k]}{\mathsf{m}(Q)} \leq C\frac{\ln|\mathbf{V}|}{k}$$

for a potentially different, but still universal C. Going through the previous proof, we see that this gives

$$\begin{aligned} &d_W\left(\frac{\mathsf{C}_+}{\mathsf{m}(Q)}, \sum_{i=2}^{k} \mathsf{Z}_i\right) \\ &\quad = O\left(k^2\mathsf{err}(Q) + k^4\mathsf{err}(Q)^2 + \frac{\ln|\mathbf{V}|}{k} + \frac{t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)}{\mathsf{m}(Q)}\right). \end{aligned} \tag{6.5}$$

To continue, we bound the term containing $t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)$ in terms of $\mathsf{err}(Q)$ [this was easier before because of the different definition of $\mathsf{err}(Q)$]. Recall from Proposition 4.4 that

$$\mathbf{P}_{\pi^{\otimes 2}}\left(M \leq t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\right) \leq \mathsf{err}(Q)^2.$$

Therefore, for all $j \in \mathbf{N}$,

$$\begin{aligned} &\mathbf{P}_{\pi^{\otimes 2}}\left(M \leq j t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\right) \\ &\quad \leq \sum_{i=1}^{j} \mathbf{P}_{\pi^{\otimes 2}}\left(M \circ \Theta_{(i-1)t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)} t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\right) \\ &\quad \leq j\,\mathsf{err}(Q)^2. \end{aligned}$$

On the other hand, taking

$$j = \left\lceil \frac{2\mathbf{E}_{\pi^{\otimes 2}}[M]}{t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)} \right\rceil,$$

we obtain

$$\mathbf{P}_{\pi^{\otimes 2}}\left(M \leq j t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)\right) \geq 1 - \frac{\mathbf{E}_{\pi^{\otimes 2}}[M]}{j t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)} \geq \frac{1}{2}.$$

Combining these two inequalities gives

$$\frac{t^Q_{\mathrm{mix}}(\mathsf{err}(Q)^2)}{\mathbf{E}_{\pi^{\otimes 2}}[M]} = O\left(\mathsf{err}(Q)^2\right).$$

This implies that the term containing $t^{Q}_{\mathrm{mix}}(\mathsf{err}(Q)^2)$ on the RHS of (6.5) can be neglected. Combining that equation with (6.4) and the Sandwich Lemma 2.2, we obtain

$$d_W\left(\frac{\mathsf{C}_1}{\mathsf{m}(Q)}, \sum_{i\geq 2}\mathsf{Z}_i\right) = O\left(k^2\mathsf{err}(Q) + k^4\mathsf{err}(Q)^2 + \frac{\ln|\mathbf{V}|}{k}\right).$$

We choose $k = \lceil(\ln|\mathbf{V}|/\mathsf{err}(Q))^{1/3}\rceil$ to finish the proof, at least if this is smaller than $1/5\sqrt{\mathsf{err}(Q)}$. But the bound in the theorem is trivial if that is not the case, so we are done. □

7. Final remarks.

- Cooper et al. [6] consider many other processes besides coalescing random walks. It is not hard to modify our analysis to study those processes over more general graphs, at least when the initial number of random walks is not too large (this restriction is also present in [6]).
- Our Theorems 3.1 and 3.2 can be used to study other problems related to hitting times. Alan Prata and the present author [16] have used these results to prove the Gumbel law for the fluctuations of cover times for a large family of graphs, including all examples where it was previously known. We have also used extensions of these results to compute the asymptotic distribution of the k last points to be visited, for any constant k: those are uniformly distributed over the graph, as conjectured by Aldous and Fill [2].

APPENDIX: PROOFS OF TECHNCAL RESULTS ON L_1 WASSERSTEIN DISTANCE

A.1. Proof of Sandwich lemma (Lemma 2.2). Notice that for all $t \in \mathbf{R}$,

$$\mathbf{P}(Z_- \geq t) \leq \mathbf{P}(Z \geq t) \leq \mathbf{P}(Z_+ \geq t).$$

By convexity, this implies

$$\begin{aligned}\left|\mathbf{P}(Z \geq t) - \mathbf{P}(W \geq t)\right| \leq{}& \left|\mathbf{P}(Z_- \geq t) - \mathbf{P}(W \geq t)\right| \\ &+ \left|\mathbf{P}(Z_+ \geq t) - \mathbf{P}(W \geq t)\right|.\end{aligned}$$

Integrate both sides to obtain the result.

A.2. Proof of conditional lemma (Lemma 2.3). First notice that the sigma field $\sigma(W_1)$ generated by W_1 is contained in $\mathcal{G}$. This implies that for all $t \in \mathbf{R}$,

$$\begin{aligned}&\mathbf{E}\left[\left|\mathbf{P}(W_2 \geq t \mid \mathcal{G}) - \mathbf{P}(Z_2 \geq t)\right|\right] \\ &\quad = \mathbf{E}\left[\mathbf{E}\left[\left|\mathbf{P}(W_2 \geq t \mid \mathcal{G}) - \mathbf{P}(Z_2 \geq t)\right| \middle| \sigma(W_1)\right]\right] \\ &\quad \geq \mathbf{E}\left[\left|\mathbf{P}\big(W_2 \geq t \mid \sigma(W_1)\big) - \mathbf{P}(Z_2 \geq t)\right|\right].\end{aligned}$$

Integrating both sides in t and applying Fubini–Tonelli gives

$$\mathbf{E}\big[d_W\big(\mathrm{Law}(W_2 \mid \mathcal{G}), \mathrm{Law}(Z_2)\big)\big] \geq \mathbf{E}\big[d_W\big(\mathrm{Law}\big(W_2 \mid \sigma(W_1)\big), \mathrm{Law}(Z_2)\big)\big].$$

Therefore it suffices to prove the theorem in the case $\mathcal{G} = \sigma(W_1)$. For simplicity, we will assume that (Z_1, Z_2, W_1, W_2) are all defined in the same probability space, with (Z_1, Z_2) independent from (W_1, W_2). Let $f : \mathbf{R} \to \mathbf{R}$ be 1-Lipschitz. We have

$$\mathbf{E}\big[f(W_1 + W_2) \mid W_1 = w_1\big] = \int f(w_1 + w_2)\mathbf{P}(W_2 \in dw_2 \mid W_1 = w_1).$$

By the duality version of d_W, we have

$$\begin{aligned}\int & f(w_1 + w_2)\mathbf{P}(W_2 \in dw_2 \mid W_1 = w_1) \\ &\leq \int f(w_1 + z_2)\mathbf{P}(Z_2 \in dz_2) + d_W\big(\mathrm{Law}(W_2 \mid W_1 = w_1), \mathrm{Law}(Z_2)\big).\end{aligned}$$

Integrating over $W_1 = w_1$ and using the fact that Z_2 is independent from W_1, we obtain

$$\mathbf{E}\big[f(W_1 + W_2)\big] \leq \mathbf{E}\big[f(W_1 + Z_2)\big] + d_W\big(\mathrm{Law}(W_2 \mid W_1), \mathrm{Law}(Z_2)\big).$$

But we also have

$$\mathbf{E}\big[f(W_1 + Z_2) \mid Z_2 = z_2\big] = \mathbf{E}\big[f(W_1 + z_2)\big] \leq \mathbf{E}\big[f(Z_1 + z_2)\big] + d_W(W_1, Z_1),$$

and the independence of Z_1, Z_2 implies

$$\mathbf{E}\big[f(W_1 + Z_2)\big] \leq \mathbf{E}\big[f(Z_1 + Z_2)\big] + d_W(W_1, Z_1).$$

We conclude

$$\begin{aligned}\mathbf{E}\big[f(W_1 + W_2)\big] \leq \mathbf{E}\big[f(Z_1 + Z_2)\big] &+ d_W(W_1, Z_1) \\ &+ d_W\big(\mathrm{Law}(W_2 \mid W_1), \mathrm{Law}(Z_2)\big).\end{aligned}$$

Since f is an arbitrary 1-Lipschitz function, we are done.

Acknowledgment. We warmly thank the anonymous referee for pointing out several typos in a previous versions of this paper.

REFERENCES

[1] ALDOUS, D. (2010). Mixing times and hitting times. Available at http://www.stat.berkeley.edu/~aldous/Talks/slides.html.

[2] ALDOUS, D. and FILL, J. A. (2001). Reversible Markov chains and random walks on graphs. Available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[3] ALDOUS, D. J. (1982). Markov chains with almost exponential hitting times. *Stochastic Process. Appl.* **13** 305–310. MR0671039

[4] ALDOUS, D. J. and BROWN, M. (1992). Inequalities for rare events in time-reversible Markov chains. I. In *Stochastic Inequalities* (*Seattle*, *WA*, 1991). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **22** 1–16. IMS, Hayward, CA. MR1228050

[5] BENJAMINI, I. and MOSSEL, E. (2003). On the mixing time of a simple random walk on the super critical percolation cluster. *Probab. Theory Related Fields* **125** 408–420. MR1967022

[6] COOPER, C., FRIEZE, A. and RADZIK, T. (2009). Multiple random walks in random regular graphs. *SIAM J. Discrete Math*. **23** 1738–1761. MR2570201

[7] COX, J. T. (1989). Coalescing random walks and voter model consensus times on the torus in $\mathbb{Z}^d$. *Ann. Probab.* **17** 1333–1366. MR1048930

[8] DURRETT, R. (2007). *Random Graph Dynamics*. Cambridge Univ. Press, Cambridge. MR2271734

[9] DURRETT, R. (2010). Some features of the spread of epidemics and information on a random graph. *Proc. Natl. Acad. Sci. USA* **107** 4491–4498.

[10] FOUNTOULAKIS, N. and REED, B. A. (2008). The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures Algorithms* **33** 68–86. MR2428978

[11] LESKOVEC, J., LANG, K. J., DASGUPTA, A. and MAHONEY, M. W. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Math*. **6** 29–123. MR2736090

[12] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[13] LIGGETT, T. M. (1985). *Interacting Particle Systems*. *Grundlehren der Mathematischen Wissenschaften* **276**. Springer, New York. MR0776231

[14] OLIVEIRA, R. I. (2012). On the coalescence time of reversible random walks. *Trans. Amer. Math. Soc*. **364** 2109–2128. MR2869200

[15] PETE, G. (2008). A note on percolation on $\mathbb{Z}^d$: Isoperimetric profile via exponential cluster repulsion. *Electron. Commun. Probab.* **13** 377–392. MR2415145

[16] PRATA, A. (2012). Stochastic processes over finite graphs. Ph.D. thesis in Mathematics. IMPA, Rio de Janeiro, Brazil.

IMPA
ESTRADA DONA CASTORINA, 110
RIO DE JANEIRO, RJ
BRAZIL
E-MAIL: rimfo@impa.br