THE NOISY VOTER MODEL WITH GENERAL INITIAL CONDITIONS

PATRIZIO CADDEO AND EYAL LUBETZKY

ABSTRACT. We study the noisy voter model with $q \geq 2$ states and noise probability θ on arbitrary bounded-degree n-vertex graphs G with subexponential growth of balls (e.g., finite subsets of $\mathbb{Z}^d$). Cox, Peres and Steif (2016) showed for the binary case $q = 2$ (and a wider class of chains) that, when starting from a worst-case initial state, this Markov chain has total variation cutoff at $t_n = \frac{1}{2\theta} \log n$. The second author and Sly (2021) analyzed faster initial conditions for Glauber dynamics for the 1D Ising model, which is the noisy voter for $q = 2$ and $G = \mathbb{Z}/n\mathbb{Z}$. They showed that the "alternating" initial state is the fastest one if $\theta \geq \frac{2}{3}$, and conjectured that this holds for all values of the noise θ.

Here we show that for every graph G as above and all θ, q and initial states x_0, the noisy voter model exhibits cutoff at an explicit function of the autocorrelation of the model started at x_0. Consequently, for $G = \mathbb{Z}/n\mathbb{Z}$ and $q = 2$ (Glauber dynamics for the 1D Ising model), we confirm the conjecture of [LS21] that the alternating initial condition is asymptotically fastest for all θ. Analogous results hold in $\mathbb{Z}_n^d$ for $q = 2$ and all $d \geq 1$ ("checkerboard" initial conditions are fastest) as well as for $d = 1$ and all $q \geq 2$ ("rainbow" initial conditions are fastest).

1. INTRODUCTION

The noisy voter model on a connected graph $G = (V, E)$, with noise parameter $0 < \theta < 1$ and $q \geq 2$ states, is the Markov chain (X_t) on the space $\Omega = \{0, \ldots, q-1\}^V$ (assigning one of q different states, or opinions, to each vertex), evolving in the following way. An independent rate-1 Poisson clock is attached to every $u \in V$. Once the clock at u rings, the value of $X_t(u)$ is updated as follows:
(i) with probability $1 - \theta$, the new value is $X_t(v)$ for a uniformly chosen neighbor $v \sim u$;
(ii) with probability θ, the new value is uniformly chosen out of $\{0, \ldots, q-1\}$.
Let μ_G be the stationary distribution of this Markov chain (ergodic but not necessarily reversible).

The classical voter model is the special case $\theta = 0$, and $q = 2$ (noiseless, binary), introduced in the 1970's independently by Clifford and Sudbury [5] and by Holley and Liggett [11] to model the evolution systems with two competing population types; see, e.g., [13] for the rich literature on it. The noisy voter model was introduced by Granovsky and Madras [10] in 1995, where properties of the equilibrium μ_G were studied (cf. the recent works [4, 21] on features of the stationary measure). Our focus is on the asymptotic time it takes (X_t) to converge to μ_G in total variation distance.

Denote the total variation mixing time to within distance ε from μ_G by

$$t_{\text{MIX}}^{x_0}(\varepsilon) = \inf \left\{ t \geq 0 \, : \, \|\mathbb{P}_{x_0}(X_t \in \cdot) - \mu_G\|_{\text{TV}} < \varepsilon \right\} \quad \text{and} \quad t_{\text{MIX}}(\varepsilon) = \max_{x_0} t_{\text{MIX}}^{x_0}(\varepsilon) \,.$$

In the context of a sequence of finite Markov chains $\{X_t^{(n)}\}$, the *cutoff phenomenon*, discovered by Aldous and Diaconis in the 1980's [1, 2, 8, 7], refers to the situation where there exists a sequence t_n such that, for every fixed $0 < \varepsilon < 1$, one has $t_{\text{MIX}}(\varepsilon)/t_n \to 1$ as $n \to \infty$. That is, a sharp transition along a $o(t_n)$ window occurs, whereby the distance from equilibrium drops from near 1 to near 0. When cutoff occurs at t_n for $t_{\text{MIX}}^{x_0}$—that is, starting from a sequence $\{x_0^{(n)}\}$ of initial conditions—the asymptotic mixing time from x_0 is well-defined as $(1+o(1))t_n$ independently of the choice of ε, and one can compare it to other sequences of initial configurations and characterize which is the fastest. We will establish cutoff for the noisy voter from any x_0 on G with subexponential growth of balls.
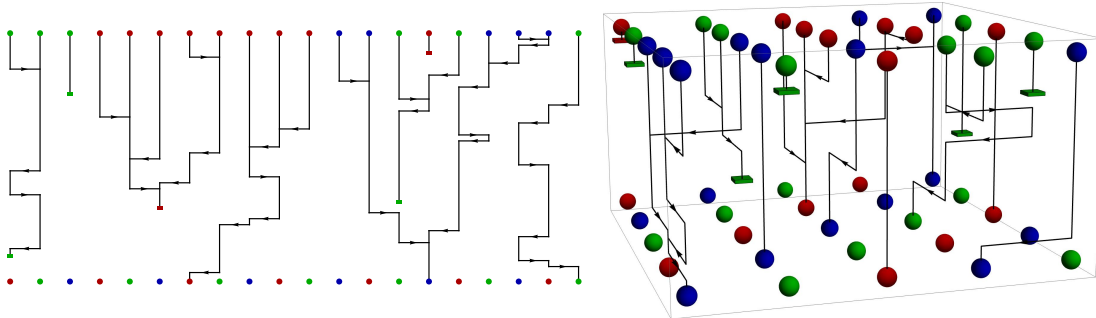
FIGURE 1. Noisy voter model for $q = 3$ on the lattice in dimensions 1 and 2, when revealing the update history (from the final (top) state X_t back in time to the initial state $X_0 = \mathsf{x}_0$).

Consider the binary case $q = 2$. Cox, Peres and Steif [6] proved in 2016 that for every sequence of bounded-degree graphs G on n vertices and fixed noise parameter $0 < \theta < 1$, the noisy voter model has $t_{\text{MIX}}(\varepsilon) = (\frac{1}{2\theta} + o(1)) \log n$ for every fixed $0 < \varepsilon < 1$; that is, when started at a worst-case initial state, there is cutoff at $t_n = \frac{1}{2\theta} \log n$ (in fact [6] allowed a wider family of rules for selecting the neighbor $v \sim u$ in the voting update of Step (i) above). Their lower bound on t_{MIX}—the asymptotically worst initial condition x_0—was realized by a monochromatic (all-0 or all-1) state.

In the special case of $q = 2$ and $G = \mathbb{Z}/n\mathbb{Z}$ (the n-cycle), the noisy voter model is well-known to be equivalent to Glauber dynamics for the Ising model on the 1D torus at inverse-temperature β, via the correspondence[1] $\theta = 1 - \tanh(2\beta)$. (NB. deleting but a single edge from G would break this equivalence, and make the stationary distribution of the noisy voter model be non-reversible.) Thus, said result of [6] recovers the 1D special case of the result [14] of the second author and Sly, where it was shown that there is cutoff for the Ising model on the torus $(\mathbb{Z}/n\mathbb{Z})^d$ in any dimension d, at high temperature (covering all β at $d = 1$ and all $\beta < \beta_c$ at $d = 2$; its companion paper [15] generalized the method to arbitrary bounded-degree graphs with sub-exponential growth rate). A later work [17] by the same authors established cutoff for the Ising model on $(\mathbb{Z}/n\mathbb{Z})^d$ for all $\beta < \beta_c$ in any dimension d, using a different approach dubbed there *information percolation* (see also [16, 18]) that carefully studies the space-time clusters of the nearest-neighbor interaction, rolled backward in time. (See Fig. 1 for an illustration for noisy voter and Section 2.1.1 for more details.) This opened the door to studying $t_{\text{MIX}}^{\mathsf{x}_0}$ for non-worst-case initial conditions x_0, where very little was previously known; e.g., it was shown there that for $d = 1$, a $(1 - o(1))$-fraction of the initial states x_0 is asymptotically as slow as the worst-case one, whereas a random i.i.d. initial condition is asymptotically twice faster.

Thereafter, the second author and Sly [19] investigated, for $q = 2$ and $G = \mathbb{Z}/n\mathbb{Z}$ (i.e., Glauber dynamics for the 1D Ising model at inverse-temperature β, equivalent to the noisy voter on $\mathbb{Z}/n\mathbb{Z}$ with $q = 2$ and $\theta = 1 - \tanh(2\beta)$), mixing from specific deterministic initial conditions, and showed:

(a) started from the alternating $\mathsf{x}_{\text{alt}} = (0, 1, 0, 1, \ldots)$, there is cutoff at $t_n = \max\{\frac{1}{4-2\theta}, \frac{1}{4\theta}\} \log n$;

(b) started from the bi-alternating $\mathsf{x}_{\text{blt}} = (0, 0, 1, 1, \ldots)$, there is cutoff at $t_n = \max\{\frac{1}{2}, \frac{1}{4\theta}\} \log n$ (matching the mixing time from x_{alt} for $\theta \leq \frac{1}{2}$ but strictly dominated by it for $\theta > \frac{1}{2}$);

(c) the initial state x_{alt} is asymptotically fastest at high enough temperature (specifically, if $\theta \geq \frac{2}{3}$).

The authors of [19] conjectured that x_{alt} is asymptotically fastest at every temperature (for all θ).

1.1. **Results.** We show that the noisy voter model for every fixed $q \geq 2$ exhibits cutoff under general starting states, in arbitrary families of graphs with a sub-exponential growth rate of balls (e.g., subsets of the lattice). Specifically, cutoff occurs at an explicit function of the autocorrelation

---

[1]In an update of Glauber dynamics for the Ising model, if a site x has a neighbors with $+$ spins and b neighbors with $-$ spins, the probability that the dynamics sets x to $+$ is $(1 + \tanh(\beta\sigma))/2$ for $\sigma = a - b$, hence the equivalence.

of the model started at x_0. Consequently, we confirm the conjecture of the second author and Sly that, in the special case of $G = \mathbb{Z}/n\mathbb{Z}$, the initial state $\mathsf{x}_{\mathrm{alt}}$ is asymptotically fastest for all θ (or β). More generally, for the noisy voter model on $(\mathbb{Z}/n\mathbb{Z})^d$ with $q = 2$ and any $d \geq 1$, the "checkerboard" initial conditions are fastest; for $q \geq 2$ and $d = 1$, the "rainbow" initial conditions are fastest.

Let $G = (V, E)$ be a finite graph on n vertices. Our assumption on its geometry is the following *sub-exponential growth of balls* property: there exist constants $\mathfrak{c}_0 > 0$ and $\alpha \in (0, 1)$ such that

$$|B_v(r)| \leq \mathfrak{c}_0 \exp(r^{1-\alpha}) \quad \text{for every } v \in V \text{ and } r \geq 1, \tag{1.1}$$

where $B_v(r)$ denotes the ball of radius r, centered at the vertex v, with respect to graph distance. Consider the noisy voter model on such a graph G with q states and noise probability $\theta \in (0, 1]$, and let x_0 be an arbitrary initial configuration. Denoting the degree of a vertex $v \in V$ by $d_G(v)$ and letting $\pi_G(v) = d_G(v)/(2|E|)$ be the stationary measure of a simple random walk on G, we define the *autocorrelation* function of the noisy voter model

$$\mathcal{A}_t^{(1)}(\mathsf{x}_0) = \sum_v \pi_G(v) \Big(\mathbb{P}_{\mathsf{x}_0}(X_t(v) = \mathsf{x}_0(v)) - \frac{1}{q} \Big), \tag{1.2}$$

as well as an L^2 analog of it which will turn out to be nothing but a time-rescaled version of $\mathcal{A}_t^{(1)}$:

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \sum_{v \in V} \pi_G(v) \sum_{j=0}^{q-1} \Big(\mathbb{P}_{\mathsf{x}_0}(X_t(v) = j) - \frac{1}{q} \Big)^2. \tag{1.3}$$

Indeed, despite their seemingly different formulation, one has $\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \mathcal{A}_{2t}^{(1)}(\mathsf{x}_0)$ (see Eq. (2.14)); hence, we will refer to both objects as the autocorrelation function (see Section 2 for more details). A key quantity for the behavior of $t_{\mathrm{MIX}}^{\mathsf{x}_0}$ is the time T_{x_0} required for $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ to drop below $\frac{1}{n}$:

$$T_{\mathsf{x}_0} = \frac{1}{2} \inf \Big\{ t \geq 0 \,:\, \mathcal{A}_t^{(1)}(\mathsf{x}_0) \leq \frac{1}{n} \Big\} = \inf \Big\{ t \geq 0 \,:\, \mathcal{A}_t^{(2)}(\mathsf{x}_0) \leq \frac{1}{n} \Big\}. \tag{1.4}$$

Our main theorem states that waiting for the maximum between T_{x_0} and $\frac{1}{4\theta} \log n$—which will correspond to a quantity T_{corr} as explained later (see Section 1.2)—is both necessary and sufficient for the model to be mixed. That is, $\max\{T_{\mathsf{x}_0}, T_{\mathrm{corr}}\}$ gives the correct asymptotic mixing time on any family of graphs with limited expansion, under arbitrary initial conditions x_0.

Theorem 1.1. *Fix $\theta \in (0, 1]$ and $q \geq 2$. Let $G = (V, E)$ be a connected graph on n vertices satisfying Eq. (1.1) for some $\mathfrak{c}_0 > 0$ and $\alpha \in (0, 1)$, and let x_0 be an arbitrary initial condition. Then the q-state noisy voter model on G and noise probability θ started from x_0 satisfies*

$$t_{\mathrm{MIX}}^{\mathsf{x}_0}(\varepsilon) = (1 + o(1)) \max \Big\{ T_{\mathsf{x}_0}, \frac{1}{4\theta} \log n \Big\}, \tag{1.5}$$

where the $o(1)$-term is $O\big(\log^{-\alpha/2} n \big)$ with a constant depending on $\varepsilon, \theta, q, \mathfrak{c}_0, \alpha$ (but not G or x_0).

In particular, for any sequence of connected graphs $G^{(n)}$ with n vertices satisfying Eq. (1.1) for the same $\alpha, \mathfrak{c}_0$, and any sequence of initial conditions $\mathsf{x}_0^{(n)}$, the q-state noisy voter model on $G^{(n)}$ with noise probability θ has cutoff at $\max\{T_{\mathsf{x}_0^{(n)}}, \frac{1}{4\theta} \log n\}$.

Theorem 1.1 shows that a dynamical phase transition in the parameter θ can occur under certain initial conditions x_0: we will see that in the case of $G = \mathbb{Z}_n^d$, if x_0 is chosen by repeating a balanced pattern of the q colors, then $T_{\mathsf{x}_0} \sim \frac{1}{2[(1-\lambda)+\lambda\theta]} \log n$ for some $\lambda < 1$ (here and in what follows, we let $f_n \sim g_n$ denote $f_n = (1 + o(1))g_n$). It follows that t_{MIX} changes from $\frac{1}{4\theta} \log n$ to $\frac{1}{2[(1-\lambda)+\lambda\theta]} \log n$ at some $\theta_0 \in (0, 1)$, implying a singularity in the analytical behavior of $t_{\mathrm{MIX}}^{\mathsf{x}_0}$ as a function of θ.
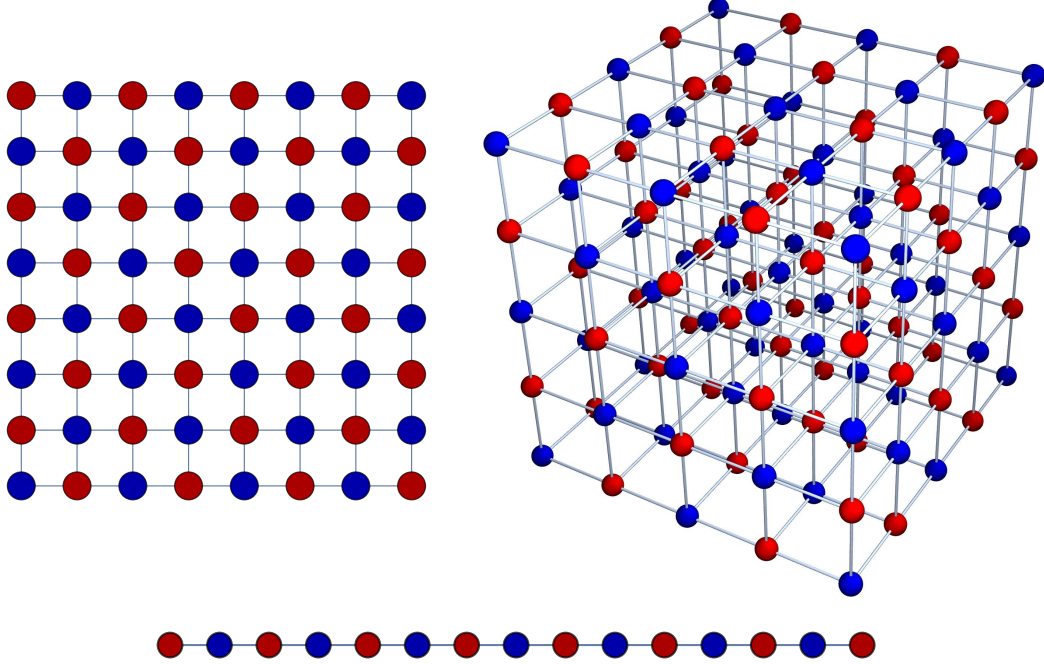
FIGURE 2. When $q = 2$, alternating/checkerboard starting configurations are asymptotically the fastest in all dimensions.

The full Eq. (1.5) was only known [19] for the 1D Ising model on $\mathbb{Z}_n$ with periodic boundary conditions and specific initial conditions x_0. While [19] showed that $t_{\text{MIX}}^{\mathsf{x}_0} \geq \frac{1-o(1)}{4-2\theta} \log n$ for all x_0 (implying, via their analysis of the alternating initial condition $\mathsf{x}_{\text{alt}} = (0, 1, 0, 1, \dots) \in \mathbb{Z}_n$ that it is asymptotically fastest for $\theta \geq \frac{2}{3}$), the fact that one also has $t_{\text{MIX}}^{\mathsf{x}_0} \geq \frac{1-o(1)}{4\theta} \log n$ for all x_0 (implying that x_{alt} is always asymptotically fastest) was left there as a conjecture. Here not only do we confirm this, but we extend the result to the noisy voter model on any bipartite graph with $n = |V|$ large satisfying Eq. (1.1) for some $\alpha, \mathfrak{c}_0$. Indeed, if G is bipartite with parts V_1, V_2, we can define

$$\mathsf{x}_{\text{alt}}(v) := \mathbb{1}_{\{v \in V_1\}} \tag{1.6}$$

(see Fig. 2 for an illustration on $\mathbb{Z}^d$) and have the following.

Corollary 1.2. *Fix $\theta \in (0, 1]$. Let $G^{(n)}$ be a sequence of bipartite graphs on n vertices satisfying Eq. (1.1) for the same $\alpha, \mathfrak{c}_0$. Let $\mathsf{x}_{\text{alt}} = \mathsf{x}_{\text{alt}}^{(n)}$ denote the alternating condition on $G^{(n)}$ as in Eq. (1.6). Then, for every fixed $0 < \varepsilon < 1$, the binary noisy voter model on $G^{(n)}$ with noise θ satisfies*

$$t_{\text{MIX}}^{\mathsf{x}_{\text{alt}}}(\varepsilon) = \frac{1 + o(1)}{\min\{4 - 2\theta, 4\theta\}} \log n, \tag{1.7}$$

where the $o(1)$-term is $O\big(\log^{-\alpha/2} n\big)$ with a constant depending on $\varepsilon, \theta, \alpha, \mathfrak{c}_0$ (but not on $G^{(n)}$). Furthermore, if $\mathsf{x}_0 = \mathsf{x}_0^{(n)}$ is any arbitrary sequence of initial conditions, we have

$$\liminf_{n \to \infty} \frac{t_{\text{MIX}}^{\mathsf{x}_0}(\varepsilon)}{t_{\text{MIX}}^{\mathsf{x}_{\text{alt}}}(\varepsilon)} \geq 1,$$

thus x_{alt} is the asymptotically fastest initial condition on $G^{(n)}$. In particular, for Glauber dynamics for the Ising model on $\mathbb{Z}_n$ (the case $G^{(n)} = \mathbb{Z}_n$) at any inverse temperature $\beta > 0$, the alternating initial condition x_{alt} is the asymptotically fastest initial condition.
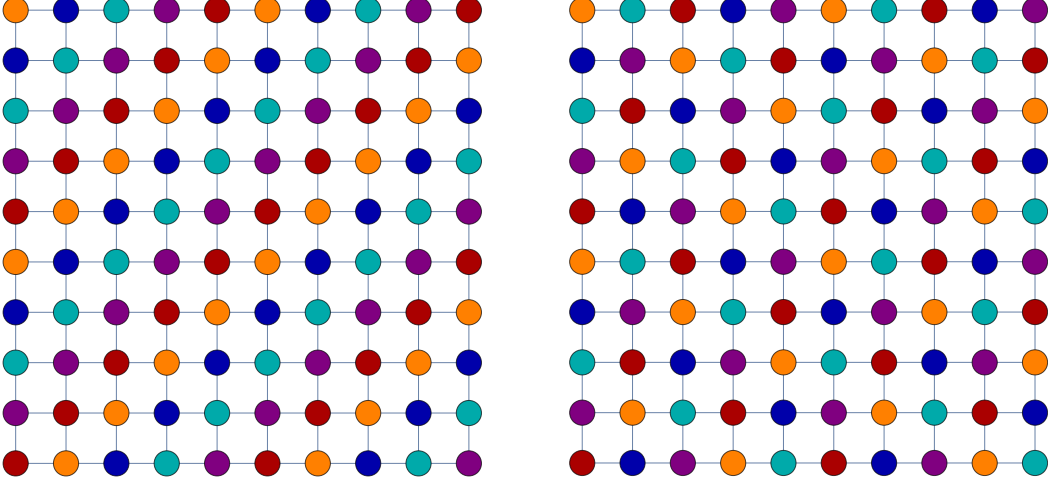
FIGURE 3. In $\mathbb{Z}_n^2$ with $q = 5$, starting from the rainbow configuration (left) leads to slower mixing compared to starting from the knight configuration (right).

FIGURE 4. Rainbow configuration on $\mathbb{Z}_n$ for $q = 5$. On $\mathbb{Z}_n$, the rainbow initial condition is asymptotically the fastest for all $q \geq 2$.

Corollary 1.2 shows that when $q = 2$, alternating between the 2 colors (when possible) gives the asymptotic fastest rate of convergence. In the case of the d-dimensional lattice $\mathbb{Z}_n^d$ for any $d \geq 1$, this results in "checkerboard" configurations being the fastest initial conditions. However, when $q > 2$, the answer can become more complicated. Consider again the d-dimensional lattice $G = \mathbb{Z}_n^d$: we can define a generalization of $\mathsf{x}_{\mathrm{alt}}$ by cycling over the q colors in all directions, creating the *rainbow* configuration (depicted in Fig. 3 on the left)

$$\mathsf{x}_{\mathrm{rbw}}(i_1, \ldots, i_d) = i_1 + \ldots + i_d \pmod{q}, \quad (i_1, \ldots, i_d) \in \mathbb{Z}_n^d. \tag{1.8}$$

It turns out that $\mathsf{x}_{\mathrm{rbw}}$ does not always achieve the fastest rate. Indeed, we will see in Section 3 that

$$T_{\mathsf{x}_{\mathrm{rbw}}} \sim \frac{d}{2\left[1 - (1 - \theta)\cos\left(\frac{2\pi}{q}\right)\right]} \log n \,,$$

while for $d = 2$ and $q = 5$, if we define the "knight" configuration (depicted in Fig. 3 on the right)

$$\mathsf{x}_{\mathrm{knt}}(i, j) = i + 2j \pmod{5}$$

(taking after a chess knight move), then we will get

$$T_{\mathsf{x}_{\mathrm{knt}}} \sim \frac{1}{1 + (1 - \theta)/4} \log n \,,$$

and in particular $T_{\mathsf{x}_{\mathrm{knt}}} < (1 - \delta)T_{\mathsf{x}_{\mathrm{rbw}}}$ with $\delta = \frac{\sqrt{5}(1-\theta)}{5-\theta} > 0$ for all $\theta < 1$ and every n in this setting. The reason for this is that when $d > 1$, there is more room to spread the colors while avoiding early repetitions (which $\mathsf{x}_{\mathrm{rbw}}$ does not do effectively in $d \geq 2$, since we can encounter the same color we are currently at after making just two moves).

Nonetheless, we can show that when $d = 1$, $\mathsf{x}_{\mathrm{rbw}} = (0, 1, \ldots, q-1, 0, 1, \ldots, q-1, 0, 1, \ldots)$ (depicted in Fig. 4) does indeed achieve the fastest possible rate (see also Fig. 5), as we next formulate.
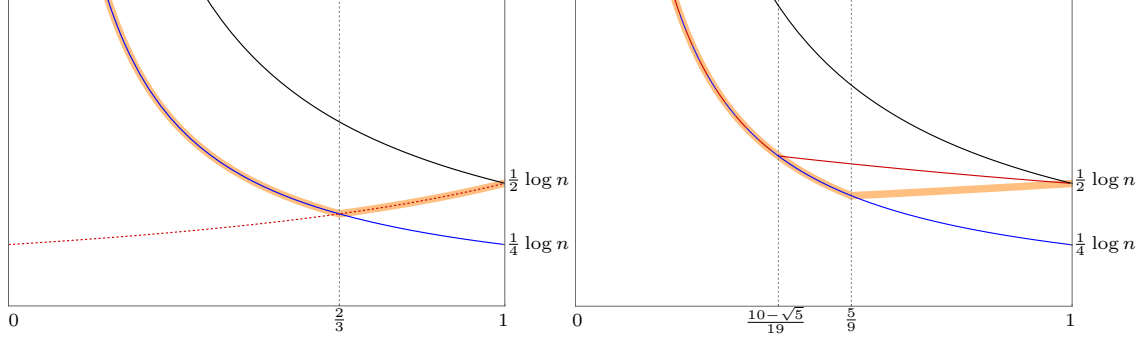
FIGURE 5. Mixing as a function of θ in $(\mathbb{Z}/n\mathbb{Z})^d$. Left: $d = 1$ and $q = 2$ (previously known lower bound in red, x_{alt} in orange). Right: $d = 2$ and $q = 5$ (x_{rbw} in red, x_{knt} in orange).

Theorem 1.3. *Let $G = \mathbb{Z}_n$. There exists $c > 0$ such that, for every $q \geq 3$, every $\theta \in (0,1]$ and every initial state x_0,*

$$\mathcal{A}_t^{(2)}(x_0) \geq \frac{c_0}{q^2} \mathcal{A}_t^{(2)}(x_{\text{rbw}}). \tag{1.9}$$

Consequently, $T_{x_0} \geq T_{x_{\text{rbw}}} + O(1)$, and thus x_{rbw} is the asymptotically fastest initial condition.

The results so far only pertained to deterministic initial states. The effect of a random initial condition was explored in [17] for the 1D Ising model: instead of starting from a deterministic state x_0, one assigns an i.i.d. uniform color to each site. This means starting from the initial configuration $\mathcal{U}$, the uniform distribution over Ω, or equivalently replacing $\mathbb{P}_{x_0}(X_t \in \cdot)$ with its annealed version $\mathbb{P}_{\mathcal{U}}(X_t \in \cdot) := \frac{1}{|\Omega|} \sum_{x_0 \in \Omega} \mathbb{P}_{x_0}(X_t \in \cdot)$. Denote by $t_{\text{MIX}}^{\mathcal{U}}(\varepsilon)$ the mixing time of the chain started at $\mathcal{U}$ (the time for $\mathbb{P}_{\mathcal{U}}(X_t \in \cdot)$ to be within ε of the equilibrium measure μ_G in total variation distance). It was shown in [17] that, for any fixed $0 < \varepsilon < 1$, Glauber dynamics for Ising on $(\mathbb{Z}/n\mathbb{Z})$ has

$$t_{\text{MIX}}^{\mathcal{U}}(\varepsilon) \sim \frac{1}{4\theta} \log n$$

(that is, the bound in Eq. (1.5) but without the autocorrelation term T_{x_0}). We can now extend this result to the noisy voter model in the same setting as Theorem 1.1. Furthermore, as the lower bound $t_{\text{MIX}}^{x_0} \geq \frac{1-o(1)}{4\theta} \log n$ in Eq. (1.5) holds for all deterministic starting states, we can deduce that the uniform initial condition is asymptotically fastest compared to all deterministic starting states.

Corollary 1.4. *Fix $\theta \in (0,1]$ and $q \geq 2$. Let $G = (V, E)$ be a connected graph on n vertices satisfying (1.1) for some $\mathfrak{c}_0 > 0$ and $0 < \alpha < 1$, and $\mathcal{U}$ be the uniform distribution on $\{0, \ldots, q-1\}^V$. Then, for every fixed $0 < \varepsilon < 1$, the q-state noisy voter model on G with noise θ started from $\mathcal{U}$ has*

$$t_{\text{MIX}}^{\mathcal{U}}(\varepsilon) = \frac{1 + o(1)}{4\theta} \log n, \tag{1.10}$$

where the $o(1)$-term is $O\left(\log^{-\alpha/2} n\right)$ with a constant depending on $\varepsilon, \theta, q, \mathfrak{c}_0, \alpha$ (but not G), and

$$t_{\text{MIX}}(\varepsilon) = \frac{1 + o(1)}{2\theta} \log n, \tag{1.11}$$

i.e., the asymptotic worst-case mixing time (attained by the monochromatic $x_0 = \mathbf{1}$) is twice slower. Moreover, if $G^{(n)}$ is a sequence graphs satisfying Eq. (1.1) for the same $\alpha, \mathfrak{c}_0$, then for any $x_0 = x_0^{(n)}$,

$$\liminf_{n \to \infty} \frac{t_{\text{MIX}}^{x_0}(\varepsilon)}{t_{\text{MIX}}^{\mathcal{U}}(\varepsilon)} \geq 1,$$

i.e., asymptotically, no deterministic initial state x_0 is faster than the uniform initial condition $\mathcal{U}$.

(NB. Eq. (1.11) extends [6, Thm 1, Eq.(1.2)]—which was valid for all bounded-degree graphs—from $q = 2$ to all $q \geq 2$, albeit for graphs that further have subexponential growth of balls.)

1.2. **Heuristics and main proof ideas.**

The autocorrelation time. As mentioned above, in the binary case $q = 2$, it was shown in [6] that the worst-case mixing time for bounded-degree n-vertex graphs is $(1 + o(1))\frac{1}{2\theta} \log n$, i.e., the same as the chain where the noise probability is 1 (no vote-copying mechanism) and sites are independently refreshed at rate θ (a random walk on the hypercube). This can be understood as follows: looking at the dynamics backward in time (as discussed in Section 2.1.1) shows that sites coalesce into clusters (arising from the copying mechanism) that receive an identical color. Under a monochromatic initial configuration x_0, say $\mathsf{x}_0 = \mathbf{1}$, a cluster roughly behaves like a single site, attaining its color by either getting a noisy update (with rate θ) or surviving the noise to accept the color (in this case, 1) assigned by x_0. In graphs with bounded degree, said clusters typically have $O(1)$ size, thus one sees cn clusters behaving as an infinite temperature (independent spins) single-site dynamics that evolve at rate θ, leading to the aforementioned mixing time.

The role of the autocorrelation function $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ is to quantify the effect of a general initial condition x_0 under the same assumption that the single-site marginals govern the mixing time. Indeed, the high temperature nature of the noisy voter model and the worst-case intuition above suggest that one could treat the latter as a product chain, and it is known that the total variation (L^1) distance of a product chain can be bounded, from below and above, by the sum of (squares of) the L^2 distances to equilibrium of its coordinates (see [15, Prop. 7]). This motivates the introduction of the function $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$, which is nothing but that L^2 quantity for the noisy voter model (see Eq. (1.3), where the sum over j is simply $q\|\mathbb{P}_{\mathsf{x}_0}(X_t(v) \in \cdot) - \nu\|_{L^2(\nu)}^2$ for ν the uniform distribution on $\{0, \ldots, q-1\}$). Hence, if we were to run n i.i.d. copies of the chain $(X_t^v)_{v \in V}$, and take the value of each $v \in V$ from its own dedicated instance $X_t^v(v)$, we would have cutoff at T_{x_0}. This approximation turns out to be valid in the true model for *sufficiently large* θ (and all q), as confirmed by Theorem 1.1 for every initial condition x_0. However, as the next subsection shows, for some choices of x_0 and small θ, the function $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ does not govern the mixing time.

The correlation time. One key feature that the heuristics of single-site marginals fails to capture is, by its very definition, the correlations between sites in the equilibrium measure μ_G. This aspect can be measured by the *coupling from the past* description of μ_G; deferring details to Sections 2.1.1 and 2.1.2, one can perfectly simulate μ_G via coalescing continuous-time random walks from every site, which move at rate $1 - \theta$ and are killed at rate θ, generating a single uniformly chosen color to their entire coalescent cluster as they die. Via this description, one can read the correlations in μ_G from the coalescent clusters (conditional on the partition to clusters, every cluster has a single uniform color, and distinct ones are independent). When attempting to generate $\mathbb{P}_{\mathsf{x}_0}(X_t \in \cdot)$ via the same recipe, there is a time constraint: once we reach time t, the coalescing process ends (surviving clusters take on the colors prescribed by x_0).

With that in mind, let $\{u \overset{>t}{\longleftrightarrow} v\}$ denote the event that the continuous-time walks from u and v (moving at rate $1 - \theta$ and being killed at rate θ) do not coalesce within time t but do so afterwards. The above discussion indicates that the quantities $\mathbb{P}(u \overset{>t}{\longleftrightarrow} v)$ must be suitably small in order to avoid the situation whereby u, v are more correlated under μ_G than under $\mathbb{P}_{\mathsf{x}_0}(X_t \in \cdot)$. We will see that, for graphs with subexponential growth of balls as per Eq. (1.1), averaging over nearest-neighbors is enough to asymptotically characterize this effect. Namely, let

$$T_{\mathrm{corr}} = \inf\left\{t \geq 0 : \sum_{uv \in E} \mathbb{P}(u \overset{>t}{\longleftrightarrow} v) < \sqrt{n}\right\}. \tag{1.12}$$

Clearly, $\mathbb{P}\left(u \overset{>t}{\leftrightsquigarrow} v\right) \leq e^{-2\theta t}$, simply by asking that the two random walks survive (no killing) separately in $[0, t]$. For our class of graphs, a matching lower bound holds up to sub-exponential corrections, thus T_{corr} will reduce (up to lower order terms) to $\frac{1}{4\theta} \log n$, as featured in Theorem 1.1.

The intuition as to why T_{corr} can govern the mixing time at *sufficiently small* θ is the following. Consider again the single-site marginals. When x_0 is monochromatic, these are only affected by the noisy updates (at rate θ), whereas in general, a random walk on the values of x_0 (at rate $1 - \theta$) can also contribute to mixing (e.g., consider a "random-like" x_0, where the colors are arranged so as to resemble the effect of a noisy update). In view of their rates, the latter mechanism can have a stronger effect when θ is small, whence the single-site marginals can mix faster than the time it takes the correlations to be created. In that scenario, $T_{\mathsf{x}_0} < T_{\text{corr}}$, which means that at time T_{x_0} the vertices are not sufficiently correlated and thus T_{x_0} is only an under-estimate of t_{MIX}. Theorem 1.1 confirms this intuition and shows that waiting for (the maximum between) T_{x_0} and T_{corr} is all that is needed for mixing to occur.

Proof ideas, upper bounds. The work [6] proved an optimal worst-case upper bound for a large class of 2-state noisy voter models, but one cannot hope for that approach to be applicable for general x_0, as we next explain. At the heart of that argument was a result from [9] on noisy trees to view the backward cluster decomposition of the noisy voter model (explained in Section 2.1.1) as a projection of an equivalent chain on "stringy trees," breaking the dependencies between sites. This elegant reduction provided a direct upper bound on the total variation distance via that of a chain with n independent bits that can only change their initial value at rate θ (i.e., a random walk on the hypercube). At that point, x_0 becomes irrelevant, as by symmetry every initial condition is equivalent to the monochromatic $\mathsf{x}_0 = \mathbf{1}$ after said reduction.

Instead, our proof follows the approach of the later work [19], which analyzed the alternating and bi-alternating initial configurations for the 1D Ising model. The strategy there was to develop the backward dynamics of the Ising model for a short time, which suffices for the noise mechanism to decouple poly-logarithmic clusters of sites; then, the model was treated as a product chain on such smaller clusters, which could later be simplified to having size $\leq L$ for some fixed $L > 0$. Thereafter, the L^1-L^2 reduction mentioned at the beginning of this section allowed the authors to conclude the proof, thanks to the regularity of the lattice and the initial states x_0 considered there.

In our setting, the sub-exponential growth of balls assumption Eq. (1.1) allows us (with mostly minor changes) to make a similar decomposition as in [19]—reducing to sub-polynomial clusters, then to clusters of size $O(1)$—thus reduce the analysis of said clusters after an L^1-L^2 reduction. However, in the second part of the proof, the potential irregularity of G and x_0 and the general $q \geq 2$ form a nontrivial obstacle. The work [19] showed that the mixing of these $O(1)$ clusters is governed by a competition between 1-point and 2-point correlations functions, which is easily resolved when the initial condition x_0 is periodic; this is no longer the case for general G and x_0. Our strategy is to not treat these competitions separately per cluster, but rather combine them, reducing the cumulative 1-point contributions to the autocorrelation function $t \to \mathcal{A}_t^{(2)}(\mathsf{x}_0)$. The latter is shown to be completely monotone in Section 2, and that allows us to identify the global dominant term out of $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ and $e^{-4\theta t}$, leading to an upper bound of either T_{x_0} or $\frac{1}{4\theta} \log n$, respectively.

Proof ideas, lower bounds. The $t_{\text{MIX}}^{\mathsf{x}_0} \gtrsim T_{\mathsf{x}_0}$ part of Eq. (1.5) is achieved by taking a distinguishing statistics that yields an expected difference of $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ under $\mathbb{P}_{\mathsf{x}_0}^t$ and μ.

The $t_{\text{MIX}}^{\mathsf{x}_0} \gtrsim \frac{1}{4\theta} \log n$ part of Eq. (1.5) is more complicated. In [19], the analogous bound for the alternating and bi-alternating conditions for 1D Ising was established using the Hamiltonian of the

model $\sum_i \sigma(i)\sigma(i+1)$ as a distinguishing statistics. The proof relied on the fact that the alternating configuration x_{alt} was a minimizer of the Hamiltonian, whereas general x_0 can lead to both smaller and larger expected values of the Hamiltonian at finite time t compared to equilibrium. Thus, this approach did not generalize to arbitrary x_0, and it was left as a conjecture that the bound would hold more generally. Here we handle this lower bound in terms of site correlations, rather than the analogous spin agreements or disagreements measured by the Hamiltonian:

(1) We consider a distinguishing statistics that measures the covariance between adjacent sites, and show that the latter are always more correlated at equilibrium than at time t, with a discrepancy of at least $\mathbb{P}\big(u \overset{>t}{\leftrightsquigarrow} v\big)$, uniformly over x_0;

(2) Using a result from [20] that bounds the escape probability of random walk from a set in terms of its conductance, we show that under the sub-exponential growth of balls assumption Eq. (1.1) the probabilities above are equal to $e^{-2\theta t + o(t)}$ for $t \sim \log n$.

1.3. **Organization of the paper.** In Section 2, we prove properties of the two autocorrelation functions $\mathcal{A}_t^{(1)}(x_0), \mathcal{A}_t^{(2)}(x_0)$ that will be needed throughout the proofs of the main theorems. In Section 3, we discuss specific applications of Theorem 1.1 on $\mathbb{Z}_n^d$, as well as prove Theorem 1.3. The proofs of Theorem 1.1 and Corollary 1.4 will be split between Section 4, where we prove the upper bounds, and Section 5, where we prove the lower bounds. Finally, Section 6 is devoted to concluding remarks on the new results, and open problems on generalizations / refinements thereof.

2. Properties of autocorrelation in the noisy voter model

2.1. **Notation and preliminaries.** We refer to the spins/opinions/states at each vertex as *colors*. So far we denoted the q colors by $\{0, \ldots, q-1\}$. In Sections 2 and 5 it will be convenient to use as labels the q-th roots of unity:

$$\mathfrak{C}_q = \left\{ \omega_j = e^{2\pi i j/q} \ : \ j = 0, \ldots, q-1 \right\}. \tag{2.1}$$

Given a graph $G = (V, E)$, a deterministic color configurations on G will be denoted as x, or x_0 when we wish to emphasize that the configuration should be thought of as an initial condition. In contrast, states of the noisy voter chain at time t will be denoted X_t when starting from a deterministic initial condition x_0, and we will often use Y_t when starting from μ_G, the stationary measure of the noisy voter model on G. We also use Y to indicate a single sample from μ_G.

2.1.1. *Duality with coalescing random walks.* We now explain the classical duality between the (noisy) voter model and a system of (killed) coalescing random walks (cf. [10]). Fix $t_0 > 0$, and view the dynamics resulting in the state X_{t_0} as a space-time slab with x_0 at the bottom and X_{t_0} at the top. Moving in space (on the graph G) will be dubbed *horizontal*, and moving in time will be dubbed *vertical*. The following description constructs $X_{t_0} = \{X_{t_0}(v)\}_{v \in V}$ by looking at the dynamics top-to-bottom ("backward in time"), rather than bottom-to-top ("forward in time").

Let us start with a single vertex $v \in V$. We can reconstruct the value $X_{t_0}(v)$ of the state of v at time t_0 as follows: starting from $v_0 = v$ at the top, move down (backward in time) *vertically* until the largest time $t_1 < t_0$ where the v_0 received an update in the (forward) dynamics. (If no such update exists, we conclude that $X_{t_0}(v) = x_0(v)$.) If said update was noisy (which occurs with probability θ) and resulted in a new color ω, we conclude that $X_{t_0}(v) = X_{t_1}(v) = \omega$ and we can stop; if instead the update was a vote-copying one (with probability $1-\theta$), then $X_t(v)$ must be equal to $X_{t_1}(v_1)$, where v_1 is the neighbor of v_0 whose color was copied. We register this as a *horizontal* move from v_0 to v_1, and proceed to recover $X_{t_1}(v_1)$ recursively (by examining the largest $t_2 < t_1$

where v_1 was updated, and so forth). Letting $s_0 = 0$ and $s_i = t_i - t_{i-1}$ $(i \geq 1)$, the process defined by $Z_s^v = v_i$ for $s \in [s_i, s_{i+1})$ is a continuous-time simple random walk on G that starts at v, moves at rate $1 - \theta$ and gets killed at rate θ. Hence, if we let ω be a uniform color from $\{0, \ldots, q-1\}$ independent of Z^v, the value $X_t(v)$ can be coupled so that

$$X_{t_0}(v) = \begin{cases} \omega & \text{if } Z_s^v \text{ gets killed before time } t_0 \\ \mathsf{x}_0(Z_{t_0}^v) & \text{if } Z_s^v \text{ survives to time } t_0 \end{cases}.$$

(Note that the probability that Z_s^v survives the the bottom is $e^{-\theta t_0}$.) All is left is to describe how the different walks $\{Z^v\}_{v \in V}$ interact with each other to determine the joint law of the $\{X_{t_0}(v)\}_{v \in V}$: for any two vertices v, w, as long as Z_s^v and Z_s^w do not meet each other, their steps (or killings) are due to updates occurring in the histories of distinct vertices and thus are independent of each other. However, if Z_s^v and Z_s^w meet for the first time at time τ (time $t_0 - \tau$ in the forward dynamics), the process described above forces them to go through the exact same path (or be killed simultaneously) from $s = \tau$ until time $s = t_0$ (time 0 in the slab). Hence, if two walks Z^v and Z^w ever meet, they proceed together from that point onward and thus *coalesce into a single walk*. In particular, if the walks from v and w coalesce before time $s = t_0$ (that is, before reaching the bottom), then v and w inevitably obtain the same color (i.e., $X_{t_0}(v) = X_{t_0}(w)$). This holds in the same way for any number of vertices, meaning that any number of walks can coalesce into a single one, and the vertices such walks started from all receive the same color (a uniform one if the single walk resulted from the merging encounters a noisy update before reaching the bottom, or the value of x_0 at the final location of the walk otherwise). See Figure 1 for an illustration of this dual description.

2.1.2. *Dependence on the initial condition and coupling from the past.* The duality explained above shows that X_t depends on the initial condition x_0 through the final locations of all the surviving walks (and their merging histories). In particular, if the walk Z_s^v started from a vertex v gets killed before time t (i.e., before reaching the bottom), then the value $X_t(v)$ (as well as the values $X_t(w_i)$ at time t of all the vertices w_i whose walks $Z_s^{w_i}$ merged with Z_s^v before the killing) is *independent* of the initial condition x_0. As a consequence, if all the walks started from V die out (with or without coalescing with the other walks) before reaching the bottom, then X_t becomes independent of x_0.

This observation gives us a recipe to perfectly simulate the stationary measure μ_G, a special case of the *coupling from the past* (CFTP) method (see [12, §25]): take the walks $\{Z_s^v\}_{v \in V}$ defined above (that is, coalescing random walks with killing, running backwards in time from all vertices), but let them run until killed (without stopping them at time t whence they would reach the bottom slab, hitting x_0). With probability 1, all of them will have died out at some point in the past (either coalescing with other walks or not). Then, the resulting $Y = \{Y(v)\}_{v \in V}$ obtained by assigning a uniform color to each cluster (each group of vertices whose associated random walks coalesced before dying) will be a perfect sample from μ_G. (From the perspective of CFTP, the coalescing random walk representation of the noisy voter model defines a grand coupling of the chains from all initial states, and if $t - T$ is a time where all walks coalesce, so do all the chains at time t.)

This description of μ_G allows us to couple $\mathbb{P}_{\mathsf{x}_0}(X_t \in \cdot)$ to μ_G by simply running the same coalescing random walks backward dynamics from time t, stopping at time 0 to realize X_t from x_0 and then continuing back in the past to realize $Y \sim \mu_G$. In doing so, X_t and Y will be identical at the vertices whose associated random walks have died before reaching time 0, but possibly not at the vertices v that reached it, since in that case $X_t(v)$ takes the x_0 value at the final location of Z_s^v while $Y(v)$ will take an independent uniform value. As observed in [17], when starting from the uniform initial state $\mathcal{U}$, the value of $X_t(v)$ can be coupled to $Y(v)$ even if Z_s^v survived to time 0, so

long as it does not get to coalesce with another particle before being killed (one can simply match the value of $\mathcal{U}$ to that of Y); this governed the behavior of $t^{\mathcal{U}}_{\mathrm{MIX}}$ in Glauber dynamics for the Ising model on $\mathbb{Z}/n\mathbb{Z}$, and will play the same role here for every $q \geq 2$ and every G as per Corollary 1.4.

2.2. Alternative representations of autocorrelation.

Let $G = (V, E)$ be a finite connected graph, and let π_G be the stationary distribution of simple random walk on G, i.e., $\pi_G(v) = \frac{d_G(v)}{2|E|}$ for $v \in V$. We consider the L^2-inner product space on $\mathbb{C}^V$ with respect to π_G, via the Hermitian inner product

$$\langle f, g \rangle_{\pi_G} := \sum_{v \in V} f(v)\overline{g(v)}\pi_G(v) \qquad \text{for } f, g : V \to \mathbb{C}. \tag{2.2}$$

Given $\mathsf{x} \in \mathfrak{C}_q^V$, we define its k-th power $\mathsf{x}^k = (\mathsf{x}(v)^k)_{v \in V}$ as the configuration obtained by taking k-powers component-wise (NB. $\mathfrak{C}_q^V$ is closed under this operation, as $\mathfrak{C}_q$ is a multiplicative group).

Proposition 2.1. *Let $G = (V, E)$ be a connected graph on n vertices. Fix an initial state $\mathsf{x}_0 \in \mathfrak{C}_q^V$. Let $\psi_1, \ldots, \psi_n$ be an $L^2(\pi_G)$-orthonormal basis of eigenfunctions of the transition matrix P of the discrete-time simple random walk on G, and let $\lambda_1, \ldots, \lambda_n$ be their corresponding eigenvalues. Further define*

$$\gamma_l = 1 - (1 - \theta)\lambda_l \qquad \text{for } l = 1, \ldots, n. \tag{2.3}$$

Then, for every $t > 0$,

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{v \in V}\pi_G(v)\big|\mathbb{E}_{\mathsf{x}_0}[X_t(v)^k]\big|^2 = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{l=1}^{n}|\langle \mathsf{x}_0^k, \psi_l \rangle_{\pi_G}|^2 e^{-2\gamma_l t}. \tag{2.4}$$

Consequently, for any $s \geq 0$, we have

$$e^{-(4-2\theta)s} \leq \frac{\mathcal{A}_{t+s}^{(2)}(\mathsf{x}_0)}{\mathcal{A}_t^{(2)}(\mathsf{x}_0)} \leq e^{-2\theta s}, \tag{2.5}$$

and in particular $\frac{q-1}{q}e^{-(4-2\theta)t} \leq \mathcal{A}_t^{(2)}(\mathsf{x}_0) \leq \frac{q-1}{q}e^{-2\theta t}$.

Proof. Recalling the definition of $\mathcal{A}_t^{(2)}$ from Eq. (1.3), one has

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \sum_{v \in V}\pi_G(v)\Big(\big(\sum_{\omega \in \mathfrak{C}_q}\mathbb{P}_{\mathsf{x}_0}\left(X_t(v) = \omega\right)^2\big) - \frac{1}{q}\Big). \tag{2.6}$$

Note that if $\tilde{X}_t$ denotes an i.i.d. copy of X_t (started from the same x_0), we have

$$\sum_{\omega \in \mathfrak{C}_q}\mathbb{P}_{\mathsf{x}_0}\left(X_t(v) = \omega\right)^2 = \sum_{\omega \in \mathfrak{C}_q}\mathbb{P}_{\mathsf{x}_0}\left(X_t(v) = \tilde{X}_t(v) = \omega\right) = \mathbb{P}_{\mathsf{x}_0, \mathsf{x}_0}(X_t(v) = \tilde{X}_t(v)).$$

Since $\sum_{k=0}^{q-1}\omega^k = q\mathbb{1}_{\{\omega=1\}}$ for $\omega \in \mathfrak{C}_q$, taking $\omega = X_t(v)\tilde{X}_t(v)^{-1} = X_t(v)\overline{\tilde{X}_t(v)}$ (recall $|\tilde{X}_t(v)| = 1$) shows that

$$\mathbb{P}_{\mathsf{x}_0, \mathsf{x}_0}(X_t(v) = \tilde{X}_t(v)) = \frac{1}{q}\sum_{k=0}^{q-1}\mathbb{E}[X_t(v)^k\overline{\tilde{X}_t(v)^k}] = \frac{1}{q}\sum_{k=0}^{q-1}\big|\mathbb{E}[X_t(v)^k]\big|^2.$$

Combining the last two displays with Eq. (2.6), the $k = 0$ summand cancels the $-\frac{1}{q}$ term, yielding

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{v \in V}\pi_G(v)\big|\mathbb{E}_{\mathsf{x}_0}[X_t(v)^k]\big|^2.$$

Recalling the coalescing random walks representation of Section 2.1.1, on the event that the continuous-time random walk from v is killed before reaching time 0, the contribution to $\mathbb{E}_{\mathsf{x}_0} X_t(v)^k$ is $\frac{1}{q}\sum_{\omega \in \mathfrak{C}_q} \omega^k = 0$. Thus, if W_t denotes a continuous-time simple random walk moving at rate $1 - \theta$ (and without killing), and H_t the semigroup associated to such a walk with rate 1, then

$$\left|\mathbb{E}_{\mathsf{x}_0}[X_t(v)^k]\right|^2 = e^{-2\theta t}\left|\mathbb{E}_v[\mathsf{x}_0(W_t)^k]\right|^2 = e^{-2\theta t}(H_{(1-\theta)t}\mathsf{x}_0^k)(v)\overline{(H_{(1-\theta)t}\mathsf{x}_0^k)(v)},$$

which, when combined with the last display, gives

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \frac{1}{q}e^{-2\theta t}\sum_{k=1}^{q-1}\langle H_{(1-\theta)t}\mathsf{x}_0^k, H_{(1-\theta)t}\mathsf{x}_0^k\rangle_{\pi_G}.$$

Recall that ψ_l is an eigenfunction of the discrete-time transition matrix P with eigenvalue λ_l; hence, it is an eigenfunction for $H_t = e^{-t(I-P)}$ with eigenvalue $e^{-(1-\lambda_l)t}$. Writing $\mathsf{x}_0^k = \sum_{l=1}^n \langle \mathsf{x}_0^k, \psi_l\rangle_{\pi_G}\psi_l$ via the orthonormal basis $\{\psi_l\}$ in $L^2(\pi_G)$, we arrive at

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{l=1}^n |\langle\mathsf{x}_0^k,\psi_l\rangle_{\pi_G}|^2 e^{-2\theta t - 2(1-\lambda_l)(1-\theta)t} = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{l=1}^n |\langle\mathsf{x}_0^k,\psi_l\rangle_{\pi_G}|^2 e^{-2(1-(1-\theta)\lambda_l)t},$$

establishing Eq. (2.4). The inequalities in Eq. (2.5) readily follow by noting that $e^{-2\gamma_l s}$ is minimized at $\lambda_l = -1$ (whence $\gamma_l = 2 - \theta$) and maximized at $\lambda_l = 1$ (whence $\gamma_l = \theta$). Finally, the upper and lower bounds on $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ follow from Eq. (2.5) (applying it for $s = t_0$ and $t = 0$ to bound $\mathcal{A}_{t_0}^{(2)}(\mathsf{x}_0)$ using that $\mathcal{A}_0^{(2)}(\mathsf{x}_0) = (q-1)/q$ by definition). $\qquad\square$

Two simple consequences of Proposition 2.1 are the following.

Remark 2.2 (T_{x_0} maximized at the monochromatic **1**). As the monochromatic initial state **1** is an eigenfunction of the discrete-time random walk with eigenvalue $\lambda = 1$, the identity in Eq. (2.4) (noting $\mathbf{1}^k = \mathbf{1}$ for all $k = 1, \ldots, q-1$), along with the bounds on $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ in Proposition 2.1, show

$$\mathcal{A}_t^{(2)}(\mathbf{1}) = \frac{q-1}{q}e^{-2\theta t} \geq \mathcal{A}_t^{(2)}(\mathsf{x}_0) \qquad \text{for every } \mathsf{x}_0 \text{ and } t > 0\,;$$

hence, by definition, $T_{\mathsf{x}_0} \leq T_{\mathbf{1}} \sim \frac{1}{2\theta}\log n$ holds for every x_0.

Remark 2.3 (T_{x_0} minimized at $\mathsf{x}_{\mathrm{alt}}$ for a bipartite G and $q = 2$). Since G is bipartite, $\mathsf{x}_{\mathrm{alt}}$ is an eigenfunction of the discrete-time random walk with eigenvalue $\lambda = -1$, thus Proposition 2.1 shows

$$\mathcal{A}_t^{(2)}(\mathsf{x}_{\mathrm{alt}}) = \frac{1}{2}e^{-(4-2\theta)t} \leq \mathcal{A}_t^{(2)}(\mathsf{x}_0) \quad \text{for every } \mathsf{x}_0 \text{ and } t > 0\,;$$

hence, by definition, $T_{\mathsf{x}_0} \geq T_{\mathsf{x}_{\mathrm{alt}}} \sim \frac{1}{4-2\theta}\log n$ holds for every x_0.

Plugging the lower bound on T_{x_0} from Remark 2.3 in the asymptotics for $t_{\mathrm{MIX}}^{\mathsf{x}_0}$ given in Theorem 1.1 establishes Corollary 1.2. Combining the upper bound on T_{x_0} from Remark 2.2 with Theorem 1.1 implies Eq. (1.11) in Corollary 1.4 for all $q \geq 3$ (the case $q = 2$ was established in [6]).

2.3. **Eigenfunctions of the noisy voter model.** As in Proposition 2.1, let $\psi_1, \ldots, \psi_n$ be an $L^2(\pi_G)$-orthonormal basis of eigenfunctions of the (1-step) transition matrix P of discrete-time simple random walk on G with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, and let $\gamma_l = 1 - (1-\theta)\lambda_l$. For every $l = 1, \ldots, n$ and $k = 1, \ldots, q-1$, define a function $\Psi_l^{(k)} : \mathfrak{C}_q^V \to \mathbb{C}$ by

$$\Psi_l^{(k)}(\mathsf{x}) = \langle\mathsf{x}^k, \psi_l\rangle_{\pi_G}. \tag{2.7}$$

With this notation, Eq. (2.4) reads

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \frac{1}{q} \sum_{l=1}^{n} \sum_{k=1}^{q-1} |\Psi_l^{(k)}(\mathsf{x}_0)|^2 e^{-2\gamma_l t} . \tag{2.8}$$

While the noisy voter model is, in general, a nonreversible Markov chain, the following result shows that $e^{-\lambda_l}$, inherited from the spectrum of the random walk on G, appear as eigenvalues for it.

Proposition 2.4. *For each $l = 1, \ldots, n$ and $k = 1, \ldots, q-1$, the function $\Psi_l^{(k)}$ is an eigenfunction of the noisy voter model, with eigenvalue $1 - \gamma_l/n$ for the discrete-time chain (each step updates a uniformly chosen vertex), or $e^{-\gamma_l}$ for the continuous-time model (each vertex is updated at rate 1).*

Proof. Let $l \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, q-1\}$, and first consider the discrete-time noisy voter model, letting $\mathscr{P}$ be the 1-step transition matrix of this chain. If $\mathsf{x} \in \mathbb{C}_q^V$ is any initial state, then

$$(\mathscr{P}\Psi_l^{(k)})(\mathsf{x}) = \mathbb{E}_\mathsf{x}\big[\Psi_l^{(k)}(X_1)\big] = \big\langle \mathbb{E}_\mathsf{x}[X_1^k], \psi_l \big\rangle_{\pi_G} = \Big\langle \Big(1 - \frac{1}{n}\Big)\mathsf{x}^k + \frac{1-\theta}{n} P\mathsf{x}^k, \psi_l \Big\rangle_{\pi_G} ,$$

where the last equality used that, in each coordinate $v \in V$ of the vector $\mathbb{E}_\mathsf{x}[X_1^k]$, the following holds: (a) with probability $1 - \frac{1}{n}$, a different vertex is selected for the update, whence $X_1(v) = \mathsf{x}(v)$; (b) with probability $\frac{\theta}{n}$, a noisy update occurs at v, contributing $\frac{1}{q}\sum_{\omega \in \mathfrak{C}_q} \omega^k = 0$ to the expectation; (c) with probability $\frac{1-\theta}{n}$, a vote-copying update occurs at v, contributing $(P\mathsf{x}^k)(v)$ to the expectation. Recalling that P is real symmetric (thus Hermitian), we have that

$$\big\langle P\mathsf{x}^k, \psi_l \big\rangle_{\pi_G} = \big\langle \mathsf{x}^k, P\psi_l \big\rangle_{\pi_G} = \lambda_l \big\langle \mathsf{x}^k, \psi_l \big\rangle_{\pi_G} ,$$

and combining the last two displays now shows that

$$(\mathscr{P}\Psi_l^{(k)})(\mathsf{x}) = \Big(1 - \frac{1}{n} + \frac{(1-\theta)\lambda_l}{n}\Big) \langle \mathsf{x}^k, \psi_l \rangle_{\pi_G} = \Big(1 - \frac{\gamma_l}{n}\Big) \Psi_l^{(k)}(\mathsf{x}) .$$

For the continuous-time noisy voter model, where the updates occur at rate 1 per vertex, the associated semigroup is $\mathscr{H}_t = \exp[-(I - \mathscr{P})nt]$, thus $\Psi_l^{(k)}$ is an eigenvector with eigenvalue

$$\exp\Big[-\Big(1 - \Big(1 - \frac{\gamma_l}{n}\Big)\Big)nt \Big] = \exp(-\gamma_l t) ,$$

as claimed. $\square$

Note that, while the eigenfunctions $\{\psi_l\}$ are $L^2(\pi_G)$-normalized, the eigenfunctions $\{\Psi_l^{(k)}\}$ are not normalized in $L^2(\mu_G)$ (their variance under μ_G is not 1). The following gives an exact expression for the variance of $\Psi_l^{(k)}$ and shows that when the vertices in G have bounded degrees (and more generally, if the ratio of maximal and minimal degree is $O(1)$), all these variances have order $\frac{1}{n}$.

Proposition 2.5. *For each $l = 1, \ldots, n$ and $k = 1, \ldots, q-1$, we have*

$$\mathrm{Var}_{\mu_G}(\Psi_l^{(k)}) = \frac{1}{\gamma_l} \sum_{v \in V} \pi_G(v)^2 |\psi_l(v)|^2 \big(1 - (1-\theta)h_G(v)\big) , \tag{2.9}$$

with $h_G(v) = \frac{1}{d_G(v)} \sum_{w \sim v} \mathbb{P}(v \leftrightsquigarrow w)$ and $\{v \leftrightsquigarrow w\}$ the event that two independent continuous-time random walks from v and w, which move at rate $1 - \theta$ and are killed at rate θ, coalesce before dying. In particular,

$$\frac{\theta}{\gamma_l} \sum_{v \in V} \pi_G(v)^2 |\psi_l(v)|^2 \leq \mathrm{Var}_{\mu_G}(\Psi_l^{(k)}) \leq \frac{1}{\gamma_l} \sum_{v \in V} \pi_G(v)^2 |\psi_l(v)|^2 \tag{2.10}$$

Furthermore, if we let $M := \max_v \mathrm{d}_G(v)/\min_v \mathrm{d}_G(v)$, then we have

$$\frac{\theta}{2M} \leq \frac{\theta}{M\gamma_l} \leq n\,\mathrm{Var}_{\mu_G}(\Psi_l^{(k)}) \leq \frac{M}{\gamma_l} \leq \frac{M}{\theta}\,. \tag{2.11}$$

Proof. Let $\mathscr{P}$ be the transition matrix of the discrete-time noisy voter model (where each step updates a uniformly chosen vertex), and recall from Proposition 2.4 that $\Psi_l^{(k)}$ is an eigenfunction of $\mathscr{P}$ with eigenvalue $1 - \frac{\gamma_l}{n}$. Let Y_0 be distributed according to the equilibrium measure μ_G, and let Y_1 be the configuration after one step of the discrete chain started from Y_0. Letting

$$\sigma^2 := \mathrm{Var}_{\mu_G}(\Psi_l^{(k)}) = \mathbb{E}_{\mu_G}\big[|\Psi_l^{(k)}(Y_0)|^2\big]$$

(using here that $\mathbb{E}_{\mu_G}[\Psi_l^{(k)}(Y_0)] = \langle \mathbb{E}[Y_0^k], \psi_l\rangle_{\pi_G} = 0$ as $Y_0(v)$ is uniform on $\mathfrak{C}_q$ for each v), we have

$$\mathbb{E}_{\mu_G}\big[|\Psi_l^{(k)}(Y_1) - \Psi_l^{(k)}(Y_0)|^2\big] = \mathbb{E}_{\mu_G}\big[|\Psi_l^{(k)}(Y_0)|^2\big] + \mathbb{E}_{\mu_G}\big[|\Psi_l^{(k)}(Y_1)|^2\big] - 2\Re\Big(\mathbb{E}_{\mu_G}\Big[\Psi_l^{(k)}(Y_0)\overline{\Psi_l^{(k)}(Y_1)}\Big]\Big)$$

$$= 2\sigma^2 - 2\Re\bigg(\mathbb{E}_{\mu_G}\Big[\Psi_l^{(k)}(Y_0)\,\mathbb{E}_{\mu_G}\big[\overline{\Psi_l^{(k)}(Y_1)} \mid Y_0\big]\Big]\bigg)\,.$$

Since $\mathbb{E}_{\mu_G}[\Psi_l^{(k)}(Y_1) \mid Y_0] = \mathbb{E}_{Y_0}[\Psi_l^{(k)}(Y_1)] = (1 - \frac{\gamma_l}{n})\Psi_l^{(k)}(Y_0)$ by Proposition 2.4, we deduce that

$$\mathbb{E}_{\mu_G}\big[|\Psi_l^{(k)}(Y_1) - \Psi_l^{(k)}(Y_0)|^2\big] = 2\sigma^2 - 2(1 - \frac{\gamma_l}{n})\sigma^2 = \frac{2\gamma_l}{n}\sigma^2\,,$$

that is,

$$\mathrm{Var}_{\mu_G}(\Psi_l^{(k)}) = \frac{n}{2\gamma_l}\mathbb{E}_{\mu_G}\big[|\Psi_l^{(k)}(Y_1) - \Psi_l^{(k)}(Y_0)|^2\big]\,. \tag{2.12}$$

To compute the latter, denote for brevity

$$\Upsilon := \Psi_l^{(k)}(Y_1) - \Psi_l^{(k)}(Y_0) = \langle Y_1^k - Y_0^k, \psi_l\rangle$$

and consider the two events $\mathcal{E}_v, \mathcal{E}_v'$ describing an update at $v \in V$ in the first step: a noisy update per $\mathcal{E}_v$ and a vote-copying update per $\mathcal{E}_v'$. Since $\Upsilon\mathbb{1}_{\mathcal{E}_v} = \big(U^k - Y_0(v)\big)\pi_G(v)\psi_l(v)$, with U independent and uniformly distributed on $\mathfrak{C}_q$, we have $|\Upsilon|^2\mathbb{1}_{\mathcal{E}_v} = |(U/Y_0(v))^k - 1|^2(\pi_G(v)\psi_l(v))^2$, thus

$$\mathbb{E}_{\mu_G}|\Upsilon|^2\mathbb{1}_{\mathcal{E}_v} = \pi_G(v)^2|\psi_l(v)|^2\frac{1}{q}\sum_{\omega \in \mathfrak{C}_q}|\omega^k - 1|^2 = 2\pi_G(v)^2|\psi_l(v)|^2\,.$$

On the event $\mathcal{E}_v'$, we have $\Upsilon\mathbb{1}_{\mathcal{E}_v'} = (Y_0(J)^k - Y_0(v)^k)\pi_G(v)\psi_l(v)$, with J independent and uniformly distributed on the neighbors of v in G, thus

$$\mathbb{E}_\mu|\Upsilon|^2\mathbb{1}_{\mathcal{E}_v'} = \pi_G(v)^2|\psi_l(v)|^2\frac{1}{d_G(v)}\sum_{w \sim v}\mathbb{E}_{\mu_G}\big[|Y_0(w)^k - Y_0(v)^k|^2\big]$$

$$= 2\pi_G(v)^2|\psi_l(v)|^2\frac{1}{d_G(v)}\sum_{w \sim v}\mathbb{P}(v \not\leftrightarrow w)\,,$$

since sampling Y_0 via CFTP (as described in Section 2.1.2) results either in $\{v \leftrightarrow w\}$, on which $Y_0(v) = Y_0(w)$, or in $\{v \not\leftrightarrow w\}$, on which $Y_0(w), Y_0(v)$ are independent and uniformly distributed on $\mathfrak{C}_q$, contributing 2 to the expectation (recall the case of $\mathcal{E}_v$). Combining the last two displays,

$$\mathbb{E}_{\mu_G}|\Upsilon|^2 = \frac{2}{n}\sum_v \pi_G(v)^2|\psi_l(v)|^2\Big(\theta + (1 - \theta)\frac{1}{d_G(v)}\sum_{w \sim v}\mathbb{P}(v \not\leftrightarrow w)\Big)$$

$$= \frac{2}{n}\sum_v \pi_G(v)^2|\psi_l(v)|^2\Big(1 - (1 - \theta)\frac{1}{d_G(v)}\sum_{w \sim v}\mathbb{P}(v \leftrightarrow w)\Big)\,,$$

which, via Eq. (2.12), yields Eq. (2.9). Since $0 \le h_G(v) \le 1$, we get Eq. (2.10), and Eq. (2.11) follows from the facts $\frac{1}{Mn} \le \pi_G(v) \le \frac{M}{n}$, $\theta \le \gamma_l \le 2 - \theta \le 2$ and $\sum_v \pi_G(v)|\psi_l(v)|^2 = \|\psi_l\|^2_{L^2(\pi_G)} = 1$. $\qquad\square$

Remark 2.6. Consider the "magnetization" eigenfunction Ψ associated to the eigenfunction $\psi = \mathbf{1}$ of the random walk transition matrix P, corresponding to $\lambda = 1$ (and $\gamma = \theta$):

$$\Psi(\mathsf{x}) = \sum_v \mathsf{x}(v)\pi_G(v)\,.$$

As noted in the proof of Proposition 2.5, we have $\mathbb{E}_{\mu_G}\Psi(Y) = 0$, and so

$$\mathrm{Var}_{\mu_G}(\Psi(Y)) = \mathbb{E}_{\mu_G}\big[|\Psi(Y)|^2\big] = \sum_{v,w} \pi_G(v)\pi_G(w)\mathbb{E}\big[Y(v)\overline{Y(w)}\big] = \sum_{v,w} \pi_G(v)\pi_G(w)\mathbb{P}(v \leftrightsquigarrow w)\,,$$

with the last identity using that $Y(v), Y(w)$ are independent and uniform on $\mathfrak{C}_q$ when $\{v \not\leftrightsquigarrow w\}$, so $\mathbb{E}Y(v)\overline{Y(w)}\mathbf{1}_{\{v \not\leftrightsquigarrow w\}} = 0$, and elsewhere $Y(v) = Y(w)$, so that $\mathbb{E}Y(v)\overline{Y(w)}\mathbf{1}_{\{v \leftrightsquigarrow w\}} = \mathbb{P}(v \leftrightsquigarrow w)$. By Proposition 2.5, we have $\mathrm{Var}_{\mu_G}(\Psi) \le (1/\theta)\sum_v \pi_G(v)^2$, thus we can conclude that

$$\sum_{v,w} \pi_G(v)\pi_G(w)\mathbb{P}(v \leftrightsquigarrow w) \le \frac{1}{\theta}\sum_v \pi_G(v)^2\,. \tag{2.13}$$

This will be used in Section 5 to bound the variance of certain statistics.

2.4. **Classical autocorrelation.** Fix an initial configuration x_0. The point-wise autocorrelation of X_t is routinely defined as

$$\mathsf{A}(0,t) = \sum_v \pi_G(v)\Big(\mathbf{1}_{\{X_t(v)=\mathsf{x}_0(v)\}} - \frac{1}{q}\Big)\,,$$

which in the case $q = 2$, where $\mathfrak{C}_q = \{\pm 1\}$, can also be written as $\frac{1}{2}\sum_v \pi_G(v)\mathsf{x}_0(v)X_t(v)$. The function $\mathcal{A}_t^{(1)}(\mathsf{x}_0)$ from Eq. (1.2) is nothing but $\mathbb{E}_{\mathsf{x}_0}[\mathsf{A}(0,t)]$, and we will now argue that

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \mathcal{A}_{2t}^{(1)}(\mathsf{x}_0)\,. \tag{2.14}$$

To see this, recall that $\mathbf{1}_{\{X_t(v)=x_0(v)\}} = \frac{1}{q}\sum_{k=0}^{q-1} X_t(v)^k\overline{\mathsf{x}_0(v)^k}$ (used in the proof of Proposition 2.1), via which we can rewrite $\mathsf{A}(0,t)$ as

$$\mathsf{A}(0,t) = \frac{1}{q}\sum_{k=1}^{q-1}\langle X_t^k, \mathsf{x}_0^k\rangle_{\pi_G}\,;$$

this then becomes, using the representation $\mathsf{x}^k = \sum_l \Psi_l^{(k)}(\mathsf{x})\psi_l$ for the orthonormal basis $\{\psi_l\}$,

$$\mathsf{A}(0,t) = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{l=1}^{n} \Psi_l^{(k)}(X_t)\overline{\Psi_l^{(k)}(\mathsf{x}_0)}\,.$$

By Proposition 2.4, we have that

$$\mathbb{E}_{\mathsf{x}_0}[\Psi_l^{(k)}(X_t)] = e^{-\gamma_l t}\Psi_l^{(k)}(\mathsf{x}_0)\,,$$

and it thus follows that

$$\mathcal{A}_t^{(1)}(\mathsf{x}_0) = \mathbb{E}_{\mathsf{x}_0}[\mathsf{A}(0,t)] = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{l=1}^{n} |\Psi_l^{(k)}(\mathsf{x}_0)|^2 e^{-\gamma_l t}\,,$$

establishing Eq. (2.14) in view of Eq. (2.8). (In particular, one can express all our results in terms of $\mathcal{A}_t^{(1)}$ rather than $\mathcal{A}_t^{(2)}$; we chose to use $\mathcal{A}_t^{(2)}$ to emphasize the role of L^2-mixing in the analysis.)

3. APPLICATIONS FOR THE LATTICE

3.1. **Lattice patterns.** In this section, we let G be the d-dimensional lattice $(\mathbb{Z}/n\mathbb{Z})^d$. Recall from Eq. (2.1) that we are identifying the q colors with the set $\mathfrak{C}_q$ of q-th roots of unity. We can define a general class of configurations that are periodic in every lattice direction as follows: for a given vector $\mathsf{v} = (\mathsf{v}_1, \ldots, \mathsf{v}_d)$ with $\mathsf{v}_i \in \{0, \ldots, q-1\}$, let

$$\mathsf{x}_\mathsf{v}(j_1, \ldots, j_d) = e^{2\pi i(j_1 \mathsf{v}_1 + \ldots + j_d \mathsf{v}_d)/q}, \qquad j = (j_1, \ldots, j_d) \in \mathbb{Z}_n^d \tag{3.1}$$

(NB. we assume that n is a multiple of q, so that the above patterns do not break at the edges). For example, taking $\mathsf{v} = \mathbf{1}$, we see that $\mathsf{x}_\mathbf{1}$ (not to be confused with the monochromatic state $\mathbf{1}$) is equivalent to the rainbow configuration from Equation (1.8):

$$\mathsf{x}_{\mathrm{rbw}}(j_1, \ldots, j_d) = \mathsf{x}_\mathbf{1}(j_1, \ldots, j_d) = e^{2\pi i(j_1 + \ldots + j_d)/q}, \qquad (j_1, \ldots, j_d) \in \mathbb{Z}_n^d. \tag{3.2}$$

Observe that for every v and k, one has $\mathsf{x}_\mathsf{v}^k = \mathsf{x}_{k\mathsf{v}}$, and that x_v^k is an eigenfunction of the discrete-time simple random walk on G (normalized in the sense that $\|x_\mathsf{v}^k\|_{L^2(\pi_G)} = 1$) with eigenvalue

$$\lambda_{k\mathsf{v}} = \frac{1}{d} \sum_{i=1}^d \cos\left(\frac{2\pi k \mathsf{v}_i}{q}\right).$$

Hence, for configurations of the form x_v, Proposition 2.1 yields the exact expression for $\mathcal{A}_t^{(2)}$:

$$\mathcal{A}_t^{(2)}(\mathsf{x}_\mathsf{v}) = \frac{1}{q} \sum_{k=1}^{q-1} e^{-2(1-(1-\theta)\lambda_{k\mathsf{v}})t}. \tag{3.3}$$

For a given v, if we let

$$\lambda_\mathsf{v}^* := \max\{\lambda_{k\mathsf{v}} \,:\, k = 1, \ldots, q-1\},$$

then by our definition of T_{x_0} in Eq. (1.4) we have that

$$T_{\mathsf{x}_\mathsf{v}} = \frac{d}{2(1-(1-\theta)\lambda_\mathsf{v}^*)} \log n + O(1),$$

and Theorem 1.1 will imply that

$$t_{\mathrm{MIX}}^{\mathsf{x}_\mathsf{v}} \sim \frac{d}{2 \min\{1-(1-\theta)\lambda_\mathsf{v}^*, 2\theta\}} \log n = \begin{cases} \frac{d}{4\theta} \log n & 0 < \theta \le \theta_\mathsf{v} := 1 - \frac{1}{2-\lambda_\mathsf{v}^*}, \\ \frac{d}{2(1-(1-\theta)\lambda_\mathsf{v}^*)} \log n & \theta_\mathsf{v} \le \theta \le 1. \end{cases} \tag{3.4}$$

This shows that as long as $\lambda_\mathsf{v}^* < 1$ (whence x_v is faster than worst-case), the behavior of $t_{\mathrm{MIX}}^{\mathsf{x}_\mathsf{v}}$ as a function of θ always changes at $\theta_\mathsf{v} = (1-\lambda_\mathsf{v}^*)/(2-\lambda_\mathsf{v}^*) \in (0,1)$ (a dynamical phase transition, marking the value of θ below which the correlation quantity T_{corr}, rather than $T_{\mathsf{x}_\mathsf{v}}$, governs $t_{\mathrm{MIX}}^{\mathsf{x}_\mathsf{v}}$).

Example 3.1 (Rainbow initial condition on $\mathbb{Z}_n^d$). For any $d \ge 1$ and $q \ge 2$, the rainbow initial condition $\mathsf{x}_{\mathrm{rbw}} = \mathsf{x}_\mathbf{1}$ for the noisy voter model on $\mathbb{Z}_n^d$ has $\lambda_\mathbf{1}^* = \cos(\frac{2\pi}{q})$, so by Eq. (3.4),

$$t_{\mathrm{MIX}}^{\mathsf{x}_{\mathrm{rbw}}} \sim \frac{d}{2 \min\left\{1-(1-\theta)\cos(\frac{2\pi}{q}), 2\theta\right\}} \log n = \begin{cases} \frac{d}{4\theta} \log n & 0 < \theta \le \theta_\mathbf{1} \\ \frac{d}{2-2(1-\theta)\cos(\frac{2\pi}{q})} \log n & \theta_\mathbf{1} \le \theta \le 1 \end{cases},$$

with $\theta_\mathbf{1} = 1 - 1/(2-\cos(\frac{2\pi}{q}))$.

In the special case $q = 2$, $\mathsf{x}_{\mathrm{rbw}}$ is the alternating condition $\mathsf{x}_{\mathrm{alt}} = \mathsf{x}_\mathbf{1}$ and has $\lambda_\mathbf{1}^* = -1$, the lowest possible value of any λ_v^*; thus $\mathsf{x}_\mathbf{1}$ achieves the optimal $T_{\mathsf{x}_\mathbf{1}} \sim \frac{d}{4-2\theta} \log n$ for $q = 2$ (as we have already established, more generally for any bipartite graph G, in Remark 2.3).

(NB. $t_{\mathrm{MIX}}^{\mathsf{x}_{\mathrm{rbw}}}$ is asymptotically optimal in dimension $d = 1$ over all choices of a (deterministic) initial state x_0, as was stated in Theorem 1.3; we will establish this in the next subsection.)

In higher dimensions that is no longer the case; e.g., as mentioned in the introduction, for $d = 2$ and $q = 5$ one can consider the "knight" initial condition x_{v} for $\mathsf{v} = (1, 2)$; there, one has $\lambda_{\mathsf{v}}^* = \frac{1}{2}(\cos(\frac{2\pi}{5}) - \cos(\frac{\pi}{5})) = -\frac{1}{4}$ (with an equal $\lambda_{k\mathsf{v}}$ for all $k = 1, 2, 3, 4$), whereas $\lambda_{\mathbf{1}}^* = \cos(\frac{2\pi}{5})$. More generally, we have the following.

Example 3.2 (Knight initial condition on $\mathbb{Z}_n^d$)**.** For every (d, q) with $d \geq 2$ and $q \geq 5$, as well as for $q = 4$ in dimensions $d \geq 3$, the knight initial condition $\mathsf{x}_{\mathrm{knt}} = \mathsf{x}_{\mathsf{v}}$ for $\mathsf{v} = (1, \ldots, 1, 2)$ has $\lambda_{\mathsf{v}}^* < \lambda_{\mathbf{1}}^*$. Thus, by Eq. (3.4), $t_{\mathrm{MIX}}^{\mathsf{x}_0}$ from an initial state $\mathsf{x}_0 = \mathsf{x}_{\mathrm{knt}}$ is asymptotically faster than from $\mathsf{x}_0 = \mathsf{x}_{\mathrm{rbw}}$.

To see this, write $c_k = \cos(\frac{2\pi k}{q})$, so that $\lambda_{k\mathbf{1}} = c_k$ and $\lambda_{\mathbf{1}}^* = c_1$. With this notation,
$$\lambda_{k\mathsf{v}} = \tfrac{d-1}{d} c_k + \tfrac{1}{d} c_{2k} \,.$$

Let $k' = 2k \pmod q$, and note that if $k' \neq 0$ and $c_{k'} \neq c_k$ then $\lambda_{k\mathsf{v}} < \max\{c_k, c_{k'}\} \leq \lambda_{\mathbf{1}}^*$. Indeed:
- The case $k' = 0$ occurs if and only if $k = q/2$ (for q even), whence $c_k = -1$ and $c_{k'} = 1$, so $\lambda_{k\mathsf{v}} = -1 + \frac{2}{d} \leq 0$ (strictly if $d \geq 3$) and $\lambda_{\mathbf{1}}^* = c_1 = \cos(\frac{2\pi}{q}) \geq 0$ (strictly if $q > 4$); thus, $\lambda_{k\mathsf{v}} < c_1$.
- The case $c_{k'} = c_k = c_1$ can occur only if $2\pi/q = \frac{2}{3}\pi$, i.e., when $q = 3$, precluded by assumption.

Note that, for $q = 4$, one can readily write the explicit values of $\lambda_{k\mathsf{v}}$ for $\mathsf{v} = (1, \ldots, 1, 2)$ (as c_k is simply 0 if k is odd and $(-1)^{k/2}$ if it is even), yielding $\lambda_{\mathsf{v}}^* = -1/d$ for $d \geq 3$ (and $\lambda_{\mathsf{v}}^* = 0$ for $d = 2$).

3.2. **Optimal initial condition in 1D.** In this section we establish Theorem 1.3, showing that the rainbow initial state $\mathsf{x}_{\mathrm{rbw}}$ is optimal in dimension $d = 1$ for all $q \geq 2$.

Proof of Theorem 1.3. Once we prove that $\mathcal{A}_t^{(2)}(\mathsf{x}_{\mathrm{rbw}}) = O(\mathcal{A}_t^{(2)}(\mathsf{x}_0))$ for every x_0 as per Eq. (1.9), the fact that $T_{\mathsf{x}_{\mathrm{rbw}}} \leq T_{\mathsf{x}_0} + O(1)$ will readily follow from Eq. (2.5) of Proposition 2.1.

From the discussion in Section 3.1 (specifically, Eq. (3.3)), we have
$$\mathcal{A}_t^{(2)}(\mathsf{x}_{\mathrm{rbw}}) = \frac{1}{q} \sum_{k=1}^{q-1} e^{-2(1-(1-\theta)\cos(\frac{2\pi k}{q}))t} \leq e^{-2(1-(1-\theta)\cos(\frac{2\pi}{q}))t} \,.$$

Hence, to prove Eq. (1.9) it remains to show that, for some absolute constant $c_0 > 0$,
$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) \geq \frac{c_0}{q^2} e^{-2(1-(1-\theta)\cos(\frac{2\pi}{q}))t} \,. \tag{3.5}$$

Note that Proposition 2.1 already gives such a bound for $q = 2$, as we have $\mathcal{A}_t^{(2)}(\mathsf{x}_0) \geq \frac{q}{q-1} e^{-(4-2\theta)t}$. Hence, we can assume $q \geq 3$ through the remainder of this proof.

We will identify the vertices of $G = \mathbb{Z}/n\mathbb{Z}$ with $0, \ldots, n-1$, and let
$$\psi_l = \left(\exp\left(\tfrac{2\pi i j l}{n}\right) \right)_{j=0}^{n-1} \qquad (l = 0, \ldots, n-1)$$

(the l-th n-dimensional Fourier vector), which form an $L^2(\pi_G)$-orthonormal basis of eigenfunctions for simple random walk on G, with corresponding eigenvalues
$$\lambda_l = \cos\left(\tfrac{2\pi l}{n}\right) \qquad (l = 0, \ldots, n-1) \,.$$

Next, let
$$\nu^{(k)} = \left(\langle \mathsf{x}_0^k, \psi_l \rangle_{\pi_G} \right)_{l=0}^{n-1} \,, \qquad (k = 0, 1, \ldots, q-1) \,, \tag{3.6}$$

i.e., $\nu^{(k)}$ is the discrete Fourier transform (DFT) of x_0^k, up to a factor of $1/n$ from π_G in $\langle \cdot, \cdot \rangle_{\pi_G}$. That is, if F is the DFT matrix, i.e., the symmetric matrix whose columns are $\overline{\psi_l}$, then $\nu^{(k)} = \frac{1}{n}\mathsf{F}\mathsf{x}_0^k$. We stress that $\nu^{(k)}$ have are unit vectors in L^2 by Parseval's identity:

$$\sum_{l=0}^{n-1}|\nu^{(k)}(l)|^2 = \sum_{l=0}^{n-1}|\langle \mathsf{x}_0^k, \psi_l \rangle_{\pi_G}|^2 = \langle \mathsf{x}_0^k, \mathsf{x}_0^k \rangle_{\pi_G} = \sum_{j=0}^{n-1}|\mathsf{x}_0^k(j)|^2\pi_G(j) = \sum_{j=0}^{n-1}\pi_G(j) = 1\,.$$

Recalling Eq. (2.8), which wrote $\mathcal{A}_t^{(2)}$ in terms of $\Psi_l^{(k)}(\mathsf{x}) = \langle \mathsf{x}^k, \psi_l \rangle_{\pi_G}$ from Eq. (2.7), we have

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \frac{1}{q}\sum_{k=1}^{q-1}\sum_{l=0}^{n-1}|\nu^{(k)}(l)|^2 e^{-2(1-(1-\theta)\lambda_l)t}\,. \tag{3.7}$$

Our goal will roughly be to show that the coefficients $|\nu^{(k)}(l)|^2$ for $l \leq (1+o(1))n/q$ have large enough mass c, so that by the monotonicity of λ_l we would get $\mathcal{A}_t^{(2)}(\mathsf{x}_0) \geq c\exp[-2(1-(1-\theta)\lambda_{n/q})t]$.

Now, let Θ_l denote the circular shift $(\Theta_l\mathsf{x})(j) = \mathsf{x}(j-l)$, and define

$$\nu_l^{(k)} := \Theta_l\nu^{(k)}\,.$$

We want to exploit the fact that the state $\mathsf{x}_{\mathrm{rbw}}$ yields an optimal basis in the following sense:

$$\left\{\nu_l^{(k)}\right\}_{\substack{k=0,\ldots,q-1 \\ l=0,\ldots,n/q-1}} = \left\{\delta_{k\frac{n}{q}+l}\right\}_{\substack{k=0,\ldots,q-1 \\ l=0,\ldots,n/q-1}} \qquad \text{if } \mathsf{x}_0 = \mathsf{x}_{\mathrm{rbw}}$$

(in the above and in what follows, δ_l is the indicator function $\mathbb{1}_{\{\cdot=l\}}$), while taking the same shifts with other initial configurations necessarily yields correlated vectors in the sector $\{\delta_0, \ldots \delta_{\frac{n}{q}}\}$ (which, as we will soon see, correspond to the coefficients that we are trying to bound from below).

To do so, let $\circledast$ denote circular convolution (i.e., $(\mathsf{x}\circledast y)(m) = \sum_{j=0}^{n-1}\mathsf{x}(j)y(m-j)$ with the index of y taken modulo n), so that

$$\nu_l^{(k)} = \nu^{(k)} \circledast \delta_l\,,$$

and recall that, if $\widehat{\mathsf{x}} = \mathsf{F}\mathsf{x}$ and $\widehat{y} = \mathsf{F}y$ are the DFTs of x and y, then $\langle \widehat{\mathsf{x}}, \widehat{y} \rangle = n\langle \mathsf{x}, y \rangle$, and $\mathsf{F}(\mathsf{x}\circledast y) = \widehat{\mathsf{x}}\cdot\widehat{y}$, the point-wise product of $\widehat{\mathsf{x}}$ and $\widehat{y}$. In the right hand above, $\nu^{(k)} = \frac{1}{n}\mathsf{F}\mathsf{x}_0^k$ and $\delta_l = \langle \psi_l, \psi_l \rangle_{\pi_G} = \frac{1}{n}\mathsf{F}\psi_l$, so $\mathsf{F}\overline{\nu^{(k)}} = \overline{\mathsf{x}_0^k}$ and $\mathsf{F}\delta_l = \overline{\psi_l}$, implying that $\mathsf{F}\overline{\nu_l^{(k)}} = \overline{\mathsf{x}_0^k}\cdot\overline{\psi_l}$. It then follows that

$$\langle \nu_{l_1}^{(k_1)}, \nu_{l_2}^{(k_2)} \rangle = \frac{1}{n}\langle \overline{\mathsf{F}\overline{\nu_{l_1}^{(k_1)}}}, \mathsf{F}\overline{\nu_{l_2}^{(k_2)}} \rangle = \frac{1}{n}\langle \overline{\mathsf{F}(\overline{\nu^{(k_1)}}\circledast\delta_{l_1})}, \mathsf{F}(\overline{\nu^{(k_2)}}\circledast\delta_{l_2}) \rangle = \frac{1}{n}\langle \mathsf{x}_0^{k_1}\cdot\psi_{l_1}, \mathsf{x}_0^{k_2}\cdot\psi_{l_2} \rangle$$

$$= \frac{1}{n}\langle \mathsf{x}_0^{k_1-k_2}, \psi_{l_2-l_1} \rangle = \nu^{(k_1-k_2)}(l_2-l_1)\,, \tag{3.8}$$

where we are reading both $k_1 - k_2$ and $l_1 - l_2$ modulo n. This shows that the Fourier coefficients $\nu^{(k)}(l)$ describe the correlations between the vectors $\{\nu_l^{(k)}\}$. To leverage this, consider the collection

$$\mathcal{B}_N := \{\nu_l^{(k)} \ : \ k = 0, 1, \ldots, q-1\,, \ l = 0, \ldots, N-1\}\,, \tag{3.9}$$

for $N \asymp n$ to be fixed later (in Eq. (3.11)). The correlations between $\mathsf{u}, \mathsf{v} \in \mathcal{B}_N$ satisfy the following:

$$\sum_{\substack{\mathsf{u},\mathsf{v}\in\mathcal{B}_N \\ \mathsf{u}\neq\mathsf{v}}}|\langle \mathsf{u},\mathsf{v} \rangle|^2 = \sum_{\substack{k_1,k_2\in\{0,\ldots,q-1\} \\ l_1,l_2\in\{0,\ldots,N-1\} \\ (k_1,l_1)\neq(k_2,l_2)}}|\nu^{(k_1-k_2)}(l_2-l_1)|^2 = \sum_{\substack{k_1,k_2\in\{0,\ldots,q-1\} \\ l_1,l_2\in\{0,\ldots,N-1\} \\ k_1\neq k_2}}|\nu^{(k_1-k_2)}(l_2-l_1)|^2$$

$$= q\sum_{k=1}^{q-1}\sum_{l=-(N-1)}^{N-1}(N-|l|)|\nu^{(k)}(l)|^2\,, \tag{3.10}$$

where the first equality is by Eq. (3.8); the second one used that $\nu^{(0)}(l) = \delta_0$ (as $\mathsf{x}_0^0 = \mathbf{1} = \psi_0$), so we only get a contribution from $k_1 = k_2$ if $l_1 = l_2$, disallowed in the first summation; finally, the transition between the lines counted the multiplicity of each $k = k_1 - k_2$ and $l = l_2 - l_1$ as follows:

- there are q pairs (k_1, k_2) realizing every $k = k_1 - k_2 \neq 0$ (for every k_1, take $k_2 = k + k_1$);
- there are $N - |l|$ pairs (l_1, l_2) realizing every $l = l_2 - l_1 \in \{-(N-1), \ldots, N-1\}$ (for $l \geq 0$, take $l_1 = 0, \ldots, N-1-l$ and $l_2 = l_1 + l$; for $l < 0$, take $l_2 = 0, \ldots, N-1-|l|$ and $l_1 = l_2 - l$).

To get a lower bound on the above coefficients, we use the following simple lemma, which is a variant of, e.g., [3, Lem. 9.1] and [22, Lem. 2] (those lemmas give bounds on $\max_{i \neq j} |\langle \mathsf{v}_i, \mathsf{v}_j \rangle|$ for unit vectors $\mathsf{v}_i \in \mathbb{R}^n$ as opposed to the version below that considers $\sum_{i \neq j} \langle |\mathsf{v}_i, \mathsf{v}_j \rangle|^2$ for $\mathsf{v}_i \in \mathbb{C}^n$, but the same proof applies to our setting here, and we include it for completeness).

Lemma 3.3. *Let $\mathsf{v}_1, \ldots, \mathsf{v}_{n+r} \in \mathbb{C}^n$ with $\|\mathsf{v}_i\| = 1$ for all i. Then $\sum_{i \neq j} |\langle \mathsf{v}_i, \mathsf{v}_j \rangle|^2 \geq r$.*

Proof. Consider the Gram matrix $G = (\langle \mathsf{v}_i, \mathsf{v}_j \rangle)_{i,j=1}^{n+r}$. As it is Hermitian with rank at most n ($G = B^* B$ for an $n \times (n+r)$ matrix B), the matrix $A = G - I$ has real eigenvalues $\{\lambda_i\}_{i=1}^{n+r}$ in which $\lambda = -1$ has multiplicity at least $(n + r) - n = r$. By the assumption that the v_i's are unit vectors, its diagonal is all-zero, whence its Frobenius norm $\|A\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(A^* A)}$ has $\|A\|_{\mathrm{F}}^2 = \sum_{i \neq j} |\langle \mathsf{v}_i, \mathsf{v}_j \rangle|^2$. At the same time, $\|A\|_{\mathrm{F}}^2 = \sum_i \lambda_i^2 \geq r$, as required. $\square$

Choose now

$$N = \frac{n}{q} + \frac{n}{15 + (1 - \theta)t}, \tag{3.11}$$

noting that

$$\frac{2\pi}{q} \leq \frac{2\pi N}{n} \leq \frac{2\pi}{q} + \frac{2\pi}{15} \leq \frac{4\pi}{5}$$

since $q \geq 3$. Then $\mathcal{B}_N$ from Eq. (3.9) has qN unit vectors in $\mathbb{C}^n$, and therefore, by Lemma 3.3,

$$\sum_{\substack{\mathsf{u}, \mathsf{v} \in \mathcal{B}_N \\ \mathsf{u} \neq \mathsf{v}}} |\langle \mathsf{u}, \mathsf{v} \rangle|^2 \geq qN - n = \frac{qn}{15 + (1 - \theta)t},$$

which, when combined with Eq. (3.10), gives

$$\sum_{k=1}^{q-1} \sum_{l=-(N-1)}^{N-1} (N - |l|) |\nu^{(k)}(l)|^2 \geq \frac{n}{15 + (1 - \theta)t}, \tag{3.12}$$

as well as, after plugging the trivial bound $N - |l| \leq N$ into the last display,

$$\sum_{k=1}^{q-1} \sum_{l=-(N-1)}^{N-1} |\nu^{(k)}(l)|^2 \geq \frac{1}{N} \cdot \frac{n}{15 + (1 - \theta)t} = \frac{q}{15 + q + (1 - \theta)t}. \tag{3.13}$$

In order to use these two estimates, we first bound $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ from Eq. (3.7) via

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) \geq \frac{1}{q} \sum_{k=1}^{q-1} \sum_{l=-(N-1)}^{N-1} |\nu^{(k)}(l)|^2 e^{-2(1 - (1 - \theta)\lambda_l)t},$$

and observe that, for each $-N < l < N$,

$$e^{-2(1 - (1 - \theta)\lambda_l)t} = e^{-2(1 - (1 - \theta)\cos(\frac{2\pi N}{n}))t} \exp\left(2\left(\cos\left(\frac{2\pi |l|}{n}\right) - \cos\left(\frac{2\pi N}{n}\right)\right)(1 - \theta)t\right)$$

$$\geq e^{-2(1 - (1 - \theta)\cos(\frac{2\pi N}{n}))t}\left(1 + 2\left(\cos\left(\frac{2\pi |l|}{n}\right) - \cos\left(\frac{2\pi N}{n}\right)\right)(1 - \theta)t\right).$$

Writing $x = 2\pi N/n \in [2\pi/q, 4\pi/5]$ and $y := 2\pi(N - |l|)/n \in [2\pi/n, x]$, in this range we have

$$\frac{\pi}{4q}y \le \cos(x - y) - \cos(x) \le y \,.$$

(The right inequality holds as $\cos(\cdot)$ is 1-Lipschitz; for the left one, when $y \le x/2$, the mean value theorem supports the lower bound $\delta_q y$ for $\delta_q = \min\{\sin \xi : \frac{\pi}{q} < \xi < \frac{4}{5}\pi\} \ge \pi/(2q)$, and if $y \ge x/2$ then we have the lower bound $\cos(x/2) - \cos x \ge \delta_q x/2 \ge \delta_q y/2$ for the same δ_q.) It follows that

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) \ge \frac{1}{q}e^{-2(1-(1-\theta)\cos(\frac{2\pi N}{n}))t} \sum_{k=1}^{q-1} \sum_{l=-(N-1)}^{N-1} |\nu^{(k)}(l)|^2 \left(1 + \frac{N - |l|}{n} \cdot \frac{\pi^2}{q}(1 - \theta)t\right)$$

$$\ge \frac{1}{q}e^{-2(1-(1-\theta)\cos(\frac{2\pi N}{n}))t} \left(\frac{q}{15 + q + (1 - \theta)t} + \frac{\pi^2}{q} \cdot \frac{(1 - \theta)t}{15 + (1 - \theta)t}\right) \,,$$

where between the lines we used Eqs. (3.12) and (3.13). Finally, as $\cos(\cdot)$ is 1-Lipschitz,

$$\cos\left(\frac{2\pi N}{n}\right) \ge \cos\left(\frac{2\pi}{q}\right) - \frac{2\pi}{15 + (1 - \theta)t} \,,$$

and therefore

$$\mathcal{A}_t^{(2)}(\mathsf{x}_0) \ge \frac{1}{q}e^{-2(1-(1-\theta)\cos(\frac{2\pi}{q}))t} \max\left\{\frac{q}{15 + q + (1 - \theta)t}, \frac{\pi^2}{q} \cdot \frac{(1 - \theta)t}{15 + (1 - \theta)t}\right\} e^{-4\pi \frac{(1-\theta)t}{15+(1-\theta)t}} \,.$$

If $(1 - \theta)t \le 1$ then the first term in the maximum is at least $1/7$, whereas if $(1 - \theta)t \ge 1$ then the second term in the maximum is at least $\pi^2/(16q)$. The last exponent is at least $e^{-4\pi}$, and when combined, this establishes Eq. (3.5) and thus concludes the proof. $\qquad\square$

4. UPPER BOUNDS

In this section, we prove the upper bound in Theorem 1.1, from which the upper bound in Corollary 1.4 will readily follow (see Remark 4.14).

Fix $\theta \in (0, 1]$ and $q \ge 2$. Let $G = (V, E)$ be a graph on n vertices satisfying Eq. (1.1) for some fixed $\alpha, \mathfrak{c}_0$, and let x_0 be an initial configuration on G. (We do not require that G to be connected for this part of the proof.) Define

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(t) := \|\mathbb{P}_{\mathsf{x}_0}(X_t \in \cdot) - \mu_G\|_{\mathrm{TV}}$$

and

$$t_0 := \max\left\{T_{\mathsf{x}_0}, \frac{1}{4\theta}\log n\right\}, \tag{4.1}$$

$$t_1 := t_0 + b_1(\log n)^{1-\frac{\alpha}{2}} \quad \text{for} \quad b_1 = 3/\theta\,, \tag{4.2}$$

$$t_2 := t_1 + b_2(\log n)^{1-\frac{\alpha}{2}} \quad \text{for} \quad b_2 = 2/\theta\,. \tag{4.3}$$

The goal will be to show that

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(t_2) \le Ce^{-\frac{1}{2}(\log n)^{1-\frac{\alpha}{2}}}\,, \tag{4.4}$$

for some $C > 0$ depending on $\theta, q, \alpha, \mathfrak{c}_0$ (but not x_0 and G).

Remark 4.1. If one wanted to write an upper bound that would hold for every n (as opposed to n large enough), and at the same time keeps the constant prefactor $C > 0$ reasonably small, the following could be done. Using the representation detailed in Section 2.2, one can show that $t_{\mathrm{MIX}}^{\mathsf{x}_0}(\varepsilon) \le \theta^{-1}(\log n + \log(1/\varepsilon))$, since we can couple X_t to the stationary distribution on the event that all the walks $(Z_s^v)_{v\in V}$ have died (occurring with probability at least $1 - ne^{-\theta t}$).

To prove Eq. (4.4), we follow the recipe of the analogous upper bound proof in [19], with some modifications due to the fact that we are working with a slightly more general model and on an arbitrary geometry, rather than on a lattice. The last part of the proof instead relies on the complete monotonicity of $\mathcal{A}_t^{(2)}(x_0)$, proven in Proposition 2.1.

Throughout this section, we will be making ample use of the duality with coalescing random walks and the couplings discussed in Sections 2.1.1 and 2.1.2. Hence, for the fixed time t_2, we can look at the dynamics on $[0, t_2]$ backwards in time and analyze the walks $\{Z_s^v\}_{v \in V, s \in [0,t_2]}$. We refer to the locations of the walks at some time s (or $t_2 - s$ in the forward dynamics) as the *histories* of the walks at time $t_2 - s$. We also refer to the collective locations of the walks that have not been killed by time s (or $t_2 - s$ in the forward dynamics) as the *surviving* histories at time $t_2 - s$.

4.1. **Reducing the total variation distance at time t_2 to a product measure at time t_1.** In this subsection we will establish the following result that, by analyze the backward dynamics in the interval $[t_1, t_2]$ from Eqs. (4.2) and (4.3), reduces $d_{\mathrm{TV}}^{x_0}(t_2)$ to the total variation distance between the projection of X_{t_1} on certain subsets V_i and the product measure of $\mu_G\!\restriction_{V_i}$.

Lemma 4.2. *Call a collection of subsets of vertices $\{V_i\}_{i=1}^m$ (for $m = m(n)$)* good *if, for all i, j,*

$$|V_i| \le \mathfrak{c}_0 \exp\left[(\log n)^{1-\frac{\alpha}{2}}\right], \quad \mathrm{dist}_G(V_i, V_j) \ge (\log n)^{1+\beta},$$

where $\beta := \frac{1}{2}\frac{\alpha^3}{1-\alpha}$. If $X_t^{(i)}$ are i.i.d. copies of X_t, and $Y^{(i)}$ are i.i.d. copies of μ_G, then

$$d_{\mathrm{TV}}^{x_0}(t_2) \le \max_{\{V_i\}_{i=1}^m \text{ good}} \left\| \prod_{i=1}^m \mathbb{P}_{x_0}\left(X_{t_1}^{(i)}(V_i) \in \cdot\right) - \prod_{i=1}^m \mathbb{P}\left(Y^{(i)}(V_i) \in \cdot\right) \right\|_{\mathrm{TV}} + O(n^{-9}).$$

Remark 4.3. We will in fact show a stronger statement, where the maximum over $\{V_i\}$ is replaced by an expectation over random collections $\{V_i\}$ that are good with an extra indicator that the surviving history at time t_1 is precisely $\bigcup_i V_i$ (see the event E defined in Lemma 4.6 and Eq. (4.5)).

Remark 4.4. We will reduce to good collections $\{V_i\}$ where each V_i is a subset of some ball $B_v(r)$ whose size is at most $\mathfrak{c}_0 \exp[(\log n)^{1-\frac{\alpha}{2}}]$. Since the proofs in later subsections will only rely on the sizes of the V_i's and their pairwise distances, we did not include this in the criteria for being good.

We begin with an elementary observation that, in order for the random walk (Z_s^v) to reach distance at least r from its origin v, its first r updates must be moving updates (having rate $1 - \theta$) as opposed to a killing one (having rate θ), the probability of which is at most $(1-\theta)^r \le e^{-\theta r}$.

Fact 4.5. *For every fixed $\kappa > 0$ and $\beta > 0$, the probability that that at least one of the n random walks $(Z_s^v)_{v \in V}$ travels a distance of at least $\kappa(\log n)^{1+\beta}$ is at most $n \exp[-\theta\kappa(\log n)^{1+\beta}]$.*

The next lemma will show that the surviving histories at time t_1 are typically good.

Lemma 4.6. *Let $\{V_i\}$ be the sets obtained by starting with a singleton per surviving walk at time t_1, and repeatedly merging V_i, V_j if $\mathrm{dist}_G(V_i, V_j) < (\log n)^{1+\beta}$. Let E be the event that $\{V_i\}$ is good. Then, we have $\mathbb{P}(E) \ge 1 - O(n^{-9})$.*

Proof. For any $v \in V$, let F_v be the event that Z_s^v (the random walk started at v in the backward dynamics) does not leave $B_v(\frac{1}{2}\log^{1+\beta} n)$ for $s \in [0, b_2(\log n)^{1-\frac{\alpha}{2}}]$, and let $F = \bigcap_{v \in V} F_v$. By Fact 4.5,

$$\mathbb{P}(F) = 1 - O(n^{-10}).$$

Next, we want to show that with high probability, there exists an annulus of logarithmic thickness around any $v \in V$ such that all walks started at the vertices in such annulus are killed by time $b_2(\log n)^{1-\frac{\alpha}{2}}$ (that is, before reaching t_1 backwards in time). For $\gamma := \frac{\alpha}{2} + \frac{\alpha^2}{2}$, define the annuli

$$A_v(k) = B_v(2k \cdot 2(\log n)^{1+\beta}) \setminus B_v((2k-1) \cdot 2(\log n)^{1+\beta}), \quad k = 1, \ldots, \tfrac{1}{4}(\log n)^\gamma.$$

Note that for fixed v, the annuli $\{A_v(k)\}_{k \geq 1}$ are at graph distance at least $2(\log n)^{1+\beta}$ from each other, and that

$$|A_v(k)| \leq |B_v(4k(\log n)^{1+\frac{\alpha}{2}})| \leq \mathfrak{c}_0 \exp\left(\left[2k \cdot 2(\log n)^{1+\beta}\right]^{1-\alpha}\right) \leq \mathfrak{c}_0 e^{(\log n)^{(1+\beta+\gamma)(1-\alpha)}} = \mathfrak{c}_0 e^{(\log n)^{1-\frac{\alpha}{2}}}$$

by Eq. (1.1) and using $k \leq \frac{1}{2}(\log n)^\gamma$ and $(1 + \beta + \gamma)(1 - \alpha) = 1 - \frac{\alpha}{2}$. For each v and k, let $G_{v,k}$ denote the event that all the walks started at $A_v(k)$ are killed by time $b_2(\log n)^{1-\frac{\alpha}{2}}$. Then, since a single walk survives in such time with probability $\exp[-\theta b_2(\log n)^{1-\frac{\alpha}{2}}]$, a union bound yields

$$\mathbb{P}(G_{v,k}^c) \leq |A_v(k)| e^{-\theta b_2 (\log n)^{1-\frac{\alpha}{2}}} \leq \mathfrak{c}_0 e^{-(\log n)^{1-\frac{\alpha}{2}}}$$

via the aforementioned bound on $|A_v(k)|$ and the choice $b_2 = 2/\theta$. It follows that

$$\mathbb{P}(G_{v,k}^c \mid F) \leq \frac{\mathbb{P}(G_{v,k}^c)}{\mathbb{P}(F)} = 2\mathfrak{c}_0 \exp\left[-(\log n)^{1-\frac{\alpha}{2}}\right].$$

for large enough n. Let $G_v = \bigcup_{k=1}^{\frac{1}{4}(\log n)^\gamma} G_{v,k}$ be the event that all the walks started from at least one of the annuli $A_v(k)$ are killed by time $b_2(\log n)^{1-\frac{\alpha}{2}}$. Then, noting the events $G_{v,k}$ are conditionally independent given F (as the random walks from different annuli cannot coalesce under F), we get

$$\mathbb{P}(G_v^c \mid F) = \prod_{k=1}^{\frac{1}{4}(\log n)^\gamma} \mathbb{P}(G_{v,k}^c \mid F) \leq \left(2\mathfrak{c}_0 e^{-(\log n)^{1-\frac{\alpha}{2}}}\right)^{\frac{1}{4}(\log n)^\gamma} = e^{-(\frac{1}{4}-o(1))(\log n)^{1+\alpha^2/2}} = O(n^{-10}),$$

where we used the fact that $1 - \alpha/2 + \gamma = 1 + \alpha^2/2$.

Hence, if $G = \bigcap_{v \in V} G_v$ is the event that every vertex $v \in V$ is surrounded by some annulus $A_v(k)$ (with $1 \leq k \leq \frac{1}{4}(\log n)^\gamma$) whose random walks are killed by time $b_2(\log n)^{1-\frac{\alpha}{2}}$, a union bound gives

$$\mathbb{P}(F^c \cup G^c) \leq \mathbb{P}(F^c) + \sum_{v \in V} \mathbb{P}(G_v^c \mid F) = O(n^{-9}).$$

On the event $F \cap G$, at time t_1 every surviving walk must be

(1) within $\frac{1}{2}(\log n)^{1+\beta}$ of its initial position, and
(2) surrounded by an annulus A of thickness $2(\log n)^{1+\beta}$, that is contained in a ball B of size at most $\mathfrak{c}_0 e^{(\log n)^{1-\frac{\alpha}{2}}}$, such that all random walks started from A were killed by time t_1.

This implies that, on $F \cap G$, every surviving vertex is within a ball of size at most $\mathfrak{c}_0 e^{(\log n)^{1-\frac{\alpha}{2}}}$ and is surrounded by an annulus, vacant from walks that survived to time t_1, of thickness at least $(2 - \frac{1}{2} - \frac{1}{2})(\log n)^{1+\beta} = (\log n)^{1+\beta}$. It follows that $F \cap G \subset E$, and so $\mathbb{P}(E) \geq 1 - O(n^{-9})$. $\square$

To reduce $\|\mathbb{P}_{x_0}(X_{t_2} \in \cdot) - \mu_G\|_{\mathrm{TV}}$ to $\|\mathbb{P}_{x_0}(X_{t_1}(\bigcup V_i) \in \cdot) - \mu_G\|_{\mathrm{TV}}$ for the well-separated surviving components $\{V_i\}$ whose existence was addressed in Lemma 4.6, we will use the *update support* notion from [14, 15]. This notion takes a much simpler form in our setting of the noisy voter model compared to the Ising model. We include the proof of the following simple lemma—taken verbatim from [15]—both for completeness, and to justify a slightly generalized version of it (for a product of multiple chains), which we will later need and follows from the same argument.

Lemma 4.7 (Special case of [15, Lem. 3.3])**.** *For any $t \geq \hat{t} \geq 0$ and any set W of vertices at time t in the space-time slab, if $\hat{W}$ denotes the (random) history of W at time $\hat{t}$, we have*

$$\left\| \mathbb{P}_{\mathsf{x}_0}(X_t(W) \in \cdot) - \mu_G \!\restriction_W \right\|_{\mathrm{TV}} \leq \mathbb{E}\left[\left\| \mathbb{P}_{\mathsf{x}_0}(X_{\hat{t}}(\hat{W}) \in \cdot) - \mu_G \!\restriction_{\hat{W}} \right\|_{\mathrm{TV}} \right],$$

where the expectation is with respect to the update sequences along $[\hat{t}, t]$. Furthermore, if $\{X_t^{(i)}\}$ are i.i.d. copies of the dynamics, and $\tau_i \in [0, t]$ are stopping times for their backward dynamics, then

$$\left\| \prod \mathbb{P}_{\mathsf{x}_0}\big(X_t^{(i)}(W_i) \in \cdot\big) - \prod \mu_G^{(i)} \!\restriction_{W_i} \right\|_{\mathrm{TV}} \leq \mathbb{E}\left[\Big\| \prod \mathbb{P}_{\mathsf{x}_0}(X_{t-\tau_i}(\hat{W}_i) \in \cdot) - \prod \mu_G \!\restriction_{\hat{W}_i} \Big\|_{\mathrm{TV}} \right].$$

Proof. Consider two instances of the dynamics, X_t and Y_t, started at x_0 and μ_G respectively. As explained in Section 2.1.2, we couple the backward dynamics of X_t, Y_t using the same update sequence. Given the cumulative set of updates $\mathcal{F}_{\hat{t},t}$ in the interval $[\hat{t}, t]$, the configurations $X_t(W)$ and $Y_t(W)$ at time t become deterministic functions $f_{\mathcal{F}_{\hat{t},t}}(X_{\hat{t}}(\hat{W}))$ and $f_{\mathcal{F}_{\hat{t},t}}(Y_{\hat{t}}(\hat{W}))$ (with the same $f_{\mathcal{F}_{\hat{t},t}}$) of the configurations at time $\hat{t}$ restricted to the surviving history $\hat{W}$ (as the killed histories are accounted for by $\mathcal{F}_{\hat{t},t}$). Hence, for any set of configurations Γ and a realization of $\mathcal{F}_{\hat{t},t}$,

$$\mathbb{P}_{\mathsf{x}_0}(X_t(W) \in \Gamma \mid \mathcal{F}_{\hat{t},t}) = \mathbb{P}_{\mathsf{x}_0}(f_{\mathcal{F}_{\hat{t},t}}(X_{\hat{t}}(\hat{W})) \in \Gamma),$$

$$\mathbb{P}_{\mu_G}(X_t(W) \in \Gamma \mid \mathcal{F}_{\hat{t},t}) = \mathbb{P}_{\mu_G}(f_{\mathcal{F}_{\hat{t},t}}(Y_{\hat{t}}(\hat{W})) \in \Gamma).$$

Thus, $\|\mathbb{P}_{\mathsf{x}_0}(X_t(W) \in \cdot) - \mathbb{P}_{\mu_G}(Y_t(W) \in \cdot)\|_{\mathrm{TV}} = \max_\Gamma \big(\mathbb{P}_{\mathsf{x}_0}(X_t(W) \in \Gamma) - \mathbb{P}_{\mu_G}(Y_t(W) \in \Gamma)\big)$ becomes

$$\max_\Gamma \int \Big(\mathbb{P}_{\mathsf{x}_0}(f_{\mathcal{F}_{\hat{t},t}}(X_{\hat{t}}(\hat{W})) \in \Gamma) - \mathbb{P}_{\mu_G}(f_{\mathcal{F}_{\hat{t},t}}(Y_{\hat{t}}(\hat{W})) \in \Gamma) \Big) \, \mathrm{d}\mathbb{P}(\mathcal{F}_{\hat{t},t})$$

$$\leq \int \max_\Gamma \Big(\mathbb{P}_{\mathsf{x}_0}(f_{\mathcal{F}_{\hat{t},t}}(X_{\hat{t}}(\hat{W})) \in \Gamma) - \mathbb{P}_{\mu_G}(f_{\mathcal{F}_{\hat{t},t}}(Y_{\hat{t}}(\hat{W})) \in \Gamma) \Big) \, \mathrm{d}\mathbb{P}(\mathcal{F}_{\hat{t},t})$$

$$\leq \int \Big\| \mathbb{P}_{\mathsf{x}_0}(X_{\hat{t}}(\hat{W}) \in \cdot) - \mathbb{P}_{\mu_G}(Y_{\hat{t}}(\hat{W}) \in \cdot) \Big\|_{\mathrm{TV}} \, \mathrm{d}\mathbb{P}(\mathcal{F}_{\hat{t},t}),$$

where in the last inequality we used that a projection can only reduce total variation distance.

For the generalization, note first that we can replace $\hat{t}$ by $t - \tau$ follows from exactly the same proof, as we may replace $\mathcal{F}_{\hat{t},t}$ everywhere with the stopped σ-algebra $\mathcal{F}_\tau$ for the backward dynamics. (Issues pertaining to the event $\{\tau = s\}$ having probability 0 can easily be avoided by running the discrete backward dynamics while tracking the cumulative associated rescaling time in tandem.) For a product chain $(X_t^{(i)})_{i=1}^m$ with individual stopping times τ_i for the backward dynamics $Z_s^{(i)}$, we expose the stopped chain $Z_{s \wedge \tau_i}^{(i)}$ in each coordinate, and the required result follows. $\square$

Proof of Lemma 4.2. Applying Lemma 4.7 with $t' = t_1$, $t = t_2$ and $W = V$, we write $W' = \bigcup V_i$ as prescribed in Lemma 4.6, and further apply that lemma to find that

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(t_2) \leq \mathbb{E}\left[\left\| \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1}\big(\bigcup V_i\big) \in \cdot\big) - \mu_G \!\restriction_{\bigcup V_i} \right\|_{\mathrm{TV}} \mathbb{1}_E \right] + O(n^{-9}), \tag{4.5}$$

and in particular,

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(t_2) \leq \max_{\{V_i\} \text{ good}} \left\| \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1}\big(\bigcup V_i\big) \in \cdot\big) - \mu_G \!\restriction_{\bigcup V_i} \right\|_{\mathrm{TV}} + O(n^{-9}). \tag{4.6}$$

Consider some realization of a good $\{V_i\}_{i=1}^m$, and let us examine the dynamics in the interval $[0, t_1]$. Running the backward dynamics from time t_1, we couple the vector $(X_{t_1}(V_i))_{i=1}^m$ with $(X_{t_1}^{(i)}(V_i))_{i=1}^m$, where the $X_t^{(i)}$ are i.i.d., by letting the random walks and noisy updates generating the two vectors be identical as long as no walks from two different V_i's meet. Should the walks from V_i and V_j

meet, we let them coalesce normally under $(X_{t_1}(V_i), X_{t_1}(V_j))$, and instead let them stay independent under $(X_{t_1}^{(i)}(V_i), X_{t_1}^{(j)}(V_j))$. Consequently, the two vectors are identical if no two walks from different V_i's meet, which happens for example if no walk travels more than $(\log n)^{1+\beta}$. Hence, by Fact 4.5,

$$\left\| \mathbb{P}_{\mathsf{x}_0}\big((X_{t_1}(V_1),\dots,X_{t_1}(V_m)) \in \cdot\big) - \mathbb{P}_{\mathsf{x}_0}\big((X_{t_1}^{(1)}(V_1),\dots,X_{t_1}^{(m)}(V_m)) \in \cdot\big)\right\|_{\mathrm{TV}} \leq O(n^{-10}).$$

For the same reason (using $\mathbb{P}_{\mu_G}$ in lieu of $\mathbb{P}_{\mathsf{x}_0}$ for the $\{Y(V_i)\}$), we get that

$$\left\| \mu_G{\restriction}_{\bigcup_{i=1}^m V_i} - (\mu_G{\restriction}_{V_1} \times \dots \times \mu_G{\restriction}_{V_m})\right\|_{\mathrm{TV}} \leq O(n^{-10}).$$

Combining the last two inequalities shows that for every good $\{V_i\}_{i=1}^m$,

$$\left\| \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1}(\bigcup V_i) \in \cdot\big) - \mu_G{\restriction}_{\bigcup V_i}\right\|_{\mathrm{TV}} \leq \left\| \prod_{i=1}^m \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1}^{(i)}(V_i) \in \cdot\big) - \prod_{i=1}^m \mathbb{P}\big(Y^{(i)}(V_i) \in \cdot\big)\right\|_{\mathrm{TV}} + O(n^{-10}).$$

which together with Eq. (4.6) concludes the proof. $\qquad\square$

4.2. L^1-L^2 **reduction at time** t_1**.** Thus far we have reduced $d_{\mathrm{TV}}^{\mathsf{x}_0}(t_2)$ to the total variation distance between two product measures, namely $(X_{t_1}^{(i)}(V_i))_{i=1}^m$ and $(\mu_G{\restriction}_{V_i})_{i=1}^m$ for a good collection $\{V_i\}_{i=1}^m$. In this subsection, we bound the latter as follows.

Lemma 4.8. *The following holds for every* good *collection* $\{V_i\}_{i=1}^m$*. Let* τ_i *be the minimum* $s \geq 0$ *where the backward dynamics, started at time* t_1 *from every* $v \in V_i$*, has its surviving history reach size at most* 3 *at time* $t_1 - s$*, or* $\tau_i = t_1$ *if the backward dynamics reaches time* 0 *before this happens. Further let* U_i *be the surviving history at time* τ_i*. Then there exists some* $C = C(q, \theta) > 0$ *such that*

$$\left\| \prod_{i=1}^m \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1}^{(i)}(V_i) \in \cdot\big) - \prod_{i=1}^m \mu_G{\restriction}_{V_i}\right\|_{\mathrm{TV}}^2 \leq C\,\mathbb{E}\left[\sum_{i=1}^m \left\| \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1-\tau_i}^{(i)}(U_i) \in \cdot\big) - \mu_G{\restriction}_{U_i}\right\|_{\mathrm{TV}}^2 \mathbb{1}_{\{\tau_i < t_1\}}\right]$$
$$+ O\big(e^{-8(\log n)^{1-\frac{\alpha}{2}}}\big).$$

Proof. We first examine the event $\{\tau_i < t_1\}$. The probability that 4 fixed vertices $v_1, \dots, v_4$ in V_i have each of the $Z_s^{v_j}$ survive to time 0 without ever coalescing is at most

$$e^{-4\theta t_1} \leq \frac{1}{n} e^{-12(\log n)^{1-\frac{\alpha}{2}}},$$

using Eq. (4.2). Since $|V_i| \leq \mathfrak{c}_0 \exp\big[(\log n)^{1-\frac{\alpha}{2}}\big]$, we thus have

$$\mathbb{P}(\tau_i = t_1) \leq \binom{|V_i|}{4} e^{-4\theta t_1} \leq \frac{\mathfrak{c}_0^4}{n} e^{-8(\log n)^{1-\frac{\alpha}{2}}},$$

and can conclude from a union bound over the $m \leq n$ sets V_i that

$$\mathbb{P}\left(\max_{1 \leq i \leq m} \tau_i < t_1\right) = 1 - O\big(e^{-8(\log n)^{1-\frac{\alpha}{2}}}\big).$$

We now apply Lemma 4.7, using the stopping time τ_i for $X_t^{(i)}$, inferring that

$$\left\| \prod_{i=1}^m \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1}^{(i)}(V_i) \in \cdot\big) - \prod_{i=1}^m \mu_G{\restriction}_{V_i}\right\|_{\mathrm{TV}} \leq \mathbb{E}\left[\left\| \prod_{i=1}^m \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1-\tau_i}^{(i)}(U_i) \in \cdot\big) - \prod_{i=1}^m \mu_G{\restriction}_{U_i}\right\|_{\mathrm{TV}}\right]$$
$$\leq \mathbb{E}\left[\left\| \prod_{i=1}^m \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1-\tau_i}^{(i)}(U_i) \in \cdot\big) - \prod_{i=1}^m \mu_G{\restriction}_{U_i}\right\|_{\mathrm{TV}} \mathbb{1}_{\{\max_i \tau_i < t_1\}}\right] + O\big(e^{-8(\log n)^{1-\frac{\alpha}{2}}}\big). \quad (4.7)$$

We now wish to use the L^1-L^2 reduction for product chains (we only need one side of it here):

Proposition 4.9 ([15, Prop. 7]). *Let $X_t = (X_t^{(1)}, \ldots, X_t^{(m)})$ for mutually independent ergodic chains with stationary distributions π_i. Let $\pi = \prod_{i=1}^m \pi_i$ and $\mathfrak{M}_t = \sum_{i=1}^m \|\mathbb{P}(X_t^{(i)} \in \cdot) - \pi_i\|_{L^2(\pi_i)}^2$. Then $\|\mathbb{P}(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \leq \sqrt{\mathfrak{M}_t}$ for any $t > 0$; in particular, if $\mathfrak{M}_t \to 0$ then $\|\mathbb{P}(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \to 0$. Conversely, if $\mathfrak{M}_t \to \infty$ then $\|\mathbb{P}(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \to 1$.*

Remark 4.10. The upper bound in Proposition 4.9 is straightforward and, as seen in that proof, holds more generally: if $\nu = \prod_{i=1}^m \nu_i$ and $\pi = \prod_{i=1}^m \pi_i$ then $\|\nu - \pi\|_{\mathrm{TV}} \leq (\sum_{i=1}^m \|\nu_i - \pi_i\|_{L^2(\pi_i)}^2)^{1/2}$.

By Proposition 4.9 (we apply its upper bound as per Remark 4.10), we infer from Eq. (4.7) along with Jensen's inequality that

$$\Big\|\prod_{i=1}^m \mathbb{P}_{\mathsf{x}_0}\big(X_{t_1}^{(i)}(V_i) \in \cdot\big) - \prod_{i=1}^m \mu_G\!\restriction_{V_i}\Big\|_{\mathrm{TV}}^2 \leq \mathbb{E}\big[\mathfrak{M}\mathbb{1}_{\{\max_i \tau_i < t_1\}}\big] + O\big(e^{-8(\log n)^{1-\frac{\alpha}{2}}}\big), \qquad (4.8)$$

for the random variable $\mathfrak{M} = \sum_{i=1}^m \|\nu_i - \pi_i\|_{L^2(\pi_i)}^2$ with $\nu_i = \mathbb{P}_{\mathsf{x}_0}(X_{t_1-\tau_i}(U_i) \in \cdot)$ and $\pi_i = \mu_G\!\restriction_{U_i}$. Observe that we deterministically have

$$\|\nu_i - \pi_i\|_{L^2(\pi_i)}^2 = \sum_{\mathsf{x} \in \mathfrak{C}_q^{U_i}} \frac{|\nu_i(\mathsf{x}) - \pi_i(\mathsf{x})|^2}{\pi_i(\mathsf{x})} \leq \frac{\|\nu_i - \pi_i\|_\infty \|\nu_i - \pi_i\|_1}{\min_{\mathsf{x}} \pi_i(\mathsf{x})} \leq 2\Big(\frac{q}{\theta}\Big)^{|U_i|} \|\nu_i - \pi_i\|_{\mathrm{TV}}^2,$$

using that $\min_{\mathsf{x}} \pi_i(\mathsf{x}) \geq (\theta/q)^{|U_i|}$, since any configuration x can be obtained by letting the first update in the backward dynamics of each vertex $v \in U_i$ be a noisy update (with probability θ) that assigns the new value $\mathsf{x}(v)$ (with probability $1/q$). As $|U_i| \leq 3$ when $\tau_i < t_1$, we get

$$\mathbb{E}\big[\mathfrak{M}\mathbb{1}_{\{\max_i \tau_i < t_1\}}\big] \leq 2\frac{q^3}{\theta^3} \mathbb{E}\left[\sum_{i=1}^m \Big\|\mathbb{P}_{\mathsf{x}_0}\big(X_{t_1-\tau_i}^{(i)}(U_i) \in \cdot\big) - \mu_G\!\restriction_{U_i}\Big\|_{\mathrm{TV}}^2 \mathbb{1}_{\{\tau_i < t_1\}}\right],$$

which, together with Eq. (4.8), completes the proof. $\square$

4.3. Reduction to the autocorrelation $\mathcal{A}_t^{(2)}$.

With Lemma 4.8 in mind, define

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, t) := \Big\|\mathbb{P}_{\mathsf{x}_0}\big(X_{t_1-\tau_i}^{(i)}(U_i) \in \cdot\big) - \mu_G\!\restriction_{U_i}\Big\|_{\mathrm{TV}}^2. \qquad (4.9)$$

The final ingredient in the proof will be to establish the following.

Lemma 4.11. *The following holds for every `good` collection $\{V_i\}_{i=1}^m$. Let U_i and τ_i be as in Lemma 4.8. Then, for some $C = C(q, \theta, \mathfrak{c}_0, \alpha) > 0$, the $d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, t)$ from Eq. (4.9) satisfy*

$$\mathbb{E}\left[\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}\right] \leq C(1 + t_1^3)e^{4(\log n)^{1-\frac{\alpha}{2}}} n(\mathcal{A}_{t_1}^{(2)}(\mathsf{x}_0) \vee e^{-4\theta t_1}) + O(n^{-9}).$$

To obtain this bound, we will aim to reduce the quantity $\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}$ to one involving single sites, rather than subsets U_i of size at most 3.

Remark 4.12. Roughly put, the main obstacle in this part of the proof is that, given the generality of x_0 and G, the function $s \mapsto d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, t_1 - s)$ can be highly irregular. Rather than treating the terms in this sum individually, our strategy will be to reduce their total to the function $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$, which, as we showed in Proposition 2.1, is completely monotone. To do so, we will need to sum $d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, \cdot)$ at identical times, and so we need to account for the fact that the times $\{\tau_i\}$ have different distributions. We overcome this by first replacing the U_i with deterministic sets, then bounding the densities of the τ_i, and finally reducing the analysis to single sites, thus to $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$.

While the sets V_i were well-separated (at distance at least $(\log n)^{1+\beta}$), the U_i might not be. However, they are so with high probability; namely, if V_i' is the $\frac{1}{4}(\log n)^{1+\beta}$-thickening of V_i, i.e.,

$$V_i' := \bigcup_{v \in V_i} B_v\left(\tfrac{1}{4}(\log n)^{1+\beta}\right),$$

then Fact 4.5 (used with $\kappa = \frac{1}{4}$) shows that every U_i is contained in V_i' with probability $1 - O(n^{-10})$. Note that because $\mathrm{dist}_G(V_i, V_j) \geq (\log n)^{1+\beta}$, the V_i' satisfy

$$\mathrm{dist}_G(V_i', V_j') \geq \tfrac{1}{2}(\log n)^{1+\beta}.$$

Then, if $U_i^s \subset V_i'$ denotes the subset of V_i' that maximizes $d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, t_1 - s)^2$ (note that U_i is not determined by τ_i, while $U_i^{\tau_i}$ is), we have

$$\mathbb{E}\left[\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}\right] \leq \sum_{i=1}^m \mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^{\tau_i}, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}\right] + O(n^{-10}). \tag{4.10}$$

Next, we would like to integrate over the τ_i and reduce to common times for every i. The probability of $\tau_i \geq s$ is at most bounded by the probability that at least 4 walks started at V_i survive up to time s, which is at most $\binom{|V_i|}{4}e^{-4\theta s}$. Then, if we are to have $\tau_i \in (s, s+h)$, it must be that along that interval we remain with exactly 4 surviving walks (a single coalescence or noisy update before τ_i occurs), which then receive at least one update. The number of updates from that point is at most a Poisson$(4h)$ variable, and hence the probability it is nonzero is at most its mean, $4h$. Therefore,

$$\mathbb{P}\left(\tau_i \in (s, s+h)\right) \leq 4\mathfrak{c}_0^4 e^{4(\log n)^{1-\frac{\alpha}{2}}} e^{-4\theta s} h. \tag{4.11}$$

Hence, if F_i denotes the distribution function of τ_i, then

$$\mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^{\tau_i}, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}\right] \leq \int_0^{t_1} d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^s, t_1 - s)^2 \mathrm{d}F_i(s)$$

$$\leq 4\mathfrak{c}_0^4 e^{4(\log n)^{1-\frac{\alpha}{2}}} \int_0^{t_1} d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^s, t_1 - s)^2 e^{-4\theta s} \mathrm{d}s \tag{4.12}$$

(note that in the latter expression, U_i^s is a deterministic set for each s given V_i'). Hence,

$$\sum_{i=1}^m \mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^{\tau_i}, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}\right] \leq 4\mathfrak{c}_0^4 e^{4(\log n)^{1-\frac{\alpha}{2}}} \int_0^{t_1} \left(\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^s, t_1 - s)^2\right) e^{-4\theta s} \mathrm{d}s. \tag{4.13}$$

We now use the following:

Lemma 4.13. *Let $\{W_i\}_{i=1}^m$ be subsets with $|W_i| \leq 3$ and $\mathrm{dist}_G(W_i, W_j) \geq \frac{1}{2}(\log n)^{1+\beta}$, and let Δ denote the maximum degree in G. Then for every $t \geq 0$,*

$$\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(W_i, t)^2 \leq 18q\Delta n\left(t^2 \mathcal{A}_t^{(2)}(\mathsf{x}_0) + (t+1)e^{-4\theta t}\right) + O(n^{-9}).$$

Proof. Let ξ_i denote the minimum $r \geq 0$ where the backwards dynamics, started at time t from every $v \in W_i$, has its surviving history shrink to a single (random) vertex v_i at time $t - r$. We have that $\mathbb{P}(\xi_i > t) \leq \binom{|W_i|}{2}e^{-2\theta t}$, and applying Lemma 4.7, we get

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(W_i, t) \leq \mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i, t - \xi_i)\mathbb{1}_{\{\xi_i \leq t\}}\right] + \mathbb{P}(\xi_i > t) \leq \mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i, t - \xi_i)\mathbb{1}_{\{\xi_i \leq t\}}\right] + 3e^{-2\theta t}.$$

Just as at the beginning of Section 4.3, if W_i' is a $\frac{1}{8}(\log n)^{1+\beta}$-thickening of W_i, that is,

$$W_i' := \bigcup_{v \in W_i} B_v\left(\tfrac{1}{8}(\log n)^{1+\beta}\right),$$

then $v_i \in W_i'$ for every i, with probability at least $1 - O(n^{-10})$. Thus, if v_i^r denotes the vertex $v \in W_i'$ that maximizes $d_{\mathrm{TV}}^{\mathsf{x}_0}(v, t - r)$, then we can bound

$$\mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i, t - \xi_i)^2 \mathbb{1}_{\{\xi_i \leq t\}}\right] \leq \mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^{\xi_i}, t - \xi_i)^2 \mathbb{1}_{\{\xi_i \leq t\}}\right] + O(n^{-10}).$$

Exactly as in Eqs. (4.11) and (4.12), we have $\mathbb{P}(\xi_i \in (s, s + h)) \leq 2h\binom{|W_i|}{2}e^{-2\theta s}$, and so, if we let G_i denote the distribution function of ξ_i, then

$$\mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^{\xi_i}, t - \xi_i)\mathbb{1}_{\{\xi_i \leq t\}}\right] \leq \int_0^t d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^r, t - r)\mathrm{d}G_i(r) \leq 6\int_0^t d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^r, t - r)e^{-2\theta r}\mathrm{d}r.$$

Combining the last 4 displays, along with $(a + b)^2 \leq 2(a^2 + b^2)$ and Jensen's inequality, we get

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(W_i, t)^2 \leq 72\left(\int_0^t d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^r, t - r)e^{-2\theta r}\mathrm{d}r\right)^2 + 18e^{-4\theta t} + O(n^{-10})$$

$$\leq 72t\int_0^t d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^r, t - r)^2 e^{-4\theta r}\mathrm{d}r + 18e^{-4\theta t} + O(n^{-10}).$$

This yields

$$\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(W_i, t)^2 \leq 72t\int_0^t \left(\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^r, t - r)^2\right)e^{-4\theta r}\mathrm{d}r + 18ne^{-4\theta t} + O(n^{-9}). \tag{4.14}$$

Since $\mathrm{dist}_G(W_i, W_j) \geq \frac{1}{2}(\log n)^{1+\beta}$, we have $\mathrm{dist}_G(W_i', W_j') > \frac{1}{4}(\log n)^{1+\beta} > 0$ by our definition of the W_i'. In particular, the vertices $\{v_i^r\}$ are all distinct for any fixed r, and so we may replace the sum over v_i^r $(i = 1, \ldots, m)$ by one running over all n vertices:

$$\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^r, t - r)^2 \leq \sum_v d_{\mathrm{TV}}^{\mathsf{x}_0}(v, t - r)^2 \leq 2|E|\sum_v \pi_G(v)d_{\mathrm{TV}}^{\mathsf{x}_0}(v, t - r)^2,$$

using that $2|E|\pi_G(v) = d_G(v) \geq 1$. Observe that $d_{\mathrm{TV}}^{\mathsf{x}_0}(v, s)$ is nothing but the total variation distance of the single-site marginal $X_t(v)$ from the uniform distribution:

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(v, s)^2 = \left(\frac{1}{2}\sum_{\omega \in \mathfrak{C}_q} \left|\mathbb{P}_{\mathsf{x}_0}(X_s(v) = \omega) - \frac{1}{q}\right|\right)^2 \leq \frac{q}{4}\sum_{\omega \in \mathfrak{C}_q} \left(\mathbb{P}_{\mathsf{x}_0}(X_s(v) = \omega) - \frac{1}{q}\right)^2,$$

which, upon recalling the definition of $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ from Eq. (1.3), implies that

$$\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(v_i^r, t - r)^2 \leq \frac{q}{2}|E|\mathcal{A}_{t-r}^{(2)}(\mathsf{x}_0).$$

Revisiting Eq. (4.14), and writing $|E| \leq \Delta n/2$ (noting in passing that the maximum degree Δ is less than $\mathfrak{c}_0 e$ by Eq. (1.1)), we get

$$\sum_{i=1}^m d_{\mathrm{TV}}^{\mathsf{x}_0}(W_i, t)^2 \leq 18q\Delta n\left(t\int_0^t \mathcal{A}_{t-r}^{(2)}(\mathsf{x}_0)e^{-4\theta r}\mathrm{d}r + e^{-4\theta t}\right) + O(n^{-9}).$$

Now, by Proposition 2.1, there exist $\alpha_l \geq 0$ with $\sum_l \alpha_l = \mathcal{A}_0^{(2)}(\mathsf{x}_0) = \frac{q-1}{q}$ and exponents $\gamma_l \in [\theta, 2-\theta]$ such that $\mathcal{A}_{t-r}^{(2)}(\mathsf{x}_0) = \sum_{l=1}^n \alpha_l e^{-2\gamma_l(t-r)}$, and so it follows that

$$\int_0^t \mathcal{A}_{t-r}^{(2)}(\mathsf{x}_0)e^{-4\theta r}\mathrm{d}r = \sum_l \alpha_l \int_0^t e^{-2\gamma_l(t-r)-4\theta r}\mathrm{d}r \leq \sum_l \alpha_l \int_0^t (e^{-2\gamma_l t} + e^{-4\theta t})\mathrm{d}r$$

$$= t\mathcal{A}_t^{(2)}(\mathsf{x}_0) + \frac{q-1}{q}te^{-4\theta t},$$

where we used $e^{-2\gamma_l(t-r)-4\theta r} \le e^{-(2\gamma_l \wedge 4\theta)t}$. This concludes the proof. □

Proof of Lemma 4.11. Using Lemma 4.13, we deduce from Eq. (4.13) that

$$\sum_{i=1}^{m} \mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^{\tau_i}, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}\right] \le Ct_1^2 e^{4(\log n)^{1-\frac{\alpha}{2}}} n \int_0^{t_1} (\mathcal{A}_{t_1-s}^{(2)}(\mathsf{x}_0) + e^{-4\theta(t_1-s)})e^{-4\theta s}\mathrm{d}s + O(n^{-9}).$$

(4.15)

Put $s = b_1(\log n)^{1-\frac{\alpha}{2}}$ for $b_1 = 3/\theta$ as per (4.2), whence $t_1 = t_0 + s$. Writing $\mathcal{A}_t^{(2)}(\mathsf{x}_0) = \sum_{l=1}^{n} \alpha_l e^{-2\gamma_l t}$ as in the above lemma, with $\alpha_l \ge 0$ and $\sum_l \alpha_l = \frac{q-1}{q} < 1$, we use that $1 + \frac{q}{q-1} \le 3$ to bound

$$\mathcal{A}_{t_1-s}^{(2)}(\mathsf{x}_0) + e^{-4\theta(t_1-s)} \le 3\sum_{l=1}^{n} \alpha_l e^{-(2\gamma_l \wedge 4\theta)(t_1-s)} = 3\sum_{l=1}^{n} \alpha_l \left(e^{-2\gamma_l t_1} \vee e^{-4\theta t_1}\right) e^{(2\gamma_l \wedge 4\theta)s}$$

$$\le 3\left(\mathcal{A}_{t_1}^{(2)}(\mathsf{x}_0) \vee ne^{-4\theta t_1}\right) e^{4\theta s}.$$

Plugging this in Eq. (4.15), the integral there is bounded from above by $3t_1(\mathcal{A}_{t_1}^{(2)}(\mathsf{x}_0) \vee e^{-4\theta t_1})$, so

$$\sum_{i=1}^{m} \mathbb{E}\left[d_{\mathrm{TV}}^{\mathsf{x}_0}(U_i^{\tau_i}, t_1 - \tau_i)^2 \mathbb{1}_{\{\tau_i < t_1\}}\right] \le C't_1^3 e^{4(\log n)^{1-\frac{\alpha}{2}}} n \left(\mathcal{A}_{t_1}^{(2)}(\mathsf{x}_0) \vee e^{-4\theta t_1}\right) + O(n^{-9}).$$

Combined with Eq. (4.10), this completes the proof. □

4.4. **Proof of upper bound in Theorem 1.1.** Combining Lemmas 4.2, 4.8 and 4.11, we get

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(t_2)^2 \le Ct_1^3 e^{4(\log n)^{1-\frac{\alpha}{2}}} n \left(\mathcal{A}_{t_1}^{(2)}(\mathsf{x}_0) \vee e^{-4\theta t_1}\right) + O\left(e^{-8(\log n)^{1-\frac{\alpha}{2}}}\right).$$

By definition of t_1 in Eq. (4.2) and T_{x_0} in Eq. (1.4), using Eq. (2.5) from Proposition 2.1 yields

$$n\left(\mathcal{A}_{t_1}^{(2)}(\mathsf{x}_0) \vee e^{-4\theta t_1}\right) \le e^{-2\theta b_1 (\log n)^{1-\frac{\alpha}{2}}}.$$

Hence, the choice $b_1 = (3/\theta)(\log n)^{1-\frac{\alpha}{2}}$ shows that

$$d_{\mathrm{TV}}^{\mathsf{x}_0}(t_2)^2 \le O\left(t_1^3 e^{-2(\log n)^{1-\frac{\alpha}{2}}}\right) < Ce^{-(\log n)^{1-\frac{\alpha}{2}}},$$

for some other $C > 0$ depending only on $\theta, q, \alpha, \mathfrak{c}_0$. This establishes Eq. (4.4) and concludes the proof of the upper bound. □

Remark 4.14. The proof holds verbatim if we replace x_0 with the uniform initial condition $\mathcal{U}$. However, as $\mathcal{A}_t^{(2)}(\mathcal{U}) = 0$, the T_{x_0} part can be dropped, thus the upper bound in Eq. (1.10) follows.

5. LOWER BOUNDS

5.1. **Autocorrelation lower bound.** In this section, we show the $t_{\mathrm{MIX}}^{\mathsf{x}_0} \gtrsim T_{\mathsf{x}_0}$ part of Eq. (1.5).

Let G be a graph satisfying Eq. (1.1) for some $\alpha, \mathfrak{c}_0$, and fix a configuration x_0. We will show a stronger bound: there exists some $C^\dagger = C^\dagger(\theta, q, \alpha, \mathfrak{c}_0) > 0$ such that, for every $\varepsilon > 0$, if

$$t^\dagger = T_{\mathsf{x}_0} - \frac{1}{2\theta}\log(1/\varepsilon) - C^\dagger$$

(5.1)

(where T_{x_0} is as in Eq. (1.4)), then

$$\|\mathbb{P}_{\mathsf{x}_0}(X_{t^\dagger} \in \cdot) - \mu_G\|_{\mathrm{TV}} > 1 - \varepsilon,$$

(5.2)

and in particular, $t_{\mathrm{MIX}}^{\mathsf{x}_0}(1 - \varepsilon) > t^\dagger$.

To establish this, define the statistics

$$\mathcal{R}_t^{\mathsf{x}_0}(\mathsf{x}) = \sum_{v \in V} \pi_G(v) \sum_{\omega \in \mathfrak{C}_q} a_{v,\omega,t} \left(\mathbb{1}_{\{\mathsf{x}(v)=\omega\}} - \frac{1}{q} \right), \tag{5.3}$$

where

$$a_{v,\omega,t} = \mathbb{P}_{\mathsf{x}_0}(X_t(v) = \omega) - \frac{1}{q} \qquad \text{for all } v \in V \text{ and } \omega \in \mathfrak{C}_q.$$

Notice that $\mathbb{E}_{\mu_G}[\mathcal{R}_t^{\mathsf{x}_0}(Y)] = 0$ (since $\mathbb{P}_{\mu_G}(Y(v) = \omega) = 1/q$ for all v, ω), whereas $\mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_t^{\mathsf{x}_0}(X_t)]$ is nothing but the autocorrelation:

$$\mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_t^{\mathsf{x}_0}(X_t)] = \sum_{v \in V} \pi_G(v) \sum_{\omega \in \mathfrak{C}_q} a_{v,\omega,t}^2 = \sum_{v \in V} \pi_G(v) \sum_{\omega \in \mathfrak{C}_q} \left(\mathbb{P}_{\mathsf{x}_0}(X_t(v) = \omega) - \frac{1}{q} \right)^2 = \mathcal{A}_t^{(2)}(\mathsf{x}_0)$$

as given in Eq. (1.3).

We next need a bound on the variance of $\mathcal{R}_t^{\mathsf{x}_0}(X_t)$:

Claim 5.1. *There exists $C_0 = C_0(\theta, q, \alpha, \mathfrak{c}_0)$ such that*

$$\mathrm{Var}_{\mathsf{x}_0}(\mathcal{R}_t^{\mathsf{x}_0}(X_t)) \le \frac{C_0}{n} \mathcal{A}_t^{(2)}(\mathsf{x}_0), \quad \mathrm{Var}_{\mu_G}(\mathcal{R}_t^{\mathsf{x}_0}(Y)) \le \frac{C_0}{n} \mathcal{A}_t^{(2)}(\mathsf{x}_0).$$

Proof. It will be convenient to work with the following formula for $\mathcal{R}_t^{\mathsf{x}_0}(X_t)$, equivalent to Eq. (5.3):

$$\mathcal{R}_t^{\mathsf{x}_0}(\mathsf{x}) = \sum_{v \in V} \pi_G(v) \sum_{\omega \in \mathfrak{C}_q} a_{v,\omega,t} \mathbb{1}_{\{\mathsf{x}(v)=\omega\}};$$

indeed, the equivalence is due to the fact that $\sum_\omega a_{v,\omega,t} = 0$ for every v and t. We thus have

$$\mathrm{Var}_{\mathsf{x}_0}(\mathcal{R}_t^{\mathsf{x}_0}(X_t)) = \sum_{u,v \in V} \sum_{\omega_1,\omega_2 \in \mathfrak{C}_q} \pi_G(u)\pi_G(v) a_{u,\omega_1,t} a_{v,\omega_2,t} \, \mathrm{Cov}_{\mathsf{x}_0}(\mathbb{1}_{\{X_t(u)=\omega_1\}}, \mathbb{1}_{\{X_t(v)=\omega_2\}})$$

$$\le \sum_{u,v} \sum_{\omega_1,\omega_2} \pi_G(u)\pi_G(v) \left(\frac{a_{u,\omega_1,t}^2 + a_{v,\omega_2,t}^2}{2} \right) \left| \mathrm{Cov}_{\mathsf{x}_0}(\mathbb{1}_{\{X_t(u)=\omega_1\}}, \mathbb{1}_{\{X_t(v)=\omega_2\}}) \right|,$$

which, by the symmetry between (u, ω_1) and (v, ω_2), is in turn equal to

$$\sum_{u,\omega_1} \pi_G(u) a_{u,\omega_1,t}^2 \sum_{v,\omega_2} \pi_G(v) \left| \mathrm{Cov}_{\mathsf{x}_0}(\mathbb{1}_{\{X_t(u)=\omega_1\}}, \mathbb{1}_{\{X_t(v)=\omega_2\}}) \right|$$

$$\le \frac{\Delta q}{n} \sum_{u,\omega_1} \pi_G(v) a_{v,\omega,t}^2 \max_{\omega_2} \sum_v \left| \mathrm{Cov}_{\mathsf{x}_0}(\mathbb{1}_{\{X_t(u)=\omega_1\}}, \mathbb{1}_{\{X_t(v)=\omega_2\}}) \right|.$$

(In the last inequality we used $\pi_G(v) \le \Delta/n$ where $\Delta \le \mathfrak{c}_0 e$ is the maximum degree of G.)

We next argue that the last sum over v is bounded by some $C(\theta, \mathfrak{c}_0, \alpha)$ for all u, ω_1, ω_2. To see this, let X_t' be a an independent copy of X_t, whence

$$\left| \mathrm{Cov}_{\mathsf{x}_0}(\mathbb{1}_{\{X_t(u)=\omega_1\}}, \mathbb{1}_{\{X_t(v)=\omega_2\}}) \right| = \left| \mathbb{P}(X_t(u) = \omega_1, X_t(v) = \omega_2) - \mathbb{P}(X_t(u) = \omega_1, X_t'(v) = \omega_2) \right|,$$

and let τ be the minimum time $s \ge 0$ in the backward dynamics where the walks (Z_s^u) and (Z_s^v) coalesce. We can couple $(X_t(u), X_t(v))$ with $(X_t(u), X_t'(v))$ in such a way that the two pairs are identical if $\tau > t$ (the walks do not meet until until the backward dynamics reaches time 0). Under this coupling, the expression for $\left| \mathrm{Cov}_{\mathsf{x}_0}(\mathbb{1}_{\{X_t(u)=\omega_1\}}, \mathbb{1}_{\{X_t(v)=\omega_2\}}) \right|$ simplifies to

$$\left| \mathbb{P}(X_t(u) = \omega_1, X_t(v) = \omega_2, \tau \le t) - \mathbb{P}(X_t(u) = \omega_1, X_t'(v) = \omega_2, \tau \le t) \right| \le \mathbb{P}(\tau \le t).$$

Note that $\mathbb{P}(\tau \le t) < \mathbb{P}(\tau < \infty) = \mathbb{P}(u \leftrightsquigarrow v)$ (the probability that the walks started at u and v meet before any of them dies). If $\mathrm{dist}_G(u, v) = r$, then in order for the walks at u and v to possibly

meet, the first r updates that either of them receive must be non-killing (with probability $1 - \theta$ each), and so $\mathbb{P}(u \leftrightsquigarrow v) \leq (1 - \theta)^r \leq e^{-\theta r}$. Hence, using Eq. (1.1),

$$\max_{\omega_2} \sum_v \left| \mathrm{Cov}_{\mathsf{x}_0}(\mathbb{1}_{\{X_t(u)=\omega_1\}}, \mathbb{1}_{\{X_t(v)=\omega_2\}}) \right| \leq \sum_{r \geq 0} |\{v \,:\, \mathrm{dist}_G(u,v) = r\}| e^{-\theta r}$$

$$\leq \mathfrak{c}_0 \sum_{r \geq 0} e^{r^{1-\alpha} - \theta r} =: C_1 \,.$$

In conclusion,

$$\mathrm{Var}_{\mathsf{x}_0}(\mathcal{R}_t^{\mathsf{x}_0}(X_t)) < \frac{\Delta C_1 q}{n} \sum_v \pi_G(v) a_{v,\omega,t}^2 = \frac{\Delta C_1 q}{n} \mathcal{A}_t^{(2)}(\mathsf{x}_0) \,.$$

The same proof works for $\mathrm{Var}_{\mu_G}(\mathcal{R}_t^{\mathsf{x}_0}(Y))$, noting directly that under the same coupling as above

$$\left| \mathrm{Cov}_{\mu_G}(\mathbb{1}_{\{Y(u)=\omega_1\}}, \mathbb{1}_{\{Y(v)=\omega_2\}}) \right| \leq \mathbb{P}(\tau < \infty) = \mathbb{P}(u \leftrightsquigarrow v) \leq \exp[-\theta \, \mathrm{dist}_G(u,v)] \,,$$

the same upper bound we used for the covariance when estimating $\mathrm{Var}_{\mathsf{x}_0}(\mathcal{R}_t^{\mathsf{x}_0}(X_t))$.   □

We are ready to prove Eq. (5.2). Recall that $\mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_t^{\mathsf{x}_0}(X_t)] = \mathcal{A}_t^{(2)}(\mathsf{x}_0) > 0 = \mathbb{E}_{\mu_G}[\mathcal{R}_t^{\mathsf{x}_0}(Y)]$, and define the set of configurations

$$E^\dagger = \left\{ \mathsf{x} \in \mathfrak{C}_q^V \,:\, \mathcal{R}_{t^\dagger}^{\mathsf{x}_0}(\mathsf{x}) < \frac{1}{2} \mathcal{A}_{t^\dagger}^{(2)}(\mathsf{x}_0) \right\} \,.$$

By Chebyshev's inequality,

$$\mathbb{P}_{\mathsf{x}_0}(X_{t^\dagger} \in E^\dagger) = \mathbb{P}_{\mathsf{x}_0}\left(\mathcal{R}_{t^\dagger}^{\mathsf{x}_0}(X_{t^\dagger}) - \mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_{t^\dagger}^{\mathsf{x}_0}(X_{t^\dagger})] < -\frac{\mathcal{A}_{t^\dagger}^{(2)}(\mathsf{x}_0)}{2} \right) \leq \frac{4 \, \mathrm{Var}_{\mathsf{x}_0}(\mathcal{R}_{t^\dagger}^{\mathsf{x}_0}(X_{t^\dagger}))}{\mathcal{A}_{t^\dagger}^{(2)}(\mathsf{x}_0)^2} \leq \frac{4 C_0}{n \mathcal{A}_{t^\dagger}^{(2)}(\mathsf{x}_0)}$$

using Claim 5.1 (with C_0 from that claim). Recalling from Eq. (5.1) that $t^\dagger = T_{\mathsf{x}_0} - \frac{1}{2\theta} \log(1/\varepsilon) - C^\dagger$, we see that if $C^\dagger = \frac{1}{2\theta} \log(8C_0)$ then the sub-multiplicativity of $\mathcal{A}^{(2)}$, as per Eq. (2.5), yields

$$\mathbb{P}_{\mathsf{x}_0}(X_{t^\dagger} \in E^\dagger) \leq \frac{4 C_0}{n \mathcal{A}_{T_{\mathsf{x}_0}}^{(2)}(\mathsf{x}_0) e^{2\theta(T_{\mathsf{x}_0} - t^\dagger)}} = \frac{\varepsilon}{2} \,,$$

with the last equality following from the definition of $\mathcal{A}^{(2)}$ in Eq. (1.3). Similarly,

$$\mathbb{P}_{\mu_G}(Y \notin E^\dagger) = \mathbb{P}_{\mu_G}\left(\mathcal{R}_{t^\dagger}^{\mathsf{x}_0}(Y) - \mathbb{E}_{\mu_G}[\mathcal{R}_{t^\dagger}^{\mathsf{x}_0}(Y)] > \frac{\mathcal{A}_{t^\dagger}^{(2)}(\mathsf{x}_0)}{2} \right) \leq \frac{4 \, \mathrm{Var}_{\mu_G}(\mathcal{R}_{t^\dagger}^{\mathsf{x}_0}(Y))}{\mathcal{A}_{t^\dagger}^{(2)}(\mathsf{x}_0)^2} \leq \frac{\varepsilon}{2} \,.$$

Combining the last two displays gives

$$\|\mathbb{P}_{\mathsf{x}_0}(X_{t^\dagger} \in \cdot) - \mu_G\|_{\mathrm{TV}} \geq \mathbb{P}_{\mu_G}(Y \in E^\dagger) - \mathbb{P}_{\mathsf{x}_0}(X_{t^\dagger} \in E^\dagger) \geq 1 - \varepsilon \,,$$

thereby establishing Eq. (5.2).   □

5.2. Correlation lower bound.

In this section, we show the $t_{\mathrm{MIX}}^{\mathsf{x}_0} \gtrsim \frac{1}{4\theta} \log n$ part of Eq. (1.5) and Eq. (1.10). Because the proofs of the two are essentially the same, we simply prove Eq. (1.5) and describe in Remark 5.3 the single modification needed in the case of a uniform initial state $\mathcal{U}$.

Let $G = (V, E)$ be a connected graph with $|V| = n$ satisfying Eq. (1.1) for some constants $\alpha, \mathfrak{c}_0$. We will show that there exists a constant $C^\ddagger > 0$ depending only on $(\theta, \alpha, \mathfrak{c}_0)$ such that if

$$t^\ddagger = \frac{1}{4\theta} \log n - C^\ddagger (\log n)^{1 - \frac{\alpha}{2}} \,, \tag{5.4}$$

then, for all initial states x_0,

$$\|\mathbb{P}_{\mathsf{x}_0}(X_t \in \cdot) - \mu_G\|_{\mathrm{TV}} \geq 1 - O(e^{-2(\log n)^{1 - \frac{\alpha}{2}}}) \,, \tag{5.5}$$

implying in particular that, for every $0 < \varepsilon < 1$ fixed, $t_{\mathrm{MIX}}^{\mathsf{x}_0}(1 - \varepsilon) > t^{\ddagger}$ provided n is large enough.

For fixed x_0, t, we consider the following distinguishing statistics:

$$\mathcal{R}_t^{\mathsf{x}_0}(\mathsf{x}) = \sum_{u \sim v} \left(\mathbb{1}_{\{\mathsf{x}(u) = \mathsf{x}(v)\}} - \mathbb{P}_{\mathsf{x}_0}\big(X_t(v) = \mathsf{x}(u)\big) \right) . \tag{5.6}$$

(Notice $\mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_t(X_t)]$ does not simplify to 0 but to $\sum_{u \sim v}[\mathbb{P}_{\mathsf{x}_0}(X_t(u) = X_t(v)) - \mathbb{P}_{\mathsf{x}_0}(X_t(u) = X_t'(v))]$ for an i.i.d. instance of the chain X_t'.) By the routine fact $\sum_{k=0}^{q-1} \omega^k = q \mathbb{1}_{\{\omega=1\}}$ for all $\omega \in \mathfrak{C}_q$, we can equivalently write $\mathcal{R}_t^{\mathsf{x}_0}(\mathsf{x})$ as

$$\mathcal{R}_t^{\mathsf{x}_0}(\mathsf{x}) = \frac{1}{q} \sum_{k=1}^{q-1} \sum_{u \sim v} \left(\mathsf{x}(u)^k \overline{\mathsf{x}(v)^k} - \mathsf{x}(u)^k \overline{\mathbb{E}_{\mathsf{x}_0}[X_t(v)^k]} \right) \tag{5.7}$$

(NB. we omitted the $k = 0$ term, as it has no contribution). Then, as $\mathbb{E}_{\mu_G}[Y(v)^k] = \frac{1}{q} \sum_{\omega \in \mathfrak{C}_q} \omega^k = 0$ for all $k = 1, \ldots, q - 1$, we have

$$\mathbb{E}_{\mu_G}[\mathcal{R}_t^{\mathsf{x}_0}(Y)] - \mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_t^{\mathsf{x}_0}(X_t)] = \frac{1}{q} \sum_{k=1}^{q-1} \sum_{u \sim v} \left(\mathrm{Cov}_{\mu_G}(Y(u)^k, Y(v)^k) - \mathrm{Cov}_{\mathsf{x}_0}(X_t(u)^k, X_t(v)^k) \right) . \tag{5.8}$$

While the left-hand is clearly real-valued (e.g., via Eq. (5.7)), a-priori the individual terms we sum over in Eq. (5.8) might not be. The next claim shows that they are nonnegative real numbers bounded from below by the probability that two random walks coalesce after time t.

Claim 5.2. *For every* $\mathsf{x}_0 \in \mathfrak{C}_q^V$, *every two vertices* $u, v \in V$ *and all* $k = 1, \ldots, q - 1$,

$$\mathrm{Cov}_{\mu_G}(Y(u)^k, Y(v)^k) - \mathrm{Cov}_{\mathsf{x}_0}(X_t(u)^k, X_t(v)^k) - \mathbb{P}\big(u \overset{>t}{\leftrightsquigarrow} v\big) \in \mathbb{R}_+ , \tag{5.9}$$

where $\big\{ u \overset{>t}{\leftrightsquigarrow} v \big\}$ *denotes the probability that two random walks* $(Z_s^u), (Z_s^v)$ *started at* u *and* v *with killing rate* θ *(and moving rate* $1 - \theta$*) do not coalesce until time* t, *and then coalesce after time* t.

Proof. Let τ be the first time that the walks started from u and v coalesce in the backward dynamics (with $\tau = +\infty$ if either walk gets killed before coalescing). If $X_t'(v)$ denotes a variable that has the same distribution as $X_t(v)$ but is independent of $X_t(u)$, we can couple $(X_t(u), X_t(v))$ with $(X_t(u), X_t'(v))$ by letting the two corresponding random walks $Z_s^v, Z_s'^v$ in the backward dynamics started at v be identical for $s \le \tau$, and then run independently for $s > \tau$. When doing so, one has $(X_t(u), X_t(v)) = (X_t(u), X_t'(v))$ when $\tau > t$, and thus

$$\begin{aligned}
\mathrm{Cov}_{\mathsf{x}_0}(X_t(u)^k, X_t(v)^k) &= \mathbb{E}_{\mathsf{x}_0}\left[X_t(u)^k \overline{X_t(v)^k} - X_t(u)^k \overline{X_t'(v)^k} \right] \\
&= \mathbb{E}_{\mathsf{x}_0}\left[\left(X_t(u)^k \overline{X_t(v)^k} - X_t(u)^k \overline{X_t'(v)^k} \right) \mathbb{1}_{\{\tau \le t\}} \right] \\
&= \mathbb{P}(\tau \le t) - \mathbb{E}_{\mathsf{x}_0}\left[X_t(u)^k \overline{X_t'(v)^k} \mathbb{1}_{\{\tau \le t\}} \right] ,
\end{aligned}$$

where the final identity follows from the fact that $X_t(u) = X_t(v)$ on $\mathbb{1}_{\{\tau \le t\}}$.

On the other hand, recall that μ_G can be perfectly simulated by letting the walks from V run forever in the past until all of them are killed, and noting that if two walks do not coalesce (because one gets killed first) they independently get a uniform spin (with 0 mean, again by $\frac{1}{q} \sum_{\omega \in \mathfrak{C}_q} \omega = 0$). Thus,

$$\mathrm{Cov}_{\mu_G}(Y(u)^k, Y(v)^k) = \mathbb{E}_{\mu_G}\left[Y(u)^k \overline{Y(v)}^k \right] = \mathbb{P}(\tau < \infty) .$$

By definition, $\mathbb{P}(t < \tau < \infty) = \mathbb{P}\big(u \overset{>t}{\leftrightsquigarrow} v\big)$, so it follows that

$$\mathrm{Cov}_{\mu_G}(Y(u)^k, Y(v)^k) - \mathrm{Cov}_{\mathsf{x}_0}(X_t(u)^k, X_t(v)^k) = \mathbb{P}\big(u \overset{>t}{\leftrightsquigarrow} v\big) + \mathbb{E}_{\mathsf{x}_0}[X_t(u)^k \overline{X_t'(v)^k} \mathbb{1}_{\{\tau \le t\}}] .$$

The proof will be finished if we prove that $\mathbb{E}_{\mathsf{x}_0}[X_t(u)^k \overline{X_t'(v)^k} \mathbb{1}_{\{\tau \leq t\}}]$ is real-valued and nonnegative. Indeed, if $\mathcal{F}_\tau$ denotes the σ-algebra associated with the stopping time τ for the backward dynamics, we have

$$\mathbb{E}_{\mathsf{x}_0}\left[X_t(u)^k \overline{X_t'(v)^k} \mathbb{1}_{\{\tau \leq t\}}\right] = \mathbb{E}_{\mathsf{x}_0}\left[\mathbb{E}\left[X_t(u)^k \overline{X_t'(v)^k} \mid \mathcal{F}_\tau\right] \mathbb{1}_{\{\tau \leq t\}}\right]$$
$$= \mathbb{E}_{\mathsf{x}_0}\left[\mathbb{E}\left[X_{t-\tau}(w_\tau)^k \overline{X_{t-\tau}'(w_\tau)^k} \mid \mathcal{F}_\tau\right] \mathbb{1}_{\{\tau \leq t\}}\right],$$

where w_τ denotes the vertex at which the walks from u and v meet at time τ (note that w_τ is $\mathcal{F}_\tau$-measurable). Finally, since $X_{t-\tau}(w_\tau)$ and $X_{t-\tau}'(w_\tau)$ are conditionally independent given $\mathcal{F}_\tau$,

$$\mathbb{E}_{\mathsf{x}_0}\left[\mathbb{E}\left[X_{t-\tau}(w_\tau)^k \overline{X_{t-\tau}'(w_\tau)^k} \mid \mathcal{F}_\tau\right] \mathbb{1}_{\{\tau \leq t\}}\right] = \mathbb{E}_{\mathsf{x}_0}\left[\left|\mathbb{E}[X_{t-\tau}(w_\tau)^k \mid \mathcal{F}_\tau]\right|^2 \mathbb{1}_{\{\tau \leq t\}}\right] \in \mathbb{R}_+,$$

as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Remark 5.3. To handle the uniform initial condition $\mathcal{U}$ rather than a deterministic x_0, redefine the statistics $\mathcal{R}_t^{\mathsf{x}_0}(\mathsf{x})$ in Eq. (5.6) simply as

$$\mathcal{R}(\mathsf{x}) = \sum_{u \sim v} \mathbb{1}_{\{\mathsf{x}(u) = \mathsf{x}(v)\}}.$$

Then, Eq. (5.8) holds true for the new statistics when we replace x_0 by $\mathcal{U}$. Moreover, in that case we have, in lieu of Claim 5.2, that

$$\mathrm{Cov}_{\mu_G}(Y(u)^k, Y(v)^k) - \mathrm{Cov}_{\mathcal{U}}(X_t(u)^k, X_t(v)^k) = \mathbb{P}\left(u \overset{>t}{\leftrightsquigarrow} v\right).$$

The remaining part of the proof in this section then holds verbatim once we replace x_0 by $\mathcal{U}$.

Applying Claim 5.2 to Eq. (5.8) shows that, for all $q \geq 2$ (replacing $\frac{q-1}{q}$ by $\frac{1}{2}$ in the lower bound),

$$\mathbb{E}_{\mu_G}[\mathcal{R}_t^{\mathsf{x}_0}(Y)] - \mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_t^{\mathsf{x}_0}(X_t)] \geq \frac{1}{2} \sum_{u \sim v} \mathbb{P}\left(u \overset{>t}{\leftrightsquigarrow} v\right). \tag{5.10}$$

In light of Eq. (5.10), we aim to show that $\mathbb{P}\left(u \overset{>t}{\leftrightsquigarrow} v\right) \approx e^{-2\theta t}$ for every $uv \in E$ (Claim 5.5 below), so that one has $\sum_{u \sim v} \mathbb{P}\left(u \overset{>t}{\leftrightsquigarrow} v\right) \approx n e^{-2\theta t}$, whereas $\mathrm{Var}_{\mathsf{x}_0}(\mathcal{R}_t^{\mathsf{x}_0}(X_t))$ and $\mathrm{Var}_{\mu_G}(\mathcal{R}_t^{\mathsf{x}_0}(Y))$ are $O(n)$ (see Claim 5.1 below); at that point, the proof will be concluded by Chebyshev's inequality.

Note that in order for the walks started from u and v to coalesce after t, both of them need to survive separately until time t, which incurs in a cost of $e^{-2\theta t}$ for survivability alone. Our next goal is to show that under the assumption of Eq. (1.1) and connectedness, for nearby vertices u, v the cost for the remaining conditions (the two walks avoiding each other until time t and coalescing later) is sub-exponential in $t \asymp \log n$. This, together with Claim 5.2, will yield the required lower bound. The key for this will be a result from [20], which bounds from below the probability of a random walk failing to escape a set S through the conductance of S.

Before applying said result in our setting, we prove the existence of "good" balls with limited expansion around every $v \in V$. For a set of vertices S, the conductance $\Phi(S)$ of S is the ratio

$$\Phi(S) = \frac{|E(S, S^c)|}{\mathrm{Vol}(S)},$$

where $|E(S, S^c)|$ is the number of edges connecting S to its complement S^c, and $\mathrm{Vol}(S)$ is the sum of the degrees of vertices in S. The next lemma shows that, for every graph $G = (V, E)$ of subexponential growth of balls as in Eq. (1.1) and every vertex $v \in V$, we can find a ball $B_v(r)$ in every scale with conductance at most $O(r^{-\alpha})$.

Lemma 5.4. *Let G be a graph satisfying Eq. (1.1) for some $\mathfrak{c}_0 > 0$ and $\alpha \in (0,1)$. Then, for every integer $\mathbf{r}_n \geq \mathfrak{c}_0$ and every $v \in V$, there exists $r \in [\mathbf{r}_n, 2\mathbf{r}_n]$ such that*

$$\frac{|\partial^+ B_v(r)|}{|B_v(r)|} \leq \frac{8}{r^\alpha},$$

where $B_v(r)$ is the ball of radius r in G centered at the vertex v, and $\partial^+ B_v(r) = B_v(r+1) \setminus B_v(r)$. In particular, its conductance has $\Phi(B_v(r)) \leq 8\Delta r^{-\alpha}$, where $\Delta \leq \mathfrak{c}_0 e$ is the maximum degree in G.

Proof. Denote $\widehat{\mathbf{r}}_n = 2\mathbf{r}_n$, and suppose by contradiction that the claim is false, so that there exist some $v \in V$ and $r \geq \mathfrak{c}_0$ such that

$$\frac{|\partial^+ B_v(r)|}{|B_v(r)|} > \frac{8}{r^\alpha}, \qquad \text{for all } r \in [\mathbf{r}_n, \widehat{\mathbf{r}}_n].$$

We can then write

$$|B_v(\widehat{\mathbf{r}}_n)| = |B_v(\widehat{\mathbf{r}}_n - 1)| \left(1 + \frac{|\partial^+ B_v(\widehat{\mathbf{r}}_n - 1)|}{|B_v(\widehat{\mathbf{r}}_n - 1)|}\right) > |B_v(\widehat{\mathbf{r}}_n - 1)| \left(1 + \frac{8}{(\widehat{\mathbf{r}}_n - 1)^\alpha}\right),$$

and iterating this argument shows that

$$|B_v(\widehat{\mathbf{r}}_n)| > |B_v(\mathbf{r}_n)| \prod_{r=\mathbf{r}_n}^{\widehat{\mathbf{r}}_n - 1} \left(1 + \frac{8}{r^\alpha}\right) \geq \mathbf{r}_n \exp\left[\sum_{r=\mathbf{r}_n}^{\widehat{\mathbf{r}}_n - 1} \log\left(1 + \frac{8}{r^\alpha}\right)\right].$$

Noting that $\log(1 + 8x) \geq 2x$ for all $0 \leq x \leq 1$ and that $\frac{1 - 2^{-(1-\alpha)}}{1-\alpha} \geq \frac{1}{2}$ for all $0 \leq \alpha \leq 1$, we find

$$e^{\sum_{r=\mathbf{r}_n}^{\widehat{\mathbf{r}}_n - 1} \log\left(1 + 8r^{-\alpha}\right)} \geq e^{2 \sum_{r=\mathbf{r}_n}^{\widehat{\mathbf{r}}_n - 1} r^{-\alpha}} \geq e^{2 \int_{\mathbf{r}_n}^{\widehat{\mathbf{r}}_n} r^{-\alpha} dr} = e^{2 \frac{1 - 2^{-(1-\alpha)}}{1-\alpha} (\widehat{\mathbf{r}}_n)^{1-\alpha}} \geq e^{(\widehat{\mathbf{r}}_n)^{1-\alpha}}.$$

Having arrived at $|B_v(\widehat{\mathbf{r}}_n)| > \mathbf{r}_n \exp[(\widehat{\mathbf{r}}_n)^{1-\alpha}]$ with $\mathbf{r}_n \geq \mathfrak{c}_0$, this contradicts Eq. (1.1) as desired. $\square$

For a set of vertices $S \subset V$, an integer k and a distribution ν on S, define

$$\rho_\nu^{\mathsf{lazy}}(S, k) = \mathbb{P}_\nu(\text{lazy random walk remains in } S \text{ for } k \text{ steps}).$$

(Here and in what follows, a lazy random walk refers to one that stays in place with probability $\frac{1}{2}$.) Oveis Gharan and Trevisan [20, Prop. 8] proved that on for any graph G, if π_S is the restriction of the stationary distribution π_G to S (i.e., $\pi_S(v) = d_G(v)/\operatorname{Vol}(S)$ for $v \in S$), then

$$\rho_{\pi_S}^{\mathsf{lazy}}(S, k) \geq \left(1 - \frac{\Phi(S)}{2}\right)^k.$$

This extends to a continuous-time random walk with rate λ: by a time-rescaling, said walk is equivalent to a continuous-time lazy random walk with update rate 2λ. There, the number of updates by time t is $N_t \sim \operatorname{Poisson}(2\lambda t)$, and given $N_t = k$ one can apply the bound on $\rho_{\pi_S}^{\mathsf{lazy}}(S, k)$. It follows that the probability $\rho_{\pi_S}^{\mathsf{cont}}(S, t, \lambda)$ that a continuous-time random walk with rate λ started from $v \in \pi_S$ stays completely inside S until time t satisfies

$$\rho_{\pi_S}^{\mathsf{cont}}(S, t, \lambda) \geq \mathbb{E}\left[(1 - \Phi(S)/2)^{N_t}\right] = e^{-\Phi(S)\lambda t}. \tag{5.11}$$

We will use this to prove the following:

Claim 5.5. *For every $t = t(n) \geq 0$, any edge $uv \in E$ and any integer $\mathbf{r}_n$ with $\mathfrak{c}_0 \leq \mathbf{r}_n \leq \frac{1}{10}\operatorname{diam}(G)$,*

$$\mathbb{P}\left(u \overset{>t}{\leftrightsquigarrow} v\right) \geq \exp\left[-2\theta t - C_1 \mathbf{r}_n - C_2 \mathbf{r}_n^{-\alpha} t\right],$$

where $C_1, C_2 > 0$ are constants depending only on $\theta, \alpha, \mathfrak{c}_0$.

Proof. The strategy will be the following: let the walks Z_s^u and Z_s^v from u and v, respectively, move quickly to two suitable locations, confine them to separate balls until time t using Eq. (5.11) for each of them and then let them quickly coalesce after time t.

Applying Lemma 5.4 for u and $\mathbf{r}_n$ yields $r \in [\mathbf{r}_n, 2\mathbf{r}_n]$ such that $S := B_u(r)$ has $\Phi(S) \le 8\Delta r^{-\alpha}$. Let δ_x denote the point mass distribution at x, and consider $\rho_{\delta_x}^{\mathsf{cont}}(S, t, 1 - \theta)$, the probability that the continuous-time random walk with moving rate $1 - \theta$, started at $x \in S$, remains in S for time t. From Eq. (5.11), we see that the vertex u_* maximizing $\rho_{\delta_x}^{\mathsf{cont}}(S, t, 1 - \theta)$ over $x \in S$ must satisfy

$$\rho_{\delta_{u*}}^{\mathsf{cont}}(S, t, 1 - \theta) \ge \rho_{\pi_S}^{\mathsf{cont}}(S, t, 1 - \theta) \ge e^{-8\Delta r^{-\alpha}(1-\theta)t} . \tag{5.12}$$

By connectedness of G and the fact that $\operatorname{diam}(G) \ge 10\mathbf{r}_n$, there exists a (shortest) path from the edge uv to a vertex at some distance $> 4\mathbf{r}_n$. Assume without loss of generality that such a path starts at v and (being a shortest path) does not visit u, and let w be a vertex along this path with $\operatorname{dist}_G(u, w) = 4\mathbf{r}_n + 1$. Applying Lemma 5.4 for w and $\mathbf{r}_n$ yields some $r' \in [\mathbf{r}_n, 2\mathbf{r}_n]$ such that $S' := B_w(r')$ has $\Phi(S) \le 8\Delta r^{-\alpha}$ (note that $S \cap S' = \emptyset$ thanks to the condition $\operatorname{dist}_G(u, w) = 4\mathbf{r}_n + 1$). As argued above for $B_u(r)$, there must exist $w_* \in S'$ such that

$$\rho_{\delta_{w*}}^{\mathsf{cont}}(S', t, 1 - \theta) \ge e^{-8\Delta(r')^{-\alpha}(1-\theta)t} . \tag{5.13}$$

We can now realize a sub-event of $\{u \overset{>t}{\leftrightsquigarrow} v\}$ as follows:

(i) force the first $\operatorname{dist}_G(v, w) + \operatorname{dist}_G(w, w_*) + \operatorname{dist}_G(u, u_*)$ $(\le (4\mathbf{r}_n + 1) + 2\mathbf{r}_n + 2\mathbf{r}_n = 8\mathbf{r}_n + 1)$ cumulative updates that Z_s^u and Z_s^v receive in the time interval $s \in [0, t]$ to move Z_s^v from v to w to w_* (while avoiding u), and then move Z_s^u from u to u_* (the probability that each update is non-killing and moves the specified random walk along the specified path is at least $(1 - \theta) \cdot \frac{1}{2} \cdot \frac{1}{\Delta}$, where Δ is the maximum degree of G);

(ii) for the remaining time until t (if not surpassed already in the previous step), force the two walks now at u_* and w_* to stay inside the disjoint $S = B_u(r)$ and $S' = B_w(r')$ and receive no killing updates (the non-escape probabilities are bounded from below by Eqs. (5.12) and (5.13), while the non-killing probability is bounded from below by $e^{-2\theta t}$);

(iii) force the next cumulative updates that Z_s^u and Z_s^v receive (after t) to make the two walks coalesce by going through the minimal path between their current locations (the distance between them is at most $8\mathbf{r}_n + 1$, and the probability that each sequential update is non-killing and makes the correct move along the minimal path is at least $(1 - \theta) \cdot \frac{1}{\Delta}$).

Hence,

$$\mathbb{P}\big(u \overset{>t}{\leftrightsquigarrow} v\big) \ge \left(\frac{1 - \theta}{2\Delta}\right)^{8\mathbf{r}_n + 1} e^{-2\theta t} e^{-8\Delta[r^{-\alpha} + (r')^{-\alpha}](1-\theta)t} \left(\frac{1 - \theta}{\Delta}\right)^{8\mathbf{r}_n + 1}$$

$$\ge e^{-2\theta t} \left(\frac{1 - \theta}{2\Delta}\right)^{20\mathbf{r}_n} e^{-16\Delta \mathbf{r}_n^{-\alpha}(1-\theta)t} ,$$

giving the desired bound for $C_1 = 20 \log\left(\frac{2\Delta}{1-\theta}\right)$ and $C_2 = 16\Delta(1 - \theta)$. □

Using Claim 5.5 with

$$\mathbf{r}_n = (\log n)^{\frac{1}{1+\alpha}} < (\log n)^{1 - \frac{\alpha}{2}} \tag{5.14}$$

(noting $\mathbf{r}_n = o(\operatorname{diam}(G))$ as every graph G with maximum degree Δ has $\operatorname{diam}(G) \ge \log_{\Delta-1}(n) - 2$), whereby $\mathbf{r}_n \asymp \mathbf{r}_n^{-\alpha} \log n$ and so both error terms in that claim will have the same order for $t = t^\ddagger$,

it follows from Eq. (5.10) that, for all $t \leq \frac{1}{4\theta} \log n$ and every sufficiently large n,

$$\mathbb{E}_{\mu_G}[\mathcal{R}_t^{\times_0}(Y)] - \mathbb{E}_{\times_0}[\mathcal{R}_t^{\times_0}(X_t)] \geq \frac{ne^{-2\theta t}}{2e^{C(\log n)^{1-\frac{\alpha}{2}}}}, \tag{5.15}$$

for some fixed $C = C(\theta, \alpha, \mathfrak{c}_0) > 0$ (e.g., take $C = C_1 + C_2/(4\theta)$ for C_1, C_2 from that claim).

The final ingredient is an upper bound on the variance of the distinguishing statistics $\mathcal{R}_t^{\times_0}$.

Claim 5.6. *Set $C_0 = 2\Delta^2/\theta$, where Δ is the maximum degree in G. Then for every $t > 0$ and $\times_0$,*

$$\mathrm{Var}_{\times_0}(\mathcal{R}_t^{\times_0}(X_t)) \leq C_0 n, \quad \mathrm{Var}_{\mu_G}(\mathcal{R}_t^{\times_0}(Y)) \leq C_0 n.$$

Proof. Denoting $\mathcal{R}_{u,v}(X) = \mathbb{1}_{\{X(u)=X(v)\}} - \mathbb{E}_{\times_0}[\mathbb{1}_{\{X(u)=X_t(v)\}}]$, we have

$$\mathrm{Var}_{\times_0}(\mathcal{R}_t^{\times_0}(X_t)) = \sum_{u_1 \sim v_1, u_2 \sim v_2} \mathrm{Cov}_{\times_0}(\mathcal{R}_{u_1,v_1}(X_t), \mathcal{R}_{u_2,v_2}(X_t)).$$

Let X_t' be an independent copy of X_t. As in the proof of Claim 5.2, we couple the random walks $(Z_s^{u_2}, Z_s^{v_2})$ with their counterparts $(Z_s'^{u_2}, Z_s'^{v_2})$ so that they are identical until time τ, the first time (if finite) that the backward dynamics has one of the random walks started from $\{u_1, v_1\}$ meet with one of the walks started from $\{u_2, v_2\}$. (Beyond that time, we run each chain independently.) In particular, the 4-tuples $(X_t(u_1), X_t(v_1), X_t(u_2), X_t(v_2))$ and $(X_t(u_1), X_t(v_1), X_t'(u_2), X_t'(v_2))$ are identical if $\tau > t$. Letting $\{x \leftrightsquigarrow y\}$ denote the event that two walks started at x and y with killing rate θ and moving rate $1 - \theta$ meet before either of them dies, we get

$$\begin{aligned}
\mathrm{Cov}_{\times_0}(\mathcal{R}_{u_1,v_1}(X_t), \mathcal{R}_{u_2,v_2}(X_t)) &= \mathbb{E}_{\times_0}\left[\mathcal{R}_{u_1,v_1}(X_t)\mathcal{R}_{u_2,v_2}(X_t) - \mathcal{R}_{u_1,v_1}(X_t)\mathcal{R}_{u_2,v_2}(X_t')\right] \\
&= \mathbb{E}_{\times_0}\left[\left(\mathcal{R}_{u_1,v_1}(X_t)\mathcal{R}_{u_2,v_2}(X_t) - \mathcal{R}_{u_1,v_1}(X_t)\mathcal{R}_{u_2,v_2}(X_t')\right)\mathbb{1}_{\{\tau \leq t\}}\right] \\
&\leq 2\mathbb{P}(\tau \leq t).
\end{aligned}$$

Relaxing $\{\tau \leq t\}$ into $\{\tau < \infty\}$, followed by a union bound, shows

$$\mathbb{P}(\tau \leq t) \leq \mathbb{P}(u_1 \leftrightsquigarrow u_2) + \mathbb{P}(u_1 \leftrightsquigarrow v_2) + \mathbb{P}(v_1 \leftrightsquigarrow u_2) + \mathbb{P}(v_1 \leftrightsquigarrow v_2).$$

and combining the last three displays yields

$$\mathrm{Var}_{\times_0}(\mathcal{R}_t^{\times_0}(X_t)) \leq 2\sum_{u_1,u_2} d_G(u_1)d_G(u_2)\mathbb{P}(u_1 \leftrightsquigarrow u_2),$$

Using Eq. (2.13), we thus find that

$$\mathrm{Var}_{\times_0}(\mathcal{R}_t^{\times_0}(X_t)) \leq \frac{2}{\theta}\sum_v d_G(v)^2 \leq \frac{2\Delta^2}{\theta}n = C_0 n.$$

As this holds for any t, we may take $t \to \infty$ and arrive at

$$\mathrm{Var}_{\mu_G}(\mathcal{R}_t^{\times_0}(Y)) \leq C_0 n,$$

as required. $\qquad\square$

We are ready to prove Eq. (5.5). Define the set of configurations

$$E^{\ddagger} = \left\{\times \in \mathfrak{C}_q^V : \mathcal{R}_{t^{\ddagger}}^{\times_0}(\times) > \left(\mathbb{E}_{\times_0}[\mathcal{R}_{t^{\ddagger}}^{\times_0}(X_{t^{\ddagger}})] + \mathbb{E}_{\mu_G}[\mathcal{R}_{t^{\ddagger}}^{\times_0}(Y)]\right)/2\right\}.$$

Recalling from Eq. (5.4) that $t^{\ddagger} = \frac{1}{4\theta} \log n - C^{\ddagger}(\log n)^{1-\frac{\alpha}{2}}$ and plugging it in Eq. (5.15), we see that

$$\mathbb{E}_{\mu_G}[\mathcal{R}_{t^{\ddagger}}^{\times_0}(Y)] - \mathbb{E}_{\times_0}[\mathcal{R}_{t^{\ddagger}}^{\times_0}(X_{t^{\ddagger}})] \geq \frac{1}{2}\sqrt{n}\exp\left[\left(2\theta C^{\ddagger} - C\right)(\log n)^{1-\frac{\alpha}{2}}\right],$$

for the constant $C(\theta, q, \alpha, \mathfrak{c}_0) > 0$ from Eq. (5.15). Letting $C^{\ddagger} = C/(2\theta) + 1$, this results in the lower bound $(1/C)\sqrt{n}\exp[(\log n)^{1-\frac{\alpha}{2}}]$. We now combine this with Claim 5.6, and deduce from Chebyshev's inequality that

$$\mathbb{P}_{\mathsf{x}_0}(X_{t^{\ddagger}} \in E^{\ddagger}) \leq \frac{4\,\mathrm{Var}_{\mathsf{x}_0}(\mathcal{R}_{t^{\ddagger}}^{\mathsf{x}_0}(X_{t^{\ddagger}}))}{(\mathbb{E}_{\mu_G}[\mathcal{R}_{t^{\ddagger}}^{\mathsf{x}_0}(Y)] - \mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_{t^{\ddagger}}^{\mathsf{x}_0}(X_{t^{\ddagger}})])^2} \leq 16C_0 \exp\left[-2(\log n)^{1-\frac{\alpha}{2}}\right].$$

Similarly,

$$\mathbb{P}_{\mu_G}(Y \notin E^{\ddagger}) \leq \frac{4\,\mathrm{Var}_{\mu_G}(\mathcal{R}_{t^{\ddagger}}^{\mathsf{x}_0}(Y))}{(\mathbb{E}_{\mu_G}[\mathcal{R}_{t^{\ddagger}}^{\mathsf{x}_0}(Y)] - \mathbb{E}_{\mathsf{x}_0}[\mathcal{R}_{t^{\ddagger}}^{\mathsf{x}_0}(X_{t^{\ddagger}})])^2} \leq 16C_0 \exp\left[-2(\log n)^{1-\frac{\alpha}{2}}\right],$$

and it follows that

$$\|\mathbb{P}_{\mathsf{x}_0}(X_t \in \cdot) - \mu_G\|_{\mathrm{TV}} \geq \mathbb{P}_{\mu_G}(Y \in E^{\ddagger}) - \mathbb{P}_{\mathsf{x}_0}(X_t \in E^{\ddagger}) \geq 1 - 32C_0 e^{-2(\log n)^{1-\alpha/2}},$$

thereby establishing Eq. (5.5). $\qquad\square$

6. Concluding remarks and open problems

Our main result, Theorem 1.1, established that the noisy voter model, for every choice of the parameters θ, q, exhibits total variation cutoff from any sequence of initial conditions $\mathsf{x}_0^{(n)}$ provided that the underlying graphs $G^{(n)}$ have subexponential growth of balls as per Eq. (1.1), (Previously, cutoff was only known for $q = 2$ and only from worst-case and a few other periodic initial states.) We characterized the fastest initial state x_0 on $\mathbb{Z}_n^d$ for all d at $q = 2$ and for $d = 1$ at every q, demonstrated that this optimum is nontrivial for large d, q (e.g., at $d = 3, q = 4$, or $d = 2, q = 5$), and showed in Corollary 1.4 that it cannot asymptotically outperform the uniform initial condition $\mathcal{U}$. It would be interesting to extend this to bounded-degree graphs that do not satisfy Eq. (1.1):

Question 6.1. Let $G^{(n)}$ be a sequence of bounded-degree connected graphs on n vertices. Does the noisy voter model on $G^{(n)}$, for all fixed $q \geq 2$ and $\theta \in (0, 1)$, exhibit cutoff from every sequence $\mathsf{x}_0^{(n)}$ of initial states (i.e., is there a function T_n so that $t_{\mathrm{MIX}}^{\mathsf{x}_0}(\varepsilon) = (1 + o(1))T_n(\mathsf{x}_0^{(n)})$ for all fixed $0 < \varepsilon < 1$)? Is the uniform initial state always asymptotically as fast as the deterministic optimum $\min_{\mathsf{x}_0} T_n(\mathsf{x}_0)$?

A concrete class of graphs to study would be a typical random d-regular graph for fixed $d \geq 3$. We remark that we used the condition in Eq. (1.1) in an essential way in both the upper bounds and the lower bounds. In the upper bound, it allowed us to reduce the surviving history of the sites, over a negligible time period of $o(\log n)$, to a collection of well-separated (good) subsets (see Lemma 4.6). This supported the coupling to a product chain in Section 4.1, and further played a role in the reduction to the autocorrelation function in Section 4.3. In the lower bound, the subexponential growth was used to control the variance of the test function in the bound $t_{\mathrm{MIX}}^{\mathsf{x}_0} \gtrsim T_{\mathsf{x}_0}$ (see Claim 5.1), and to identify sets S with a small conductance $\Phi(S)$ so as to bound the escape probability of the random walk from them en route to the bound $t_{\mathrm{MIX}}^{\mathsf{x}_0} \gtrsim \frac{1}{4\theta}\log n$ (see Lemma 5.4).

Another interesting open problem is the cutoff window. The work [6] showed cutoff with an $O(1)$ window from a worst-case starting state on any graph G when $q = 2$ (as did [17] when $G = \mathbb{Z}_n$, the 1D Ising model); i.e., $t_{\mathrm{MIX}}(\varepsilon) - t_{\mathrm{MIX}}(1 - \varepsilon) = O(1)$ with an implicit constant depending on ε. Our new results on a general x_0 only focused on the asymptotics of $t_{\mathrm{MIX}}^{\mathsf{x}_0}$ (as did those of [19]). For example, on $G = \mathbb{Z}_n^d$, our proof arguments will only show that $t_{\mathrm{MIX}}(\varepsilon) - t_{\mathrm{MIX}}(1 - \varepsilon) = O(\log\log n)$.

Question 6.2. Let $G^{(n)}$ be a sequence of connected graphs on n vertices satisfying Eq. (1.1). Does the noisy voter model on $G^{(n)}$, for every fixed $\theta \in (0, 1)$ and $q \geq 2$, and every sequence of initial conditions $\mathsf{x}_0^{(n)}$, satisfy $t_{\mathrm{MIX}}^{\mathsf{x}_0}(\varepsilon) - t_{\mathrm{MIX}}^{\mathsf{x}_0}(1 - \varepsilon) = O(1)$ with an implicit constant depending on ε?

We conclude with an open problem on the noisy voter model on a class of graphs $G = G^{(n)}$ in which $\max_v d_G(v)/\min_v d_G(v) \to \infty$ as $n \to \infty$. (We only considered G with $\max_v d_G(v) = O(1)$, and even the worst-case initial state analysis of [6] required $\max_v d_G(v)/\min_v d_G(v) = O(1)$.) Namely, as we mentioned the random d-regular in the context of Question 6.1, it is natural to ask what the behavior would be on its counterpart, the Erdős–Rényi random graph $\mathcal{G}(n, p)$ for $p = d/n$. Note that with an unbounded degree ratio, Eq. (2.11) no longer gives that $\mathrm{Var}_{\mu_G}(\Psi_l^{(k)}) \asymp 1/n$, and it is then plausible that the correct generalization of Theorem 1.1 would be given in terms of the variant of $\mathcal{A}_t^{(2)}(\mathsf{x}_0)$ where the functions $\Psi_l^{(k)}$ (as per Eq. (2.8)) are normalized in $L^2(\mu_G)$.

Question 6.3. Address the analogue of Question 6.1 for noisy voter model on a typical instance of an Erdős–Rényi random graph $\mathcal{G}(n, p)$ with $p = d/n$, for all fixed $d > 0$, $q \geq 2$ and $\theta \in (0, 1)$.

References

[1] D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.

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

[3] N. Alon. Perturbed identity matrices have high rank: proof and applications. *Combin. Probab. Comput.*, 18(1-2):3–15, 2009.

[4] I. Benjamini, H. Helman Tov, and M. Zhukovskii. Global information from local observations of the noisy voter model on a graph. *Ann. Probab.*, 51(5):1963–1992, 2023.

[5] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60:581–588, 1973.

[6] J. T. Cox, Y. Peres, and J. E. Steif. Cutoff for the noisy voter model. *Ann. Appl. Probab.*, 26(2):917–932, 2016.

[7] P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Nat. Acad. Sci. USA*, 93(4):1659–1664, 1996.

[8] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[9] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman. Broadcasting on trees and the ising model. *Annals of Applied Probability*, pages 410–433, 2000.

[10] B. L. Granovsky and N. Madras. The noisy voter model. *Stochastic Process. Appl.*, 55(1):23–43, 1995.

[11] R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probability*, 3(4):643–663, 1975.

[12] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. 2nd Ed. With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[13] T. M. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes*, volume 324 of *Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin, 1999.

[14] E. Lubetzky and A. Sly. Cutoff for the Ising model on the lattice. *Invent. Math.*, 191(3):719–755, 2013.

[15] E. Lubetzky and A. Sly. Cutoff for general spin systems with arbitrary boundary conditions. *Comm. Pure Appl. Math.*, 67(6):982–1027, 2014.

[16] E. Lubetzky and A. Sly. An exposition to information percolation for the Ising model. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):745–761, 2015.

[17] E. Lubetzky and A. Sly. Information percolation and cutoff for the stochastic Ising model. *J. Amer. Math. Soc.*, 29(3):729–774, 2016.

[18] E. Lubetzky and A. Sly. Universality of cutoff for the Ising model. *Ann. Probab.*, 45(6A):3664–3696, 2017.

[19] E. Lubetzky and A. Sly. Fast initial conditions for Glauber dynamics. *Probab. Theory Related Fields*, 181(1-3):647–667, 2021.

[20] S. Oveis Gharan and L. Trevisan. Approximating the expansion profile and almost optimal local graph clustering. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science—FOCS 2012*, pages 187–196. IEEE Computer Soc., Los Alamitos, CA, 2012.

[21] R. Pymar and N. Rivera. Asymptotic behaviour of the noisy voter model density process. *Ann. Appl. Probab.*, 34(5):4554–4594, 2024.

[22] T. Tao. Cheap Kabatjanskii–Levenstein bound, 2013. https://terrytao.wordpress.com/2013/07/18/a-cheap-version-of-the-kabatjanskii-levenstein-bound-for-almost-orthogonal-vectors/.

P. CADDEO
COURANT INSTITUTE, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012, USA.
  Email address: patrizio.caddeo@courant.nyu.edu

E. LUBETZKY
COURANT INSTITUTE, NEW YORK UNIVERSITY, 251 MERCER STREET, NEW YORK, NY 10012, USA.
  Email address: eyal@courant.nyu.edu