MIT Open Access Articles

Attracting random walks

Attracting random walks

Julia Gaudio*   Yury Polyanskiy†

Abstract

This paper introduces the Attracting Random Walks model, which describes the dynamics of a system of particles on a graph with n vertices. At each step, a single particle moves to an adjacent vertex (or stays at the current one) with probability proportional to the exponent of the number of other particles at a vertex. From an applied standpoint, the model captures the *rich get richer* phenomenon. We show that the Markov chain exhibits a phase transition in mixing time, as the parameter governing the attraction is varied. Namely, mixing time is $O(n \log n)$ when the temperature is sufficiently high and $\exp(\Omega(n))$ when temperature is sufficiently low. When $\mathcal{G}$ is the complete graph, the model is a projection of the Potts model, whose mixing properties and the critical temperature have been known previously. However, for any other graph our model is non-reversible and does not seem to admit a simple Gibbsian description of a stationary distribution. Notably, we demonstrate existence of the dynamic phase transition without decomposing the stationary distribution into phases.

1 Introduction

In this paper, we introduce the Attracting Random Walks (ARW) model. The motivation of the model is to understand the formation of wealth disparities in an economic network. Consider a network of economic agents, each with a certain number of coins representing their wealth. At each time step, one coin is selected uniformly at random, and moves to a neighbor of its owner with a probability that depends on how wealthy the neighbors are. Those who are well-connected and initially wealthy will tend to accumulate more wealth. We refer to particles instead of coins in what follows.

---

*Massachusetts Institute of Technology, United States of America. E-mail: jgaudio@mit.edu.

This is a flexible model based on a few principles: There are a fixed number of particles moving around on a graph. Movements are asynchronous, and particles make choices about where to move based on their local environment. The model can encompass a variety of situations. Further, the model can be extended by allowing for multiple particle types, with intra– and inter–group attraction parameters, though we do not consider this extension in this paper. There are many more applications beyond the economic application. As an interacting particle system, it could be relevant for physics or chemistry applications.

This paper analyzes the Attracting Random Walks model and establishes phase transition properties. The difficulty in bounding mixing times, particularly in finding lower bounds, is due to the fact that the stationary distribution cannot be simply formulated. Additionally, the model is not reversible unless the graph is complete (Theorem 2.5), meaning that familiar techniques do not apply.

We establish the existence of phase transition in mixing time as the attraction parameter, β, is varied. Slow mixing for β large enough is established by relating the mixing time to a suitable hitting time. Fast mixing for β small enough is proven by a path coupling approach that relates the Attracting Random Walks chain to the simple (non-interacting) random walk on the same graph (i.e. with $\beta = 0$). As a corollary of our main results, we establish properties of the Cheeger cut for the stationary distribution. We find it interesting that even though the stationary distribution is not known analytically for general graphs, we have shown that it undergoes a phase transition (i.e. develops an exponentially small Cheeger cut) by arguing indirectly via mixing times.

The rest of the paper is structured as follows. We describe the dynamics of the model in Section 2, along with some possible applications. The remainder of the paper is focused on properties of the Markov chain governing the dynamics. In Section 2.2 we discuss a link to the Potts model. Section 3 proves the existence of phase transition in mixing time for general graphs, and is the main theoretical contribution of this work. In Section 4, we collect partial results on the version of the model in which particles repel each other instead of attracting, a model we call "Repelling Random Walks."

2 The model

2.1 Definitions and main results

The model is a discrete time process on a simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices and $\mathcal{E}$ is the set of undirected edges. We assume throughout that $\mathcal{G}$ is connected. We write $i \sim j$ if $(i, j) \in \mathcal{E}$. Let $k = |\mathcal{V}|$. Initially, n indistinguishable particles are placed on the vertices of $\mathcal{G}$ in some configuration. Let $x(i)$ be the number of particles at vertex i. The particle configuration is updated in two stages, according to a fixed parameter β:

1. Choose a particle uniformly at random. Let i be the location of that particle.
2. Move the particle to a vertex $j \sim i, j \neq i$, with probability $\frac{1}{Z} \exp\left(\frac{\beta}{n} x(j)\right)$. Keep the particle at vertex i with probability $\frac{1}{Z} \exp\left(\frac{\beta}{n}\left(x(i) - 1\right)\right)$, where Z is the normalization constant.

Let P be the transition probability matrix of the resulting Markov chain. Let e_i denote the ith standard basis vector in $\mathbb{R}^k$. Then for two configurations x and y such that $y = x - e_i + e_j$ for $i \sim j$ or $i = j$, we have

$$P(x, y) = \begin{cases} \frac{x(i)}{n} \frac{\exp\left(\frac{\beta}{n} x(j)\right)}{Z} & \text{if } i \sim j \\ \frac{x(i)}{n} \frac{\exp\left(\frac{\beta}{n}\left(x(i)-1\right)\right)}{Z} & \text{if } i = j \end{cases},$$

with $Z = \sum_{l \sim i} \exp\left(\frac{\beta}{n} x(l)\right) + \exp\left(\frac{\beta}{n}(x(i) - 1)\right)$.

The probabilities are a function of the numbers of particles at each vertex, excluding the particle that is to move. This modeling choice means that the moving particle is neutral toward itself, and relates the ARW model to the Potts model, as will be explained below.

When β is positive (ferromagnetic dynamics), the particle is more likely to travel to a vertex that has more particles. Greater β encourages stronger aggregation of the particles. On the other hand, taking $\beta < 0$ (antiferromagnetic dynamics) encourages particles to spread. Note that $\beta = 0$ corresponds to the case of independent (lazy) random walks.

For an application with $\beta < 0$, consider an ensemble of identical gas particles in a container. We can discretize the container into blocks. Each block becomes a vertex in our graph. Vertices are connected by an edge whenever the corresponding blocks share a face. Note that depending on the type of gas, particles may primarily repel each other, in which case $\beta < 0$, which discourages particles from occupying the same block, does become reasonable. The focus of this paper is the case of $\beta > 0$, though. we collect some results on the $\beta < 0$ case as well.

To get an idea of the effect of β, Figure 1 displays some instances of the Attracting Random Walks model run for 10^5 steps for different values of β. The graph is the 8×8 grid graph, with $n = 320$, for an average of 5 particles per vertex.
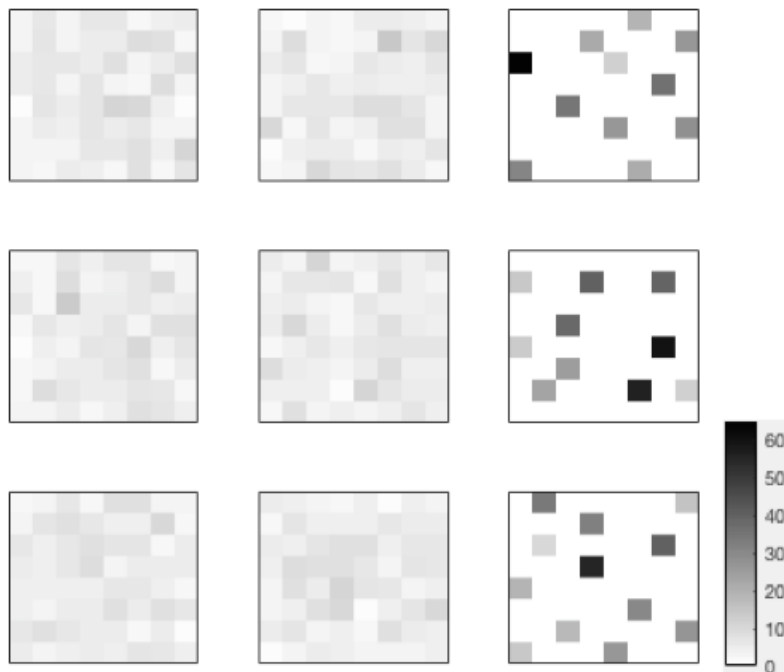

Figure 1: Simulation of the Attracting Random Walks model on a grid graph. From left to right: $\beta = 0$, $\beta = 300$, and $\beta = 500$. Three trials are shown for each.

We now state our main results regarding the phase transition in mixing time. We let $\|P - Q\|_{\mathrm{TV}}$ denote the total variation distance between two discrete probability measures P and Q, and let $d(X, t) \triangleq \max_{x \in \mathcal{X}} \|P^t(x, \cdot) - \pi\|_{\mathrm{TV}}$ be the worst-case (with respect to the initial state) total variation distance for a chain $\{X_t\}$ with stationary distribution π. Let $t_{\mathrm{mix}}(X, \epsilon) \triangleq \min\{t : d(X, t) \leq \epsilon\}$ denote the *mixing time* of a chain $\{X_t\}$.

Theorem 2.1. *For any graph $\mathcal{G}$, there exists $\beta_+ > 0$ such that if $\beta > \beta_+$, the mixing time of the ARW model is $\exp(\Omega(n))$.*

Theorem 2.2. *For any graph $\mathcal{G}$, there exists $\beta_- > 0$ such that if $0 \leq \beta < \beta_-$, the mixing time of the ARW model is $O(n \log n)$.*

Note that we do not prove that one value $\beta_+ = \beta_-$ satisfies both statements.

Through our analysis of mixing time, we establish a transition in the dynamics of the chain. By standard results, this also indirectly implies that the stationary distribution develops multiple almost disjoint phases for $\beta > \beta_+$, while this is not the case for $\beta < \beta_-$. More precisely, we have the following corollary.

Definition 2.3 (Cheeger constant [6])**.** *Let P be the transition matrix of a Markov chain that is irreducible and aperiodic. Let $\mathcal{X}$ denote the state space of the chain, and let π be the stationary distribution. Define the* edge measure Q *by*

$$Q(x, y) \triangleq \pi(x)P(x, y).$$

For two sets $A, B \subset \mathcal{X}$, let $Q(A, B) = \sum_{x \in A, y \in B} Q(x, y)$. For $S \subset \mathcal{X}$, let

$$\Phi(S) \triangleq \frac{Q(S, S^c)}{\pi(S)}.$$

Finally, the Cheeger constant *is defined as*

$$\Phi_* \triangleq \min_{S : \pi(S) \leq \frac{1}{2}} \Phi(S).$$

Our results on fast and slow mixing allow us to indirectly bound the Cheeger constant of the Attracting Random Walks chain on a given graph. We obtain the following corollary of Theorems 2.1 and 2.2.

Corollary 2.4. *Fix a graph $\mathcal{G}$, and let P be the transition probability matrix of the Attracting Random Walks chain on G. Let Φ_* be the Cheeger constant of P. Then if $0 \leq \beta < \beta_-$ we have $\Phi_* = \frac{1}{O(n \log n)}$. If $\beta > \beta_+$ then $\Phi_* = \exp(-\Omega(n))$.*

2.2 Connection to the Potts model

In the case where $\mathcal{G}$ is the complete graph, the Attracting Random Walks model is a projection of Glauber dynamics of the Curie–Weiss Potts model. The Potts model is a multicolor generalization of the Ising model, and the Curie–Weiss version considers a complete graph. In the Curie–Weiss Potts model, the vertices of a complete graph are assigned a color from $[q] = \{1, \ldots, q\}$. Setting $q = 2$ corresponds to the Ising model.

Let $s(i)$ be the color of vertex i for each $1 \leq i \leq n$. Define

$$\delta\left(s(i), s(j)\right) \triangleq \begin{cases} 1, & \text{for } s(i) = s(j) \\ 0, & \text{for } s(i) \neq s(j) \end{cases}.$$

The stationary distribution of the Potts model, with no external field, is

$$\pi(s) = \frac{1}{Z} \exp\left(\frac{\beta}{n} \sum_{(i,j), i \neq j} \delta\left(s(i), s(j)\right)\right).$$

The Glauber dynamics for the Curie–Weiss Potts model are as follows:

1. Choose a vertex i uniformly at random.
2. Update the color of vertex i to color $k \in [q]$ with probability proportional to $\exp\left(\frac{\beta}{n} \sum_{j \neq i} \delta\left(k, s(j)\right)\right)$.

Observe that the summation $\sum_{j \neq i} \delta(k, s(j))$ is equal to the number of vertices, apart from vertex i, that have color k. Therefore if each vertex in the Potts model corresponds to a particle in the ARW model, and each color in the Potts model corresponds to a vertex in the ARW model, then the ARW model is a projection of the Glauber dynamics for the Potts model. The correspondence is illustrated in Figure 2. Under the correspondence, the ARW chain is exactly the "vector of proportions" chain in the Potts model.
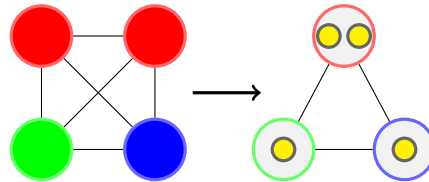

Figure 2: Correspondence of the Curie–Weiss Potts model to the Attracting Random Walks model. A Potts configuration is drawn on the left, and the corresponding ARW configuration is drawn on the right.

Let $v(i)$ be the vertex location of the ith particle in the ARW model, for $1 \leq i \leq n$. By the correspondence, we show that the stationary distribution of the ARW model is

$$\pi(x) = \frac{1}{Z} \binom{n}{x(1), x(2), \ldots, x(k)} \exp \left(\frac{\beta}{n} \sum_{(i,j), i \neq j} \delta(v(i), v(j)) \right)$$

$$= \frac{1}{Z} \binom{n}{x(1), x(2), \ldots, x(k)} \exp \left(\frac{\beta}{2n} \sum_{i=1}^{n} (x(v(i)) - 1) \right)$$

$$= \frac{1}{Z} \binom{n}{x(1), x(2), \ldots, x(k)} \exp \left(\frac{\beta}{2n} \sum_{i=1}^{k} x(i)^2 - \frac{\beta}{2} \right)$$

$$= \frac{1}{Z'} \binom{n}{x(1), x(2), \ldots, x(k)} \exp \left(\frac{\beta}{2n} \sum_{i=1}^{k} x(i)^2 \right).$$

Observe that the $\exp \left(\frac{\beta}{2n} \sum_i x(i)^2 \right)$ factor encourages particle aggregation, while the multinomial encourages particle spread.

The reader is encouraged to refer to [2] for a detailed study of the mixing time of the Curie–Weiss Potts model, for different values of β. For instance, [2] shows that there exists $\beta_s(q)$ such that if $\beta < \beta_s(q)$, the mixing time is $\Theta(n \log n)$, and if $\beta > \beta_s(q)$, the mixing time is exponential in n. In the ARW context, these results hold with q replaced by k. On the other hand, when $\mathcal{G}$ is not the complete graph, the correspondence to the Potts model is lost. In fact, the following can be shown:

Theorem 2.5. *For $n \geq 3$, the ARW Markov chain is reversible for all β if and only if the graph $\mathcal{G}$ is complete.*

The non-reversibility can be shown by applying Kolmogorov's cycle criterion, demonstrating a cycle of states (configurations) that violates the criterion.

Lemma 2.6 (Kolmogorov's criterion)**.** *A finite state space Markov chain associated with the transition probability matrix P is reversible if and only if for all cyclic sequences of states $i_1, i_2, \ldots, i_{l-1}, i_l, i_1$ it holds that*

$$\left(\prod_{j=1}^{l-1} P(i_j, i_{j+1}) \right) P(i_l, i_1) = P(i_1, i_l) \left(\prod_{j=0}^{l-2} P(i_{l-j}, i_{l-j-1}) \right).$$

In other words, the forward product of transition probabilities must equal the reverse product, for all cycles of states.
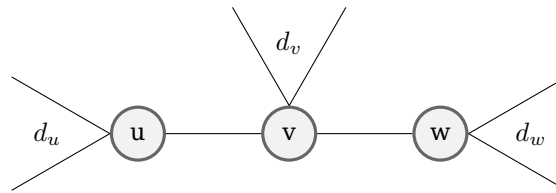

Figure 3: Initial state of a cycle that breaks Kolmogorov's criterion.

Proof of Theorem 2.5. First, if the graph is complete, then the chain is a projection of Glauber dynamics, which is automatically reversible. Now suppose $\mathcal{G}$ is not complete. We apply Kolmogorov's cycle criterion. In the ARW model, a state is a particle configuration. A cycle of states is then a sequence of particle configurations such that

1. Subsequent configurations differ by the movement of a single particle.
2. The first and last configurations are the same.

If $\mathcal{G}$ is not a complete graph, then it is straightforward to show that there exist three vertices $u \sim v \sim w$ such that $u \nsim w$. Now we demonstrate a cycle of states that breaks Kolmogorov's criterion. We have the following situation, illustrated by Figure 3. The values d_u, d_v, and d_w indicate the degrees of the vertices, excluding the named vertices. Place $n - 2$ particles at u and 2 particles at v. The particle movements are as follows: $v \to u$, $v \to w$, $u \to v$, $w \to v$.

For clarity, let $f(z) = \exp\left(\frac{\beta}{n} z\right)$. The forward transition probabilities are

$$\left(\frac{2}{n} \frac{f(n-2)}{f(n-2) + f(1) + 1 + d_v}\right) \left(\frac{1}{n} \frac{1}{f(n-1) + 1 + 1 + d_v}\right) \left(\frac{n-1}{n} \frac{1}{f(n-2) + 1 + d_u}\right) \left(\frac{1}{n} \frac{f(1)}{f(1) + 1 + d_w}\right).$$

The reverse transition probabilities are

$$\left(\frac{2}{n} \frac{1}{f(n-2) + f(1) + 1 + d_v}\right) \left(\frac{1}{n} \frac{f(n-2)}{f(n-2) + 1 + f(1) + d_v}\right) \left(\frac{1}{n} \frac{1}{1 + 1 + d_w}\right) \left(\frac{n-1}{n} \frac{f(1)}{f(n-2) + f(1) + d_u}\right).$$

Canceling factors that appear in both products, we are left comparing

$$(f(n-1) + 1 + 1 + d_v)\,(f(n-2) + 1 + d_u)\,(f(1) + 1 + d_w)$$

to

$$(f(n-2) + 1 + f(1) + d_v)\,(1 + 1 + d_w)\,(f(n-2) + f(1) + d_u).$$

Observe that $f(z_1)f(z_2) = f(z_1 + z_2)$. Taking leading terms, the first product is therefore a degree-$(2n - 2)$ polynomial in e^β. Since $n - 2 \geq 1$, the second is a degree-$(2n - 4)$ polynomial in e^β. These polynomials have a finite number of solutions for e^β, and therefore β itself. Therefore the Markov chain is not reversible. $\qquad\square$

3 Mixing time on general graphs

In this section, we show the existence of phase transition in mixing time in the ARW model when β is varied, for a general fixed graph. First, we show exponentially slow mixing for β suitably large, namely prove Theorem 2.1 by relating mixing times to hitting times. Next, we show polynomial time mixing for small values of β. The proof is by an adaptation of path coupling. We use definitions and notations on Markov chains from [6].

3.1 Slow mixing

The idea of the proof of slow mixing is to show that with substantial probability, the chain takes an exponential time to access a constant portion of the state space. We now outline the proof, deferring the proofs of the lemmas. First we state a helper lemma.

Lemma 3.1. *For any graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, there exists a vertex $v \in \mathcal{V}$ such that for the set of configurations $S_v \triangleq \{x : x(v) = \max_w x(w)\}$, it holds that $\pi(S_v) \geq 1/k$. In other words, the states where v has the greatest number of particles contribute at least $1/k$ to the stationary probability mass.*

By Lemma 3.1, there exists a vertex v such that $\pi(S_v) \geq 1/k$. Choose any other vertex u. Whenever $x(u) > n/2$, we can be sure that v is not the maximizing vertex, and therefore that a set of states having at least $1/k$ mass under the stationary measure has not been reached. It therefore suffices to lower bound the time until vertex u has lost sufficient particles for vertex v to have the maximum number of particles.

Let $T_x \triangleq \inf\{t : X_t(u) \leq \frac{1}{2}n, X_0 = x\}$. If the probability that $\{X_t\}$ has reached the set $\{x \in \Omega : x(0) \leq n/2\}$ by time t is less than some p, then the total variation distance at time t is at least $(1 - p)\frac{1}{k}$. Therefore we get the following relationship between the mixing time and hitting time:

Proposition 3.2.

$$t_{mix}\left(X, (1-p)\frac{1}{k}\right) \geq \inf\left\{t : \min_x \mathbb{P}\left(T_x \leq t\right) \geq p\right\}.$$

The problem now reduces to lower bounding this hitting time. The idea is that when particles leave vertex u, there is a strong drift back to u. However, controlling the hitting times of a multidimensional Markov chain is challenging, and direct comparison is difficult to establish. We instead reason by comparison to another Markov chain, Z, which lower-bounds the particle occupancy at vertex u.

Let $l(w)$ be the length of the shortest path connecting vertex u to vertex w. Let $\tilde{X}_t$ be a projection of the X_t chain defined by $\tilde{X}_t(d) \triangleq \sum_{w:l(w)=d} X_t(w)$, and let $\tilde{\Omega}$ be its state space. In other words, the dth coordinate of the projected chain counts the number of particles that are a distance d away from vertex u. Note that $\tilde{X}_t(0) = X_t(u)$. We let F denote this projection, writing, $\tilde{X} = F(X)$. For any $0 < \delta < 1/2$, define

$$T_x(\delta) \triangleq \inf\{t : X_t(u) \leq (1-\delta)n, X_0 = x\} = \inf\{t : \tilde{X}_t(0) \leq (1-\delta)n, \tilde{X}_0 = F(x)\}.$$

For some $\delta > 0$ to be determined, let

$$S \triangleq \{x \in \tilde{\Omega} : x(0) > (1-\delta)n\} \text{ and } S^c \triangleq \tilde{\Omega} \setminus S.$$

We now build a chain Z on $\tilde{\Omega}$ coupled to $\tilde{X}$ such that as long as $\tilde{X}_t \in S$, $Z_t(0) \overset{\text{st}}{\leq} \tilde{X}_t(0)$. Then $T_x(\delta) \overset{\text{st}}{\geq} \inf_t\{Z_t \in S^c\}$. The remainder the proof of slow mixing is as follows.

1. Construct a lower-bounding comparison chain Z satisfying $Z_t(0) \overset{\text{st}}{\leq} \tilde{X}_t(0)$ when $t \leq T_x(\delta)$.
2. Compute $\mathbb{E}_{\pi_Z}[Z(0)]$ and use a concentration bound to show that $Z(0) \sim \pi_Z(0)$ places exponentially little mass on the set S^c.
3. Comparing the chain X to Z, show that X takes exponential time to achieve $X(u) \leq (1-\delta)n$. The result is complete by $1 - \delta > 1/2$.

We now define the lower-bounding comparison chain Z, which is a chain on n independent particles. These particles move on the discrete line with points $\{0, 1, \ldots, D\}$,

where $D = diam(\mathcal{G})$. We first describe the case $D \geq 2$. Since the comparison needs to hold only when $\tilde{X}_t(0) \geq (1-\delta)n$, we assume that $\tilde{X}_t(0) \geq (1-\delta)n$. The idea is to identify a uniform constant lower bound on the probability of a particle moving closer to u under this assumption, which tells us that once the particle is at u, there is a high probability of remaining there.

Let $\mathcal{N}(u)$ denote the neighbourhood of u, i.e. $\mathcal{N}(u) = \{w : w \sim u\}$. In the X chain, when a particle is at a vertex $w \notin \{u\} \cup \mathcal{N}(u)$, its probability of moving to any one of its neighbors is at least

$$p \triangleq \frac{1}{e^{\beta\delta} + \Delta},$$

where Δ is the maximum degree of the graph. This is because the lowest probability when β is large corresponds to placing all δn movable particles at some other neighbor of w. When a particle is at a vertex u, it stays there with probability at least

$$q \triangleq \frac{\exp\left(\beta(1-\delta) - \frac{\beta}{n}\right)}{\exp\left(\beta(1-\delta) - \frac{\beta}{n}\right) + e^{\beta\delta} + \Delta - 1},$$

When a particle is at a vertex $w \in \mathcal{N}(u)$, it moves to u with probability at least

$$\frac{\exp\left(\beta(1-\delta)\right)}{\exp\left(\beta(1-\delta)\right) + e^{\beta\delta} + \Delta - 1} > q.$$

Note that $q > p$.

The transitions of the Z chain are chosen in order to maintain comparison. At each time step, a particle is selected uniformly at random. When the chosen particle is located at $d \notin \{0,1\}$, the particle moves to $d-1$ with probability p and moves to $\min\{d+1, D\}$ with probability $(1-p)$. When the chosen particle is located at $d \in \{0,1\}$, it moves to 0 with probability q, and moves to $d+1$ with probability $1-q$. The transition probabilities for single particle movements are depicted in Figure 4. When $D = 1$ (i.e., $\mathcal{G}$ is the complete graph), we instead have the transitions depicted by Figure 5. Lemma 3.3 establishes the comparison.

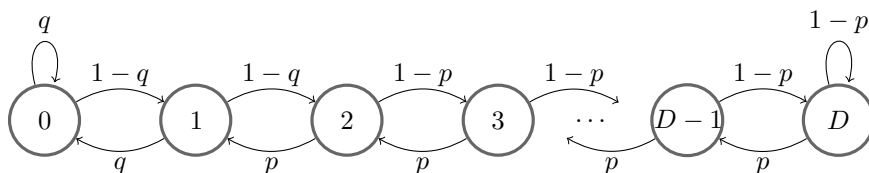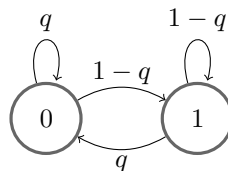Figure 4: Single-particle Markov chain from the Z chain ($D \geq 2$)

Figure 5: Single-particle Markov chain from the Z chain ($D = 1$)

Let π_Z denote the stationary distribution of the Z chain, and let $\lambda(w)$ be the probability according to π_Z of a particular particle being located at vertex w in the line graph. The following results about the Z chain are required to complete the proof.

Lemma 3.3. *For a configuration $x \in \Omega$, set $Z_0 = \tilde{X}_0 = F(x)$. As long as $t \leq T_x(\delta)$, the chain Z_t satisfies*

$$\sum_{r=0}^{d} Z_t(r) \overset{\text{st}}{\leq} \sum_{r=0}^{d} \tilde{X}_t(r)$$

for all $d \in \{0, 1, \ldots, D\}$ and $t \in \{0, 1, 2, \ldots\}$. In particular, $Z_t(0) \overset{\text{st}}{\leq} \tilde{X}_t(0)$.

Lemma 3.4. *Recall that $D = diam(\mathcal{G})$. Let $\delta = \frac{1}{3D}$ and fix $0 < \bar{\epsilon} < \delta/2$. For all β large enough, $\mathbb{E}_{\pi(Z)}[Z(0)] \geq (1 - \delta + \bar{\epsilon})n$. Moreover,*

$$\mathbb{P}_{\pi(Z)}\left(Z(0) \leq \mathbb{E}_{\pi(Z)}[Z(0)] - \bar{\epsilon}n\right) \leq 2\exp\left(-2\bar{\epsilon}^2 n\right),$$

which implies

$$\mathbb{P}_{\pi(Z)}\left(Z(0) \leq (1-\delta)n\right) \leq 2\exp\left(-2\bar{\epsilon}^2 n\right).$$

Proof of Theorem 2.1. Recall the choices of u and v above. Lemma 3.4 tells us that the Z chain places exponentially little stationary mass on the set S^c. We now combine this fact with the comparison established in Lemma 3.3.

Recall $T_x(\delta) = \inf\{t : X_t(u) \leq (1-\delta)n, X_0 = x\} = \inf\{t : \tilde{X}_t(0) \leq (1-\delta)n, \tilde{X}_0 = F(x)\}$. Applying Proposition 3.2 with $p = 1/2$,

$$t_{\text{mix}}\left(X, \frac{1}{2k}\right) \geq \min\left\{t : \min_x \mathbb{P}\left(T_x\left(\frac{1}{2}\right) \leq t\right) \geq \frac{1}{2}\right\}.$$

Since $1/2 < 1 - \delta$, it also holds that

$$\begin{aligned}
t_{\text{mix}}\left(X, \frac{1}{2k}\right) &\geq \min\left\{t : \min_x \mathbb{P}\left(T_x(\delta) \leq t\right) \geq \frac{1}{2}\right\} \\
&= \min\left\{t : \mathbb{P}(T_x(\delta) \leq t) \geq \frac{1}{2}, \forall x \in \Omega\right\} \\
&= \min\left\{t : \mathbb{P}(T_x(\delta) \leq t) \geq \frac{1}{2}, \forall x \in S\right\}.
\end{aligned} \tag{3.1}$$

The last equality is due to the fact that $\mathbb{P}(T_x(\delta) \leq t) = 1$ for all x in S^c.

Additionally define

$$T_x^Z \triangleq \inf\left\{t : Z_t \in \mathcal{S}^c, Z_0 = F(x)\right\}.$$

Now because Z_t is a lower-bounding chain, it holds that

$$\mathbb{P}\left(T_x(\delta) \leq t\right) \leq \mathbb{P}\left(T_x^Z \leq t\right)$$

for all $x \in S$ and $t \geq 0$. Therefore,

$$t_{\text{mix}}\left(X, \frac{1}{2k}\right) \geq \min\left\{t : \mathbb{P}\left(T_x^Z \leq t\right) \geq \frac{1}{2}, \forall x \in S\right\}.$$

Finally, from Lemma 3.4 we know that $\pi_Z(S^c) \leq 2\exp\left(-2\bar{\epsilon}^2 n\right)$. Suppose that Z_0 is distributed according to π_Z and consider the hitting time $T_{\pi_z}^Z$. It holds that

$$T_{\pi_z}^Z \overset{\text{st}}{\geq} \text{Geom}\left(2\exp\left(-2\bar{\epsilon}^2 n\right)\right).$$

Therefore, $t = e^{\Theta(n)}$ time is required for $\mathbb{P}\left(T_{\pi_Z}^Z \leq t\right) \geq \frac{1}{2}$. The same is true when $Z_0 = x$, for some $x \in S$. Therefore $\min\left\{t : \mathbb{P}\left(T_x^Z \leq t\right) \geq \frac{1}{2}, \forall x \in S\right\} = e^{\Theta(n)}$ and $t_{\text{mix}}\left(X, \frac{1}{2k}\right) = e^{\Omega(n)}$, which proves Theorem 2.1. $\square$

We now provide the deferred proofs.

Proof of Lemma 3.1. By the Union Bound,

$$\sum_{v \in \mathcal{V}} \pi(S_v) \geq \mathbb{P}_\pi \left(\cup_{v \in \mathcal{V}} \left\{ x(v) = \max_w x(w) \right\} \right) = 1. \qquad \square$$

Proof of Lemma 3.3. We show that there exists a coupling $(\tilde{X}_t, Z_t)$ satisfying

$$\sum_{r=0}^{d} Z_t(r) \leq \sum_{r=0}^{d} \tilde{X}_t(r) \qquad (3.2)$$

for all $d \in \{0, 1, \dots, D\}$ and $t \leq T_x(\delta)$. Since $Z_0 = \tilde{X}_0$, we can pair up the particles at time $t = 0$ and design a synchronous coupling, i.e. when a certain particle is chosen in the $\tilde{X}$ process, its copy is chosen in the Z chain. We design the coupling so that for each particle, the $\tilde{X}$–copy is at least as close to 0 as the Z–copy, for all $t \leq T_x(\delta)$. Note that this implies (3.2) for all $d \in \{0, 1, \dots, D\}$ and $t < T_x(\delta)$. The uniformity of p and q over all configurations in S ensures that the coupling will maintain the requirement (3.2), which is established by induction on t. The following analysis applies to both $D \geq 2$ and $D = 1$ by considering the relevant cases.

The base case $(t = 0)$ holds since $Z_0 = \tilde{X}_0$. Suppose that at time $t < T_x(\delta)$, each particle in the $\tilde{X}$ chain is at least as close to 0 as its copy in the Z chain. We will show that the same property holds for time $t + 1$. First consider a particle located at 0 in the Z chain. By the inductive hypothesis, its copy must be located at 0 in the $\tilde{X}$ process also, and the corresponding particle in the X chain must be at u. The probability of the particle staying at 0 in the Z chain is smaller than the probability of the corresponding particle staying at u in the X chain, since q is a uniform lower bound on the probability of staying at u. Therefore in this case, the property is maintained in the next time step.

Next consider a particle located at vertex $d \neq 0$ in the Z chain and suppose its copy is located at vertex d' in the $\tilde{X}$ process. By the inductive hypothesis, $d' \leq d$. If $d' < d - 1$, then clearly the property is maintained in the next step. It remains to consider the cases $d' = d$ and $d' = d - 1$. Consider the case $d = d'$. We couple the particles so that if the particle in the Z chain moves left to vertex $d-1$, then the particle in the $\tilde{X}$ process makes the same transition. This coupling is possible by the uniformity of p and q. Otherwise, the particle in the Z chain moves right, and the property is maintained.

Next consider the case $d' = d - 1$. It suffices to design a coupling such that if the particle in the $\tilde{X}$ process moves right, then so does the particle in the Z chain. If $d \geq 3$, this is possible due to the fact that $1 - p$ is a uniform upper bound on the probability of moving right from these states. Next suppose that $d = 2$ and $d' = 1$. The particle in the $\tilde{X}$ process moves right with probability upper-bounded by $1 - q$, which is smaller than $1 - p$ for δ sufficiently small and n sufficiently large. Therefore we can ensure the property in the next step. Finally, suppose $d = 1$ and $d' = 0$. Due to the fact that $1 - q$ is a uniform upper bound on the probability of moving right from these states, we can again construct a coupling that maintains the property. $\qquad \square$

To prove Lemma 3.4, we need the stationary probability $\lambda(0)$.

Proposition 3.5. *It holds*

$$\lambda(0) = \begin{cases} q & \text{if } D = 1 \\ q \left[1 + \frac{(1-q)^2}{p} \left(\frac{p}{1-p} \right)^{2-D} \left(\frac{1 - \left(\frac{p}{1-p} \right)^{D-1}}{1 - \frac{p}{1-p}} \right) \right]^{-1} & \text{if } D \geq 2 \end{cases}.$$

The proof of Proposition 3.5 is deferred to the appendix.

Proof of Lemma 3.4. When $D = 1$, we have $\lambda(0) = q$. Since $q(\beta) \to 1$ as $\beta \to \infty$, it holds that $\lambda(0) \geq 1 - \delta + \bar{\epsilon}$ for β large enough. Next, for $D \geq 2$ we have

$$
\begin{aligned}
\lambda(0) &= \frac{q}{1 + \frac{(1-q)^2}{p} \left(\frac{p}{1-p}\right)^{2-D} \left(\frac{1 - \left(\frac{p}{1-p}\right)^{D-1}}{1 - \frac{p}{1-p}}\right)} \\
&= \frac{q}{1 + \frac{(1-q)^2}{p} \left(\frac{p}{1-p}\right)^{1-D} \left(\frac{\frac{p}{1-p} - \left(\frac{p}{1-p}\right)^{D}}{1 - \frac{p}{1-p}}\right)} \\
&\geq \frac{q}{1 + \frac{(1-q)^2}{p} \left(\frac{p}{1-p}\right)^{1-D} \left(\frac{\frac{p}{1-p}}{1 - \frac{p}{1-p}}\right)} \\
&= \frac{q}{1 + \frac{(1-q)^2}{1-2p} \left(\frac{p}{1-p}\right)^{1-D}}.
\end{aligned}
\tag{3.3}
$$

We show that $\lambda(0) \geq 1 - \delta + \bar{\epsilon}$ for β and n large enough. Again, for β large enough, we have $q \geq 1 - \delta + 2\bar{\epsilon}$. Next,

$$
\begin{aligned}
1 - q &= \frac{e^{\beta\delta} + \Delta - 1}{\exp\left(\beta(1-\delta) - \frac{\beta}{n}\right) + e^{\beta\delta} + \Delta - 1} \\
&\leq \frac{e^{\beta\delta} + \Delta - 1}{\exp\left(\beta(1-\delta) - \frac{\beta}{n}\right)} \\
&= \exp\left(\beta\left(\frac{2}{3D} - 1 + \frac{1}{n}\right)\right) + (\Delta - 1)\exp\left(\beta\left(\frac{1}{3D} - 1 + \frac{1}{n}\right)\right) \\
&\leq \exp\left(\beta\left(\frac{1}{D} - 1\right)\right),
\end{aligned}
\tag{3.4}
$$

where the last inequality holds for β and n large enough, since the first term of (3.4) dominates. Since $p < \frac{1}{\Delta+1}$, we have $1 - 2p > \frac{\Delta-1}{\Delta+1}$. Finally,

$$
\left(\frac{p}{1-p}\right)^{1-D} = \left(e^{\beta\delta} + \Delta - 1\right)^{D-1} = \left(e^{\beta/(3D)} + \Delta - 1\right)^{D-1} \leq \left(e^{\beta/(2D)}\right)^{D-1} < e^{\beta/2},
$$

where the first inequality holds for β large enough. Substituting into (3.3), we obtain for β and n large enough

$$
\lambda(0) \geq \frac{1 - \delta + 2\bar{\epsilon}}{1 + \frac{\Delta+1}{\Delta-1} \exp\left(2\beta\left(\frac{1}{D} - 1\right) + \frac{\beta}{2}\right)} = \frac{1 - \delta + 2\bar{\epsilon}}{1 + \frac{\Delta+1}{\Delta-1} \exp\left(\beta\left(\frac{2}{D} - \frac{3}{2}\right)\right)} \geq 1 - \delta + \bar{\epsilon},
$$

where the second inequality holds for β large enough, due to $^2/_D - {}^3/_2 < 0$ for $D \geq 2$.

We conclude that the expectation is linearly separated from the boundary:

$$
\mathbb{E}_{\pi_Z}[Z(0)] = \lambda(0)n \geq (1 - \delta + \bar{\epsilon})\, n.
$$

Next we show concentration. Label all the particles, and define $U_i = 1$ if particle i is at vertex 0 in the line graph, and $U_i = 0$ otherwise. Then $Z(0) = \sum_i U_i$, and U_i is independent of U_j for all $i \neq j$. Applying Hoeffding's inequality,

$$
\mathbb{P}_{\pi_Z}\left(\left|Z(0) - \mathbb{E}_{\pi(Z)}[Z(0)]\right| \geq cn\right) \leq 2\exp\left(-\frac{2(cn)^2}{n}\right) = 2\exp\left(-2c^2 n\right).
$$

for $c > 0$. Let $c = \bar{\epsilon}$. Then the above implies

$$
\begin{aligned}
\mathbb{P}_{\pi_Z}\left(Z(0) \leq \mathbb{E}_{\pi_Z}[Z(0)] - \bar{\epsilon}n\right) &\leq 2\exp\left(-2\bar{\epsilon}^2 n\right) \\
\implies \mathbb{P}_{\pi_Z}\left(Z(0) \leq (1-\delta)\, n\right) &\leq 2\exp\left(-2\bar{\epsilon}^2 n\right). \qquad \square
\end{aligned}
$$

3.2 Fast mixing

The proof is by a modification of path coupling, which is a method to find an upper bound on mixing time through contraction of the Wasserstein distance. [1] The following definition can be found in [6], pp. 189.

Definition 3.6 (Transportation metric). *Given a metric ρ on a state space Ω, the associated transportation metric ρ_T for two probability distributions μ and ν is defined as*

$$\rho_T(\mu, \nu) \triangleq \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[\rho(X, Y)]$$

where the infimum is over all couplings of μ and ν on $\Omega \times \Omega$.

Definition 3.7 (Wasserstein distance). *Let P be the transition probability matrix of a Markov chain on a state space Ω, and let ρ be a metric on Ω. The Wasserstein distance $W_\rho^P(x, y)$ of two states $x, y \in \Omega$ with respect to P and ρ is defined as follows:*

$$W_\rho^P(x, y) \triangleq \rho_T\left(P(x, \cdot), P(y, \cdot)\right) = \inf_{X_1 \sim P(x, \cdot), Y_1 \sim P(y, \cdot)} \mathbb{E}_{X_1, Y_1}\left[\rho(X_1, Y_1)\right].$$

In other words, the Wasserstein distance is the transportation metric distance between the next state distributions from initial states x and y.

The following lemma is the path coupling result which can be found in [7] and [6]. Given a Markov chain on state space Ω with transition probability matrix P, consider a connected graph $\mathcal{H} = (\Omega, \mathcal{E}_\mathcal{H})$, i.e. the vertices of $\mathcal{H}$ are the states in Ω and the edges are $\mathcal{E}_\mathcal{H}$. Let l be a "length function" for the edges of $\mathcal{H}$, which is an arbitrary function $l : \mathcal{E}_\mathcal{H} \to [1, \infty)$. For $x, y \in \Omega$, define $\rho(x, y)$ to be the path metric, i.e. $\rho(x, y)$ is the length of the shortest path from x to y in terms of l and $\mathcal{H}$.

Lemma 3.8 (Path Coupling). *Under the above construction, if there exists $\delta > 0$ such that for all x, y that are connected by an edge in $\mathcal{H}$ it holds that*

$$W_\rho^P(x, y) \leq (1 - \delta)\rho(x, y),$$

then

$$d(X, t) \leq (1 - \delta)^t \, diam(\Omega),$$

where $diam(\Omega) = \max_{x, y \in \Omega} \rho(x, y)$ is the diameter of the graph $\mathcal{H}$ with respect to ρ.

Our proof of rapid mixing for small enough β relies on rapid mixing of a single random walk. The following lemma demonstrates the existence of a contracting metric for a single random walk. It is possible that such a result appears elsewhere, but we are not aware of a published proof.

Lemma 3.9. *Consider a random walk on $\mathcal{G}$ which makes a uniform choice among staying or moving to any of the neighbors and denote by Q its transition matrix. Let $d(x, y)$ be the expected meeting time of two independent copies of a random walk on a graph started from states x and y. Then $d(x, y)$ is a metric and Q contracts the respective Wasserstein distance. In particular,*

$$W_d^Q(x, y) \leq \left(1 - \frac{1}{d_{max}}\right) d(x, y),$$

where $d_{max} = \max_{x, y} d(x, y)$. Furthermore, if $x \sim y$, then

$$W_d^Q(x, y) \leq \left(1 - \frac{1}{d'_{max}}\right) d(x, y), \tag{3.5}$$

where $d'_{max} = \max_{x, y : x \sim y} d(x, y)$. [2]

---

[1] An alternative prove of fast mixing is to use a variable-length path coupling, as introduced in [5]. For further details, see [4].

[2] The statement for $x \sim y$ was pointed out by the reviewer.

Remark 3.10. In fact, we can show a stronger result (i.e. with a smaller value in the place of $d_{\max}$): we can allow arbitrary Markovian coupling between two copies of the random walk and define $d(x, y)$ to be the meeting time under that coupling.

In order to apply path coupling, we let $\mathcal{H} = (\Omega, \mathcal{E}_\mathcal{H})$ be a graph on particle configurations, where $(x, y) \in \mathcal{E}_\mathcal{H}$ whenever $y = x - e_i + e_j$ for some pair of distinct vertices i and j in $\mathcal{G}$. In other words, x and y differ by the position of a single particle. Note that i and j need not be neighboring vertices in $\mathcal{G}$. For such a pair of neighboring configurations (x, y), let $l(x, y) = d(i, j)$. Clearly, $l(x, y) \geq \mathbb{1}\{x \neq y\}$. Now for any two configurations $x, y \in \Omega$, let $\rho(x, y)$ denote the path metric induced by $\mathcal{H}$ and $l(\cdot, \cdot)$. We show that $\rho(x, y) = l(x, y)$ for neighboring configurations.

Proposition 3.11. *For any two configurations x, y such that $y = x - e_i + e_j$, it holds that $\rho(x, y) = l(x, y)$.*

Let $P_x(i, \cdot)$ be the probability distribution of the next location of the selected particle, when it is initially located at vertex $i \in \mathcal{V}$ in configuration x. Recall that $Q(i, \cdot)$ is the probability distribution of the next location of a simple random walk on $\mathcal{G}$, initially located at vertex i. Note that when $\beta = 0$, it holds that $P_x(i, \cdot) = Q(i, \cdot)$. When β is small, $P_x(i, \cdot) \approx Q(i, \cdot)$. Lemma 3.12 quantifies this statement.

Lemma 3.12. *For all configurations x and vertices $i \in \mathcal{G}$, it holds that*

$$\|P_x(i, \cdot) - Q(i, \cdot)\|_{TV} \leq \frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}.$$

Next, consider two neighbouring configurations x and y. Because only the position of one particle is different between the two configurations, $P_x(v, \cdot) \approx P_y(v, \cdot)$. The following lemma makes this precise.

Lemma 3.13. *Let x and y be neighbouring configurations. Recall that Δ is the maximum degree of the vertices in $\mathcal{V}$. The following holds:*

$$\|P_x(v, \cdot) - P_y(v, \cdot)\|_{TV} \leq \frac{\Delta + 1}{n}\beta.$$

With these results stated, we prove Theorem 2.2.

Proof of Theorem 2.2. Suppose $d(i, j) \geq \mathbb{1}\{i \neq j\}$ is a metric on $\mathcal{G}$ such that a single-particle random walk's kernel Q satisfies

$$W_d^Q(i, j) \leq (1 - \delta)d(i, j) \tag{3.6}$$

for all $i \neq j$ and $d(i, j) \leq d_{\max}$. Note that the existence of such a metric $d(\cdot, \cdot)$ was established in Lemma 3.9 with an estimate of $\delta = 1/d_{\max}$.

Now we wish to bound $W_\rho^P(x, y)$ for all neighboring particle configurations x and y related by $y = x - e_i + e_j$. We may choose any coupling in order to obtain an upper bound. The coupling will be synchronous: the choice of particle to be moved will be coordinated between the chains. Namely, if the "extra" particle is chosen in configuration x, then so too will the "extra" particle be chosen in configuration y. Similarly, if some other particle is chosen in x, than a particle at the same vertex will be chosen in y. For an illustration, see Figure 6.

Let $X_1 \sim P(x, \cdot)$ and $Y_1 \sim P(y, \cdot)$ denote the coupled random variables corresponding to the next configurations. Let $p^\star$ be the "extra" particle. Let $\tilde{p}$ be a random variable that denotes the uniformly selected particle. Since our coupling gives an upper bound, we can write

$$W_\rho^P(x, y) \leq \frac{1}{n}\mathbb{E}\left[\rho(X_1, Y_1)|\tilde{p} = p^\star\right] + \frac{n - 1}{n}\mathbb{E}\left[\rho(X_1, Y_1)|\tilde{p} \neq p^\star\right]. \tag{3.7}$$
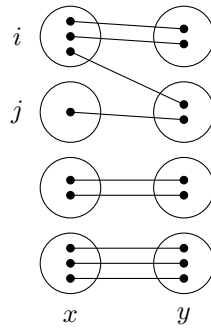
Figure 6: Pairing of particles in the coupling. The edges between vertices are omitted.

First, suppose the "extra" particle, $p^\star$, is chosen in both chains. This happens with probability $\frac{1}{n}$. By Lemma 3.12, we can couple the distributions $P_x(i, \cdot)$ and $P_y(j, \cdot)$ to $Q(i, \cdot)$ and $Q(j, \cdot)$ respectively with probability at least $1 - (e^{\beta/2}-1)/(e^{\beta/2}+1)$. In that case, we get contraction by a factor of $(1 - \delta)$. With the remaining probability, we assume the worst-case distance of $d_{\max}$. Therefore, the conditional Wasserstein distance is upper bounded as follows:

$$\mathbb{E}\left[\rho(X_1, Y_1)|\tilde{p} = p^\star\right] \leq \left(1 - \frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)(1 - \delta)d(i, j) + \left(\frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)d_{\max}. \tag{3.8}$$

Next, suppose some other particle (located at v) is chosen in both chains. This happens with probability $\frac{n-1}{n}$. We claim

$$\mathbb{E}\left[\rho(X_1, Y_1)|\tilde{p} \neq p^\star\right] \leq \rho(x, y) + 2d_{\max}\frac{\Delta + 1}{n}\beta. \tag{3.9}$$

Indeed, by Lemma 3.13, we can couple particle $\tilde{p}$ so that it moves to the same vertex in both chains with probability at least

$$1 - \frac{\Delta + 1}{n}\beta.$$

By Proposition 3.11, it holds that $\rho(X_1, Y_1) = d(i, j) = \rho(x, y)$ in the case that the particle $\tilde{p}$ moves to the same vertex in both chains. Otherwise, an additional distance of at most $2d_{\max}$ is incurred.

Finally, we substitute the bounds (3.8) and (3.9) into (3.7).

$$W_\rho^P(x, y)$$
$$\leq \rho(x, y) + \frac{1}{n}\left(-\rho(x, y) + \left(1 - \frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)(1 - \delta)\rho(x, y) + \left(\frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)d_{\max}\right)$$
$$+ \frac{n-1}{n}\left(\frac{\Delta + 1}{n}2\beta d_{\max}\right)$$
$$= \rho(x, y)\left[1 - \frac{1}{n}\left(1 - \left(1 - \frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)(1 - \delta) - \left(\frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)\frac{d_{\max}}{\rho(x, y)}\right.\right.$$
$$\left.\left. - \frac{n-1}{n}\left(\frac{\Delta + 1}{\rho(x, y)}2\beta d_{\max}\right)\right)\right]$$
$$\leq \rho(x, y)\left[1 - \frac{1}{n}\left(1 - \left(1 - \frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)(1 - \delta) - \left(\frac{e^{\beta/2} - 1}{e^{\beta/2} + 1}\right)d_{\max} - (\Delta + 1)2\beta d_{\max}\right)\right]$$
$$\tag{3.10}$$

where the last inequality is due to $\rho(x,y) \geq 1$ and $\frac{n-1}{n} < 1$. In order to show contraction, it is sufficient that the expression multiplying $\frac{1}{n}$ be positive:

$$1 - \left(1 - \frac{e^{\beta/2}-1}{e^{\beta/2}+1}\right)(1-\delta) - \left(\frac{e^{\beta/2}-1}{e^{\beta/2}+1}\right)d_{\max} - (\Delta+1)2\beta d_{\max} > 0$$

$$\Longleftrightarrow \frac{e^{\beta/2}-1}{e^{\beta/2}+1} < \frac{\delta - (\Delta+1)2\beta d_{\max}}{d_{\max} + \delta - 1}.$$

For an example of a satisfying β, choose β so that

$$(\Delta+1)2\beta d_{\max} < \frac{\delta}{2} \quad \text{and} \quad \frac{e^{\beta/2}-1}{e^{\beta/2}+1} = \tanh\left(\frac{\beta}{4}\right) < \frac{\delta}{2(d_{\max}+\delta-1)}.$$

Therefore, we can choose

$$0 < \beta_- < \min\left\{\frac{\delta}{4d_{\max}(\Delta+1)}, 4\tanh^{-1}\left(\frac{\delta}{2(d_{\max}+\delta-1)}\right)\right\}.$$

Substituting $\beta = \beta_-$ into (3.10), we obtain for some $\delta' > 0$

$$W_\rho^P(x,y) \leq \rho(x,y)\left(1 - \frac{1}{n}\delta'\right).$$

Applying the path coupling lemma (Lemma 3.8), we obtain

$$d(X,t) \leq \left(1 - \frac{1}{n}\delta'\right)^t diam(\Omega) \leq \left(1 - \frac{1}{n}\delta'\right)^t nd_{\max}.$$

Setting the right hand side to be less than $\epsilon > 0$ in order to bound $t_{\mathrm{mix}}(X,\epsilon)$,

$$\left(1 - \frac{1}{n}\delta'\right)^t nd_{\max} \leq \epsilon \iff t \geq \frac{\log\left(\frac{\epsilon}{nd_{\max}}\right)}{\log\left(1 - \frac{1}{n}\delta'\right)} \iff t \geq \frac{\log\left(\frac{nd_{\max}}{\epsilon}\right)}{\log\left(\frac{n}{n-\delta'}\right)}.$$

Since

$$\log\left(\frac{n}{n-\delta'}\right) = \log\left(1 + \frac{\delta'}{n-\delta'}\right) \geq \frac{\frac{\delta'}{n-\delta'}}{1 + \frac{\delta'}{n-\delta'}},$$

we have

$$\frac{\log\left(\frac{nd_{\max}}{\epsilon}\right)}{\log\left(\frac{n}{n-\delta'}\right)} \leq \log\left(\frac{nd_{\max}}{\epsilon}\right)\left(1 + \frac{\delta'}{n-\delta'}\right)\frac{n-\delta'}{\delta'} = O(n\log n).$$

Therefore, $t_{\mathrm{mix}}(X,\epsilon) = O(n\log n)$, which completes the proof of Theorem 2.2. $\qquad\square$

Remark 3.14. Arguably, a more natural approach to show fast mixing would be through a more traditional path coupling approach: Let $\mathcal{H}$ have an edge between configurations x and $y = x - e_i + e_j$ if i and j are adjacent vertices in $\mathcal{G}$. Set $l(x,y) = 1$ for adjacent configurations. However, this approach does not yield contraction in the Wasserstein distance, which we show at the end of this section.

We now provide the deferred proofs.

Proof of Lemma 3.9. First we verify that $d(x,y)$ is a metric. It holds that $d(x,y) = d(y,x)$, and $d(x,y) \geq 0$ with equality if and only if $x = y$. To show the triangle inequality, start three random walks from vertices x, y, z and let $\tau(x,y)$ be the meeting time of the walks started from x and y. The three random walks are advanced according to

the independent coupling, and if a pair of walks collides, they are advanced identically starting from that time. Under this coupling, observe that

$$\tau(x, z) \leq \max\{\tau(x, y), \tau(y, z)\} \leq \tau(x, y) + \tau(y, z)$$

and take expectations. Next we show that $W_\rho^Q(x, y) \leq d(x, y) - 1$ for $x \neq y$. We can choose any coupling of $X_1 \sim P(x, \cdot)$ and $Y_1 \sim P(y, \cdot)$ to show an upper bound. Letting $X_1 \sim P(x, \cdot)$ and $Y_1 \sim P(y, \cdot)$ be independent, we have

$$W_\rho^Q(x, y) \leq \mathbb{E}\left[\tau(X_1, Y_1)\right] = \sum_{a,b} Q(x, a)Q(y, b)\mathbb{E}[\tau(a, b)]$$

and

$$d(x, y) = \mathbb{E}[\tau(x, y)] = 1 + \sum_{a,b} Q(x, a)Q(y, b)\mathbb{E}[\tau(a, b)].$$

These two equations imply $W_\rho^Q(x, y) \leq d(x, y) - 1$. Finally, $d(x, y) - 1 \leq d(x, y)\left(1 - \frac{1}{d_{\max}}\right)$. If $x \sim y$, then we conclude $d(x, y) - 1 \leq d(x, y)\left(1 - \frac{1}{d'_{\max}}\right)$. □

Proof of Proposition 3.11. Consider any path from x to y: $(x = x_0, x_1, \ldots, x_{m-1}, x_m = y)$, where $x_{r+1} = x_r - e_{i_r} + e_{j_r}$ for $r \in \{0, 1, \ldots, m-1\}$. Then we have

$$\sum_{r=0}^{m-1} l(x_r, x_{r+1}) = \sum_{r=0}^{m-1} d(i_r, j_r).$$

We claim that we can rearrange this summation to be of the form

$$d(i, l_1) + \sum_{r=1}^{m-2} d(l_r, l_{r+1}) + d(l_{m-1}, j)$$

for some sequence $l_1, \ldots, l_{m-1}$. Indeed, let $\mathcal{I} = \{i_r : 0 \leq r \leq m - 1\}$ and $\mathcal{J} = \{j_r : 0 \leq r \leq m - 1\}$ be the multisets that collect the "outbound" and "inbound" particle transfers, respectively. The value i must appear one more time in $\mathcal{I}$ than in $\mathcal{J}$. Similarly, the value j must appear one more time in $\mathcal{J}$ than in $\mathcal{I}$. All other values appear an equal number of times in $\mathcal{I}$ and $\mathcal{J}$. By choosing terms $d(i_r, j_r)$ in order, beginning with $d(i, l_1)$, it is possible to rearrange the sum into the given form. By the triangle inequality for $d(\cdot, \cdot)$,

$$d(i, l_1) + \sum_{r=1}^{m-2} d(l_r, l_{r+1}) + d(l_{m-1}, j) \geq d(i, j) = l(x, y).$$

Therefore, the shortest distance between x and y is along the edge connecting them, and we conclude that $\rho(x, y) = l(x, y)$ for neighboring configurations. □

To prove Lemma 3.12, we state the following proposition.

Proposition 3.15. *The set of distributions $\{P_x(i, \cdot) : x \in \Omega\}$ parametrized by the configuration x is contained within the convex set*

$$P_\beta \triangleq \left\{ (p_0, \ldots, p_d) : \frac{p_i}{p_j} \leq e^\beta \,\forall i, j; \sum_{i=0}^{d} p_i = 1; p_i \geq 0 \,\forall i \right\}.$$

Proof. To show this claim, we compute the ratio $\frac{P_x(i, j_1)}{P_x(i, j_2)}$ when $j_1, j_2 \in \mathcal{N}(i) \cup \{i\}$ and $j_1 \neq j_2$, and show that it is upper bounded by e^β. There are three cases to consider.

1. The case $j_1 = i$.

$$\frac{P_x(i, j_1)}{P_x(i, j_2)} = \frac{\exp\left(\frac{\beta}{n}(x(j_1) - 1)\right)}{\exp\left(\frac{\beta}{n}x(j_2)\right)} = \exp\left(\frac{\beta}{n}(x(j_1) - x(j_2) - 1)\right).$$

Since $x(j_1) - x(j_2) - 1 \le n - 1 < n$, it holds that $\frac{P_x(i,j_1)}{P_x(i,j_2)} < e^\beta$.

2. The case $j_2 = i$.

$$\frac{P_x(i, j_1)}{P_x(i, j_2)} = \exp\left(\frac{\beta}{n}(x(j_1) - x(j_2) + 1)\right).$$

Since $j_2 = i$, we have $j_2 \ge 1$. Therefore, again $\frac{P_x(i,j_1)}{P_x(i,j_2)} < e^\beta$.

3. The case $j_1, j_2 \ne i$.

$$\frac{P_x(i, j_1)}{P_x(i, j_2)} = \exp\left(\frac{\beta}{n}(x(j_1) - x(j_2))\right) \le e^\beta. \qquad \square$$

Proof of Lemma 3.12. Recall that $\mathcal{N}(i)$ is the neighbor set of vertex i in graph $\mathcal{G}$. Let $d = |\mathcal{N}(i)|$. We have

$$P_x(i, j) = \begin{cases} \frac{\exp\left(\frac{\beta}{n}x(j)\right)}{\sum_{l \sim i} \exp\left(\frac{\beta}{n}x(l)\right) + \exp\left(\frac{\beta}{n}(x(i) - 1)\right)} & \text{if } i \sim j \\ \frac{\exp\left(\frac{\beta}{n}(x(i) - 1)\right)}{\sum_{l \sim i} \exp\left(\frac{\beta}{n}x(l)\right) + \exp\left(\frac{\beta}{n}(x(i) - 1)\right)} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

and

$$Q(i, j) = \begin{cases} \frac{1}{d+1} & \text{if } i \sim j \\ \frac{1}{d+1} & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

Using Proposition 3.15,

$$\max_x \|P_x(i, \cdot) - Q(i, \cdot)\|_{\mathrm{TV}} \le \sup_{p \in P_\beta} \|p - Q(i, \cdot)\|_{\mathrm{TV}} = \frac{e^\beta}{d + e^\beta} - \frac{1}{d + 1}. \tag{3.11}$$

The inequality is due to the fact that $\{P_x(i, \cdot) : x \in \Omega\} \subset P_\beta$ and the equality is due to the fact that the maximum of a convex function over a closed and bounded convex set is achieved at an extreme point, namely $\left(\frac{e^\beta}{d+e^\beta}, \frac{1}{d+e^\beta}, \ldots, \frac{1}{d+e^\beta}\right)$. To maximize the right hand side of (3.11), let $f(d) = \frac{e^\beta}{d+e^\beta} - \frac{1}{d+1}$. Then

$$\begin{aligned} f'(d) &= -\frac{e^\beta}{(d + e^\beta)^2} + \frac{1}{(d + 1)^2} \\ &= \frac{(d + e^\beta)^2 - e^\beta(d + 1)^2}{(d + e^\beta)^2 (d + 1)^2} \\ &= \frac{\left(d + e^\beta - e^{\beta/2}(d + 1)\right)\left(d + e^\beta + e^{\beta/2}(d + 1)\right)}{(d + e^\beta)^2 (d + 1)^2}. \end{aligned}$$

Setting $f'(d) = 0$ we obtain the solutions $d = \pm e^{\beta/2}$. The solution $d = e^{\beta/2}$ is the maximizer. Substituting $d = e^{\beta/2}$ into (3.11),

$$\max_{x,i} \|P_x(i, \cdot) - Q(i, \cdot)\|_{TV} \le \frac{e^\beta}{e^{\beta/2} + e^\beta} - \frac{1}{e^{\beta/2} + 1} = \frac{e^{\beta/2}}{e^{\beta/2} + 1} - \frac{1}{e^{\beta/2} + 1} = \frac{e^{\beta/2} - 1}{e^{\beta/2} + 1},$$

which completes the proof. $\qquad \square$

Proof of Lemma 3.13. First,

$$\|P_x(v, \cdot), P_y(v, \cdot)\|_{\text{TV}} = \frac{1}{2} \sum_{w \in \mathcal{N}(v) \cup \{v\}} |P_x(v, w) - P_y(v, w)|.$$

We will show that each term is upper bounded by $\frac{2\beta}{n}$. Since there are at most $\Delta + 1$ terms, the bound follows.

We compute $\max_{x,y:x \sim y} |P_x(v, w) - P_y(v, w)|$ for $w \in \mathcal{N}(v) \cup \{v\}$. Since x and y are interchangeable, we can drop the absolute value.

$$\max_{x,y:x \sim y} |P_x(v, w) - P_y(v, w)| = \max_{x,y:x \sim y} P_x(v, w) - P_y(v, w).$$

First consider the case that $v \neq w$. Then

$$\max_{x,y:x \sim y} P_x(v, w) - P_y(v, w) = \max_{x,y:x \sim y} \frac{\exp\left(\frac{\beta}{n} x(w)\right)}{\exp\left(\frac{\beta}{n}(x(v) - 1)\right) + \sum_{u \sim v} \exp\left(\frac{\beta}{n} x(u)\right)}$$

$$- \frac{\exp\left(\frac{\beta}{n} y(w)\right)}{\exp\left(\frac{\beta}{n}(y(v) - 1)\right) + \sum_{u \sim v} \exp\left(\frac{\beta}{n} y(u)\right)}.$$

Let

$$A(z) = \exp\left(\frac{\beta}{n}(z(v) - 1)\right) + \sum_{u \sim v} \exp\left(\frac{\beta}{n} z(u)\right).$$

Note that $A(y) \leq e^{\beta/n} A(x)$ for $x \sim y$. We have

$$\max_{x,y:x \sim y} P_x(v, w) - P_y(v, w) = \max_{x,y:x \sim y} \frac{\exp\left(\frac{\beta}{n} x(w)\right)}{A(x)} - \frac{\exp\left(\frac{\beta}{n} y(w)\right)}{A(y)}$$

$$\leq \max_{x,y:x \sim y} \frac{\exp\left(\frac{\beta}{n} x(w)\right)}{A(x)} - \frac{\exp\left(\frac{\beta}{n} y(w)\right)}{e^{\beta/n} A(x)}$$

$$= \max_{x,y:x \sim y} \frac{\exp\left(\frac{\beta}{n} x(w)\right) - \exp\left(\frac{\beta}{n}(y(w) - 1)\right)}{A(x)}$$

$$\leq \max_x \frac{\exp\left(\frac{\beta}{n} x(w)\right) - \exp\left(\frac{\beta}{n}(x(w) - 2)\right)}{A(x)}$$

$$= \max_x \frac{\exp\left(\frac{\beta}{n} x(w)\right)\left(1 - e^{-2\beta/n}\right)}{A(x)}$$

$$\leq 1 - e^{-2\beta/n}$$

$$\leq \frac{2\beta}{n}.$$

Next, we consider the case $v = w$. We have

$$\max_{x,y:x \sim y} P_x(v, v) - P_y(v, v) = \max_{x,y:x \sim y} \frac{\exp\left(\frac{\beta}{n}(x(v) - 1)\right)}{A(x)} - \frac{\exp\left(\frac{\beta}{n}(y(v) - 1)\right)}{A(y)}$$

$$\leq \max_{x,y:x \sim y} \frac{\exp\left(\frac{\beta}{n}(x(v) - 1)\right)}{A(x)} - \frac{\exp\left(\frac{\beta}{n}(y(v) - 1)\right)}{e^{\beta/n} A(x)}$$

$$= \max_{x,y:x \sim y} \frac{\exp\left(\frac{\beta}{n}(x(v) - 1)\right) - \exp\left(\frac{\beta}{n}(y(v) - 2)\right)}{A(x)}$$

$$\leq \max_x \frac{\exp\left(\frac{\beta}{n}(x(v)-1)\right) - \exp\left(\frac{\beta}{n}(x(v)-3)\right)}{A(x)}$$

$$= \max_x \frac{\exp\left(\frac{\beta}{n}(x(v)-1)\right)\left(1 - e^{-2\beta/n}\right)}{A(x)}$$

$$\leq 1 - e^{-2\beta/n}$$

$$\leq \frac{2\beta}{n}. \qquad\qquad \square$$

We now show that the approach for proving Theorem 2.2 based on the natural one-step path coupling does not yield the required contraction.

Theorem 3.16. *Let $\mathcal{H}$ have an edge between configurations x and $y = x - e_i + e_j$ whenever i and j are adjacent vertices in $\mathcal{G}$. Let $l(x,y) = 1$ for adjacent configurations. There exists a graph $\mathcal{G}$ such that for $\beta = 0$,*

$$W_\rho^P(x,y) \geq 1$$

for some adjacent configurations x, y.

Proof. Let $\mathcal{G}$ be the 4-vertex path graph. Label the vertices $1, 2, 3, 4$ in order along the path, and consider x and y related by $y = x - e_2 + e_3$ so that the two configurations differ by a transfer from one middle vertex to the other. When $\beta = 0$, the transition probabilities are simple: given that a particle is chosen at vertex v, it moves to vertex $w \in \mathcal{N}(v) \cup \{v\}$ with probability $\frac{1}{\deg(v)+1}$. The optimal coupling of $P(x,\cdot)$ and $P(y,\cdot)$ may be expressed as an optimal solution of a linear program, as follows. Write $x' \sim x$ if x' is adjacent to x in $\mathcal{H}$ or $x' = x$. For each $x' \sim x$ and $y' \sim y$, let $z(x',y')$ be the probability of the next states being x' and y' in a coupling. The constraints require the collection of z variables to be a valid coupling, and the objective function calculates the expected distance under the coupling.

$$\min \sum_{x' \sim x, y' \sim y} z(x',y')\rho(x',y')$$

$$\text{s.t.} \sum_{y' \sim y} z(x',y') = P(x,x') \qquad\qquad \forall x' \sim x$$

$$\sum_{x' \sim x} z(x',y') = P(y,y') \qquad\qquad \forall y' \sim y$$

$$z \geq 0$$

This linear program is known as a Kantorovich problem. Our goal is to show that the optimal objective value is at least 1. We will first write down the dual problem. By weak duality, any feasible solution to the dual problem gives a lower bound to the optimal value of the primal problem. Next we will construct a primal solution with objective value equal to 1, and apply the complimentary slackness condition to help us construct a dual solution whose objective value is also equal to 1. Finally we will conclude that the optimal solution to the primal problem is equal to 1, by strong duality. For a reference to linear programming duality, see e.g. Chapter 4 of [1].

First we take the dual of the linear program, introducing dual variables $u(x')$ for $x' \sim x$ and $v(y')$ for $y' \sim y$:

$$\max \sum_{x' \sim x} u(x')P(x,x') + \sum_{y' \sim y} v(y')P(y,y')$$

$$\text{s.t.} \ u(x') + v(y') \leq \rho(x',y') \qquad\qquad \forall x' \sim x, y' \sim y$$

This linear program is a Kantorovich dual problem. By weak duality, if there exists a dual solution with objective value Z, then the optimal solution of the primal is at least Z. Therefore our goal is to find a dual solution with objective value at least 1.

For $x' = x - e_a + e_b$ with $a, b \in \{1, 2, 3, 4\}$, $P(x, x') = \frac{x(a)}{n(\deg(a)+1)}$. Similarly, for $y' = y - e_a + e_b$, $P(y, y') = \frac{y(a)}{n(\deg(a)+1)}$. The value of ρ is given by

$$\rho(x', y') = \begin{cases} 0 & \text{if } [x' = y' = x] \text{ or } [x' = y' = y] \\ 1 & \text{if } [x' = y, y' \neq y] \text{ or } [y' = x, x' \neq x] \\ 1 & \text{if } [x' = x - e_a + e_b, y' = y - e_a + e_b, a \neq b] \\ 2 & \text{if } [x' = x, y' \notin \{x, y\}] \text{ or } [y' = y, x' \notin \{x, y\}] \\ 3 & \text{otherwise} \end{cases}.$$

There exists a primal solution with objective value 1: Set

$$z(x - e_a + e_b, y - e_a + e_b) = \min\left\{ P(x, x - e_a + e_b), P(y, y - e_a + e_b) \right\},$$

$$z(x - e_2 + e_1, y - e_3 + e_2) = \frac{1}{3n},$$

and

$$z(x - e_2 + e_3, y - e_3 + e_4) = \frac{1}{3n}.$$

Other values of $z(x', y')$ are set to zero. In other words, z describes a synchronous coupling according to the pairing in Figure 6, with particles moving in the same direction always. Now supposing this is an optimal solution, we apply complementary slackness to identify candidate dual optimal solutions. The complementary slackness condition states that if z and (u, v) are optimal primal and dual solutions, then it holds that for all $x' \sim x, y' \sim y$,

$$z(x', y') \left[\rho(x', y') - u(x') - v(y') \right] = 0.$$

If our primal solution z is optimal, then whenever $z(x', y') \neq 0$, we need $u(x') + v(y') = \rho(x', y')$. These additional constraints help us construct the following dual feasible solution:

$$u(x) = 1, u(x - e_1 + e_2) = 0, u(x - e_2 + e_1) = 2, u(x - e_2 + e_3) = 0,$$
$$u(x - e_3 + e_2) = 2, u(x - e_3 + e_4) = 0, u(x - e_4 + e_3) = 0$$
$$v(y) = 0, v(y - e_1 + e_2) = 1, v(y - e_2 + e_1) = -1, v(y - e_2 + e_3) = 1,$$
$$v(y - e_3 + e_2) = -1, v(y - e_3 + e_4) = 1, v(y - e_4 + e_3) = 1.$$

We find that the objective value of this solution is equal to 1. By strong duality, we conclude that the optimal value of the primal problem is equal to 1, and therefore there does not exist a contractive coupling. $\qquad\square$

Remark 3.17. The argument in the proof of Theorem 3.16 should apply to all graphs $\mathcal{G}$ that contain the a four-vertex path graph as a subgraph, and possibly to other graphs as well.

3.3 Bounding the Cheeger constant

The following results will be useful in proving Corollary 2.4.

Lemma 3.18 ([6]). *Let Φ_* be the Cheeger constant of an aperiodic and irreducible Markov chain, and let t_{mix} be its mixing time. Then*

$$t_{mix}\left(\frac{1}{4}\right) \geq \frac{1}{4\Phi_*}.$$

The following result follows directly from Equation 2.28 in [3].

Lemma 3.19. *Let P be the transition probability matrix of an aperiodic and irreducible Markov chain on state space $\mathcal{X}$, satisfying $P(x,x) \geq \frac{1}{2}$ for all $x \in \mathcal{X}$. Let π denote the stationary distribution, and let $\pi_{min} = \min_{x \in \mathcal{X}} \pi(x)$. Let Φ_* be the Cheeger constant of this chain, and let t_{mix} be its mixing time. Then*

$$t_{mix}(\epsilon) \leq \frac{2 \log\left(\frac{1}{2\epsilon\sqrt{\pi_{min}}}\right)}{\log\left(\frac{2}{2-\Phi_*^2}\right)}.$$

Proof of Corollary 2.4. We first prove the lower bound. Let $0 \leq \beta < \beta_-$. By Lemma 3.18 and Theorem 2.2, we have

$$\frac{1}{4\Phi_*} = O\left(n \log n\right) \implies \Phi_* = \frac{1}{O\left(n \log n\right)}.$$

We next prove the upper bound. Let $\overline{P} = \frac{1}{2}(P+I)$ denote the lazy version of the ARW chain. Note that π is also the stationary distribution of the lazy chain. Let $\overline{\Phi}$ denote its Cheeger constant. Observe that for all $S \subset \Omega$, it holds that $\Phi(S) = 2\overline{\Phi}(S)$. Therefore,

$$\Phi_* = \min_{S:\pi(S)\leq\frac{1}{2}} \Phi(S) = 2\min_{S:\pi(S)\leq\frac{1}{2}} \overline{\Phi}(S) = 2\overline{\Phi}_*,$$

and so it suffices to show $\overline{\Phi}_* = e^{-\Omega(n)}$.

Fix $\beta > \beta^+$. We claim that the mixing time of the lazy chain is $e^{\Omega(n)}$. To show this, we modify the proof of Theorem 2.1 as follows. We define $\overline{Z}$ to be the lazy version of the Z chain. That way, we can couple the $\overline{Z}$ chain to the chain advancing according to $\overline{P}$. Observe that the lower-bounding property still holds. Furthermore, the hitting time of the $\overline{Z}$ chain to the set $\{x \in \tilde{\Omega} : x(0) \leq (1-\delta)n\}$ is greater than the hitting time for the Z chain. Finally, since the stationary distributions of the $\overline{Z}$ and Z chains coincide, the rest of the proof follows identically.

We now apply Lemma 3.19 to the lazy chain. We need a lower bound on $\pi_{\min}$. Observe that if $t \in \mathbb{N}$ and $p \in [0,1]$ are such that $P^t(x,y) \geq p$ for all $x,y \in \Omega$, then $\pi_{\min} \geq p$. Set $t = n \cdot diam(\mathcal{G})$. There exists at least one sequence of possible transitions in t steps to get from x to y. Each step in the sequence has probability at least $\frac{1}{n(\Delta+e^\beta)}$. Therefore,

$$\pi_{\min} \geq \left(\frac{1}{n(\Delta+e^\beta)}\right)^{n\cdot diam(\mathcal{G})}.$$

By Lemma 3.19, we have

$$\frac{2\log\left(\frac{1}{2\epsilon\sqrt{\pi_{\min}}}\right)}{\log\left(\frac{2}{2-\overline{\Phi}_*^2}\right)} = e^{\Omega(n)}.$$

Since $\log(x) \geq 1 - \frac{1}{x}$ for $x > 0$, we have

$$\frac{2\log\left(\frac{1}{2\epsilon\sqrt{\pi_{\min}}}\right)}{1 - \frac{2-\overline{\Phi}_*^2}{2}} = e^{\Omega(n)}$$

$$\overline{\Phi}_*^2 = 4e^{-\Omega(n)}\log\left(\frac{1}{2\epsilon\sqrt{\pi_{\min}}}\right)$$

$$\overline{\Phi}_*^2 = e^{-\Omega(n)}\log\left(\frac{1}{\pi_{\min}}\right).$$

Substituting the lower bound for $\pi_{\min}$, we obtain

$$\overline{\Phi}_*^2 = e^{-\Omega(n)} \log\left(\left(2n(\Delta + e^\beta)\right)^{n \cdot diam(\mathcal{G})} \right) = n\log(n)e^{-\Omega(n)} = e^{-\Omega(n)},$$

which implies $\overline{\Phi}_* = e^{-\Omega(n)}$. □

4 Repelling random walks

Throughout our analysis, we have only considered $\beta \geq 0$. However, the case $\beta < 0$ ("Repelling Random Walks") is theoretically and practically interesting to study also. Simulations confirm the intuition that the particles behave like independent random walks when β is close to zero, and spread evenly when β is very negative (see Figure 7). We conjecture that there are not any hard-to-escape subsets of the state space for all $\beta < 0$.
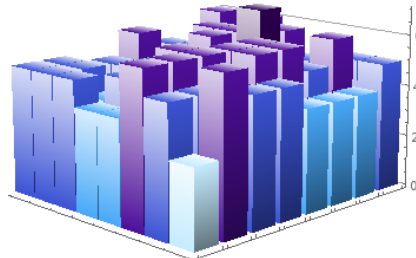

Figure 7: Simulation of the Attracting Random Walks model on an 8×8 grid graph after 10^6 steps for $n = 320$, $\beta = -500$.

Conjecture 4.1. *For all $\beta < 0$ and any graph, the mixing time of the ARW model is polynomial in n.*

We consider two cases: the extreme case of $\beta = -\infty$, and the case where $\mathcal{G}$ is the complete graph, for certain values of β.

4.1 The case $\beta = -\infty$

Theorem 4.2. *When $\beta = -\infty$, the mixing time of the Attracting Random Walks model is $O(n^2)$.*

Proof. When $\beta = -\infty$, the dynamics are simplified. Suppose a particle is chosen at vertex i. Let A be the set of vertices corresponding to the minimal value(s) of $\{x(i) - 1\} \cup \{x(j) : j \sim i\}$. The chosen particle moves to a vertex among those in A, uniformly at random.

Our goal is to show that the set

$$C \triangleq \left\{ x : x(v) \in \left\{ \left\lfloor \frac{n}{k} \right\rfloor, \left\lfloor \frac{n}{k} \right\rfloor + 1 \right\} \forall v \in V \right\} \cap \Omega$$

satisfies the following three properties: (1) It is absorbing, meaning that once the chain enters C, it cannot escape C; (2) The chain enters C in polynomial time; (3) Within C, the chain mixes in constant time with respect to n.

We claim that the maximum particle occupancy cannot increase, and the minimum particle occupancy cannot decrease. We now show that the maximum particle occupancy, $M_t \triangleq \max_v X_t(v)$, is monotonically non-increasing over time. Suppose that at time t, a particle at vertex i is selected and moves to vertex j. There are five cases:

1. $i = j$. The maximum does not change.
2. $i \neq j$, and both are maximizers. This case is not possible, since $x(j) > x(i) - 1$.
3. $i \neq j$, i is a maximizer, and j is not. The new maximum value is at most M_t, in the case that $X_t(j) = X_t(i) - 1$.
4. $i \neq j$, i is not a maximizer, and j is. This case is not possible, since $x(j) > x(i) - 1$.
5. $i \neq j$, i and j are not maximizers. The new maximum value is at most M_t, in the case that $X_t(j) = X_t(i) - 1$.

Therefore $M_{t+1} \leq M_t$. A similar argument shows that the minimum particle occupancy is monotonically non-decreasing over time. Together, they imply Property (1).

Next, we show Property (2). Assume $X_t \notin C$. Let $\mathcal{M}_t$ be the set of maximizing vertices at time t. We claim there exists at least one vertex $u \in \mathcal{M}_t$ such that there exists a path of distinct vertices $u = i_1 \sim i_2 \sim \cdots \sim i_p$ satisfying $x_{i_2} = x_{i_3} = \cdots = x_{i_{p-1}} = M_t - 1$ and $x_{i_p} \leq M_t - 2$ (allowing $p = 2$). In other words, there is a walkable path from $u = i_1$ to i_p. The maximum length of the path is $k - 1$. The probability that a particle is transferred along this path before any other events happen is therefore lower bounded by

$$\left(\frac{M_t}{n} \cdot \frac{1}{\Delta + 1} \right)^{k-1} \geq \left(\frac{1}{k} \cdot \frac{1}{\Delta + 1} \right)^{k-1}.$$

Therefore the probability that such a transfer happens within T_1 trials is at least

$$p \triangleq 1 - \left(1 - \left(\frac{1}{k} \cdot \frac{1}{\Delta + 1} \right)^{k-1} \right)^{T_1}.$$

If there had been at least two maximizing vertices to start, the number of maximizing vertices would have fallen by 1. If there had been only one maximizing vertex to start, the maximum value itself would have fallen by 1.

We see that there are two types of "good" events: reducing the number of maximizing vertices while the maximum value stays the same, or reducing the maximum value. We claim that the number of "good" events that happen before the chain enters the set C is upper bounded by n^2. Indeed, imagine that the particles at each vertex are stacked vertically. A particle movement from vertex i to vertex j is interpreted as a particle moving from the top of the stack at vertex i to the top of the stack at vertex j. Observe that the height of a particle cannot increase. Further, each particle's height can fall by at most $n - 1$ units over time, and can therefore drop at most $n - 1$ times. Since all good events require a particle's height to drop, the number of good events is at most $n(n-1) < n^2$. Let $T_2 = \lceil 2n^2 \frac{1}{p} \rceil$ be the number of trials of length T_1 each. Let N be the number of successes during the T_2 trials. By the Hoeffding inequality,

$$\mathbb{P} \left(|N - \mathbb{E}[N]| \geq n^2 \right) \leq 2 \exp \left(-\frac{2n^4}{T_2} \right) \leq 2 \exp \left(-\frac{2n^4}{2n^2 \frac{1}{p} + 1} \right).$$

Since $\mathbb{E}[N] = p \lceil 2n^2 \frac{1}{p} \rceil \geq 2n^2$,

$$\mathbb{P} \left(N \leq n^2 \right) \leq 2 \exp \left(-\frac{2n^4}{2n^2 \frac{1}{p} + 1} \right) \leq 2 \exp \left(-\frac{1}{2} p n^2 \right).$$

Therefore the probability that the chain is in C after $T_1 \times (k-1) \times T_2$ steps is at least $1 - 2\exp(-\frac{1}{2}pn^2)$. For an example, we can even set $T_1 = 1$. Then

$$p = \left(\frac{1}{k} \cdot \frac{1}{\Delta + 1} \right)^{k-1} \quad \text{and} \quad T_2 \leq 1 + 2n^2 \left(\frac{1}{k} \cdot \frac{1}{\Delta + 1} \right)^{1-k},$$

Therefore, within $O(n^2)$ steps, the chain is in C with high probability.

Finally, we show Property (3). Once the chain is in C, there are two types of vertices: those that have $\lfloor \frac{n}{k} \rfloor$ particles, and those that have $\lfloor \frac{n}{k} \rfloor + 1$ particles. Note that there are always $\tilde{k} \triangleq n - k\lfloor \frac{n}{k} \rfloor$ vertices with the higher number of particles. Therefore it is equivalent to study an exclusion process with just $\tilde{k}$ particles on the graph $\mathcal{G}$. With probability $\lfloor \frac{n}{k} \rfloor \cdot \frac{k-\tilde{k}}{n}$, an unoccupied vertex is selected, and the chain stays in place. With the remaining probability, an occupied vertex is chosen uniformly at random. Its particle then moves to a neigboring empty vertex or stays where it is, uniformly at random. Equivalently, the chain is lazy with probability $\lfloor \frac{n}{k} \rfloor \cdot \frac{k-\tilde{k}}{n}$, and otherwise one of the $\tilde{k}$ particles is chosen, and either stays or moves to a neighbor. Since the number of particles $\tilde{k}$ can be upper and lower bounded by constants ($0 \leq \tilde{k} \leq k$), the mixing time within C is independent of n. Therefore, we conclude that the overall mixing time is $O(n^2)$. $\qquad\square$

4.2 The complete graph case

Note that the complete graph case for $\beta < 0$ is equivalent to the vector of proportions chain in the antiferromagnetic Curie–Weiss Potts model.

Theorem 4.3. *On the complete graph with k vertices, the mixing time is $O\left(n \log n\right)$ for all β satisfying $-\frac{k}{10} < \beta \leq 0$.*

The proof relies on the following two lemmas.

Lemma 4.4. *Let $(X_t, t \geq 0)$ be the ARW chain for any $\beta < 0$ and let $(Y_t, t \geq 0)$ be a chain of independent particles ($\beta = 0$). Set $X_0 = Y_0$. For every vertex v and time t,*

$$\left| X_t(v) - \frac{n}{k} \right| \overset{\text{st}}{\leq} \left| Y_t(v) - \frac{n}{k} \right|.$$

For $\lambda \geq 0$, let $C(\lambda) \triangleq \left\{ x : \left| x(v) - \frac{n}{k} \right| \leq \lambda n \right\}$.

Lemma 4.5. *On the complete graph, if $y = x - e_i + e_j$ and $x, y \in C(\lambda)$, then*

$$\| P_x(v, \cdot) - P_y(v, \cdot) \|_{TV} \leq \frac{-5\beta/n}{2 + (k-2)e^{2\lambda\beta}}$$

for

$$n \geq \frac{-3\beta}{\log\left(\frac{5}{4}\right)}.$$

The proof of Lemma 4.4 appears later in this section, and the proof of Lemma 4.5 is deferred to the appendix due to its technical nature.

Proof of Theorem 4.3. We may assume that n is large enough so that

$$n \geq \frac{-3\beta}{\log\left(\frac{5}{4}\right)} \iff e^{-3\beta/n} \leq \frac{5}{4}.$$

Let $\{Y(v), v \in \mathcal{V}\}$ be a random variable distributed according to the stationary distribution of the $\{Y_t(v), v \in \mathcal{V}, t \geq 0\}$ chain at stationarity. At stationarity, the vertex occupancies are strongly concentrated around their means. By the Hoeffding Inequality, for every $\lambda > 0$,

$$\mathbb{P}\left(\left| Y(v) - \frac{n}{k} \right| \geq \lambda n \right) \leq 2e^{-2\lambda^2 n},$$

for every vertex v.

Fix $\epsilon > 0$. We wish to upper bound $t_{\text{mix}}(X, \epsilon)$. Note that the mixing time of the Y chain is $O(n \log n)$. To see this, consider a synchronous coupling. The expected amount of time

to select all the particles is $O(n \log n)$, and whenever a particle is selected, it moves to a uniformly random location, which is coupled. Now, for all ϵ', $T_1 \triangleq t_{\mathrm{mix}}(Y, \epsilon') = O(n \log n)$. Therefore at time T_1, for every $\lambda > 0$,

$$\mathbb{P}\left(\left|Y_{T_1}(v) - \frac{n}{k}\right| \geq \lambda n\right) \leq 2e^{-2\lambda^2 n} + \epsilon',$$

for every vertex v. By Lemma 4.4, it also holds that for every $\lambda > 0$,

$$\mathbb{P}\left(\left|X_{T_1}(v) - \frac{n}{k}\right| \geq \lambda n\right) \leq 2e^{-2\lambda^2 n} + \epsilon'$$

for every vertex v. Recall that $C(\lambda) = \left\{x : \left|x(v) - \frac{n}{k}\right| \leq \lambda n\right\}$. Then by the Union Bound,

$$\mathbb{P}\left(X_{T_1} \notin C(\lambda)\right) \leq k\left(2e^{-2\lambda^2 n} + \epsilon'\right),$$

for every λ and v. We observe that for n large enough, there is always an ϵ' small enough so that

$$k\left(2e^{-2\lambda^2 n} + \epsilon'\right) \leq \frac{\epsilon}{2}.$$

Then with probability at least $1 - \epsilon/2$, X_{T_1} belongs to $C(\lambda)$.

Next, we establish that for every $\beta < 0$, there exists λ_β such that (1) once the chain enters $C(\lambda_\beta)$, it takes exponential time to leave $C(2\lambda_\beta)$, with high probability; (2) we can applying path coupling within $C(2\lambda_\beta)$. The first claim is due to comparison with the $\beta = 0$ chain, as established above.

We now demonstrate the required contraction for path coupling within $C(2\lambda)$. Recall that we need to define the edges of the graph $\mathcal{H} = (\Omega, \mathcal{E}_\mathcal{H})$ and choose a length function on the edges. Let $(x, y) \in \mathcal{E}_\mathcal{H}$ if $y = x - e_i + e_j$ for some $i \neq j$, and let $l(x, y) = 1$. Consider any pair of neighboring configurations x and y. We employ a synchronous coupling, as in Figure 6. Namely, the "extra" particle at vertex i in configuration x is paired to the "extra" particle at vertex j in configuration y. All other particles are paired by vertex location. When a particle is selected to be moved in the x configuration, the particle that it is paired to in the y configuration is also selected to be moved.

With probability $\frac{n-1}{n}$, one of the $(n-1)$ pairs that has the same vertex location is chosen. Suppose it is located at vertex v. We couple the transitions in the two chains according to the coupling achieving the total variation distance $\|P_x(v, \cdot) - P_y(v, \cdot)\|_{\mathrm{TV}}$.

By Lemma 4.5, when one of the $(n-1)$ particles paired by vertex location is chosen, we can couple them so that they move to the same vertex with probability at least

$$1 - \frac{-5\beta/n}{2 + (k-2)e^{4\lambda\beta}}.$$

With the remaining probability, the distance increases by at most 2.

With the remaining $\frac{1}{n}$ probability, the "extra" particle is chosen in both chains. The chains can then equalize with probability 1 because $P_x(i, \cdot) = P_y(j, \cdot)$ on the complete graph. Therefore, we can bound the Wasserstein distance as follows:

$$W_\rho^P(x, y) \leq 1 + 2\frac{-5\beta/n}{2 + (k-2)e^{4\lambda\beta}} - \frac{1}{n} = 1 - \frac{1}{n}\left(1 + \frac{10\beta}{2 + (k-2)e^{4\lambda\beta}}\right)$$

Therefore, in order to achieve contraction, it suffices that

$$1 + \frac{10\beta}{2 + (k-2)e^{4\lambda\beta}} > 0 \iff -10\beta < 2 + (k-2)e^{4\lambda\beta} \tag{4.1}$$

Fix $0 < \delta < 1$, and let $\lambda_\beta = \frac{1}{4\beta} \log(1 - \delta) > 0$. Then substituting $\lambda = \lambda_\beta$, we obtain the condition

$$-10\beta < 2 + (k - 2)(1 - \delta) = k + \delta(2 - k). \tag{4.2}$$

When $\beta \le 0$ is such that $-10\beta < k$, there exists $\delta > 0$ small enough so that the condition (4.2) holds. We conclude that contraction holds for $-k/10 < \beta \le 0$.

To summarize the argument, we have shown that in time $O(n \log n)$, the chain enters $C(\lambda_\beta)$. After that, the chain leaves the larger set, $C(2\lambda_\beta)$, with exponentially small probability, which can be disregarded. Within $C(2\lambda_\beta)$, the Wasserstein distance with respect to the chosen $\mathcal{H}$ and ρ contracts by a factor of $\left(1 - \theta\left(\frac{1}{n}\right)\right)$, so an additional $O\left(n \log n\right)$ steps are sufficient. Therefore, the overall mixing time is $O\left(n \log n\right)$. $\qquad\square$

Proof of Lemma 4.4. We claim that there exists a coupling of $\{X_t, Y_t\}$ such that for all v and t, $\left|X_t(v) - \frac{n}{k}\right| \le \left|Y_t(v) - \frac{n}{k}\right|$. Let $\tilde{X}_t(v) = \left|X_t(v) - \frac{n}{k}\right|$ and define $\tilde{Y}_t(v)$ similarly. We claim that for all configurations x and vertices v, if $x(v) \ne \frac{n}{k}$, then

$$\mathbb{P}\left(\tilde{X}_{t+1}(v) = \tilde{X}_t(v) + 1 | X_t = x\right) \le \mathbb{P}\left(\tilde{Y}_{t+1}(v) = \tilde{Y}_t(v) + 1 | Y_t = x\right) \tag{4.3}$$

and

$$\mathbb{P}\left(\tilde{X}_{t+1}(v) = \tilde{X}_t(v) - 1 | X_t = x\right) \ge \mathbb{P}\left(\tilde{Y}_{t+1}(v) = \tilde{Y}_t(v) - 1 | Y_t = x\right). \tag{4.4}$$

If $x(v) = \frac{n}{k}$, then

$$\mathbb{P}\left(X_{t+1}(v) = \frac{n}{k} + 1 | X_t = x\right) \le \mathbb{P}\left(Y_{t+1}(v) = \frac{n}{k} + 1 | Y_t = x\right) \tag{4.5}$$

and

$$\mathbb{P}\left(X_{t+1}(v) = \frac{n}{k} - 1 | X_t = x\right) \le \mathbb{P}\left(Y_{t+1}(v) = \frac{n}{k} - 1 | Y_t = x\right). \tag{4.6}$$

In other words, the inequalities (4.3)–(4.6) state that the X chain is less likely to move in the absolute value–increasing direction, and more likely to move in the absolute value–decreasing direction. These inequalities, along with the fact that $X_0 = Y_0$, suffice to prove the lemma.

The transitions for the $Y_t(v)$ process are $+1$ with probability $\left(1 - \frac{Y_t(v)}{n}\right)\frac{1}{k}$, and -1 with probability $\frac{Y_t(v)}{n}\frac{k-1}{k}$. With the remaining probability, $Y_{t+1}(v) = Y_t(v)$. Suppose $x(v) \ne \frac{n}{k}$. There are two cases to analyze when $x(v) \ne \frac{n}{k}$:

1. $X_t(v) < \frac{n}{k}$. The probability that $X_{t+1}(v) = X_t(v) - 1$ is upper bounded by $\frac{X_t(v)}{n}\frac{k-1}{k}$, because vertex v is a more likely than average destination. In other words, it is harder to lose a particle from vertex v that has fewer than the average number of particles when $\beta < 0$, compared to when $\beta = 0$. Formally,

$$\frac{X_t(v)}{n}\left(1 - \frac{\exp\left(\frac{\beta}{n}(X_t(v) - 1)\right)}{\exp\left(\frac{\beta}{n}(X_t(v) - 1)\right) + \sum_{w \ne v} \exp\left(\frac{\beta}{n}X_t(w)\right)}\right) < \frac{X_t(v)}{n}\left(1 - \frac{1}{k}\right).$$

For the same reason, the probability that $X_{t+1}(v) = X_t(v) + 1$ is lower bounded by

$$\left(1 - \frac{X_t(v)}{n}\right)\frac{1}{k}.$$

Therefore, inequalities (4.3) and (4.4) hold in this case.

2. $X_t(v) > \frac{n}{k}$. This time, v is a less likely than average destination. The probability that $X_{t+1}(v) = X_t(v) - 1$ is lower bounded by

$$\frac{X_t(v)}{n}\frac{k-1}{k}.$$

The probability that $X_{t+1}(v) = X_t(v) + 1$ is upper bounded by

$$\left(1 - \frac{X_t(v)}{n}\right)\frac{1}{k}.$$

Therefore, inequalities (4.3) and (4.4) hold in this case also.

Finally, suppose $x(v) = \frac{n}{k}$. Then the probability of losing a particle is upper bounded by $\frac{1}{k}\frac{k-1}{k}$, and the probability of gaining a particle is upper bounded by $\frac{k-1}{k}\frac{1}{k}$. Therefore, inequalities (4.5) and (4.6) hold.

We conclude that such a coupling exists, and therefore the stochastic dominance holds. □

5 Conclusion

In this paper we have introduced a new interacting particle system model. We have shown that for any fixed graph, the mixing time of the Attracting Random Walks Markov chain exhibits phase transition. We have also partially investigated the Repelling Random Walks model, and we conjecture that model is always fast mixing. Beyond theoretical results, it is our hope that the model will find practical use.

A Appendix

Proof of Proposition 3.5. To compute the stationary probabilities $\lambda(r), r \in \{0, 1, \ldots, D\}$, note that we can disregard the initial uniform particle choice, and simply consider a Markov chain on a graph with $(D+1)$ nodes as in Figure 4 or 5.

When $D = 1$, we have $\lambda(0) = q\lambda(0) + q\lambda(1) \implies \lambda(0) = q$.

Next, consider $D = 2$. We have

$$\lambda(2) = (1-p)\lambda(2) + (1-q)\lambda(1) \implies \lambda(1) = \frac{p}{1-q}\lambda(2)$$

$$\lambda(0) = q\lambda(0) + q\lambda(1) \implies \lambda(0) = \frac{pq}{(1-q)^2}\lambda(2).$$

Since $\lambda(0) + \lambda(1) + \lambda(2) = 1$, we have

$$\left(\frac{pq}{(1-q)^2} + \frac{p}{1-q} + 1\right)\lambda(2) = 1,$$

and so

$$\lambda(0) = \frac{\frac{pq}{(1-q)^2}}{\frac{pq}{(1-q)^2} + \frac{p}{1-q} + 1} = \frac{q}{1 + \frac{(1-q)^2}{p}}.$$

Finally, consider the case $D \geq 3$. We solve the equations for the stationary distribution.

$$\lambda(D) = (1-p)\lambda(D) + (1-p)\lambda(D-1)$$

$$\implies \lambda(D-1) = \frac{p}{1-p}\lambda(D)$$

$$\lambda(D-1) = p\lambda(D) + (1-p)\lambda(D-2)$$

$$\implies \lambda(D-2) = \frac{1}{1-p}\left(\frac{p}{1-p}\lambda(D) - p\lambda(D)\right) = \left(\frac{p}{1-p}\right)^2 \lambda(D)$$

$$\cdots$$

$$\lambda(D-i) = \left(\frac{p}{1-p}\right)^i \lambda(D) \text{ for } 0 \le i \le D-2 \tag{A.1}$$

$$\lambda(2) = p\lambda(3) + (1-q)\lambda(1) \tag{A.2}$$

$$\lambda(0) = q\lambda(1) + q\lambda(0) \tag{A.3}$$

Using Equations (A.1)–(A.3),

$$\lambda(1) = \frac{1}{1-q}\left(\left(\frac{p}{1-p}\right)^{D-2}\lambda(D) - p\left(\frac{p}{1-p}\right)^{D-3}\lambda(D)\right)$$

$$= \frac{p}{1-q}\left(\frac{p}{1-p}\right)^{D-2}\lambda(D)$$

$$\lambda(0) = \frac{pq}{(1-q)^2}\left(\frac{p}{1-p}\right)^{D-2}\lambda(D). \tag{A.4}$$

Since $\sum_{i=0}^{D}\lambda(i) = 1$,

$$\lambda(D)\left(\frac{pq}{(1-q)^2}\left(\frac{p}{1-p}\right)^{D-2} + \frac{p}{1-q}\left(\frac{p}{1-p}\right)^{D-2} + \sum_{i=0}^{D-2}\left(\frac{p}{1-p}\right)^i\right) = 1$$

$$\lambda(D) = \frac{1}{\frac{p}{(1-q)^2}\left(\frac{p}{1-p}\right)^{D-2} + \left(\frac{1-\left(\frac{p}{1-p}\right)^{D-1}}{1-\frac{p}{1-p}}\right)}.$$

Substituting into (A.4)

$$\lambda(0) = \frac{q}{1 + \frac{(1-q)^2}{p}\left(\frac{p}{1-p}\right)^{2-D}\left(\frac{1-\left(\frac{p}{1-p}\right)^{D-1}}{1-\frac{p}{1-p}}\right)}. \qquad \Box$$

Proof of Lemma 4.5. Let

$$B(x) = \sum_{u \notin \{i,j,v\}} \exp\left(\frac{\beta}{n}x(u)\right) + \mathbb{1}_{v \notin \{i,j\}}\exp\left(\frac{\beta}{n}(x(v)-1)\right)$$

$$C(x) = \mathbb{1}_{v \ne i}\exp\left(\frac{\beta}{n}x(i)\right) + \mathbb{1}_{v=i}\exp\left(\frac{\beta}{n}(x(i)-1)\right) + \mathbb{1}_{v \ne j}\exp\left(\frac{\beta}{n}x(j)\right)$$

$$+ \mathbb{1}_{v=j}\exp\left(\frac{\beta}{n}(x(j)-1)\right), \text{ and}$$

$$D(x) = \mathbb{1}_{v \ne i}\exp\left(\frac{\beta}{n}(x(i)-1)\right) + \mathbb{1}_{v=i}\exp\left(\frac{\beta}{n}(x(i)-2)\right) + \mathbb{1}_{v \ne j}\exp\left(\frac{\beta}{n}(x(j)+1)\right)$$

$$+ \mathbb{1}_{v=j}\exp\left(\frac{\beta}{n}x(j)\right).$$

Then we can write

$$P_x(v,w) = \frac{\mathbb{1}_{v \ne w}\exp\left(\frac{\beta}{n}x(w)\right) + \mathbb{1}_{v=w}\exp\left(\frac{\beta}{n}(x(w)-1)\right)}{B(x) + C(x)}$$

and

$$P_y(v,w) = \frac{\mathbb{1}_{v \ne w}\exp\left(\frac{\beta}{n}y(w)\right) + \mathbb{1}_{v=w}\exp\left(\frac{\beta}{n}(y(w)-1)\right)}{B(x) + D(x)}.$$

To check the sign of $P_x(v, w) - P_y(v, w)$, it is equivalent to check the sign of

$$\exp\left(\frac{\beta}{n}x(w)\right)(B(x) + D(x)) - \exp\left(\frac{\beta}{n}y(w)\right)(B(x) + C(x)).$$

Next we show that for fixed v, the sign of $P_x(v, w) - P_x(v, w)$ is the same for all $w \notin \{i, j\}$. Suppose $w \notin \{i, j\}$. Then $\exp\left(\frac{\beta}{n}x(w)\right) = \exp\left(\frac{\beta}{n}y(w)\right)$, and is equivalent to check the sign of the expression $D(x) - C(x)$. Since this expression does not depend on w, we conclude that the sign is the same for all $w \notin \{i, j\}$.

If $P_x(v, w) - P_x(v, w) \geq 0$ for all $w \notin \{i, j\}$, then

$$\|P_x(v, \cdot) - P_y(v, \cdot)\|_{\text{TV}} = \max\{P_y(v, i) - P_x(v, i), 0\} + \max\{P_y(v, j) - P_x(v, j), 0\}.$$

Similarly, if $P_x(v, w) - P_x(v, w) < 0$ for all $w \notin \{i, j\}$, then

$$\|P_x(v, \cdot) - P_y(v, \cdot)\|_{\text{TV}} = \max\{P_x(v, i) - P_y(v, i), 0\} + \max\{P_x(v, j) - P_y(v, j), 0\}.$$

Therefore,

$$\|P_x(v, \cdot) - P_y(v, \cdot)\|_{\text{TV}} \leq |P_x(v, i) - P_y(v, i)| + |P_x(v, j) - P_y(v, j)|.$$

Consider the ratio of denominators of $P_x(v, w)$ and $P_y(v, w)$. We have

$$e^{\beta/n} \leq \frac{B(x) + C(x)}{B(x) + D(x)} \leq e^{-\beta/n}.$$

We first bound $|P_x(v, i) - P_y(v, i)|$. If $v \neq i$, we obtain

$$|P_x(v, i) - P_y(v, i)|$$
$$\leq \frac{1}{B(x) + C(x)}\max\left\{\left|\exp\left(\frac{\beta}{n}x(i)\right) - \exp\left(\frac{\beta}{n}(x(i) - 1)\right)e^{\beta/n}\right|,\right.$$
$$\left.\left|\exp\left(\frac{\beta}{n}x(i)\right) - \exp\left(\frac{\beta}{n}(x(i) - 1)\right)e^{-\beta/n}\right|\right\}$$
$$= \frac{\exp\left(\frac{\beta}{n}x(i)\right)}{B(x) + C(x)}\left(e^{-2\beta/n} - 1\right).$$

Similarly, if $v = i$, we obtain

$$|P_x(v, i) - P_y(v, i)| \leq e^{-\beta/n}\frac{\exp\left(\frac{\beta}{n}x(i)\right)}{B(x) + C(x)}\left(e^{-2\beta/n} - 1\right).$$

We similarly bound $|P_x(v, j) - P_y(v, j)|$. If $v \neq j$, we obtain

$$|P_x(v, j) - P_y(v, j)|$$
$$\leq \frac{1}{B(x) + C(x)}\max\left\{\left|\exp\left(\frac{\beta}{n}x(j)\right) - \exp\left(\frac{\beta}{n}(x(j) + 1)\right)e^{\beta/n}\right|,\right.$$
$$\left.\left|\exp\left(\frac{\beta}{n}x(j)\right) - \exp\left(\frac{\beta}{n}(x(j) + 1)\right)e^{-\beta/n}\right|\right\}$$
$$= \frac{\exp\left(\frac{\beta}{n}x(j)\right)}{B(x) + C(x)}\left(1 - e^{2\beta/n}\right).$$

If $v = j$, we obtain

$$|P_x(v, j) - P_y(v, j)| \leq e^{-\beta/n}\frac{\exp\left(\frac{\beta}{n}x(j)\right)}{B(x) + C(x)}\left(1 - e^{2\beta/n}\right).$$

For any choice of v,

$$C(x) \geq \exp\left(\frac{\beta}{n}x(i)\right) + \exp\left(\frac{\beta}{n}x(j)\right).$$

Therefore,

$$
\begin{aligned}
&|P_x(v,i) - P_y(v,i)| + |P_x(v,j) - P_y(v,j)| \\
&\leq \frac{e^{-\beta/n}}{B(x) + \exp\left(\frac{\beta}{n}x(i)\right) + \exp\left(\frac{\beta}{n}x(j)\right)} \left(\exp\left(\frac{\beta}{n}x(i)\right)\left(e^{-2\beta/n} - 1\right)\right. \\
&\hspace{6cm} \left. + \exp\left(\frac{\beta}{n}x(j)\right)\left(1 - e^{2\beta/n}\right)\right) \\
&= \frac{e^{-\beta/n}\left(1 - e^{2\beta/n}\right)}{B(x) + \exp\left(\frac{\beta}{n}x(i)\right) + \exp\left(\frac{\beta}{n}x(j)\right)} \left(\exp\left(\frac{\beta}{n}x(i)\right)e^{-2\beta/n} + \exp\left(\frac{\beta}{n}x(j)\right)\right) \\
&\leq \frac{e^{-3\beta/n}\left(1 - e^{2\beta/n}\right)}{B(x) + \exp\left(\frac{\beta}{n}x(i)\right) + \exp\left(\frac{\beta}{n}x(j)\right)} \left(\exp\left(\frac{\beta}{n}x(i)\right) + \exp\left(\frac{\beta}{n}x(j)\right)\right).
\end{aligned}
$$

Recall that $x \in C(\lambda)$. We upper bound by setting $x(i)$ and $x(j)$ to their lower bounds, and $x(u)$ to its upper bound for $u \notin \{i, j\}$.

$$
\begin{aligned}
&|P_x(v,i) - P_y(v,i)| + |P_x(v,j) - P_y(v,j)| \\
&\leq \frac{e^{-3\beta/n}\left(1 - e^{2\beta/n}\right)}{(k-2)\exp\left(\frac{\beta}{n}\left(\frac{n}{k} + \lambda n\right)\right) + 2\exp\left(\frac{\beta}{n}\left(\frac{n}{k} - \lambda n\right)\right)} 2\exp\left(\frac{\beta}{n}\left(\frac{n}{k} - \lambda n\right)\right) \\
&= \frac{2e^{-3\beta/n}\left(1 - e^{2\beta/n}\right)}{(k-2)e^{2\lambda\beta} + 2} \\
&\leq \frac{2e^{-3\beta/n}\left(-2\beta/n\right)}{(k-2)e^{2\lambda\beta} + 2} \\
&\leq \frac{-5\beta/n}{(k-2)e^{2\lambda\beta} + 2},
\end{aligned}
$$

where in the second-last inequality we have used the fact that $1 + z \leq e^z$ and the last inequality holds when $e^{-3\beta/n} \leq 5/4$. $\qquad\square$

References

[1] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*, Athena Scientific, Dynamic Ideas, 1997.

[2] P. Cuff, J. Ding, O. Louidor, E. Lubetzky, Y. Peres, and A. Sly, *Glauber dynamics for the mean-field Potts model*, Journal of Statistical Physics **149** (2012), 432–477. MR-2992796

[3] J. A. Fill, *Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the Exclusion Process*, The Annals of Applied Probability **1** (1991), no. 1, 62–87. MR-1097464

[4] Julia Gaudio, *Investigations in applied probability and high-dimensional statistics*, Ph.D. thesis, Massachusetts Institute of Technology, 2020.

[5] Thomas P. Hayes and Eric Vigoda, *Variable length path coupling*, Random Structures and Algorithms **31** (2007), no. 3, 251–272. MR-2352176

[6] David A. Levin and Yuval Peres, *Markov chains and mixing times*, 2 ed., American Mathematical Society, 2017. MR-3726904

[7] R. Bubley and M. E. Dyer, *Path coupling: A technique for proving rapid mixing in Markov chains*, 38th Annual Symposium on Foundations of Computer Science (1975).

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm