

Mixing Times of Markov Chains for Self-Organizing Lists and Biased Permutations *

Prateek Bhakta [†] Sarah Miracle [‡] Dana Randall [§] Amanda Pascoe Streib [¶]

Abstract

We study the mixing time of a Markov chain \mathcal{M}_{nn} on biased permutations, a problem arising in the context of self-organizing lists. In each step, \mathcal{M}_{nn} chooses two adjacent elements k , and ℓ and exchanges their positions with probability $p_{\ell,k}$. Here we define two general classes and give the first proofs that the chain is rapidly mixing for both. We also demonstrate that the chain is not always rapidly mixing.

Keywords: biased permutations, Markov chains, inversion tables, self-organizing lists, ASEP

1 Introduction

Sampling from the permutation group S_n is one of the most fundamental problems in probability theory. A natural Markov chain that has been studied extensively is a symmetric chain, \mathcal{M}_{nn} , that iteratively makes nearest neighbor transpositions on adjacent elements. We are given a set of input probabilities $\mathbf{P} = \{p_{i,j}\}$ for all $1 \leq i, j \leq n$ with $p_{i,j} = 1 - p_{j,i}$. At each step, the Markov chain \mathcal{M}_{nn} uniformly chooses a pair of adjacent elements, i and j , and puts i ahead of j with probability $p_{i,j}$, and j ahead of i with probability $p_{j,i} = 1 - p_{i,j}$.

The problem of biased permutations arises naturally from the Move-Ahead-One (MA1) list update algorithm and was considered by Fill [9, 10]. In the MA1 protocol, elements are chosen according to some underlying distribution and they move up by one in a linked list after each request is serviced, if possible. Thus, the most frequently requested elements will move toward the front of the list and will require less access time. If we consider a pair of adjacent elements i and j , the probability of performing a transposition that moves i ahead of j is proportional to i 's request frequency, and similarly the probability of moving j ahead of i is proportional to j 's frequency, so the transposition rates vary depending on i and j and we are always more likely to put things in order (of their request frequencies) than out of order. Fill asked for which $\mathbf{P} = \{p_{i,j}\}$ the chain is rapidly mixing.

Despite the simplicity of the model, the mixing times of only a few special cases are known. After a series of papers [8, 6], Wilson [20] showed that in the unbiased case when $p_{i,j} = 1/2$ for all i, j the mixing time is $\Theta(n^3 \log n)$, with upper and lower bounds within a factor of two.

*A preliminary version of this paper appeared in the *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp 1-15.

[†]College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0765; pbhakta@gatech.edu. Supported in part by NSF CCF-0830367 and a Georgia Institute of Technology ARC Fellowship.

[‡]Computer and Information Sciences, University of St. Thomas, Saint Paul, MN 55105; sarah.miracle@stthomas.edu.

[§]College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0765; randall@cc.gatech.edu. Supported in part by NSF CCF-0830367 and CCF-0910584.

[¶]Center for Computing Sciences, Bowie, MD 20715-4300; ampasco@super.org.

Subsequently Benjamini et al. [2] considered a constant bias version of this chain, where we are given a fixed parameter $0 \leq p \leq 1$ such that $p \neq 1/2$ and $p_{i,j} = p$ for all $i < j$ and $p_{i,j} = 1 - p$ for $i > j$. They relate this biased shuffling Markov chain to a chain on an asymmetric simple exclusion process (ASEP) and showed that they both converge in $\Theta(n^2)$ time. These bounds were matched by Greenberg et al. [11] who also generalized the result on ASEPs to sampling biased surfaces in two and higher dimensions in optimal $\Theta(n^d)$ time. Note that when the bias is a constant for all $i < j$ there are other methods for sampling from the stationary distribution, but studying the Markov chain \mathcal{M}_{nn} is of independent interest, partly because of the connection to ASEPs and other combinatorial structures. Finally, we also have polynomial bounds on the mixing time when each of the $p_{i,j}$ for $i < j$ is equal to $1/2$ or 1 ; in this case we are sampling linear extensions of a partial order over the set $\{1 \dots n\}$, and the chain \mathcal{M}_{nn} was shown by Bubley and Dyer [5] to mix in $O(n^3 \log n)$ time.

It is easy to see that \mathcal{M}_{nn} is not always rapidly mixing. Consider, for example, n elements $1 \dots n$ such that $p_{i,j} = 1$ for all $1 \leq i < j \leq n - 1$, $p_{n,i} = .9$ for $i \leq n/2$ and $p_{i,n} = .9$ for $i > n/2$. Then the first $n - 1$ elements will stay in order once they become ordered. All n places where the last element can be placed have nonzero stationary probability, but the configurations that have this last element at the beginning or end of the permutation will have exponentially larger stationary probability than the configuration that has this last element near the middle of the permutation. This defines an exponentially small cut in the state space and we can conclude that the nearest neighbor transposition chain must be slowly mixing for this choice of \mathbf{P} .

To avoid such situations, we restrict our attention to the *positively biased* setting where for all $i < j$, we have $1/2 \leq p_{i,j} \leq 1$. Note that any transposition that puts elements in the proper order has probability at least $1/2$, so starting at any permutation, we can always perform a series of transpositions to move to the ordered permutation $1, 2, \dots, n$ without ever decreasing the stationary probability. It is also worth noting that the classes for which the chain is known to mix rapidly are all positively biased. Fill [9, 10] conjectured that when \mathbf{P} is positively biased and also satisfies a monotonicity condition where $p_{i,j} \leq p_{i,j+1}$ and $p_{i,j} \geq p_{i+1,j}$ for all $1 \leq i < j \leq n$, then the chain is always rapidly mixing. In fact, he conjectured that the spectral gap is minimized when $p_{i,j} = 1/2$ for all i, j , a problem he refers to as the “gap problem.” Fill verified the conjecture for $n = 4$ and gave experimental evidence for slightly larger n .

In this paper, we make progress on the question of determining for which values of \mathbf{P} the chain \mathcal{M}_{nn} is rapidly mixing. First, we show that restricting \mathbf{P} to be positively biased is not sufficient to guarantee fast convergence to equilibrium. Our example uses a reduction to ASEPs and biased lattice paths. The construction is motivated by models in statistical physics that exhibit a phase transition arising from a “disordered phase” of high entropy and low energy, an “ordered phase” of high energy and low entropy, and a bad cut separating them that is both low energy and entropy. We note that this example does not satisfy the monotonicity condition of Fill, thus leaving his conjecture open, but does give insight into why bounding the mixing rate of the chain in more general settings has proven quite challenging.

In addition, we identify two new classes of input probabilities \mathbf{P} for which we can prove that the chain is rapidly mixing. It is important to note that these classes are not necessarily monotone. The first, which we refer to as “Choose Your Weapon,” we are given a set of input parameters $1/2 \leq r_1, \dots, r_{n-1} < 1$ representing each player’s ability to win a duel with his or her weapon of choice. When a pair of neighboring players are chosen to compete, the dominant player gets to choose the weapon, thus determining his or her probability of winning the match. In other words, we set $p_{i,j} = r_i$ when $i < j$. We show that the nearest neighbor transposition chain \mathcal{M}_{nn} is rapidly mixing for any choice of $\{r_i\}$. The second class, which we refer to as “League Hierarchies,” is defined by a binary tree with n leaves labeled $1, \dots, n$. We are given q_1, \dots, q_{n-1} with $1/2 \leq q_i < 1$

for all i , each associated with a distinct internal node in the tree. We then set $p_{i,j} = q_{i \wedge j}$ for all $i < j$. We imagine that the two subtrees under the root represent two different leagues, where each player from one league have a fixed advantage over each player from the other. Moreover, each league is subdivided into two sub-leagues, and each player from one has a fixed advantage over a player from the other, and so on recursively. We prove that there is a Markov chain based on transpositions (not necessarily nearest neighbors) that is always rapidly mixing for positively biased \mathbf{P} defined as League Hierarchies. Moreover, if the $\{q_i\}$ additionally satisfy “weak monotonicity” (i.e., $p_{i,j} \leq p_{i,j+1}$ if $j > i$) then the nearest neighbor chain \mathcal{M}_{nn} is also rapidly mixing. Note that both the choose-your-weapon and the tree-hierarchy classes are generalizations of the constant bias setting, which can be seen by taking all parameters r_i or q_i to be constant.

Our proofs rely on various combinatorial representations of permutations, including Inversion Tables and families of ASEPs. In each case there is a natural Markov chain based on (non necessarily adjacent) transpositions for which we can more easily bound the mixing time in the new context. We then interpret these new moves in terms of the original permutations in order to derive bounds on the mixing rate of the nearest neighbor transposition via comparison methods. These new chains that allow additional, but not necessarily all, transpositions are also interesting in the context of permutations and these related combinatorial families. Finally, we note that the choose-your-weapon class is actually a special case of the league-hierarchy class, but the proofs bounding the mixing rate of \mathcal{M}_{nn} are simpler and yield faster mixing times, so we present these proofs separately in Sections 4 and 5.

2 The Markov Chains \mathcal{M}_{nn} and \mathcal{M}_{tr}

We begin by formalizing the nearest neighbor and transposition Markov chains. Let $\Omega = S_n$ be the set of all permutations $\sigma = (\sigma(1), \dots, \sigma(n))$ of n integers. We consider Markov chains on Ω whose transitions transpose two elements of the permutation. Recall we are given a set \mathbf{P} , consisting of $p_{i,j} \in [0, 1]$ for each $1 \leq i \neq j \leq n$, where $p_{j,i} = 1 - p_{i,j}$. In this paper we only consider sets \mathbf{P} which are positively biased and bounded away from 1. Specifically, for any $i < j$, $1/2 \leq p_{i,j} < 1$. The Markov chain \mathcal{M}_{nn} will sample elements from Ω as follows.

The Nearest Neighbor Markov chain \mathcal{M}_{nn}

Starting at any permutation σ_0 , repeat:

- At time t , select index $i \in [n - 1]$ uniformly at random (u.a.r).
 - Exchange the elements $\sigma_t(i)$ and $\sigma_t(i + 1)$ with probability $p_{\sigma_t(i+1), \sigma_t(i)}$ to obtain σ_{t+1} .
 - With probability $p_{\sigma_t(i), \sigma_t(i+1)}$ do nothing so that $\sigma_{t+1} = \sigma_t$.

The chain \mathcal{M}_{nn} connects the state space, since every permutation σ can move to the ordered permutation $(1, 2, \dots, n)$ (and back) using the bubble sort algorithm. Since \mathcal{M}_{nn} is also aperiodic, this implies that \mathcal{M}_{nn} is ergodic. For an ergodic Markov chain with transition probabilities \mathcal{P} , if some assignment of probabilities π satisfies the *detailed balance condition* $\pi(\sigma)\mathcal{P}(\sigma, \tau) = \pi(\tau)\mathcal{P}(\tau, \sigma)$ for every $\sigma, \tau \in \Omega$, then π is the stationary distribution of the Markov chain [14]. It is easy to see that for \mathcal{M}_{nn} , the distribution

$$\pi(\sigma) = \left(\prod_{i < j: \sigma(i) < \sigma(j)} \frac{p_{i,j}}{p_{j,i}} \right) Z^{-1},$$

where Z is the normalizing constant $\sum_{\sigma \in \Omega} \pi(\sigma)$, satisfies detailed balance, and is thus the stationary distribution.

Next, we define the Markov chain \mathcal{M}_{tr} which can make any transposition at each step, while maintaining the stationary distribution π . The transition probabilities of \mathcal{M}_{tr} can be quite complicated, since swapping two distant elements in the permutation consists of many transitions of \mathcal{M}_{nn} , each with different probabilities.

The All-Transposition Markov chain \mathcal{M}_{tr}

Starting at any permutation σ_0 , repeat:

- At time t , select indices $i, j \in [n-1]$ u.a.r.
 - Swap the elements $\sigma_t(i), \sigma_t(j)$ with probability $p_{\sigma_t(j), \sigma_t(i)} \prod_{i < k < j} \frac{p_{\sigma_t(j,k)}}{p_{\sigma_t(i,k)}}$ to obtain σ_{t+1} .
 - Otherwise, do nothing so that $\sigma_{t+1} = \sigma_t$.

The transition probabilities of \mathcal{M}_{tr} are chosen so that the distribution π will satisfy detailed balance, and therefore be the stationary distribution of \mathcal{M}_{tr} . In the following sections, we will introduce two other Markov chains whose transitions are a subset of those of \mathcal{M}_{tr} , but for which we can describe the transition probabilities succinctly.

The relevant measure of the number of times we need to repeat steps of a Markov chain \mathcal{M} so that we are close (within total variation distance ϵ of stationarity) is the *mixing time* $\tau(\epsilon)$. The *total variation distance* between the stationary distribution π and the distribution of the Markov Chain at time t is

$$\|\mathcal{P}^t, \pi\|_{tv} = \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |\mathcal{P}^t(x, y) - \pi(y)|,$$

where $\mathcal{P}^t(x, y)$ is the t -step transition probability. For all $\epsilon > 0$, we define

$$\tau(\epsilon) = \min\{t : \|\mathcal{P}^t, \pi\|_{tv} \leq \epsilon, \forall t' \geq t\}.$$

We say that a Markov chain is *rapidly mixing* if the mixing time is bounded above by a polynomial in n and $\log(\epsilon^{-1})$ and *slowly mixing* if the mixing time is bounded below by an exponential in n , where n is the size of each configuration in Ω .

3 A Positively Biased \mathbf{P} that is Slowly Mixing

We begin by presenting an example that is positively biased yet takes exponential time to mix. In particular, we show that there are positively biased \mathbf{P} for which the chains \mathcal{M}_{nn} and even \mathcal{M}_{tr} require exponential time to converge to equilibrium. The key component used in the construction of these \mathbf{P} values is an example of slow mixing which was discovered by Pascoe and Randall [16] in the context of tile-based self-assembly models and is of independent interest in this setting.¹ We use a mapping from biased permutations to multiple particle ASEP configurations with n zeros and n ones. The resulting ASEPs are in bijection with staircase walks [11], which are sequences of

¹In [3] we give a simpler example also based on work by Pascoe and Randall in the context of tile-based self-assembly models. However, this example requires $p_{i,j} = 1$ for $i < j \leq n$ or $n < i < j$ (for permutations on $2n$ numbers), ensuring that once the elements $1, 2, \dots, n$ get in order, they stay in order (and similarly for the elements $n+1, n+2, \dots, 2n$). The example presented here, although more complex, allows all $p_{i,j}$ values to be bounded away from 1.

n ones and n zeros, that correspond to paths on the Cartesian lattice from $(0, n)$ to $(n, 0)$, where each 1 represents a step to the right and each 0 represents a step down (see Figure 1b). In [11], Greenberg et al. examined the Markov chain which attempts to swap a neighboring $(0, 1)$ pair, which equivalently adds or removes a unit square from the region below the walk, with probability depending on the position of that unit square. Each unit square (x, y) is assigned a bias $\lambda_{x,y}$ and the probability of each staircase walk w is proportional to the product of the bias assigned to each square below the walk. More formally, the stationary weight of a walk w is $\pi(w) = Z^{-1} \prod_{xy < w} \lambda_{x,y}$, where $xy < w$ whenever the square at (x, y) lies underneath the walk w and Z is the normalizing constant. For example, the walk w in Figure 2a has weight

$$\pi(w) = Z^{-1} \lambda_{1,5} \lambda_{1,6} \lambda_{2,6} \lambda_{3,6} \lambda_{1,7} \lambda_{2,7} \lambda_{3,7} \lambda_{4,7} \lambda_{1,8} \lambda_{2,8} \lambda_{3,8} \lambda_{4,8}$$

We show that there are settings of the $\{\lambda_{x,y}\}$ which cause the chain to be slowly mixing from any starting configuration (or walk). Consider the $\{\lambda_{x,y}\}$ given in Figure 1a where roughly all squares have bias $1/2$ except those in the upper right corner which have bias 1. If you start at a configuration that never touches the upper right corner (for example, the walk from $(0, n)$ to $(0, 0)$ to $(n, 0)$), then with at most polynomially many steps you can expect to stay within \sqrt{n} of the diagonal from $(0, n)$ to $(n, 0)$ and never discover the walks with high weight that go through the upper right corner. More specifically, with these bias values, we show that at stationarity the most likely configurations will be concentrated near the diagonal from $(0, n)$ to $(n, 0)$ (the high entropy, low energy states) or they will extend close to the point (n, n) (the high energy, low entropy states) but it will be unlikely to move between these sets of states because there is a bottleneck that has both low energy and low entropy. Finally, we give a map from biased permutations to biased lattice paths to produce a positively biased set of probabilities \mathbf{P} for which \mathcal{M}_{nn} also requires exponential time to mix.

Suppose, for ease of notation, that we are sampling permutations with $2n$ entries (having an odd number of elements will not cause qualitatively different behavior). We begin by setting $p_{i,j} = 1$ when $i < j \leq n$ or $n < i < j$, ensuring that once the elements $1, 2, \dots, n$ get in order, they stay in order (and similarly for the elements $n + 1, n + 2, \dots, 2n$). Since the smallest (largest) n elements of the biased permutation never change order once they get put in increasing order, permutations with these elements out of order have zero stationary probability. Hence, we can represent the smallest n numbers as ones and the largest n numbers as zeros, assuming that within each class the elements are in increasing order. Thus, in this setting, we have a bijection between staircase walks and permutations and it suffices to show a set of $\{\lambda_{x,y}\}$, staircase walk bias values, for which the Markov chain on staircase walks is slow. We do this first and then show how we can extend this to the more interesting case where we set $p_{i,j} = 1 - \beta$ when $i < j \leq n$ or $n < i < j$, and we have a more complex map between permutations and staircase walks.

3.1 Slow Mixing of Biased Staircase Walks

First we define a set of $\{\lambda_{x,y}\}$, staircase walk bias values, for which the Markov chain on staircase walks is slow. Let $M = 2n^{2/3}$, $0 < \delta < \frac{1}{2}$ be a constant, $\epsilon = \beta = 1/n^2$. We will define the $\{\lambda_{x,y}\}$ as follows (see Figure 1a):

$$\lambda_{x,y} = \begin{cases} 1 - \delta & \text{if } x + y > n + M; \\ \frac{1}{2} + \epsilon & \text{otherwise.} \end{cases} \quad (3.1)$$

We identify sets W_1, W_2, W_3 such that $\pi(W_2)$ is exponentially smaller than both $\pi(W_1)$ and $\pi(W_3)$, but to get between W_1 and W_3 , the Markov chain on staircase walks must pass through W_2 , the cut. We prove that for the set $\{\lambda_{x,y}\}$ defined above, the Markov chain on staircase walks has a

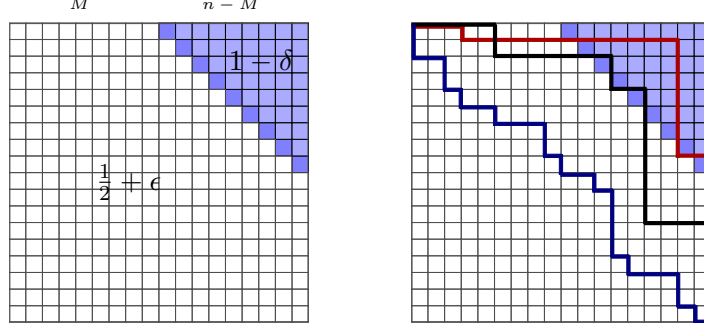


Figure 1: An example of fluctuating bias with exponential mixing time and staircase walks in S_1 , S_2 , and S_3 .

bad cut. Then we use the *conductance* to prove it is slowly mixing. For an ergodic Markov chain with distribution π , the conductance is defined as

$$\Phi = \min_{\substack{S \subseteq \Omega \\ \pi(S) \leq 1/2}} \sum_{s_1 \in S, s_2 \in \bar{S}} \pi(s_1) \mathcal{P}(s_1, s_2) / \pi(S).$$

We will show that the bad cut (S_1, S_2, S_3) implies that the conductance Φ is exponentially small. The following theorem relates the conductance and mixing time (see, e.g., [12, 18]).

Theorem 3.1: For any Markov chain with conductance Φ and mixing time $\tau(\epsilon)$, for all $\epsilon > 0$ we have

$$\tau(\epsilon) \geq \left(\frac{1}{4\Phi} - 1/2 \right) \log \left(\frac{1}{2\epsilon} \right)$$

.

For a staircase walk w , define the *height of w_i* as $\sum_{j \leq i} w_j$, and let $\max(w)$ be the maximum height of w_i over all $1 \leq i \leq 2n$. Let W_1 be the set of walks w such that $\max(w) < n + M$, W_2 the set of walks such that $\max(w) = n + M$, and W_3 the set of walks such that $\max(w) > n + M$. That is, W_1 is the set of walks that never reach the dark blue diagonal in Figure 1b, W_2 is the set whose maximum peak is on the dark blue line, and W_3 is the set which crosses that line and contains squares in the light blue triangle. Define $\gamma = (1/2 + \epsilon)/(1/2 - \epsilon)$, which is the ratio of two configurations that differ by swapping a $(0, 1)$ pair with probability $\frac{1}{2} + \epsilon$. First we notice that since the maximal staircase walk is in W_3 ,

$$\pi(W_3) \geq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}}$$

. Also, $\pi(W_1) = \frac{1}{Z} \sum_{w \in W_1} \gamma^{A(w)}$, where $A(w)$ is the number of unit squares below w . We have

that

$$\begin{aligned}
\pi(W_1) &= \frac{1}{Z} \sum_{w \in W_1} \gamma^{A(w)} \\
&\leq \frac{1}{Z} \sum_{w \in W_1} \gamma^{n^2 - \frac{(n-M)^2}{2}} \\
&\leq \frac{1}{Z} \binom{2n}{n} \gamma^{n^2 - \frac{(n-M)^2}{2}} \\
&\leq \frac{1}{Z} (2e)^n \gamma^{n^2 - \frac{(n-M)^2}{2}} \\
&\leq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}} \\
&\leq \pi(W_3)
\end{aligned}$$

for large enough n , since $1/\delta > 2$ is a constant. Hence $\pi(W_1) \leq \pi(W_3)$. We will show that $\pi(W_2)$ is exponentially small in comparison to $\pi(W_1)$ (and hence also to $\pi(W_3)$).

$$\pi(W_2) = \frac{1}{Z} \sum_{\sigma \in W_2} \gamma^{A(w)} \leq \frac{\gamma^{n^2} |W_2|}{Z}.$$

We bound $|W_2|$ as follows. The unbiased Markov chain is equivalent to a simple random walk $w_{2n} = X_1 + X_2 + \dots + X_{2n} = 0$, where $X_i \in \{+1, -1\}$ and where a $+1$ represents a step to the right and a -1 represents a step down. We call this random walk *tethered* since it is required to end at 0 after $2n$ steps. Compare walk w_{2n} with the untethered simple random walk $w'_{2n} = X'_1 + X'_2 + \dots + X'_{2n}$.

$$\begin{aligned}
P\left(\max_{1 \leq t \leq 2n} w_t \geq M\right) &= P\left(\max_{1 \leq t \leq 2n} w'_t \geq M \mid w'_{2n} = 0\right) \\
&= \frac{P(\max_{1 \leq t \leq 2n} w'_t \geq M)}{P(w'_{2n} = 0)} \\
&= \frac{2^{2n}}{\binom{2n}{n}} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \\
&\approx \sqrt{\pi n} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right).
\end{aligned}$$

Since the $\{X'_i\}$ are independent, we can use Chernoff bounds to see that

$$P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \leq 2nP(w'_{2n} \geq M) \leq 2ne^{-\frac{M^2}{2n}}.$$

Together these show that

$$P\left(\max_{1 \leq t \leq 2n} w_t \geq M\right) < e^{-n^{1/3}},$$

by definition of M . Therefore we have

$$\begin{aligned}
\pi(W_2) &\leq \frac{1}{Z} \gamma^{n^2} |W_2| \leq \frac{\gamma^{n^2}}{Z} \binom{2n}{n} e^{-n^{1/3}} \\
&\leq \frac{1}{Z} \binom{2n}{n} e^{-n^{1/3}+1} (1 - e^{-n^{1/3}}) \\
&\leq \frac{1}{Z} |W_1| e^{-n^{1/3}+1} \\
&\leq e^{-n^{1/3}+1} \pi(W_1),
\end{aligned}$$

as desired. Thus, $\pi(W_2)$ is exponentially smaller than $\pi(W_1)$ for every value of δ and the conductance satisfies

$$\begin{aligned}
\Phi &\leq \sum_{x \in W_1} \frac{\pi(x)}{\pi(W_1)} \sum_{y \in W_2} P(x, y) \\
&\leq \sum_{x \in S_1} \frac{\pi(x)}{\pi(W_1)} \pi(W_2) \\
&\leq e^{-n^{1/3}+1} \pi(W_1) \leq \frac{e^{-n^{1/3}+1}}{2}.
\end{aligned}$$

Using Theorem 3.1, we can use the bound on conductance to bound the mixing time.

3.2 Extending Slow Mixing of Staircase Walks to Biased Permutations

Next, we extend the above proof to the more general case where we set $p_{i,j} = 1 - \beta$ when $i < j \leq n$ or $n < i < j$, and we have a more complex map between permutations and staircase walks. Again, let $M = 2n^{2/3}$, $0 < \delta < \frac{1}{2}$ be a constant, $\epsilon = \beta = 1/n^2$. For $i < j \leq n$ or $n < i < j$, $p_{i,j} = 1 - \beta$, ensuring that the elements $1, 2, \dots, n$ are likely to be in order (and similarly for the elements $n + 1, n + 2, \dots, 2n$). The remaining $p_{i,j}$ values are defined as follows:

$$p_{i,j} = \begin{cases} 1 - \beta & i < j \leq n \text{ or } n < i < j; \\ 1 - \delta & \text{if } i + 2n - j + 1 \geq n + M; \\ \frac{1}{2} + \epsilon & \text{otherwise.} \end{cases} \quad (3.2)$$

We again identify sets S_1, S_2, S_3 such that $\pi(S_2)$ is exponentially smaller than both $\pi(S_1)$ and $\pi(S_3)$, but to get between S_1 and S_3 , \mathcal{M}_{nn} and \mathcal{M}_{tr} must pass through S_2 , the cut. In order to do this, we will define a map from permutations to staircase walks by representing the smallest n numbers as ones and the largest n numbers as zeros. More precisely, given a permutation σ , let $f(\sigma)$ be a sequence of ones and zeros, where $f(\sigma)_i = 1$ if $i \leq n$ and 0 otherwise. For example, the permutation $\sigma = (5, 1, 7, 8, 4, 3, 6, 2)$ maps to $f(\sigma) = (0, 1, 0, 0, 1, 1, 0, 1)$. If the first n and last n elements were always in order then, the probability that an adjacent 1 and a 0 swap in \mathcal{M}_{nn} depends on how many ones and zeros occur before that point in the permutation. Specifically, if element i is a 0 and element $i + 1$ is a 1 then we swap them with probability $\frac{1}{2} + \epsilon$ if the number of ones occurring before position x plus the number of zeros occurring after $i + 1$ is less than $n + M - 1$. Otherwise, they swap with probability $1 - \delta$. Equivalently, the probability of adding a unit square at position $v = (x, y)$ is $\frac{1}{2} + \epsilon$ if $x + y \leq n + M$, and $1 - \delta$ otherwise; see Figure 1b. We will show that in this case, the Markov chain is slow. The idea is that in the stationary distribution, there is a good chance that the ones and zeros will be well-mixed, since this is a high entropy situation.

However, the identity permutation also has high weight, and the parameters are chosen so that the entropy of the well-mixed permutations balances with the energy of the maximum (identity) permutation, and that to get between them is not very likely (low entropy and low energy). We prove that for the set \mathbf{P} defined above, \mathcal{M}_{nn} and \mathcal{M}_{tr} have a bad cut.

Given a staircase walk w , define σ_w to be the highest weight permutation σ such that $f(\sigma) = w$. Notice that σ_w is the permutation where elements $1, 2, \dots, n$ and elements $n + 1, n + 2, \dots, 2n$ are each in order (for example, $\sigma_{10110010} = (1, 5, 2, 3, 6, 7, 4, 8)$). First, we will show how the combined weight of all permutations that map to w relates to $\pi(\sigma_w)$. We will assign bias to the squares as follows. Each square (i, j) is given weight $\lambda_{i,j} = p_{i,j}/p_{j,i}$, where the squares are numbered as shown in Figure 2a. Thus the weight of σ_w satisfies

$$\pi(\sigma_w) = \frac{(\beta^{-1} - 1)^{n^2 - n} \prod_{xy < w} \lambda_{x,y}}{Z}.$$

The factor $(\beta^{-1} - 1)^{n^2 - n}$ comes from having the first and last n elements in order. For any staircase walk w , define $\Pi(w) = \sum_{\sigma: f(\sigma)=w} \pi(\sigma)$, the sum of the weights of the permutations which map to w . We will show that with our choice of β , we have $\pi(w)$ within a factor of 2 of the weight of σ_w .

Lemma 3.2: Given the set of probabilities \mathbf{P} defined in Equation 3.2, for all staircase walks w , the weight $\Pi(w)$ satisfies the following,

$$\pi(\sigma_w) < \Pi(w) < 2\pi(\sigma_w).$$

PROOF: First, notice that since $f(\sigma_w) = w$ and $\Pi(w) = \sum_{\sigma: f(\sigma)=w} \pi(\sigma)$, we trivially have that $\Pi(w) > \pi(\sigma_w)$. Given any permutation σ , let $h_1(\sigma)$ be the number of inversions between the first n numbers, specifically, pairs $(i, j) : 1 \leq i, j \leq n, i < j, \sigma_i > \sigma_j$. Similarly, let $h_2(\sigma)$ be the number of inversions in the second n numbers, specifically, pairs $(i, j) : n < i, j \leq 2n, i < j, \sigma_i > \sigma_j$. We will start by showing that

$$\frac{\pi(\sigma)}{\pi(\sigma_w)} \leq \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)} \right)^{h_1(\sigma) + h_2(\sigma)}. \quad (3.3)$$

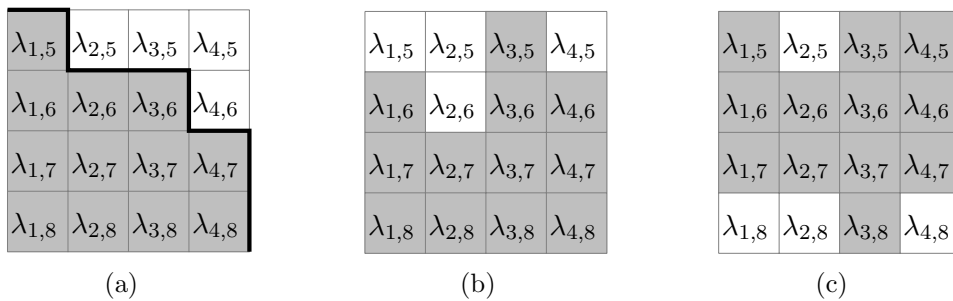


Figure 2: (a) Permutation 15236478 (b) Permutation 35416278 (c) Permutation 38415276

Given a walk w , consider any permutation $\sigma : f(\sigma) = w$ and let σ_1 be the sub-permutation corresponding to the first n integers. Similarly let σ_2 be the sub-permutation corresponding to the last n integers. We start by studying the effect of inversions within σ_1 on the weight of σ relative to the weight of σ_w . If σ_1 and σ_2 contain no inversions (i.e. $h_1(\sigma) = h_2(\sigma) = 0$) then $\pi(\sigma) = \pi(\sigma_w)$. First, assume $h_2(\sigma) = 0$, the effect of inversions between the first n elements is to reorder the columns of the walk. For example in Figure 2b, $\sigma_1(1) = 3$ so each inversion $(1, i)$ is replaced by a $(3, i)$ inversion and $\lambda_{1,i}$ is replaced with $\lambda_{3,i}$ in the weight. This can be visualized as moving

the shaded squares (those included in the weight) from column i to column $\sigma(i)$ (see Figure 2b). If $h_1(\sigma) = k, h_2(\sigma) = 0$ the weight is now the product of the bias of the shaded squares (after rearranging the columns as specified due to σ_1) times $(\beta^{-1} - 1)^{n^2 - n - k} / Z$. We want to determine if σ_1 has a inversions how much can this change the weight of the σ . Notice that if we shift a column by i since the boundary between the region where p_{ij} is $1 - \delta$ and where it is $1/2 + \epsilon$ is a diagonal (see Figure 1), we can increase the weight of the permutation by at most $\left(\frac{\delta^{-1} - 1}{\gamma}\right)^i$. Each column gets shifted by $\sigma(i) - i$. We are only interested in the case where $\sigma(i) - i > 0$, otherwise, the weight decreases. Let $I(i)$ be the number of inversions associated with i , specifically $j > i : \sigma(j) < \sigma(i)$. We will show that $\sigma(i) - i \leq I(i)$. To see this consider any permutation σ . Consider the permutations $\sigma' = \sigma(1)\sigma(2)\sigma(3)\dots\sigma(i-1)r_1, r_2, r_3, \dots$ where r_1, \dots, r_k are the remaining integers not included in $\sigma(1)\dots\sigma(i-1)$. Assume $\sigma(i) = r_j$ then $I(i) = j$ and since every number less than $\sigma(i)$ occurs before $\sigma(i)$ in σ' , it follows that $\sigma_i \leq i + I(i)$ implying that $\sigma(i) - i \leq I(i)$ as desired. If $h_2(\sigma) \neq 0$, we can use the exact same argument to bound the increase in weight due to inversion between the last n elements. These inversions correspond to switching rows of the staircase walk instead of columns. For example, see Figure 2c. Similarly, each row i moves a distance of $\sigma(i) - i$ so we can bound these in the exact same way. Combining these gives us the following,

$$\begin{aligned} \frac{\pi(\sigma)}{\pi(\sigma_w)} &\leq (\beta^{-1} - 1)^{-(h_1(\sigma) + h_2(\sigma))} \left(\frac{\delta^{-1} - 1}{\gamma}\right)^{\sum_{i:1 \leq i \leq n, \sigma(i) > i} \sigma(i) - i + \sum_{i:n < i \leq 2n, \sigma(i) > i} \sigma(i) - i} \\ &\leq (\beta^{-1} - 1)^{-(h_1(\sigma) + h_2(\sigma))} \left(\frac{\delta^{-1} - 1}{\gamma}\right)^{\sum_{i=0}^n I(i) + \sum_{i=n+1}^{2n} I(i)} \\ &\leq (\beta^{-1} - 1)^{-(h_1(\sigma) + h_2(\sigma))} \left(\frac{\delta^{-1} - 1}{\gamma}\right)^{h_1(\sigma) + h_2(\sigma)} \\ &\leq \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)}\right)^{h_1(\sigma) + h_2(\sigma)}. \end{aligned}$$

Next, notice that there are most n^{i+j} permutations $\sigma : f(\sigma) = w, h_1(\sigma) = i, h_2(\sigma) = j$. This is because we can think of this as first choosing a permutation of the first n elements with i inversions and then choosing a permutation of the next n elements with j inversions. The number of permutations of n elements with i inversions is upper bounded by n^i . This is straightforward to see in the context of the bijection with inversion tables discussed in Section 4.1. Combining this with Equation 3.3 gives the following:

$$\begin{aligned} \Pi(w) &= \sum_{i=0}^n \sum_{j=0}^n \sum_{\sigma: f(\sigma)=w, h_1(\sigma)=i, h_2(\sigma)=j} \pi(\sigma) \\ &\leq \sum_{i=0}^n \sum_{j=0}^n \sum_{\sigma: f(\sigma)=w, h_1(\sigma)=i, h_2(\sigma)=j} \pi(\sigma_w) \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)}\right)^{i+j} \\ &\leq \pi(\sigma_w) \sum_{i=0}^n \sum_{j=0}^n n^{i+j} \left(\frac{\delta^{-1} - 1}{\gamma(\beta^{-1} - 1)}\right)^{i+j} \\ &< 2\pi(\sigma_w) \quad \blacksquare \end{aligned}$$

We are now ready to prove the main theorem of the section.

Theorem 3.3: There exists a positively biased preference set \mathbf{P} for which the mixing time $\tau(\epsilon)$ of the Markov chain \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = \Omega\left(e^{n^{1/3}} \log(\epsilon^{-1})\right).$$

PROOF: As in the earlier example, for a staircase walk w , define the *height of w_i* as $\sum_{j \leq i} w_j$, and let $\max(w)$ be the maximum height of w_i over all $1 \leq i \leq 2n$. Again let W_1 be the set of walks w such that $\max(w) < n + M$, W_2 the set of walks such that $\max(w) = n + M$, and W_3 the set of walks such that $\max(w) > n + M$. Let S_1 be the set of permutations σ such that $f(\sigma) \in W_1$, S_2 the permutations such that $f(\sigma) \in W_2$ and S_3 the permutations such that $f(\sigma) \in W_3$. That is, W_1 is the set of walks that never reach the dark blue diagonal in Figure 1b, W_2 is the set whose maximum peak is on the dark blue line, and W_3 is the set which crosses that line and contains squares in the light blue triangle. Define $\gamma = (1/2 + \epsilon)/(1/2 - \epsilon)$, which is the ratio of two configurations that differ by swapping a $(0, 1)$ pair with probability $\frac{1}{2} + \epsilon$. First we notice that since the maximum weight permutation (which maps to the maximal tiling) is in S_3 ,

$$\pi(S_3) \geq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n}.$$

Using Lemma 3.2, $\pi(S_1) \leq \frac{2}{Z} \sum_{w \in W_1} \gamma^{A(w)} (\beta^{-1} - 1)^{n^2 - n}$, where $A(w)$ is the number of unit squares below w . We have that

$$\begin{aligned} \pi(S_1) &\leq \frac{2}{Z} \sum_{w \in W_1} \gamma^{A(w)} (\beta^{-1} - 1)^{n^2 - n} \\ &\leq \frac{2}{Z} \sum_{w \in W_1} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\ &\leq \frac{2}{Z} \binom{2n}{n} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\ &\leq \frac{1}{Z} (2e)^{n+1} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\ &\leq \frac{1}{Z} \gamma^{n^2 - \frac{(n-M)^2}{2}} (\delta^{-1} - 1)^{\frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n} \\ &\leq \pi(S_3) \end{aligned}$$

for large enough n , since $1/\delta > 2$ is a constant. Hence $\pi(S_1) \leq \pi(S_3)$. We will show that $\pi(S_2)$ is exponentially small in comparison to $\pi(S_1)$ (and hence also to $\pi(S_3)$).

$$\pi(S_2) \leq \frac{2}{Z} \sum_{\sigma \in W_2} \gamma^{A(w)} (\beta^{-1} - 1)^{n^2 - n} \leq \frac{2\gamma^{n^2} (\beta^{-1} - 1)^{n^2 - n} |W_2|}{Z}.$$

We bound $|W_2|$ in exactly the same way as before. The unbiased Markov chain is equivalent to a simple random walk $w_{2n} = X_1 + X_2 + \dots + X_{2n} = 0$, where $X_i \in \{+1, -1\}$ and where a $+1$ represents a step to the right and a -1 represents a step down. We call this random walk *tethered* since it is required to end at 0 after $2n$ steps. Compare walk w_{2n} with the untethered simple random

walk $w'_{2n} = X'_1 + X'_2 + \dots + X'_{2n}$.

$$\begin{aligned}
P\left(\max_{1 \leq t \leq 2n} w_t \geq M\right) &= P\left(\max_{1 \leq t \leq 2n} w'_t \geq M \mid w'_{2n} = 0\right) \\
&= \frac{P(\max_{1 \leq t \leq 2n} w'_t \geq M)}{P(w'_{2n} = 0)} \\
&= \frac{2^{2n}}{\binom{2n}{n}} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \\
&\approx \sqrt{\pi n} P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right).
\end{aligned}$$

Since the $\{X'_i\}$ are independent, we can use Chernoff bounds to see that

$$P\left(\max_{1 \leq t \leq 2n} w'_t \geq M\right) \leq 2nP(w'_{2n} \geq M) \leq 2ne^{-\frac{M^2}{2n}}.$$

Together these show that

$$P\left(\max_{1 \leq t \leq 2n} W_t \geq M\right) < e^{-n^{1/3}},$$

by definition of M . Therefore we have

$$\begin{aligned}
\pi(S_2) &\leq \frac{2}{Z} \gamma^{n^2} (\beta^{-1} - 1)^{n^2 - n} |W_2| \leq \frac{2\gamma^{n^2} (\beta^{-1} - 1)^{n^2 - n}}{Z} \binom{2n}{n} e^{-n^{1/3}} \\
&\leq \frac{(\beta^{-1} - 1)^{n^2 - n}}{Z} \binom{2n}{n} e^{-n^{1/3} + 1} (1 - e^{-n^{1/3}}) \\
&\leq \frac{(\beta^{-1} - 1)^{n^2 - n}}{Z} |S_1| e^{-n^{1/3} + 1} \\
&\leq e^{-n^{1/3} + 1} \pi(S_1),
\end{aligned}$$

as desired. Thus, $\pi(S_2)$ is exponentially smaller than $\pi(S_1)$ for every value of δ and the conductance satisfies

$$\begin{aligned}
\Phi &\leq \sum_{x \in S_1} \frac{\pi(x)}{\pi(S_1)} \sum_{y \in S_2} P(x, y) \\
&\leq \sum_{x \in S_1} \frac{\pi(x)}{\pi(S_1)} \pi(S_2) \\
&\leq e^{-n^{1/3} + 1} \pi(S_1) \leq \frac{e^{-n^{1/3} + 1}}{2}.
\end{aligned}$$

Hence, by Theorem 3.1, $\tau(\epsilon)$, the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) \geq \frac{1}{2} \left(e^{n^{1/3} - 1} - 1 \right) \log \left(\frac{1}{2\epsilon} \right). \quad \blacksquare$$

In fact, this proof can be extended to the more general Markov chain where we can swap any 1 with any 0, as long as we maintain the correct stationary distribution. This is easy to see, because any move that swaps a single 1 with a single 0 can only change the maximum height by at most

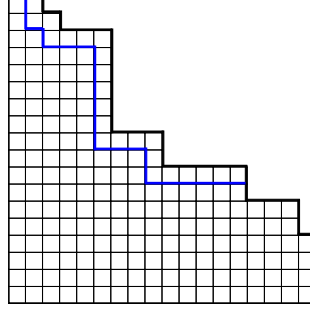


Figure 3: A move that swaps an arbitrary $(1,0)$ pair.

2 (see Figure 3). If we expand S_2 to include all configurations with maximum height $n + M$ or $n + M + 1$, $\pi(S_2)$ is still exponentially smaller than $\pi(S_1) \leq \pi(S_3)$. Hence the Markov chain that swaps an arbitrary $(1,0)$ pair still takes exponential time to converge.

Next, we show that there exists a value of δ for which $\pi(S_3) = \pi(S_1)$, which will imply that $\pi(S_2)$ is also exponentially smaller than $\pi(S_3)$, and hence the set S_2 forms a bad cut, regardless of which state the Markov chain begins in.

Lemma 3.4: There exist a constant δ , $\frac{1}{65} < \delta < \frac{1}{2}$, such that for this choice of δ , $\pi(S_3) = \pi(S_1)$.

PROOF: To find this value of δ , we will rely on the continuity of the function $f(\xi) = Z\pi(S_3) - Z\pi(S_1)$ with respect to $\xi = (1 - \delta)/\delta$. Let $a(\sigma)$ be the number of non-inversions in σ between i and j , $i < j$ such that $p_{ij} = 1/2 + \epsilon$ (for any highest weight configuration σ_w this corresponds to the number of tiles above the diagonal M in w) and let $b(\sigma)$ be the number of non-inversions in σ between i and j such that $p_{ij} = 1 - \delta$ (the number of tiles below the diagonal M). Notice that $Z\pi(S_1)$ is constant with respect to ξ and $Z\pi(S_3) = \sum_{\sigma \in S_3} \gamma^{b(\sigma)} \xi^{a(\sigma)} (\beta^{-1} - 1)^{n^2 - n - h_1(\sigma) - h_2(\sigma)}$ is just a polynomial in ξ . Therefore $Z\pi(S_3)$ is continuous in ξ and hence $f(\xi)$ is also continuous with respect to ξ . Moreover, when $\xi = \gamma$, clearly $Z\pi(S_3) < Z\pi(S_1)$, so $f(\gamma) < 0$. We will show that $f(4e^2) > 0$, and so by continuity we will conclude that there exists a value of ξ satisfying $\gamma < \xi < 4e^2$ for which $f(\xi) = 0$ and $Z\pi(S_3) = Z\pi(S_1)$. Clearly this implies that for this choice of ξ , $\pi(S_3) = \pi(S_1)$, as desired. To obtain the corresponding value of δ , we notice that $\delta = 1/(\xi + 1)$. In particular, δ is a constant satisfying $\frac{1}{65} < \delta < \frac{1}{2}$.

Thus it remains to show that $f(4e^2) > 0$. First we notice that since the maximal tiling is in S_3 , $\pi(S_3) \geq Z^{-1} \gamma^{n^2 - \frac{(n-M)^2}{2}} \xi^{\frac{(n-M)^2}{2}} (\beta^{-1} - 1)^{n^2 - n}$. Also,

$$\pi(S_1) = Z^{-1} \sum_{\sigma \in S_1} \gamma^{a(\sigma)} (\beta^{-1} - 1)^{n^2 - n - h_1(\sigma) - h_2(\sigma)} < Z^{-1} \binom{2n}{n} \gamma^{n^2 - \frac{(n-M)^2}{2}} 2(\beta^{-1} - 1)^{n^2 - n}.$$

Therefore

$$\pi(S_1)/\pi(S_3) < \frac{2 \binom{2n}{n}}{\xi^{\frac{(n-M)^2}{2}}} \leq (2e)^n \xi^{-n/2} = 1$$

since $\xi = 4e^2$. Hence $f(4e^2) = Z\pi(S_3) - Z\pi(S_1) > Z\pi(S_3) - Z\pi(S_3) = 0$, as desired. \blacksquare

Remark: In the setting of biased staircase walks, if the bias on each unit square (x, y) satisfies $\lambda_{x,y} \geq 2$, Pascoe and Randall give polynomial bounds on the mixing time of the Markov chain which adds or removes a unit square from under the walk [16]. Using a similar reduction to the one used in the proof of Theorem 3.3, in the biased permutations setting, these results can provide a class of positively biased \mathbf{P} for which there are $O(n^2)$ input parameters and \mathcal{M}_{nn} is rapidly mixing.

4 Choose Your Weapon

Despite the slow mixing example outlined in the previous section, there are many cases for which the chain will be rapidly mixing. We define two new classes for which we can rigorously demonstrate this and we provide the proofs in the next two sections.

For the first class, imagine a community of n people, each with a unique combative talent. Each member has his or her weapon of choice, and a competition with any other member of the community using this weapon affords that person a fixed advantage. When two people are chosen to compete, they each prefer using their own weapon of choice, so we resolve this by letting the person with the higher rank (e.g., age, seniority, etc.) choose the weapon they both will use. At any point in time our competitors are ordered and nearest neighbors are randomly selected to compete, where upon the winner is moved in front of the loser in the ordering.

To formalize the “Choose Your Weapon” scenario, we are given $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$ and the set \mathbf{P} satisfies $p_{i,j} = r_i$, if $i < j$ and $p_{i,j} = 1 - p_{j,i}$ if $j < i$. The moves of the nearest neighbor Markov chain \mathcal{M}_{nn} formalize the competitions, and our goal is to bound the mixing rate of this chain. Notice that this class includes the constant bias case studied by Benjamini et al. as a special case, and indeed our analysis yields an independent and simpler proof that the nearest neighbor Markov chain \mathcal{M}_{nn} is rapidly mixing in that context.

We shall show that the chain \mathcal{M}_{nn} is always rapidly mixing for probabilities \mathbf{P} defined in this way. Our proof relies on a bijection between permutations and *Inversion Tables* [13, 19] that, for each element i , record how many elements $j > i$ come before i in the permutation. We consider a Markov chain \mathcal{M}_{inv} that simply increments or decrements a single element of the inversion table in each step; using the bijection with permutations this corresponds to transpositions of elements that are not necessarily nearest neighbors to the Markov chain \mathcal{M}_{nn} . Remarkably, this allows \mathcal{M}_{inv} to decompose into a product of simple one-dimensional random walks and bounding the convergence time is very straightforward. Finally, we use comparison techniques [7, 17] to bound the mixing time of the nearest neighbor chain \mathcal{M}_{nn} .

4.1 The inversion table representation.

The Markov chain \mathcal{M}_{inv} acts on the *inversion table* for the permutation [13, 19], which has an entry for each $i \in [n]$ counting the number of inversions involving i ; that is, the number of values $j > i$ where j comes before i in the permutation (see Figure 4). It is easy to see that the i th element of the inversion table is an integer between 0 and $n - i$. In fact, the function I is a bijection between the set of permutations and the set \mathcal{I} of all possible inversion tables (all sequences $X = (x_1, x_2, \dots, x_n)$ where $0 \leq x_i \leq n - i$ for all $i \in [n]$). To see this, we will construct a permutation from any inversion table $X \in \mathcal{I}$. Place the element 1 in the $(x_1 + 1)$ st position of the permutation. Next, there are $n - 1$ slots remaining. Among these, place the element 2 in the $(x_2 + 1)$ st position remaining (ignoring the slot already filled by 1). Continuing, after placing $i - 1$ elements into the permutation, there are $n - i + 1$ slots remaining, and we place the element i into the $(x_i + 1)$ st position among the remaining slots. This proves that I is a bijection from S_n to \mathcal{I} .

Given this bijection, a natural algorithm for sampling permutations is the following local Markov chain on inversion tables: select a position $i \in [n]$ and attempt to either add one or subtract one

$$\begin{array}{rcl} \sigma & = & 8 \ 1 \ 5 \ 3 \ 7 \ 4 \ 6 \ 2 \\ I(\sigma) & = & 1 \ 7 \ 2 \ 3 \ 1 \ 2 \ 1 \ 0 \end{array}$$

Figure 4: The inversion table for a permutation.

from x_i , according to the appropriate probabilities. In terms of permutations, this amounts to adding or removing an inversion involving i without affecting the number of inversions involving any other integer, and is achieved by swapping the element i with an element $j > i$ such that every element in between is smaller than both i and j . If i moves ahead of j , this move happens with probability $p_{i,j}$ because for each k originally between i and j , $p_{k,i} = r_k = p_{k,j}$ (since $k < i$ and $k < j$), so the net effect of the move is neutral. The detailed balance condition ensures that π is the correct stationary distribution. Formally, the Markov chain \mathcal{M}_{nn} is defined as follows.

The Inversion Markov chain \mathcal{M}_{inv}

Starting at any permutation σ_0 , repeat:

- With probability $1/2$ let $\sigma_{t+1} = \sigma_t$.
- Otherwise, select $(i, b) \in [n] \times \{-1, +1\}$ u.a.r.
 - If $b = +1$ let j be the first element after i in σ_t such that $j > i$ (if such a j does not exist let $\sigma_{t+1} = \sigma_t$). With probability $p_{j,i}$, obtain σ_{t+1} from σ_t by swapping i and j .
 - If $b = -1$ let j be the last element before i in σ_t such that $j > i$ (if such a j does not exist let $\sigma_{t+1} = \sigma_t$). With probability $p_{i,j}$, obtain σ_{t+1} from σ_t by swapping i and j .

This Markov chain contains the moves of \mathcal{M}_{nn} (and therefore also connects the state space). Although elements can jump across several elements, it is still fairly local compared with the general transposition chain \mathcal{M}_{tr} which has $\binom{n}{2}$ choices at every step, since \mathcal{M}_{inv} has at most $2n$.

4.2 Rapid mixing of \mathcal{M}_{inv} .

The inversion Markov chain \mathcal{M}_{inv} can be viewed as a product of n independent processes. The i th process is a one-dimensional random walk bounded between 0 and $n - i$ that moves up by one with probability r_i and down by one with probability $1 - r_i$; its mixing time is $O(n^2 \log n)$, unless r_i is bounded away from $1/2$, in which case its mixing time is $O(n)$. We make moves in each chain with probability $1/n$, since we update one random walk at a time. The main tool we use for proving rapid mixing of \mathcal{M}_{inv} is coupling. A *coupling* is a stochastic process $(X_t, Y_t)_{t=0}^\infty$ on $\Omega \times \Omega$ with the properties:

1. Each of the processes X_t and Y_t is a faithful copy of \mathcal{M} (given initial states $X_0 = x$ and $Y_0 = y$).
2. If $X_t = Y_t$, then $X_{t+1} = Y_{t+1}$.

The coupling theorem bounds the mixing time in terms of the expected time of coalescence of any coupling. For initial states x, y let $T^{x,y} = \min\{t : X_t = Y_t | X_0 = x, Y_0 = y\}$, and define the *coupling time* to be $T = \max_{x,y} \mathbb{E}[T^{x,y}]$. The following result which relates the mixing time to the coupling time is well-known (see e.g. [1]).

Theorem 4.1: $\tau(\epsilon) \leq \lceil T \epsilon \ln \epsilon^{-1} \rceil$.

For the general case where the r_i 's are not bounded about from $1/2$, we will first bound the mixing time of each one-dimensional walk by using the following Lemma due to Luby, Randall and Sinclair [15] to bound the coupling time.

Lemma 4.2: Let d be an integer valued metric defined on $\Omega \times \Omega$ which takes values in $[0, B]$, and $d(x, y) = 0$ iff $x = y$. Let \mathcal{M} be a Markov chain on Ω and let (X_t, Y_t) be a coupling of \mathcal{M} , with $d_t = d(X_t, Y_t)$. Suppose the coupling satisfies $\mathbb{E}[\Delta d_t | X_t, Y_t] \leq 0$ and, whenever $d_t > 0$, $\mathbb{E}[(\Delta d_t)^2 | X_t, Y_t] \geq V$. Then the expected coupling time from initial states x, y satisfies

$$\mathbb{E}T^{x,y} \leq \frac{d_0(2B - d_0)}{V}.$$

Next, we will use the following theorem, which relates the mixing time of a product of independent Markov chains to the mixing time of each component to bound the mixing time of \mathcal{M}_{inv} . Similar results have been proved before in other settings (i.e., see [2, 4] and Corollary 12.12 of [14]). The proof is given in section 6.

Theorem 4.3: Suppose the Markov chain \mathcal{M} is a product of N independent Markov chains $\{\mathcal{M}_i\}$, where \mathcal{M} updates each \mathcal{M}_i with probability p_i . If $\tau_i(\epsilon)$ is the mixing time for \mathcal{M}_i and $\tau_i(\epsilon) \geq 4 \ln \epsilon$ for each i , then

$$\tau(\epsilon) \leq \max_{i=1,2,\dots,N} \frac{2}{p_i} \tau_i\left(\frac{\epsilon}{2N}\right).$$

Now we are ready to prove the following theorem, bounding the mixing time of \mathcal{M}_{inv} .

Theorem 4.4: Given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$, let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. The mixing time of \mathcal{M}_{inv} with preference set \mathbf{P} satisfies $\tau(\epsilon) = O(n^3 \log(n\epsilon^{-1}))$.

PROOF: To prove our theorem, we need to analyze the one-dimensional process $\mathcal{M}(r, k)$, bounded between 0 and k , which chooses to move up with probability $r/2 \geq 1/4$, down with probability $(1-r)/2$ and does nothing with probability $1/2$ at each step, if possible. This simple random walk is well-studied; we include the proof for completeness. We use Lemma 4.2. First, we define a natural distance metric $d(X_t, Y_t) = d_t$ on pairs X_t, Y_t of walks where d_t is the distance between the two walks at time t . We construct a coupling on the two lazy walks where with probability $r/2$ the walk X_t moves up (Y_t does nothing), with probability $(1-r)/2$ the walk X_t moves down (Y_t does nothing), similarly with probability $r/2$ the walk Y_t moves up (X_t does nothing) and with probability $(1-r)/2$ the walk Y_t moves down (X_t does nothing). The directions for Y_t and X_t are chosen independently. Once the walks collide, they make the exact same moves. Since the two walks never move at the same time, they will never jump over each other. Without loss of generality, assume that X_t is above Y_t then there are at most two cases where the distance is increase. Namely, if X_t moves up which happens with probability $r/2$ or if Y_t moves down which happens with probability $(1-r)/2$. The distance is decreased if X_t moves down or Y_t moves up which happens with probabilities $(1-r)/2$ and $r/2$ respectively. Thus, $\mathbb{E}[\Delta d_t | X_t, Y_t] \leq r/2 + (1-r)/2 - r/2 - (1-r)/2 = 0$. Similarly, whenever $d_t > 0$, $\mathbb{E}[(\Delta d_t)^2 | X_t, Y_t] \geq 1(r/2) + 1((1-r)/2) + 1((1-r)/2) \geq 1/2 = V$. Notice that $0 \leq d_0 \leq n = B$. Applying Lemma 4.2 with these choices of B, V and d_0 gives the following:

$$\mathbb{E}T^{x,y} \leq \frac{d_0(2B - d_0)}{V} \leq \frac{n(2n - 0)}{1/2} = 4n^2.$$

To bound the mixing time using our bound on the coupling time we apply Theorem 4.1 as follows:

$$\tau(\epsilon) \leq \lceil Te \ln \epsilon^{-1} \rceil = \lceil 4n^2 e \ln \epsilon^{-1} \rceil.$$

Next we use Theorem 4.3 to bound the mixing time of \mathcal{M}_{inv} as follows:

$$\tau(\epsilon) \leq \frac{2}{1/n} (\lceil 4n^2 e \ln \epsilon^{-1} \rceil) = O(n^3 \log(n\epsilon^{-1})). \quad \blacksquare$$

When each r_i is bounded away from $1/2$ and 1 , by using the path coupling theorem we obtain a stronger result. We use the following version due to Greenberg, Pascoe and Randall [11].

Theorem 4.5 (Path Coupling): Let $d : \Omega \times \Omega \rightarrow \mathbb{R}^+ \cup \{0\}$ be a metric that takes on finitely many values in $\{0\} \cup [1, B]$. Let U be a subset of $\Omega \times \Omega$ such that for all $(X_t, Y_t) \in \Omega \times \Omega$ there exists a path $X_t = Z_0, Z_1, \dots, Z_r = Y_t$ such that $(Z_i, Z_{i+1}) \in U$ for $0 \leq i < r$ and $\sum_{i=0}^{r-1} d(Z_i, Z_{i+1}) = d(X_t, Y_t)$. Let \mathcal{M} be a lazy Markov chain on Ω and let (X_t, Y_t) be a coupling of \mathcal{M} , with $d_t = d(X_t, Y_t)$. Suppose there exists a $\beta < 1$ such that, for all $(X_t, Y_t) \in U$, $\mathbb{E}[d_{t+1}] \leq \beta d_t$. Then, the mixing time satisfies

$$\tau(\epsilon) \leq \frac{\ln(B\epsilon^{-1})}{1 - \beta}.$$

Theorem 4.6: Given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$ and a positive constant c such that $1/2 + c < r_i < 1 - c$ for $1 \leq i \leq n - 1$, let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. The mixing time of \mathcal{M}_{inv} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^2 \ln(n\epsilon^{-1})).$$

PROOF: Here we use path coupling. We use the natural coupling on inversion tables where we choose the same element i for both $X = X_t = (x_1, \dots, x_n)$ and $Y = Y_t = (y_1, \dots, y_n)$ for each step t . A pair of inversion tables (X, Y) is in $U \subset I \times I$ if Y can be obtained from X by adding or subtracting 1 from a single x_i . We define a distance metric $d(X, Y)$ on pairs X, Y of inversion tables as a sum over the distances between the entries x_i and y_i . Let $\alpha_i = 1/(2(1 - r_i))$ and define

$$d(X, Y) = \sum_{i=1}^n \sum_{j=\min\{x_i, y_i\}}^{\max\{x_i, y_i\}-1} \alpha_i^j.$$

Let $(X, Y) \in U$ and suppose Y is obtained from X by adding 1 to x_i . Then as before, any move of \mathcal{M}_{inv} whose smaller index is not i succeeds or fails with the same probability in X and Y . There are two moves that decrease the distance: adding 1 to x_i , which happens with probability $(1 - r_i)/(4n)$, or subtracting 1 from y_i , which happens with probability $r_i/(4n)$. Both of these moves decrease the distance by $\alpha_i^{x_i}$. On the other hand, \mathcal{M}_{inv} proposes adding 1 to y_i with probability $(1 - r_i)/(4n)$, which increases the distance by $\alpha_i^{x_i+1}$, and \mathcal{M}_{inv} proposes subtracting 1 from x_i and succeeds with probability $r_i/(4n)$, increasing the distance by $\alpha_i^{x_i-1}$. Thus the expected change in distance is

$$\begin{aligned} & E[d(X_{t+1}, Y_{t+1}) - d(X_t, Y_t)] \\ &= \frac{1}{4n} \left(-\alpha_i^{x_i} + (1 - r_i)\alpha_i^{x_i+1} + r_i\alpha_i^{x_i-1} \right) \\ &= \frac{\alpha_i^{x_i-1}}{4n} \left(-\alpha_i + (1 - r_i)\alpha_i^2 + r_i \right) \\ &= \frac{\alpha_i^{x_i-1}}{4n} \left(r_i - \frac{1}{4(1 - r_i)} \right) \\ &= \frac{\alpha_i^{x_i}}{2n} \cdot \frac{-(2r_i - 1)^2}{4} \\ &= \frac{-d_t(2r_i - 1)^2}{8n}, \end{aligned}$$

since $d_t = \alpha_i^{x_i}$. Hence $E[d_{t+1}] \leq d_t(1 - (2r_i - 1)^2/(8n))$. Moreover, the maximum distance between any two inversion tables is

$$B = \sum_{i=1}^{n-1} 1 + \alpha_i + \dots + \alpha_i^{n-i} = \sum_{i=1}^{n-1} \frac{\alpha_i^n - 1}{\alpha_i - 1} = O(n\alpha_{\max}^n),$$

where $\alpha_{\max} = \max_i \{\alpha_i\} = \frac{1}{2(1-r_{\max})}$. Hence

$$\ln(B\epsilon^{-1}) = O(n \log(n\epsilon^{-1})),$$

so by Theorem 4.5, we have

$$\tau(\epsilon) = O(n^2 \log(n\epsilon^{-1})). \quad \blacksquare$$

Remark: The proofs of Theorem 4.4 and Theorem 4.6 also apply to the case where the probability of swapping i and j depends on the object with lower rank (i.e., we are given r_2, \dots, r_n and we let $p_{i,j} = r_j$ for all $i < j$). This case is related to a variant of the MA1 list update algorithm, where if a record is requested, we try to move the associated record x ahead of its immediate predecessor in the list, if it exists. If it has higher rank than its predecessor, then it always succeeds, while if its rank is lower we move it ahead with probability $f_x = r_x/(1 + r_x) \leq 1$.

4.3 Comparing \mathcal{M}_{inv} with \mathcal{M}_{nn} .

The comparison method can be used to infer the mixing time of one chain given the mixing time of another, similar chain. If P' and P are the transition matrices of two reversible Markov chains on the same state space Ω with the same stationary distribution π , the comparison method [7, 17] allows us to relate the mixing times of these two chains. Let $E(P) = \{(\sigma, \beta) : P(\sigma, \beta) > 0\}$ and $E(P') = \{(\sigma, \beta) : P'(\sigma, \beta) > 0\}$ denote the sets of edges of the two graphs, viewed as directed graphs. For each σ, β with $P'(\sigma, \beta) > 0$, define a path $\gamma_{\sigma\beta}$ using a sequence of states $\sigma = \sigma_0, \sigma_1, \dots, \sigma_k = \beta$ with $P(\sigma_i, \sigma_{i+1}) > 0$, and let $|\gamma_{\sigma\beta}|$ denote the length of the path. Let $\Gamma(v, \omega) = \{(\sigma, \beta) \in E(P') : (v, \omega) \in \gamma_{\sigma\beta}\}$ be the set of paths that use the transition (v, ω) of P . Finally, let $\pi_* = \min_{\rho \in \Omega} \pi(\rho)$ and define

$$A = \max_{(v, \omega) \in E(P)} \frac{1}{\pi(v)P(v, \omega)} \sum_{\Gamma(v, \omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta).$$

We use the following formulation of the comparison method [17].

Theorem 4.7: Given two Markov chains each with stationary distribution π , transition matrices P and P' and mixing times $\tau(\epsilon)$ and $\tau'(\epsilon)$, respectively. Define A and π_* as above, then for $0 < \epsilon < 1$, we have

$$\tau(\epsilon) \leq \frac{4 \log(1/(\epsilon\pi_*))}{\log(1/2\epsilon)} A \tau'(\epsilon).$$

First, we show that the two Markov chains \mathcal{M}_{inv} and \mathcal{M}_{nn} have the same stationary distribution. Then we will use Theorem 4.7 to infer a bound on the mixing time of \mathcal{M}_{nn} from the bounds on the mixing time of \mathcal{M}_{inv} shown in Theorem 4.4 and Theorem 4.6.

Theorem 4.8: Given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1$ and a positive constant c such that $r_i < 1 - c$ for $1 \leq i \leq n - 1$, let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$.

1. If $\forall 1 \leq i \leq n-1, c+1/2 < r_i$, the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

2. Otherwise, the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

PROOF: In order to apply Theorem 4.7, we need to define, for any transition $e = (\sigma, \beta)$ of the Markov chain \mathcal{M}_{inv} , a sequence of transitions of \mathcal{M}_{nn} . Let e be a transition of \mathcal{M}_{inv} which performs a transposition on elements $\sigma(i)$ and $\sigma(j)$, where $i < j$. Recall \mathcal{M}_{inv} can only swap $\sigma(i)$ and $\sigma(j)$ if all the elements between them are smaller than both $\sigma(i)$ and $\sigma(j)$. To obtain a sufficient bound on the congestion along each edge, we ensure that in each step of the path, we do not decrease the weight of the configuration. This is easy to do; in the first stage, move $\sigma(i)$ to the right, one step at a time, until it swaps with $\sigma(j)$. This removes an inversion of the type $(\sigma(i), \sigma(k))$ for every $i < k < j$, so clearly we have not decreased the weight of the configuration at any step. Next, move $\sigma(j)$ to the left, one step at a time, until it reaches position i . This completes the move e , and at each step, we are adding back an inversion of the type $(\sigma(j), \sigma(k))$ for some $i < k < j$. Since $\sigma(k) = \min\{\sigma(j), \sigma(k)\} = \min\{\sigma(i), \sigma(k)\}$, we have $p_{\sigma(i), \sigma(k)} = p_{\sigma(j), \sigma(k)}$ for every $i < k < j$, so in this stage we restore all the inversions destroyed in the first stage, for a net change of $p_{\sigma(i), \sigma(j)}$. See Figure 5.

Given a transition (v, ω) of \mathcal{M}_{nn} we must upper bound the number of canonical paths $\gamma_{\sigma\beta}$ that use this edge, which we do by bounding the amount of information needed in addition to (v, ω) to determine σ and β uniquely. For moves in the first stage, all we need to remember is $\sigma(i)$, because we know $\sigma(j)$ (it is the element moving forward). We also need to remember where $\sigma(j)$ came from. Given this information along with v and ω we can uniquely recover (σ, β) . Thus there are at most n^2 paths which use any edge (v, ω) . Also, notice that the maximum length of any path is $2n$.

Next we bound the quantity A which is needed to apply Theorem 4.7. Let $\lambda = \max_{i < j} p_{i,j}/p_{j,i}$. Recall that we have guaranteed that $\pi(\sigma) \leq \max\{\pi(v), \pi(\omega)\}$. Assume first that $\pi(\sigma) \leq \pi(v)$. Then

$$\begin{aligned} A &= \max_{(v, \omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v, \omega)} \sum_{\Gamma(v, \omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &\leq \max_{(v, \omega) \in E(P)} \sum_{\Gamma(v, \omega)} 2n \frac{P'(\sigma, \beta)}{P(v, \omega)} \\ &\leq \max_{(v, \omega) \in E(P)} \sum_{\Gamma(v, \omega)} 2n \frac{1/(2n)}{\frac{1}{(1+\lambda)(n-1)}} = O(n^3). \end{aligned}$$

5	2	3	7
2	5	3	7
2	3	5	7
2	3	7	5
2	7	3	5
7	2	3	5

Figure 5: The canonical path for transposing 5 and 7.

If, on the other hand, $\pi(\sigma) \leq \pi(\omega)$, then we use detailed balance to obtain:

$$\begin{aligned}
A &= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v,\omega)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\
&= \max_{(v,\omega) \in E(P)} \left\{ \frac{1}{\pi(\omega)P(\omega, v)} \sum_{\Gamma(v,\omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\
&\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{P'(\sigma, \beta)}{P(\omega, v)} \\
&\leq \max_{(v,\omega) \in E(P)} \sum_{\Gamma(v,\omega)} 2n \frac{1/(2n)}{(1+\lambda)(n-1)} = O(n^3).
\end{aligned}$$

In either case, we have $A = O(n^3)$. Then $\pi_* = \min_{\rho \in \Omega} \pi(\rho) \geq (\lambda \binom{n}{2} n!)^{-1}$ where λ is defined as above, so $\log(1/(\epsilon\pi_*)) = O(n^2 \log \epsilon^{-1})$, since λ is bounded from above by a positive constant. Appealing to Theorem 4.7 and Theorem 4.6, if the r_i 's are bounded away from $1/2$, we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

Similarly, appealing to Theorem 4.7 and Theorem 4.4 we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})). \quad \blacksquare$$

Remark: If the input probabilities are not bounded away from 1 by a constant but instead by some function of n , then using the same proof as Theorem 4.8 we can obtain a bound on the mixing time. Specifically, given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1 - 1/f(n)$ let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. Then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 f(n) \log(n\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

If additionally, there exists a positive constant c such that $r_i > 1/2 + c$ for all $1 \leq i \leq n-1$, then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 f(n) \log(nf(n)\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

Note that in this case it is necessary to also modify the proof of Theorem 4.6.

5 League Hierarchies

We now introduce a second general class of input probabilities \mathbf{P} for which we show \mathcal{M}_{nn} is always rapidly mixing. Imagine a sporting franchise consisting of an A-league with stronger players and a B-league with weaker players. We assume that any player from the A-league has a fixed advantage over any player from the B-league, representing his or her probability of winning in a matchup. Within each of these leagues we have tier-1 and tier-2 players, where again a player from the stronger tier has a fixed probability of winning a competition against a tier-2 player. Likewise for the tiers in the other league, but of course the fixed advantage there can be different. This

partition of each tier into stronger and weaker players continues recursively. To formalize the class of “League Hierarchies,” let T be a proper rooted binary tree with n leaf nodes, labeled $1, \dots, n$ in sorted order. Each non-leaf node v of this tree is labeled with a value $\frac{1}{2} \leq q_v < 1$. For $i, j \in [n]$, let $i \wedge j$ be the lowest common ancestor of the leaves labeled i and j . We say that \mathbf{P} has *league structure* T if $p_{i,j} = q_{i \wedge j}$. For example, Figure 6a shows a set \mathbf{P} such that $p_{1,4} = .8$, $p_{4,9} = .9$, and $p_{5,8} = .7$. We define matches by pairing up adjacent players in the current ranking and then we promote the winners, thus simulating \mathcal{M}_{nn} .

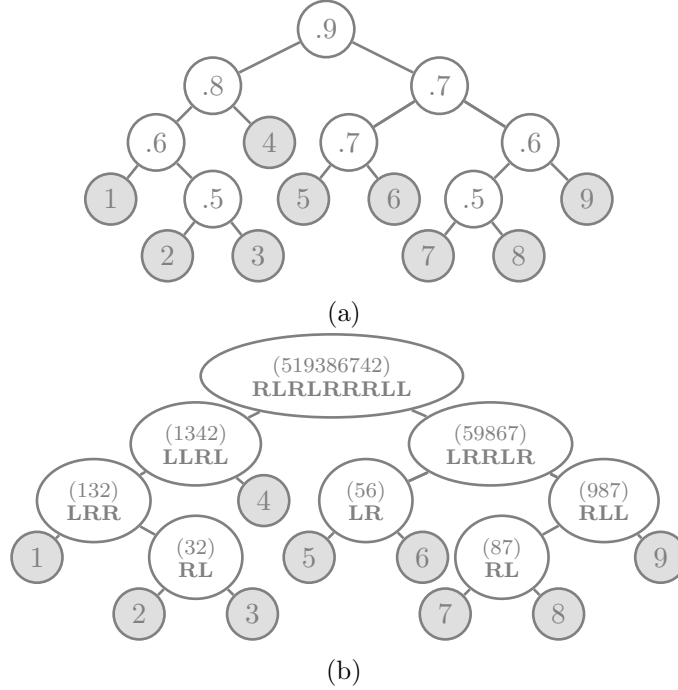


Figure 6: A set \mathbf{P} with tree structure, and the corresponding tree-encoding of the permutation 519386742.

To show \mathcal{M}_{nn} is rapidly mixing for any input probabilities in the League Hierarchy class, we introduce a new combinatorial representation of each permutation that will be useful for the proofs. This representation associates a bit string b_v to each node v of a binary tree with n leaves. Specifically, $b_v \in \{L, R\}^{\ell_v}$ where ℓ_v is the number of leaves in t_v , the subtree rooted at v , and for each element i of the sub-permutation corresponding to the leaves of t_v , $b_v(i)$ records whether i lies under the left or the right branch of v (see Figure 6b). The set of these bit strings is in bijection with the permutations. We consider a chain $\mathcal{M}_{tree}(T)$ that allows transpositions when they correspond to a nearest neighbor transposition in exactly one of the bit strings. Thus, the mixing time of $\mathcal{M}_{tree}(T)$ decomposes into a product of $n - 1$ ASEP chains and we can conclude that the chain $\mathcal{M}_{tree}(T)$ is rapidly mixing using results in the constant bias case [2, 11]. Again, we use comparison techniques to conclude that \mathcal{M}_{nn} is also rapidly mixing when we have weak monotonicity, although we show that $\mathcal{M}_{tree}(T)$ is always rapidly mixing.

5.1 The Markov chain $\mathcal{M}_{tree}(T)$.

We define the Markov chain $\mathcal{M}_{tree}(T)$ over permutations, given set \mathbf{P} with league structure T .

The Markov chain $\mathcal{M}_{tree}(T)$

Starting at any permutation σ_0 , repeat:

- Select distinct $a, b \in [n]$ with $a < b$ u.a.r.
- If every number between a and b in the permutation σ_t is not a descendant in T of $a \wedge b$, obtain σ_{t+1} from σ_t by placing a, b in order with probability $p_{a,b}$, and out of order with probability $1-p_{a,b}$, leaving all elements between them fixed.
- Otherwise, $\sigma_{t+1} = \sigma_t$.

First, we show that this Markov chain samples from the same distribution as \mathcal{M}_{nn} .

Lemma 5.1: The Markov chain $\mathcal{M}_{tree}(T)$ has the same stationary distribution as \mathcal{M}_{nn} .

PROOF: Let π be the stationary distribution of \mathcal{M}_{nn} , and let (σ_1, σ_2) be any transition in $\mathcal{M}_{tree}(T)$ such that $P_{tree}(\sigma_1, \sigma_2) > 0$ where P_{tree} is the transition matrix of $\mathcal{M}_{tree}(T)$. It suffices to show that the detailed balance condition holds for this transition with the stationary distribution π . Specifically we will show that

$$\pi(\sigma_1)P_{tree}(\sigma_1, \sigma_2) = \pi(\sigma_2)P_{tree}(\sigma_2, \sigma_1).$$

Recall that we may express $\pi(\sigma) = \left(\prod_{i < j: \sigma(i) < \sigma(j)} \frac{p_{i,j}}{p_{j,i}} \right) Z^{-1}$ where $Z = \sum_{\sigma \in \Omega} \pi(\sigma)$. The transition (σ_1, σ_2) transposes some two elements $a <_{\sigma_1} b$, where every element between a and b in σ_1 (note that these are the same as the elements between a and b in σ_2) is not a descendant of $a \wedge b$ in T . Although swapping arbitrary non-adjacent elements could potentially change the weight of the permutation dramatically, for any element c that is not a descendant in T of $a \wedge b$ the relationship between a and c is the same as the relationship between b and c . Thus, the league structure ensure that swapping a and b only changes the weight by a multiplicative factor of $p_{a,b}/p_{b,a}$. Let $X = \{x_1, \dots, x_k\}$ be the elements between a and b in σ_1 . Thus, the path from a or b to x_i in T must pass through $a \wedge b$ and go to another part of the tree. For every such element x_i , $a \wedge x_i = (a \wedge b) \wedge x_i = b \wedge x_i$.

From the observation, we see from the league structure that $p_{a,x_i} = p_{b,x_i}$ for every x_i between a and b . Also, we see that either both $a < x_i, b < x_i$ or $a > x_i, b > x_i$, since all numbers between a, b are necessarily descendants of $a \wedge b$. Define $S = \{x \in X : x < a, b\}$ and $B = \{x \in X : x > a, b\}$.

Therefore,

$$\frac{\pi(\sigma_1)}{\pi(\sigma_2)} = \frac{p_{a,b} \prod_{x \in S} (p_{x,b}/p_{b,x}) \prod_{x \in B} (p_{a,x}/p_{x,a})}{p_{b,a} \prod_{x \in S} (p_{x,a}/p_{a,x}) \prod_{x \in B} (p_{b,x}/p_{x,b})} = \frac{p_{a,b}}{p_{b,a}}.$$

This is exactly the ratio of the transition probabilities in $\mathcal{M}_{tree}(T)$

$$\frac{\pi(\sigma_1)}{\pi(\sigma_2)} = \frac{p_{a,b}}{p_{b,a}} = \frac{P_{tree}(\sigma_2, \sigma_1)}{P_{tree}(\sigma_1, \sigma_2)},$$

thus the detailed balance condition is satisfied and $\mathcal{M}_{tree}(T)$ also has stationary distribution π . ■

The key to the proof that $\mathcal{M}_{tree}(T)$ is rapidly mixing is again to decompose the chain into $n - 1$ independent Markov chains, $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_{n-1}$, one for each non-leaf node of the tree T . We introduce an alternate representation of a permutation as a set of binary strings arranged like the tree T . We use the characters L and R for our binary representation instead of 0 and 1 for convenience. For each non-leaf node v in the tree T , let $L(v)$ be its left descendants, and $R(v)$ be its right descendants. We now do the following:

- For each non-leaf node v do the following:

- List each descendant x of v in the order we encounter them in the permutation σ . These are parenthesized in Figure 6b.
- For each listed element x , write a L if $x \in L(v)$ and a R if $x \in R(v)$. This is the final encoding in Figure 6b.

We see that any σ will lead to an assignment of binary strings at each non-leaf node v with $L(v)$ L 's and $R(v)$ R 's. This is a bijection between the set of permutations and the set of assignments of binary strings to the tree T . Given any such assignment of binary strings, we can recursively reconstruct the permutation σ as follows:

- For each leaf node i , let its string be the string “ i ”.
- For any node n with binary string b ,
 - Determine the strings of its two children. Call these s_L, s_R .
 - Interleave the elements of s_L with s_R , choosing an element of s_L for each L in b , and an element of s_R for each R .

With this bijection, we first analyze $\mathcal{M}_{tree}(T)$'s behavior over tree representations and later extend this analysis to permutations. The Markov chain $\mathcal{M}_{tree}(T)$, when proposing a swap of the elements a and b , will only attempt to swap them if a, b correspond to some *adjacent* L and R in the string associated with $a \wedge b$. Swapping a and b does not affect any other string, so each non-leaf node v represents an independent exclusion process with $L(v)$ L 's and $R(v)$ R 's. These exclusion processes have been well-studied [5, 20, 2, 11]. We use the following bounds on the mixing times of the symmetric and asymmetric simple exclusion processes.

Theorem 5.2: Let \mathcal{M} be the exclusion process with parameter $1/2 \leq p < 1$ on a binary string of length k .

1. If $p > 1/2 + c$ for some positive constant c , then $\tau(\epsilon) = O(k^2 \log(\epsilon^{-1}))$. [11]
2. Otherwise, $\tau(\epsilon) = O(k^3 \log(k/\epsilon))$. [5, 11, 20]

The bounds in Theorem 5.2 refer to the exclusion process which selects a position at random and swaps the two elements in that position with the appropriate probability.

Since each exclusion process \mathcal{M}_i operates independently, the overall mixing time will be roughly n times the mixing time of each piece, slowed down by the inverse probability of selecting that process. Each \mathcal{M}_i has a different size, and a different mixing time relative to its size. To employ the bounds from Theorem 5.2, we will use Theorem 4.3, which relates the mixing time of a product of independent Markov chains to the mixing time of each component. The proof of Theorem 4.3 is given in section 6. Finally, we can prove that $\mathcal{M}_{tree}(T)$ is rapidly mixing.

Theorem 5.3: Given input probabilities $1/2 \leq q_1, q_2, \dots, q_{n-1} < 1$ let $\mathbf{P} = \{p_{i,j} = q_{i \wedge j}\}$.

1. If $\forall 1 \leq i \leq n-1, c + 1/2 < q_i$ for some positive constant c , the mixing time of $\mathcal{M}_{tree}(T)$ with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^3 \log(n\epsilon^{-1})).$$

2. Otherwise, the mixing time of $\mathcal{M}_{tree}(T)$ with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^4 \log(n\epsilon^{-1})).$$

PROOF: In order to apply Theorem 4.3 to the Markov chain $\mathcal{M}_{tree}(T)$, we note that for a node whose associated bit string has length k , the probability of selecting a move that corresponds to two neighboring bits in the string is $\frac{k-1}{\binom{n}{2}} = \frac{k-1}{2n(n-1)}$. Combining the first bound from Theorem 5.2 with Theorem 4.3 where $N = n - 1$ gives the following result,

$$\tau(\epsilon) = O\left(\frac{n(n-1)}{k-1}k^3 \log(2(n-1)k/\epsilon)\right) = O(n^4 \log(n\epsilon^{-1})).$$

If all of the chains have probabilities that are bounded away from $1/2$, then we can use the second bound from Theorem 5.2 to obtain

$$\tau(\epsilon) = O\left(\frac{n(n-1)}{k-1}k^2 \log(2(n-1)/\epsilon)\right) = O(n^3 \log(n\epsilon^{-1})). \quad \blacksquare$$

5.2 Comparing $\mathcal{M}_{tree}(T)$ with \mathcal{M}_{nn} .

Next, we show that \mathcal{M}_{nn} is rapidly mixing when \mathbf{P} has league structure and is *weakly monotone*:

Definition 5.1: The set \mathbf{P} is *weakly monotone* if properties 1 and *either* 2 or 3 are satisfied.

1. $p_{i,j} \geq 1/2$ for all $1 \leq i < j \leq n$, and
2. $p_{i,j+1} \geq p_{i,j}$ for all $1 \leq i < j \leq n - 1$ or
3. $p_{i-1,j} \geq p_{i,j}$ for all $2 \leq i < j \leq n$.

We note that if \mathbf{P} satisfies all three properties then it is *monotone*, as defined by Jim Fill [10].

The comparison proof in this setting is similar to the comparison proof in Section 4.3, except we allow elements between $\sigma(i)$ and $\sigma(j)$ that are larger or smaller than both i and j . This poses a problem, because we may not be able to move $\sigma(j)$ towards $\sigma(i)$ without greatly decreasing the weight. However, we can resolve this if \mathbf{P} is weakly monotone. Specifically, we are now ready to prove the following theorem.

Theorem 5.4: Given input probabilities $1/2 \leq q_1, q_2, \dots, q_{n-1} < 1$ and a positive constant c such that $r_i < 1 - c$ for $1 \leq i \leq n - 1$ and $\mathbf{P} = \{p_{i,j} = q_{i \wedge j}\}$ is weakly monotone.

1. If $\forall 1 \leq i \leq n - 1, c + 1/2 < r_i$ the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

2. Otherwise, the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

PROOF: Throughout this proof we assume that \mathbf{P} satisfies properties 1 and 2 of the weakly monotone definition. If instead \mathbf{P} satisfies property 3, then the proof is almost identical. In order to apply Theorem 4.7 to relate the mixing time of \mathcal{M}_{nn} to the mixing time of $\mathcal{M}_{tree}(T)$ we need to define for each transition of $\mathcal{M}_{tree}(T)$ a canonical path using transitions of \mathcal{M}_{nn} . Let $e = (\sigma, \beta)$ be a transition of $\mathcal{M}_{tree}(T)$ which performs a transposition on elements $\sigma(i)$ and $\sigma(j)$. If there are no elements between $\sigma(i)$ and $\sigma(j)$ then e is already a transition of \mathcal{M}_{nn} and we are done. Otherwise, σ contains the string $\sigma(i), \sigma(i+1), \dots, \sigma(j-1), \sigma(j)$ and β contains $\sigma(j), \sigma(i+1), \dots, \sigma(j-1), \sigma(i)$. From the definition of $\mathcal{M}_{tree}(T)$ we know that for each $\sigma(k), k \in [i+1, j-1]$, either $\sigma(k) > \sigma(i), \sigma(j)$

or $\sigma(k) < \sigma(i), \sigma(j)$. Define $S = \{\sigma(k) : \sigma_k < \sigma(i), \sigma(j)\}$ and $B = \{\sigma(k) : \sigma_k > \sigma(i), \sigma(j)\}$. To obtain a good bound on the congestion along each edge we must ensure that the weight of the configurations on the path are not smaller than the weight of σ . Thus, we define three stages in our path from σ to β . In the first, we shift the elements of S to the left, removing an inversion with each element of B . In the second stage we move $\sigma(i)$ next to $\sigma(j)$ and in the third stage we move $\sigma(j)$ to $\sigma(i)$'s original location. Finally, we shift the elements of S to the right to return them to their original locations. See Figure 7.

Stage 1: At a high-level in this stage we are shifting the elements in S to the left in order to remove an inversion with every element in B . First if $\sigma(j-1) \in B$, shift $\sigma(j)$ to the left until an element from S is immediately to the left of $\sigma(j)$. Next, starting at the right-most element in S and moving left, for each $\sigma(k) \in S$ such that $\sigma(k-1) \in B$, move $\sigma(k)$ to the left one swap at a time until $\sigma(k)$ has an element from S or $\sigma(i)$ on its immediate left (see Figure 8). Notice that for each element $\sigma(l) \in B$ we have removed exactly one $(\sigma(l), \sigma(k))$ inversion where $\sigma(k) \in S$.

Stage 2: Next perform a series of nearest neighbor swaps to move $\sigma(i)$ to the right until it is in the position occupied by $\sigma(j)$ at the end of Stage 1 (see Figure 8). While we have created an $(\sigma(k), \sigma(i))$ inversion for each element $\sigma(k) \in B$ (if originally $\sigma(j-1) \notin B$ then this stage will not create an inversion for every element in B) the weight has not decreased from the original weight because in Stage 1 we removed an $(\sigma(k), \sigma(l))$ inversion (or an $(\sigma(k), \sigma(j))$ inversion) and $(\sigma(k), \sigma(l)) > (\sigma(k), \sigma(j))$ and $(\sigma(k), \sigma(j)) = (\sigma(k), \sigma(i))$ because the \mathbf{P} are weakly monotone. For each $\sigma(k) \in S$ we also removed a $(\sigma(k), \sigma(j))$ inversion.

Stage 3: Perform a series of nearest neighbor swaps to move $\sigma(j)$ to the left until it is in the same position $\sigma(i)$ was originally. While we created an $(\sigma(k), \sigma(j))$ inversion for each $\sigma(k) \in S$, these

Stage 1	5	8	9	<u>2</u>	10	<u>3</u>	<u>4</u>	<u>1</u>	7
Stage 2	5	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7
		<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	5 7
Stage 3	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7	5
Stage 4	7	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	5
	7	8	9	<u>2</u>	10	<u>3</u>	<u>4</u>	<u>1</u>	5

Figure 7: The stages in the canonical path for transposing 5 and 7. Notice that the elements in S are underlined.

5	8	9	<u>2</u>	10	<u>3</u>	<u>4</u>	<u>1</u>	7
5	8	9	<u>2</u>	<u>3</u>	10	<u>4</u>	<u>1</u>	7
5	8	<u>2</u>	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7
5	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7
5	<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7
<u>2</u>	5	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7
<u>2</u>	8	5	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7
				⋮				
<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	5	7
<u>2</u>	8	9	<u>3</u>	10	<u>4</u>	<u>1</u>	7	5

Figure 8: Stages 1 and 2 of the canonical path for transposing 5 and 7.

inversions have the same weight as the $(\sigma(i), \sigma(k))$ inversion we removed in Stage 2. In addition we have removed an $(\sigma(l), \sigma(j))$ inversion for each $\sigma(l) \in B$.

Stage 4: Finally we want to return the elements in S and B to their original position. Starting with the left-most element in S that was moved in Stage 1, perform the nearest neighbor swaps to the right necessary to return it to its original position. Finally, if originally $\sigma(j-1) \in B$, then move $\sigma(i)$ to the original location of $\sigma(j)$. Note that if originally $\sigma(j-1) \notin B$, then $\sigma(i)$ was placed in the original location of $\sigma(j)$ at the end of Stage 2. It's clear from the definition of the stages that the weight of a configuration never decreases below the weight of $\min(\pi(\sigma), \pi(\beta))$.

Given a transition (v, ω) of \mathcal{M}_{nn} we must upper bound the number of canonical paths $\gamma_{\sigma\beta}$ that use this edge. Thus, we analyze the amount of information needed in addition to (z, w) to determine σ and β uniquely. First we record whether (σ, β) is already a nearest neighbor transition or which stage we are in. Next for any of the 4 stages we record the original location of $\sigma(i)$ and $\sigma(j)$. Given this information, along with v and ω , we can uniquely recover (σ, β) . Hence, there are at most $4n^2$ paths through any edge (v, ω) . Also, note that the maximum length of any path is $4n$.

Next we bound the quantity A which is needed to apply Theorem 4.7. Recall that we have guaranteed that $\pi(\sigma) \leq \max\{\pi(v), \pi(\omega)\}$. Assume that $\pi(\sigma) \leq \pi(v)$. Let $\lambda = \max_{i < j} p_{i,j}/p_{j,i}$. Then

$$\begin{aligned} A &= \max_{(v, \omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v, \omega)} \sum_{\Gamma(v, \omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &\leq \max_{(v, \omega) \in E(P)} \sum_{\Gamma(v, \omega)} 2n \frac{P'(\sigma, \beta)}{P(v, \omega)} \\ &\leq \max_{(v, \omega) \in E(P)} \sum_{\Gamma(v, \omega)} 2n \frac{1/\binom{n}{2}}{(1+\lambda)(n-1)} = O(n^2). \end{aligned}$$

If, on the other hand, $\pi(\sigma) \leq \pi(\omega)$, then we use detailed balance to obtain:

$$\begin{aligned} A &= \max_{(v, \omega) \in E(P)} \left\{ \frac{1}{\pi(v)P(v, \omega)} \sum_{\Gamma(v, \omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &= \max_{(v, \omega) \in E(P)} \left\{ \frac{1}{\pi(\omega)P(\omega, v)} \sum_{\Gamma(v, \omega)} |\gamma_{\sigma\beta}| \pi(\sigma) P'(\sigma, \beta) \right\} \\ &\leq \max_{(v, \omega) \in E(P)} \sum_{\Gamma(v, \omega)} 2n \frac{P'(\sigma, \beta)}{P(\omega, v)} \\ &\leq \max_{(v, \omega) \in E(P)} \sum_{\Gamma(v, \omega)} 2n \frac{1/\binom{n}{2}}{(1+\lambda)(n-1)} = O(n^2). \end{aligned}$$

In either case, we have $A = O(n^2)$. Then $\pi_* = \min_{\rho \in \Omega} \pi(\rho) \geq (\lambda \binom{n}{2} n!)^{-1}$ where λ is defined as above, so $\log(1/(\epsilon \pi_*)) = O(n^2 \log \epsilon^{-1})$, as above.. Appealing to Theorem 4.7 and the first bound from Theorem 5.3 if the q_i 's are bounded away from $1/2$, we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^7 \log(n\epsilon^{-1}) \log(\epsilon^{-1})).$$

Similarly, appealing to Theorem 4.7 and the second bound from Theorem 5.3 we have that the mixing time of \mathcal{M}_{nn} satisfies

$$\tau(\epsilon) = O(n^8 \log(n\epsilon^{-1}) \log(\epsilon^{-1})). \quad \blacksquare$$

Remark: By repeating Stage 1 of the path a constant number of times, it is possible to relax the weakly monotone condition slightly if we are satisfied with a polynomial bound on the mixing time.

Remark: If the input probabilities are not bounded away from 1 by a constant but instead by some function of n , then using the same proof as Theorem 5.4 we can obtain a bound on the mixing time. Specifically, given input probabilities $1/2 \leq r_1, r_2, \dots, r_{n-1} < 1 - 1/f(n)$ let $\mathbf{P} = \{p_{i,j} = r_{\min\{i,j\}}\}$. Then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^8 f(n) \log(n\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

If additionally, there exists a positive constant c such that $r_i > 1/2 + c$ for all $1 \leq i \leq n-1$, then the mixing time of \mathcal{M}_{nn} with preference set \mathbf{P} satisfies

$$\tau(\epsilon) = O(n^7 f(n) \log(n\epsilon^{-1}) \log(f(n)\epsilon^{-1})).$$

6 Bounding the Mixing Time of a Product of Markov Chains (Theorem 4.3)

In order to prove Theorem 4.4 and Theorem 5.3, we use combinatorial bijections to express the mixing time of \mathcal{M}_{inv} and $\mathcal{M}_{tree}(T)$ as a product of independent, smaller Markov chains. Our bounds on the mixing rate rely on relating the mixing time of the larger and smaller chains. While there exist many results relating the mixing time of a product of Markov chains (see, for example, references in [2, 4]), these assume that the smaller chains defining the product are of comparable size. These theorems would yield weaker results in our case where the smaller Markov chains can be of vastly different size, so we include a proof of the more tailored theorem here for completeness.

We now prove Theorem 4.3, which states that if the Markov chain \mathcal{M} is a product of M independent Markov chains $\mathcal{M}_1, \mathcal{M}_2, \dots, \mathcal{M}_M$, each with mixing time $\tau_i(\epsilon)$, and \mathcal{M} updates \mathcal{M}_i with probability p_i , then the mixing time of \mathcal{M} is

$$\tau(\epsilon) \leq \max_{i=1,2,\dots,M} \max \left\{ \frac{2}{p_i} \tau_i \left(\frac{\epsilon}{4M} \right), \frac{8}{p_i} \ln \left(\frac{\epsilon}{8M} \right) \right\}.$$

In particular, if each $\tau_i(\epsilon) \geq 4 \ln(\epsilon)$ then

$$\tau(\epsilon) \leq \max_{i=1,2,\dots,M} \frac{2}{p_i} \tau_i \left(\frac{\epsilon}{4M} \right).$$

PROOF: Suppose the Markov chain \mathcal{M} has transition matrix P , and each \mathcal{M}_i has transition matrix P_i and state space Ω_i . Let $B_i = p_i P_i + (1 - p_i)I$, where I is the identity matrix of the same size as P_i , be the transition matrix of \mathcal{M}_i , slowed down by the probability p_i of selecting \mathcal{M}_i . First we show that the total variation distance satisfies

$$1 + 2d_{tv}(P^t, \pi) \leq \prod_i (1 + 2d_{tv}(B_i^t, \pi_i)).$$

To show this, notice that for $x = (x_1, x_2, \dots, x_M), y = (y_1, y_2, \dots, y_M) \in \Omega$, $P^t(x, y) = \prod_i B_i^t(x_i, y_i)$. Let $\epsilon_i(x_i, y_i) = B_i^t(x_i, y_i) - \pi_i(y_i)$ and for any $x_i \in \Omega_i$,

$$\epsilon_i(x_i) = \sum_{y_i \in \Omega_i} |\epsilon_i(x_i, y_i)| \leq 2d_{tv}(B_i^t, \pi_i).$$

Then,

$$\begin{aligned} d_{tv}(P^t, \pi) &= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| \\ &= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \left| \prod_i B_i^t(x_i, y_i) - \prod_i \pi_i(y_i) \right| \\ &= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \left| \prod_i (\epsilon_i(x_i, y_i) + \pi_i(y_i)) - \prod_i \pi_i(y_i) \right| \\ &= \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} \epsilon_i(x_i, y_i) \prod_{i \notin S} \pi_i(y_i) \right| \\ &\leq \max_{x \in \Omega} \frac{1}{2} \sum_{y \in \Omega} \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} |\epsilon_i(x_i, y_i)| \prod_{i \notin S} |\pi_i(y_i)| \\ &= \max_{x \in \Omega} \frac{1}{2} \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} \sum_{y_i \in \Omega_i} |\epsilon_i(x_i, y_i)| \prod_{i \notin S} \sum_{y_i \in \Omega_i} |\pi_i(y_i)| \\ &= \max_{x \in \Omega} \frac{1}{2} \sum_{S \subseteq [M], S \neq \emptyset} \prod_{i \in S} \epsilon_i(x_i) \prod_{i \notin S} 1 \\ &= \max_{x \in \Omega} \frac{1}{2} \prod_i (1 + \epsilon_i(x_i)) - 1/2 \\ &\leq \frac{1}{2} \prod_i (1 + 2d_{tv}(B_i^t, \pi_i)) - 1/2, \end{aligned}$$

as desired. Thus to show $d_{tv}(P^t, \pi) \leq \epsilon$, it suffices to show $d_{tv}(B_i^t, \pi_i) \leq \epsilon/(2M)$ for each i , as

$$\begin{aligned} 1 + 2d_{tv}(P^t, \pi) &\leq \prod_i (1 + 2d_{tv}(B_i^t, \pi_i)) \\ &\leq \prod_i (1 + 2\epsilon/(2M)) \\ &\leq e^\epsilon \leq 1 + 2\epsilon. \end{aligned}$$

Hence it suffices to show $d_{tv}(B_i^t, \pi_i) \leq \epsilon/(2M)$ for each i .

Let $q_i = 1 - p_i$. Since

$$B_i^t = (p_i P_i + q_i I)^t = \sum_{j=0}^t \binom{t}{j} p_i^j q_i^{t-j} P_i^j I,$$

we have

$$\begin{aligned}
d_{tv}(B_i^t, \pi_i) &= \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} |B_i^t(x_i, y_i) - \pi_i(y_i)| \\
&= \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \left| \sum_{j=0}^t \binom{t}{j} p_i^j q_i^{t-j} P_i^j(x_i, y_i) - \pi_i(y_i) \right| \\
&= \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \left| \sum_{j=0}^t \binom{t}{j} p_i^j q_i^{t-j} (P_i^j(x_i, y_i) - \pi_i(y_i)) \right| \\
&\leq \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} \sum_{j=0}^t \binom{t}{j} p_i^j q_i^{t-j} |P_i^j(x_i, y_i) - \pi_i(y_i)| \\
&= \sum_{j=0}^t \binom{t}{j} p_i^j q_i^{t-j} \max_{x_i \in \Omega_i} \frac{1}{2} \sum_{y_i \in \Omega_i} |P_i^j(x_i, y_i) - \pi_i(y_i)| \\
&= \sum_{j=0}^t \binom{t}{j} p_i^j q_i^{t-j} d_{tv}(P_i^j, \pi_i).
\end{aligned}$$

Let $t_i = \tau_i(\epsilon/(4M))$. Now, for $j \geq t_i = \tau_i(\epsilon/(4M))$, we have that $d_{tv}(P_i^j, \pi_i) < \epsilon/(4M)$. For all j , we have $d_{tv}(P_i^j, \pi_i) \leq 2$, so if X is a binomial random variable with parameters t and p_i with $q_i = 1 - p_i$, we have

$$\begin{aligned}
d_{tv}(B_i^t, \pi_i) &\leq \sum_{j=0}^t \binom{t}{j} p_i^j q_i^{t-j} d_{tv}(P_i^j, \pi_i) \\
&= \sum_{j=0}^{t_i-1} \binom{t}{j} p_i^j q_i^{t-j} d_{tv}(P_i^j, \pi_i) + \sum_{j=t_i}^t \binom{t}{j} p_i^j q_i^{t-j} d_{tv}(P_i^j, \pi_i) \\
&< 2 \sum_{j=0}^{t_i-1} \binom{t}{j} p_i^j q_i^{t-j} + \sum_{j=t_i}^t \binom{t}{j} p_i^j q_i^{t-j} \epsilon/(2M) \\
&= 2P(X < t_i) + \epsilon/(2M).
\end{aligned}$$

By Chernoff bounds, $P(X < (1 - \delta)tp_i) \leq e^{-tp_i\delta^2/2}$. Setting $\delta = 1 - t_i/(tp_i)$, then for all $t > 2t_i/p_i$, $\delta^2 \geq 1/4$ and we have

$$P(X < t_i) \leq e^{-tp_i\delta^2/2} \leq e^{-tp_i/8} \leq \epsilon/(8M)$$

as long as $t \geq 8 \ln(\epsilon/(8M))/p_i$. Therefore for $t \geq \max\{8 \ln(\epsilon/(8M))/p_i, 2t_i/p_i\}$,

$$\begin{aligned}
d_{tv}(B_i^t, \pi_i) &= 2P(X < t_i) + \epsilon/(4M) \\
&\leq 2\epsilon/(8M) + \epsilon/(4M) = \epsilon/(2M).
\end{aligned}$$

Hence by time t the total variation distance satisfies $d_{tv}(P^t, \pi) \leq \epsilon$. ■

7 Conclusions

In this paper, we introduced new classes of positively biased probability distributions \mathbf{P} for which the nearest neighbor transposition chain is provably rapidly mixing, the “Choose your Weapon” and “League Hierarchies” classes. Both classes represent linear families of input parameters defining \mathbf{P} , greatly generalizing the single parameter constant bias case studied previously. The only cases in which we know the chain is rapidly mixing for a quadratic family of inputs is when all of the $p_{i,j}$ are 0 or 1 and the problem reduces to linear extensions of a partial order or under conditions which reduce to the case of biased staircase walks that are rapidly mixing (see Remark 3.1). It would be interesting to bound the mixing rate for more general quadratic families of input probabilities.

It is also worth noting that the counterexample from Section 3 showing that there exist positively biased distributions for which the chain mixes slowly does not satisfy the monotonicity condition in Fill’s conjecture, so this conjecture certainly is worthy of further consideration. Moreover, the new classes for which the chain always converges quickly do not necessarily satisfy monotonicity, so there may be another condition that characterizes a larger class of input probabilities \mathbf{P} for which the chain always converges quickly.

Acknowledgments. We thank Jim Fill for sharing his earlier work on this problem and for several useful conversations.

References

- [1] D. Aldous. Random walk on finite groups and rapidly mixing markov chains. *In Seminaire de Probabilites XVII*, pages 243–297, 1983.
- [2] I. Benjamini, N. Berger, C. Hoffman, and E. Mossel. Mixing times of the biased card shuffling and the asymmetric exclusion process. *Trans. Amer. Math. Soc.*, 2005.
- [3] P. Bhakta, S. Miracle, A.P. Streib, and D. Randall. Mixing times of Markov chains for self-organizing lists and biased permutations. In *Proceedings of the 24th ACM/SIAM Symposium on Discrete Algorithms*, SODA '13, pages 1–15, 2013.
- [4] N. Bhatnagar and D. Randall. Torpid mixing of simulated tempering on the potts model. In *Proceedings of the 15th ACM/SIAM Symposium on Discrete Algorithms*, SODA '04, pages 478–487, 2004.
- [5] R. Bubley and M. Dyer. Faster random generation of linear extensions. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*, SODA '98, 1998.
- [6] P. Diaconis and L. Saloff-Coste. Comparison techniques for random walks on finite groups. *The Annals of Applied Probability*, 21:2131–2156, 1993.
- [7] P. Diaconis and L. Saloff-Coste. Comparison theorems for reversible markov chains. *The Annals of Applied Probability*, 3:696–730, 1993.
- [8] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 57:159–179, 1981.
- [9] J. Fill. Background on the gap problem. *Unpublished manuscript*, 2003.
- [10] J. Fill. An interesting spectral gap problem. *Unpublished manuscript*, 2003.
- [11] S. Greenberg, A. Pascoe, and D. Randall. Sampling biased lattice configurations using exponential metrics. In *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, 2009.
- [12] M. Jerrum and A. Sinclair. Approximate counting, uniform generation and rapidly mixing markov chains. *Information and Computation*, 82:93–133, 1989.
- [13] D. Knuth. *The Art of Computer Programming*, volume 3: Sorting and Searching. Addison Wesley, 1973.
- [14] D. Levin, Y. Peres, and E. Wilmer. *Markov chains and mixing times*. American Mathematical Society, 2006.
- [15] M. Luby, D. Randall, and A.J. Sinclair. Markov chain algorithms for planar lattice structures. *SIAM Journal on Computing*, 31:167–192, 2001.
- [16] A. Pascoe and D. Randall. Self-assembly and convergence rates of heterogenous reversible growth processes. In *Foundations of Nanoscience*, 2009.
- [17] D. Randall and P. Tetali. Analyzing glauber dynamics by comparison of Markov chains. *Journal of Mathematical Physics*, 41:1598–1615, 2000.
- [18] A. Sinclair. *Algorithms for random generation and counting*. Progress in theoretical computer science. Birkhäuser, 1993.
- [19] S. Turrini. Optimization in permutation spaces. *Western Research Laboratory Research Report*, 1996.
- [20] D. Wilson. Mixing times of lozenge tiling and card shuffling markov chains. *The Annals of Applied Probability*, 1:274–325, 2004.