A Coupling Argument for the Random Transposition Walk

Olena Blumberg

Abstract

This paper explores the mixing time of the random transposition walk on the symmetric group S_n. While it has long been known that this walk mixes in $O(n \log n)$ time, this result has not previously been attained using coupling. A coupling argument showing the correct order mixing time is presented. This is accomplished by first projecting to conjugacy classes, and then using the Bubley-Dyer path coupling construction. In order to obtain appropriate bounds on the time it takes the path coupling to meet, ideas from Schramm's paper "Compositions of Random Transpositions" are used.

1 Introduction

This paper studies the random transposition walk on the symmetric group S_n – in card shuffling terms, the possible permutations of a deck of n cards. Here's a description of the random walk: lay n cards out in a row, and pick one card uniformly with your right hand, and another card independently uniformly with your left hand (note that you may have picked the same card.) Then, swap the cards – this is an extremely simple shuffling scheme for n cards.

Below, we study the *mixing time* of the above random walk: that is, the number of shuffles that it takes to thoroughly mix up the deck (see Section 2 for a precise definition.) To be more specific, a coupling argument demonstrating that the mixing time of the random transposition walk is $O(n \log n)$ is presented. Coupling is an intuitive probabilistic technique that bounds mixing time in the following way: define a process $(X_t, Y_t)_{t \geq 0}$ such that both $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ are Markov chains with the same transition matrix, but with X_t starting at x and Y_t starting at y. As will be described more precisely in Section 2 below, the goal is to have the two chains meet: by the time that this has happened with high probability for every choice of x and y, it can be shown that the Markov chain has mixed. This technique is usually traced back to Doeblin [9]; two good reference books which illustrate its many uses are Lindvall's "Lectures on the coupling method" [15] and Thorisson's "Coupling, stationarity, and regeneration" [20].

The existence of a coupling argument showing an $O(n \log n)$ mixing time is a long-standing open problem. Due to its simplicity and symmetry, the random transposition walk was one of the first ones considered in burgeoning field of

Markov chains mixing times. As noted in [8], the mixing time of this walk was first bounded by Aldous in 1980, who showed that it must be between order n and n^2 and conjectured that it must be of order $n \log n$. This was proved in 1981 in "Generating a random permutation with random transpositions" by Diaconis and Shahshahani [8]. This paper uses Fourier analysis to show that the walk experiences a cut-off, mixing in a window of order n around time $\frac{1}{2} n \log n$.

The beautiful proof in [8] uses the tools of representation theory and Fourier analysis, and hence is non-probabilistic. While a purely probabilistic strong stationary time proof for an $O(n \log n)$ mixing time was discovered by Broder in 1985 [3], a coupling argument proved to be more elusive. The main difficulty is due to the fact that a Markovian coupling cannot succeed; indeed, Lemma 8 below shows that such an approach can never prove a bound of order better than n^2. It has been shown by Griffeath [10] and then Pitman [18] that a maximal coupling must exist, but it evidently has to be non-Markovian. There has been continued interest in finding such a coupling – for example, Peres named it as an interesting open problem in [17]. This paper resolves this problem. (Another approach for finding such a non-Markovian coupling can be seen in the preprint "Mixing times via super-fast coupling" [13].)

This question is approached here by first projecting the random transposition walk to conjugacy classes. T his projection is also a Markov chain, called a split-merge random walk [19]. Using the fact that the random transposition walk started from the identity is constant on conjugacy classes, it suffices to find the mixing time of the split-merge random walk. The path coupling technique of Bubley and Dyer [4] is used to examine the split-merge random walk. However, this is not straightforward, since in the worst case scenario, the split-merge random walk contracts by only $1 - \frac{1}{n^2}$.

It is shown here that on average, the split-merge random walk does indeed contract by $1 - \frac{1}{n}$, enabling the use of path coupling to conclude that the walk mixes in $O(n \log n)$ time. This argument does not, however, show cut-off: indeed, as noted in Remark 37 below, the constant in front of the $n \log n$ is very large. To show that the contraction coefficient is of the right order, the techniques of Schramm from his paper "Compositions of random transpositions" [19] are used. He shows that large cycles in the random transposition walk emerge after time $\frac{n}{2}$, and then proves the law for the scaled cycles. Methods from "Compositions of random transpositions" have given rise to the wonderful paper "Mixing times for random k-cycles and coalescence-fragmentation chains" by Berestycki, Schramm, and Zeitouni [2], which uses probabilistic techniques to get the right answer for a generalization of the random transposition walk.

2   Background and Definitions

Before stating the main result of this paper, a number of definitions are necessary. If μ and ν are two probability distributions on a finite state space Ω, then the total variation distance between μ and ν is defined to be $\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|$. For a Markov chain with transition probabilities $Q(x, y)$

and stationary distribution π, the total variation distance at time t is defined to be $d(t) = \|Q^t(x, \cdot) - \pi\|_{TV}$ and the mixing time is

$$\tau_{\mathrm{mix}}(\epsilon) = \min \{t \mid d(t) \leq \epsilon\}$$

Conventionally, τ_{mix} is defined to be $\tau_{\mathrm{mix}}(1/4)$.

A coupling of a pair of Markov chains both with transition matrix Q is a process $(X_t, Y_t)_{t \geq 0}$ such that both $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ are Markov chains with transition matrix Q, but which might have different starting distributions. The coupling inequality (Corollary 5.3 in [14]) states that if (X_t, Y_t) is a coupling of a pair of Markov chains such that $X_0 = x$ and $Y_0 = y$, and $T_{x,y}$ is a random time at which the chains have met, then

$$d(t) \leq \max_{x,y} \mathbb{P}\{T_{x,y} > t\}$$

The above inequality allows coupling to be used to bound mixing times. It is now possible to state the main result of this paper:

Theorem 1. *There exists a coupling argument that shows that the random transposition walk on S_n mixes in time of order $n \log n$: that is, it demonstrates that there exists a constant C such that*

$$\tau_{\mathrm{mix}} \leq Cn \log n$$

Before launching into the proof, it is instructive to consider the many ways an $O(n \log n)$ mixing time has been obtained for this walk, as well as the uses of the result. This bound was first obtained by Diaconis and Shahshahani in [8]. This result is beautiful and extremely precise; however, the scope of the technique is limited as it requires fully diagonalizing the random walk. While this is possible for a number of walks, including walks that are not random walks on groups, this is a drawback to the method. This result is also extremely useful for comparison theory. As shown by Diaconis and Saloff-Coste in [7], the Dirichlet form can be used to compare all the eigenvalues of the chain, resulting in good bounds for a variety of walks. For example, Jonasson uses this result in [12] to show that the overlapping cycle shuffle mixes in $O(n^3 \log n)$ time.

As noted above, the first probabilistic proof of the result was by Broder [3] and used *strong stationary times*: stopping times T such that the conditional distribution of X_T given T is stationary. Since the stationary distribution for the random transposition walk is uniform, this is equivalent to stating that for all $\sigma \in S_n$ and all positive integers k,

$$\mathbb{P}(X_T = \sigma \mid T = k) = \frac{1}{n!}$$

The following is Broder's strong stationary time argument, as summarized in Chapter 9 of [14]. Let R_t and L_t be the cards chosen by the right and left hand, respectively. Start the process with no marked cards, and use the following marking scheme: at each step, mark a card R_t if R_t is unmarked, and either (a)

L_t is marked or (b) $R_t = L_t$. Define the stopping time T to be the first time all n cards are marked. It is easy to show that this is indeed a strong stationary time, and that T is around $2n \log n$. This argument provides an $O(n \log n)$ mixing time, but not the correct constant. It was improved by Matthews [16] in 1988 by creating a more complicated rule for marking the cards. This argument showed a cut-off for the walk at time $\frac{1}{2}n \log n$. These arguments are probabilistic and intuitive, and elucidate the reasons for the mixing time in a way that Fourier analysis does not. However, they are heavily reliant on the symmetry of the random transposition walk and as such are difficult to generalize.

The recent paper by Berestycki, Schramm and Zeitouni [2] uses a different approach. Their technique provides the correct answer for the following generalization of the Markov chain: instead of using a uniformly chosen random transposition at each step, a random k-cycle is used. This paper obtains the correct $\frac{1}{k}n \log n$ answer for any fixed k. Like this paper, they begin by projecting the walk to conjugacy classes and then make use of the results of Schramm in [19]. The tools of both this result and Schramm's original paper are graph theoretic: for example, a transposition is considered to be an edge in a random graph process on n vertices. Unfortunately, this exciting method again requires considerable symmetry, since the projection to conjugacy classes has to be a Markov chain. This is also a drawback of the coupling approach which is presented here.

Another intriguing technique explored by Burton and Kovchegov [13] uses non-Markovian coupling. While I have found the ideas in this paper difficult, the approximate approach is that the standard coupling argument by Aldous which results in $O(n^2)$ bound can be improved by 'looking into the future.' A non-Markovian argument with a somewhat similar flavor has previously been implemented for the coloring chain by Hayes and Vigoda [11]. Here's a very approximate sketch of the idea for random transpositions: say that a pair (σ, τ) in S_n currently differs in the cards labeled i and j. The standard coupling for this pair transposes the cards with the same labels in both σ and τ, unless the next transposition is (i, j). However, it is possible to do something different: if the next step transposes cards labeled i and k in σ, the next step in τ can transpose either cards labeled i and k or cards labeled j and k. If the coupling is Markovian, then the choice makes no difference; however, 'looking into the future' can substantially improve the bounds. In work stemming from an unrelated project, I hope to show this for a number of different walks in an upcoming paper.

The argument in this paper proceeds by projecting the walk to conjugacy classes. It is a well-known result that the conjugacy classes of S_n are indexed by partitions of n. Recall that a partition of n is an m-tuple $(a_1, a_2, \ldots, a_m)$ of positive integers that sum to n, where m can be any integer, and $a_1 \geq a_2 \geq \cdots \geq a_m$. Let $\mathcal{P}_n$ be the set of partitions of n. The projection of the random transposition walk on S_n to conjugacy classes is also a Markov chain, called a *split-merge* random walk. It is often referred to as a coagulation-fragmentation chain, and it has been extensively studied – see [6] for some references.

Definition 2. *Assume the random walk is currently at partition $(a_1, \ldots, a_m)$. Then, there are three possibilities for the next move: either merge a pair of parts, split a part into two pieces, or stay in place. (All of these moves are followed by rearranging the new parts to be in non-decreasing order.)*

- **Split:** *A pair a_i can be replaced by the pair $(r, a_i - r)$. For each r between 1 and $a_i - 1$, the probability of this particular split is $\frac{a_i}{n^2}$.*

  Note that this phrasing takes the order into account: here, a more convenient phrasing is the following: for each $r < \frac{a_i}{2}$, split a_i into $\{r, a_i - r\}$ with probability $\frac{2a_i}{n^2}$. If a_i is even and $r = \frac{a_i}{2}$, split a_i into $\{r, a_i - r\}$ with probability $\frac{a_i}{n^2}$.

- **Merge:** *Replace the parts a_i and a_j by $a_i + a_j$. This is done with probability $\frac{2a_i a_j}{n^2}$.*

- **Stay in Place:** *Stay at the partition $(a_1, a_2, \ldots, a_m)$ with probability $\frac{1}{n}$.*

Example 3. Here is an example of the split-merge random walk. Let $n = 5$, and assume the walk is currently at $(4, 1)$. Then, the next step X_1 is distributed as follows:

$$
X_1 = \begin{cases}
(5) & \text{with probability } \frac{8}{25} \\
(4, 1) & \text{with probability } \frac{1}{5} \\
(3, 1, 1) & \text{with probability } \frac{8}{25} \\
(2, 2, 1) & \text{with probability } \frac{4}{25}
\end{cases}
$$

The primary walk under consideration is the split-merge random walk, but for some of the proofs, the original transposition walk is needed. With that in mind, make the following two definitions:

Definition 4. *For $\alpha \in S_n$, define $\mathrm{Cyc}(\alpha)$ to be the partition corresponding to the cycle type of α. For $\sigma \in \mathcal{P}_n$, let*

$$
\mathrm{Perm}(\sigma) = \{\alpha \in S_n \mid \mathrm{Cyc}(\alpha) = \sigma\}
$$

be the set of all permutations with cycle type σ.

Definition 5. *Let $(X_t)_{t \geq 0}$ denote the split-merge random walk, and let $(\bar{X}_t)_{t \geq 0}$ denote the random transposition walk, so that for all t,*

$$
X_t = \mathrm{Cyc}(\bar{X}_t)
$$

Furthermore, let P and π be the transition matrix and stationary distribution for $(X_t)_{t \geq 0}$, respectively, and define $\bar{P}$ and $\bar{\pi}$ analogously for $(\bar{X}_t)_{t \geq 0}$.

The next argument shows it suffices to consider the split-merge random walk. The following proof take a little bit of space to write down, but is actually very simple – the key idea is that the random transposition walk started at the identity is always uniformly distributed over each conjugacy class. (This also follows from a more general result – see Chapter 3F of [5].)

Lemma 6. *Let $P, \bar{P}, \pi$ and $\bar{\pi}$ be defined as in Definition 5 above. Then,*

$$\max_{\alpha \in S_n} \left\| \bar{P}^t(\alpha, \cdot) - \bar{\pi} \right\|_{TV} \leq \max_{\sigma \in \mathcal{P}_n} \left\| P^t(\sigma, \cdot) - \pi \right\|_{TV}$$

Proof: Since the random transposition walk is a random walk on a group, it's vertex transitive. Therefore, for all $\alpha \in S_n$,

$$\left\| \bar{P}^t(\alpha, \cdot) - \bar{\pi} \right\|_{TV} = \left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV}$$

where id is the identity permutation. Thus, it suffices to show that

$$\left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV} \leq \max_{\sigma \in \mathcal{P}_n} \left\| P^t(\sigma, \cdot) - \pi \right\|_{TV}$$

Now, let $\sigma_0 = \text{Cyc}(id) = (1, 1, \ldots, 1)$. It suffices to show that

$$\left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV} = \left\| P^t(\sigma_0, \cdot) - \pi \right\|_{TV} \tag{2.1}$$

Since the split-merge random walk is a projection of the random transposition walk, for $\sigma \in \mathcal{P}_n$,

$$\pi(\sigma) = \sum_{\alpha \in \text{Perm}(\sigma)} \bar{\pi}(\alpha) = \frac{|\text{Perm}(\sigma)|}{n!} \tag{2.2}$$

since $\bar{\pi}$ is the uniform distribution on S_n. Similarly,

$$P^t(\sigma_0, \sigma) = \sum_{\alpha \in \text{Perm}(\sigma)} \bar{P}^t(id, \alpha)$$

Furthermore, note that both the identity permutation and the the random transposition walk are symmetric with respect to $\{1, 2, \ldots, n\}$. Hence for any α_1, α_2 with the same cycle structure, $\bar{P}^t(id, \alpha_1) = \bar{P}^t(id, \alpha_2)$ for all t. Combining this with the equation above shows that for $\alpha \in \text{Perm}(\sigma)$,

$$P^t(\sigma_0, \sigma) = |\text{Perm}(\sigma)| \, \bar{P}^t(id, \alpha) \tag{2.3}$$

Using Equations (2.2) and (2.3),

$$\sum_{\alpha \in \text{Perm}(\sigma)} \left| \bar{P}^t(id, \alpha) - \frac{1}{n!} \right| = \left| P^t(\sigma_0, \sigma) - \pi(\sigma) \right|$$

Finally, putting all this together,

$$2 \left\| \bar{P}^t(id, \cdot) - \bar{\pi} \right\|_{TV} = \sum_{\alpha \in S_n} \left| \bar{P}^t(id, \alpha) - \frac{1}{n!} \right| = \sum_{\sigma \in \mathcal{P}_n} \sum_{\alpha \in \text{Perm}(\sigma)} \left| \bar{P}^t(id, \alpha) - \frac{1}{n!} \right|$$

$$= \sum_{\sigma \in \mathcal{P}_n} \left| P^t(\sigma_0, \sigma) - \pi(\sigma) \right| = 2 \left\| P^t(\sigma_0, \cdot) - \pi \right\|_{TV}$$

which proves Equation (2.1), as desired. $\qquad\square$

6

Remark 7. Although it is not needed, it is very easy to use the triangle inequality to prove the opposite inequality to the one in Lemma 2.1. Hence, the two quantities are actually equal.

Before proceeding to sketch the upcoming proof, it is shown that a Markovian coupling for the random transposition walk cannot hope to give an $O(n \log n)$ mixing time.

Lemma 8. *A Markovian coupling $(\bar{X}_t, \bar{Y}_t)$ of the random transposition walk takes at least $\Omega(n^2)$ time to meet.*

Proof: It easy to check that wherever the two random transposition walks currently are, if $\bar{X}_t \neq \bar{Y}_t$, then

$$\mathbb{P}\left(\bar{X}_{t+1} = \bar{Y}_{t+1}\right) \leq \frac{6}{n^2}$$

To verify this, note that if $\bar{X}_t$ and $\bar{Y}_t$ differ only in the transposition (i, j), then the only way to meet is to transpose i and j in one of them, and to stay in place in the other one; similar arguments hold if $\bar{X}_t$ and $\bar{Y}_t$ are two transpositions apart, and in all other cases, the probability of meeting at the next step is 0. Combining the above inequality with the Markov property leads to the desired result. $\qquad\qquad\square$

Turn next to an explanation of the idea behind the coupling. The argument uses path coupling – that is, coupling a pair of split-merge random walks started at a neighboring pair of elements. This technique was invented by Bubley and Dyer in [4]; a good reference is Chapter 14 of [14]. To be precise, endow the state space Ω with a connected graph structure: that is, select a set of edges E' between elements of Ω, such that for any $u, v \in \Omega$, there exists a path between u and v only using the edges in E'. It is then only necessary to define a coupling for $(x, y) \in E'$.

Assign lengths $l(x, y) \geq 1$ to each edge $(x, y) \in E'$, and define a path metric ρ on Ω by

$$\rho(z, w) = \min\left\{\sum_{i=0}^{n-1} l(x_i, x_{i+1}) \mid x_0 = z, x_n = w, (x_i, x_{i+1}) \in E' \text{ for all } i\right\}$$

Furthermore, define the diameter of the set Ω in the usual way as $\text{diam}(\Omega) = \max_{u,v \in \Omega} \rho(u, v)$ The following theorem is the basic path coupling bound.

Theorem 9. *Let $(X_t)_{t \geq 0}$ be a Markov chain on a set Ω, and let E', l and ρ be defined as above. Let (X_1, Y_1) be the first step of a coupling started at $(x, y) \in E'$. Then, if there is a $\kappa < 1$ such that for every $(x, y) \in E'$,*

$$\mathbb{E}\left[\rho(X_1, Y_1)\right] \leq \kappa \rho(x, y) \tag{2.4}$$

then for all $t \geq 1$,

$$d(t) \leq \text{diam}(S)\kappa^t$$

Returning to the random walk under consideration, define neighboring pairs of partitions to be precisely the pairs which are one step away in the split-merge random walk. Then, define a coupled process (X_t, Y_t) such that $X_0 = \sigma$ and $Y_0 = \tau$, making sure that the distance between X_t and Y_t at each step is at most 1. Here are some useful definitions.

Definition 10. *For σ and τ partitions of n, define $\rho(\sigma, \tau)$ to be the distance between σ and τ induced by the split-merge random walk; that is, $\rho(\sigma, \tau)$ is the number of split-merge steps it takes to get from σ to τ.*

The next definition is useful for finding a lower bound on the probability of coupling at each step given the current location of the two walks.

Definition 11. *Let σ and τ be partitions of n such that $\rho(\sigma, \tau) = 1$. Then σ and τ are exactly one merge away, so rearranging parts appropriately and without loss of generality letting σ be the partition with more parts,*

$$\begin{aligned} \sigma &= (a_1, a_2, \ldots, a_m, b, c) \\ \tau &= (a_1, a_2, \ldots, a_m, b + c) \end{aligned} \tag{2.5}$$

where $b \leq c$. Then, define

$$s(\tau, \sigma) = s(\sigma, \tau) = b \quad and \quad m(\tau, \sigma) = m(\sigma, \tau) = c \tag{2.6}$$

That is, since σ and τ differ in the parts b, c and $b + c$, $s(\sigma, \tau)$ is the smallest part in which they differ, and $m(\sigma, \tau)$ is the medium part in which they differ.
For later use, define $m(\sigma, \sigma) = n$ and $s(\sigma, \sigma) = \frac{n}{2}$.

In the next section, the coupling is given along with the following lemma:

Lemma 12. *Assume that $(X_t, Y_t) = (\sigma, \tau)$, for σ and τ such that $\rho(\sigma, \tau) = 1$. Then, $\rho(X_{t+1}, Y_{t+1}) \leq 1$, and*

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4s(\sigma, \tau)}{n^2}$$

That is, the chain stays at most distance 1 apart, and gives the above lower bound for the probability of coupling.

After proving the above lemma, it is shown below that after order n steps, $s(X_t, Y_t)$ is on average of order n. The lemma then implies that the probability of coupling at each step is of order $\frac{1}{n}$, which will show that there is a high probability of coupling after order n steps. Using the fact that the diameter of the set of partitions is no greater than n, Theorem 9 shows that the random transposition walk mixes in $O(n \log n)$ time.

3 The Coupling

This section defines the coupling for neighboring pairs for the split-merge random walk, and proves Lemma 12. The coupling is defined in such a way that the distance between X_t and Y_t at each step is at most 1 for all t. As usual, once the two chains meet, they are run together.

Definition 13. *Consider the next step (X_1, Y_1) of a coupling which is currently at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) = 1$ and*

$$\sigma = (a_1, a_2, \ldots, a_m, b, c)$$
$$\tau = (a_1, a_2, \ldots, a_m, b + c)$$

where $b \leq c$. There are a number of cases, considered in the following order: go through the possible moves in σ, then provide corresponding moves in τ.

- **Operations only using the** a_i**:** *If a_i and a_j are merged in σ for any i and j, perform the same operation in τ. Similarly, if a_i is split in σ into $\{r, a_i - r\}$, do the same for a_i in τ. Then,*

$$X_1 = (a'_1, \ldots, a'_k, b, c)$$
$$Y_1 = (a'_1, \ldots, a'_k, b + c)$$

  for the appropriate $\{a'_1, a'_2, \ldots, a'_k\}$.

- **Merging b or c and** a_i**:** *If b and a_i are merged in σ, merge $b + c$ and a_i in τ. If c and a_i are merged in σ, also merge $b + c$ and a_i in τ. In the first case,*

$$X_1 = (a'_1, \ldots, a'_{m-1}, b + a_i, c)$$
$$Y_1 = (a'_1, \ldots, a'_{m-1}, b + c + a_i)$$

  where $\{a'_1, a'_2, \ldots, a'_{m-1}\} = \{a_1, a_2, \ldots, a_m\}/\{a_i\}$. The case where c and a_i are merged in σ is analogous.

- **Splitting b or c:** *If b is split in σ into $\{r, b - r\}$ where $r \leq \frac{b}{2}$, then split $b + c$ in τ into $\{r, b + c - r\}$. Similarly, if c is split in σ into $\{r, c - r\}$ where $r \leq \frac{c}{2}$, then split $b + c$ in τ into $\{r, b + c - r\}$. The first case results in*

$$X_1 = (a_1, \ldots, a_m, r, b - r, c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b + c - r)$$

  The second case, where c is split into $\{r, c - r\}$, is analogous.

- **Staying in place:** *If the walk stays in place in σ, it is coupled with either staying in place in τ or with splitting $b + c$ in τ into $\{b, c\}$. Since splitting $b + c$ into $\{b, c\}$ may have already been coupled with splitting c into $\{b, c - b\}$, let p be the remaining probability of splitting $b + c$ into $\{b, c\}$. Then, couple staying in place in σ with splitting $b + c$ into $\{b, c\}$ in τ with probability $\min\left(p, \frac{1}{n}\right)$. This results in*

$$X_1 = (a_1, \ldots, a_m, b, c)$$
$$Y_1 = (a_1, \ldots, a_m, b, c)$$

  That is, the chains will couple.

  Couple staying in place in σ to staying in place in τ the rest of the time – that is, with probability $\frac{1}{n} - \min\left(p, \frac{1}{n}\right)$.

- **Merging b and c:** *Couple merging b and c in σ to any remaining possibilities in τ. It is easy to check that these are either staying in place or splitting $b + c$ into $\{r, b + c - r\}$. The first leads to the chains coupling; the second leads to*

$$X_1 = (a_1, \ldots, a_m, b + c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b + c - r)$$

for some r.

Example 14. As this coupling looks fairly complicated, here are a couple of examples. The possible pairs for (X_1, Y_1) are listed, as well as the probability of each pair.

1. Let $(X_0, Y_0) = (\sigma, \tau) = ((2, 3), (5))$. Here, there are no a_i, $b = 2$, $c = 3$, and $b + c = 5$. A description is provided for each pair of moves: the first move corresponds to σ, the second to τ.

$$(X_1, Y_1) : \begin{cases} ((1,1,3),(1,4)), p = \frac{2}{25} & \text{split 2 as } \{1,1\}, \text{ split 5 as } \{1,4\} \\ ((1,2,2),(1,4)), p = \frac{6}{25} & \text{split 3 as } \{1,2\}, \text{ split 5 as } \{1,4\} \\ ((2,3),(2,3)), p = \frac{5}{25} & \text{stay at } \sigma, \text{ split 5 as } \{2,3\} \\ ((5),(1,4)), p = \frac{2}{25} & \text{merge 2 and 3, split 5 as } \{1,4\} \\ ((5),(2,3)), p = \frac{5}{25} & \text{merge 2 and 3, split 5 as } \{2,3\} \\ ((5),(5)), p = \frac{5}{25} & \text{merge 2 and 3, stay at } \tau \end{cases}$$

2. Let $(X_0, Y_0) = (\sigma, \tau) = ((2, 1, 3), (2, 4))$, written with the above convention that the parts σ and τ disagree on are written last. Here, $a_1 = 2$, $b = 1$, $c = 3$, and $b + c = 4$, and the first move again corresponds to σ, while the second corresponds to τ.

$$(X_1, Y_1) : \begin{cases} ((1,1,1,3),(1,1,4)), p = \frac{2}{36} & \text{split 2 as } \{1,1\} \text{ in both} \\ ((3,3),(6)), p = \frac{4}{36} & \text{merge 2 and 1, merge 2 and 4} \\ ((1,5),(6)), p = \frac{12}{36} & \text{merge 2 and 3, merge 2 and 4} \\ ((2,1,1,2),(2,1,3)), p = \frac{6}{36} & \text{split 3 as } \{1,2\}, 4 \text{ as } \{1,3\} \\ ((2,1,3),(2,1,3)), p = \frac{2}{36} & \text{stay at } \sigma, \text{ split 4 as } \{1,3\} \\ ((2,4),(2,2,2)), p = \frac{4}{36} & \text{merge 1 and 3, split 4 as } \{2,2\} \\ ((2,4),(2,4)), p = \frac{2}{36} & \text{merge 1 and 3, stay at } \tau \\ ((2,1,3),(2,4)), p = \frac{4}{36} & \text{stay at } \sigma \text{ and } \tau \end{cases}$$

Going back to the general case, here is a check that the above definition provides the correct marginal distribution for Y_1. Note that given the way that the coupling was defined, it clearly provides the correct distribution for X_1.

Lemma 15. *The coupling in Definition 13 has the correct marginal distribution for Y_1.*

Proof: Since σ and τ share the parts a_i, the operations only using the a_i are distributed identically in both and hence pose no problem. Furthermore,

$$\mathbb{P}(\text{Merge } b \text{ and } a_i \text{ in } \sigma) + \mathbb{P}(\text{Merge } c \text{ and } a_i \text{ in } \sigma) = \frac{2ba_i}{n^2} + \frac{2ca_i}{n^2} = \frac{2(b+c)a_i}{n^2}$$
$$= \mathbb{P}(\text{Merge } b+c \text{ and } a_i \text{ in } \tau)$$

Thus, all the operations involving any a_i work properly.

Consider next operations that only involve b and c in σ. Splitting b into $\{r, b-r\}$ where $r \leq \frac{b}{2}$ in σ is coupled with splitting $b+c$ into $\{r, b+c-r\}$ in τ, and similarly for c. It needs to be checked that this is possible – that is, that the probability of splitting $b+c$ into $\{r, b+c-r\}$ in τ is sufficiently large to accommodate all these moves in σ.

There are a number of possibilities. First of all, if $r \leq \frac{b}{2}$, then clearly $r < \frac{b+c}{2}$, and hence according to Definition 2,

$$\mathbb{P}(\text{Split } b+c \text{ into } \{r, b+c-r\} \text{ in } \tau) = \frac{2(b+c)}{n^2} = \frac{2b}{n^2} + \frac{2c}{n^2}$$
$$\geq \mathbb{P}(\text{Split } b \text{ into } \{r, b-r\} \text{ in } \sigma) +$$
$$\mathbb{P}(\text{Split } c \text{ into } \{r, c-r\} \text{ in } \sigma)$$

In this case, the probability of splitting $b+c$ into $\{r, b+c-r\}$ in τ is sufficiently large.

x

Now, if $\frac{b}{2} < r \leq \frac{c}{2}$, the procedure couples splitting $b+c$ into $\{r, b+c-r\}$ with splitting c into $\{r, c-r\}$. Thus, since in this case r is still less than $\frac{b+c}{2}$,

$$\mathbb{P}(\text{Split } b+c \text{ into } \{r, b+c-r\} \text{ in } \tau) = \frac{2(b+c)}{n^2} \geq \frac{2c}{n^2}$$
$$\geq \mathbb{P}(\text{Split } c \text{ into } \{r, c-r\} \text{ in } \sigma)$$

which again works.

Finally, if $r > \frac{c}{2}$, splitting $b+c$ into $\{r, b+c-r\}$ is not coupled to splitting either b or c in σ, which obviously does not pose a problem. None of the other moves considered in Definition 13 could be an issue, and hence the marginal distribution of Y_1 under this definition is correct. $\square$

The next step proves Lemma 12. This states that the coupled chains stay at most one step apart, and that

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4s(X_t, Y_t)}{n^2}$$

Proof of Lemma 12: It should be clear from Definition 13 that the coupling stays at most one step apart for all t. To show that if $(X_t, Y_t) = (\sigma, \tau)$, where $\rho(\sigma, \tau) = 1$, then

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4s(\sigma, \tau)}{n^2}$$

11

let

$$\sigma = (a_1, \ldots, a_m, b, c)$$
$$\tau = (a_1, \ldots, a_m, b + c)$$

where $b \leq c$. Then by Definition 11, $s(\sigma, \tau) = b$.

From Definition 13, the chains can couple either if σ stays in place, or if b and c are merged in σ. Consider those two cases separately.

Staying in place in σ: The chains will couple if σ stays in place and $b + c$ is split in τ into $\{b, c\}$. As noted in the definition, these are coupled together with probability $\min\left(p, \frac{1}{n}\right)$, where p is the remaining probability of splitting $b + c$ into $\{b, c\}$ in τ – the probability that this split hasn't already been coupled to something else. To find a lower bound on p, first note that splitting $b + c$ into $\{b, c\}$ in τ couldn't have been coupled with any splits of b in σ. However, it might have been coupled with a split of c in σ. Consider two cases: $c < 2b$ and $c \geq 2b$.

If $c < 2b$, then splitting c into $\{c - b, b\}$ in σ is coupled to splitting $b + c$ into $\{c - b, 2b\}$ in τ since $c - b < b$. This means that nothing is coupled to splitting $b + c$ into $\{b, c\}$, and therefore

$$p = \mathbb{P}(\text{Splitting } b + c \text{ into } \{b, c\} \text{ in } \tau) \geq \frac{b + c}{n^2} \geq \frac{2b}{n^2} \tag{3.1}$$

If $c \geq 2b$, then splitting $b + c$ into $\{b, c\}$ in τ is indeed coupled with splitting c into $\{b, c - b\}$ in σ. In this case, clearly $b \neq c$, and hence

$$\mathbb{P}(\text{Splitting } b + c \text{ into } \{b, c\} \text{ in } \tau) = \frac{2(b + c)}{n^2}$$

Therefore,

$$p \geq \mathbb{P}(\text{Splitting } b + c \text{ into } \{b, c\} \text{ in } \tau) - \mathbb{P}(\text{Splitting } c \text{ into } \{b, c - b\} \text{ in } \sigma)$$
$$\geq \frac{2(b + c)}{n^2} - \frac{2c}{n^2} = \frac{2b}{n^2} \tag{3.2}$$

Equations (3.1) and (3.2) give $p \geq \frac{2b}{n^2}$. Furthermore, note that $b \leq c$, and $b + c \leq n$, and hence $b \leq \frac{n}{2}$. Therefore,

$$\min\left(p, \frac{1}{n}\right) \geq \min\left(\frac{2b}{n^2}, \frac{1}{n}\right) \geq \frac{2b}{n^2} \tag{3.3}$$

Hence,

$$\mathbb{P}(\text{Coupling if staying in place in } \sigma) = \min\left(p, \frac{1}{n}\right) \geq \frac{2b}{n^2} \tag{3.4}$$

12

Merging b and c in σ: Next, consider the probability of coupling if b and c are merged in σ. Clearly, this would need to be coupled with staying in place in τ. The only other thing that staying in place in τ could have been coupled with so far is staying in place in σ. As noted in Definition 13,

$$\mathbb{P}(\text{Both } \sigma \text{ and } \tau \text{ stay in place}) = \frac{1}{n} - \min\left(p, \frac{1}{n}\right)$$

for the same p used above. Thus,

$$\mathbb{P}(\text{Coupling if merging } b \text{ and } c \text{ in } \sigma) = \mathbb{P}(b \text{ and } c \text{ merged in } \sigma, \tau \text{ stayed})$$

$$= \frac{1}{n} - \mathbb{P}(\text{Both } \sigma \text{ and } \tau \text{ stayed}) \qquad (3.5)$$

$$= \min\left(p, \frac{1}{n}\right) \geq \frac{2b}{n^2}$$

using Equation (3.3) above.

Finally, combining Equations (3.4) and (3.5),

$$\mathbb{P}(X_{t+1} = Y_{t+1}) \geq \frac{4b}{n^2} = \frac{4s(\sigma, \tau)}{n^2} = \frac{4s(X_t, Y_t)}{n^2}$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Continuing with the proof, as sketched out earlier, the rest of this paper will be concerned with showing that $s(X_t, Y_t)$ is of order n after $O(n)$ time. The next section shows how that proves Theorem 1, and provides a summary of the proof.

4 Proof of Main Theorem Using $\mathbb{E}\left[s(X_t, Y_t)\right]$

As described above, one of the main tools of this paper is the following theorem:

Theorem 16. *There exist constants α and β such that for all $t \geq \alpha n$,*

$$\mathbb{E}\left[s(X_t, Y_t)\right] \geq \beta n$$

This section uses the above result to prove Theorem 1. To start, prove the following easy lemma:

Lemma 17. *Let (X_t, Y_t) be defined as in Definition 13, where as usual $\rho(X_0, Y_0)$ is equal to 1. Let α and β be the constants in Theorem 16 above. Then,*

$$\mathbb{P}\left(X_{\alpha n + \frac{n}{2}} = Y_{\alpha n + \frac{n}{2}}\right) \geq \beta$$

Proof: Since by Lemma 12, $\mathbb{P}(X_t = Y_t)$ is non-decreasing, if $\mathbb{P}(X_t = Y_t) \geq \beta$ for any $t \leq \alpha n + \frac{n}{2}$, the argument is complete. Thus, assume that

$$\mathbb{P}(X_t = Y_t) \leq \beta \qquad (4.1)$$

for all $t \leq \alpha n + \frac{n}{2}$.

Clearly,

$$\mathbb{P}(X_{t+1} = Y_{t+1}) = \mathbb{P}(X_t = Y_t) + \mathbb{P}(X_{t+1} = Y_{t+1} \mid X_t \neq Y_t)\mathbb{P}(X_t \neq Y_t)$$

Rearranging, and using Lemma 12,

$$\mathbb{P}(X_{t+1} = Y_{t+1}) - \mathbb{P}(X_t = Y_t) \geq \mathbb{E}\left[\frac{4s(X_t, Y_t)}{n^2}\Big| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t)$$

$$= \frac{4}{n^2}\mathbb{E}\left[s(X_t, Y_t)\Big| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t)$$

A lower bound is now needed for the right-hand side. Assume that $t \geq \alpha n$, and hence that $\mathbb{E}\left[s(X_t, Y_t)\right] \geq \beta n$ by Lemma 16. Then,

$$\mathbb{E}\left[s(X_t, Y_t)\Big| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t) = \mathbb{E}\left[s(X_t, Y_t)\right] -$$

$$\mathbb{E}\left[s(X_t, Y_t)\Big| X_t = Y_t\right]\mathbb{P}(X_t = Y_t)$$

$$\geq \beta n - \frac{n}{2}\mathbb{P}\left(X_t = Y_t\right)$$

since if $X_t = Y_t$, $s(X_t, Y_t) = \frac{n}{2}$. Furthermore, using Equation (4.1),

$$\mathbb{E}\left[s(X_t, Y_t)\Big| X_t \neq Y_t\right]\mathbb{P}(X_t \neq Y_t) \geq \frac{\beta n}{2}$$

Combining this with Equation (4.2),

$$\mathbb{P}(X_{t+1} = Y_{t+1}) - \mathbb{P}(X_t = Y_t) \geq \frac{2\beta}{n}$$

for all $\alpha n \leq t \leq \alpha n + \frac{n}{2}$. Adding up these inequalities for all t in $[\alpha n, \alpha n + \frac{n}{2}]$,

$$\mathbb{P}\left(X_{\alpha n + \frac{n}{2}} = Y_{\alpha n + \frac{n}{2}}\right) \geq \beta$$

as required. $\square$

For path coupling, a lemma about the diameter of P_n under the split-merge random walk is needed.

Lemma 18. *The diameter of P_n under the split-merge random walk is at most n.*

Proof: Proceed by induction on n. This statement is clearly true for $n = 1$. Now, assume it's true for all $m \leq n - 1$, and show it for n. Let $\sigma = (a_1, \ldots, a_k)$ and $\tau = (b_1, \ldots, b_l)$ be two partitions of n. Without loss of generality, assume that $a_1 \geq b_1$.

If $a_1 = b_1$, create a path from σ to τ by just changing the parts $(a_2, \ldots, a_k)$ to $(b_2, \ldots, b_l)$. Since $(a_2, \ldots, a_k)$ is a partition of $n - a_1$, by the inductive hypothesis,

$$\rho(\sigma, \tau) \leq n - a_1 \leq n - 1.$$

so this case follows.

Otherwise, $a_1 > b_1$. Let σ_1 be σ with a_1 split into $(b_1, a_1 - b_1)$. Then, σ_1 and τ match on the part b_1, and hence by the argument above,

$$\rho(\sigma_1, \tau) \leq n - 1$$

Since σ is a neighbor of σ_1, this implies that $\rho(\sigma, \tau) \leq n$, completing the proof. □

Theorem 1 is now proved using path coupling. It shows an $O(n \log n)$ bound on the split-merge random walk, and hence on the random transposition walk.

Proof of Theorem 1. Let $t_1 = \alpha n + \frac{n}{2}$. Consider the walk $(\tilde{X}_k)_{k \geq 1}$, where each step consists of making t_1 steps of the split-merge random walk. Let $(\tilde{X}_k, \tilde{Y}_k)$ be the coupling on this new walk induced by the current coupling (X_t, Y_t). Now, Proposition 17 shows that if $(\tilde{X}_0, \tilde{Y}_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) = 1$, then

$$\mathbb{E}\left[\rho(\tilde{X}_1, \tilde{Y}_1)\right] = \mathbb{E}\left[\rho(X_{t_1}, Y_{t_1})\right] = \mathbb{P}(X_{t_1} \neq Y_{t_1})$$
$$\leq (1 - \beta)\rho(\sigma, \tau)$$

using the fact that $\rho(X_t, Y_t)$ is always either 0 or 1. Therefore, if $\tilde{d}(k)$ is defined to be the distance from stationarity of $(\tilde{X}_k, \tilde{Y}_k)$, then from Theorem 9,

$$\tilde{d}(k) \leq \text{diam}(P_n)\left(1 - \beta\right)^k$$

Since neighboring pairs are pairs that are one step apart in the split-merge random walk, Proposition 18 implies that $\text{diam}(P_n) \leq n$. Also using the fact that $1 - x \leq e^{-x}$,

$$\tilde{d}(k) \leq ne^{-\beta k}$$

Thus, if $k = \frac{\log n}{2\beta}$, then $\tilde{d}(k) \leq e^{-2} < \frac{1}{4}$. But it's clear from the definition of the new walk that

$$d(kt_1) = \tilde{d}(k)$$

Thus,

$$d\left(\left(\frac{\alpha}{2\beta} + \frac{1}{4\beta}\right)n \log n\right) = d(kt_1) < \frac{1}{4}$$

which means that the walk has mixed by time $\left(\frac{\alpha}{2\beta} + \frac{1}{4\beta}\right)n \log n$, completing the proof. □

5 Proving $\mathbb{E}\left[s(X_t, Y_t)\right]$ is large

Let us now summarize the rest of the proof. The remainder of this paper will be devoted to proving Theorem 16, which states that after $O(n)$ time, the expected value of $s(X_t, Y_t)$ is of order n.

The proof will be structured as follows: it is shown that in $O(n)$ time, $s(X_t, Y_t)$ will have a high probability of being at least order $n^{1/3}$. Then it is shown that it takes another $o(n)$ time for $s(X_t, Y_t)$ to have a high probability of being of order n. This will clearly suffice to show that that after $O(n)$ time, $\mathbb{E}\left[s(X_t, Y_t)\right]$ is of order n. Section 6 below will be concerned with growing $s(X_t, Y_t)$ to order $n^{1/3}$, while Section 7 will be concerned with growing it to order n.

Before stating the theorems and sketching their proofs, a number of useful definitions are needed. Note that some of these definitions are asymmetrical: they are defined in terms of $\bar{X}_t$ and not $\bar{Y}_t$. This is an arbitrary choice; since the pair (X_t, Y_t) is only a step apart, it doesn't make any difference.

Definition 19. *For $v \in \{1, 2, \ldots, n\}$, define $C_t(v)$ to be the cycle of $\bar{X}_t$ containing v. Furthermore, for a number x, define*

$$V_t(x) = \{v \in \{1, 2, \ldots, n\} \mid |C_t(v)| \geq x\}$$

Thus, $V_t(x)$ is the union of all cycles of size at least x.

Remark 20. Note that if $X_t = (a_1, a_2, \ldots, a_m)$, then

$$|V_t(x)| = \sum_{a_i \geq x} a_i$$

Thus, the size of $V_t(x)$ is a function of X_t.

The first proposition that grows $s(X_t, Y_t)$ to order $n^{1/3}$ is now stated.

Proposition 21. *Let (X_t, Y_t) be the usual coupling started at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) \leq 1$. Then, for n sufficiently large and $t \geq 9n$,*

$$\mathbb{P}\left\{s(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \geq \frac{1}{2}$$

Remark 22. Here, the choice of $n^{1/3}$ is in some sense arbitrary – any n^α, where $\alpha < \frac{1}{2}$, would have done just as well.

A few other definitions which are needed for the statement of the theorem growing $s(X_t, Y_t)$ from order $n^{1/3}$ to order n. Indeed, a more general theorem is proved. Fix constants ϵ and δ: then, if $s(X_t, Y_t)$ starts by being of size 2^{j+1} (where j can be a function of n), after a certain amount of time q, $s(X_{t+q}, Y_{t+q})$ has a high probability of being least $\epsilon\delta n$. The following definition introduces some notation necessary for stating the theorem; it currently looks completely inexplicable, but will be justified in Section 7.

Definition 23. *Assume ϵ and δ are fixed constants, and j is a number (possibly a function of n). Then, define*

$$K = \lceil \log_2(\epsilon\delta n) \rceil \tag{5.1}$$

Furthermore, for r between j and K define

$$a_r = \lceil 2\delta^{-1}2^{-r}n(\log_2 n - r)\rceil \quad and \quad \tau_r = \sum_{i=j}^{r-1} a_i \tag{5.2}$$

where as usual, $\lceil \cdot \rceil$ stands for the ceiling function.

The following proposition proves that $s(X_t, Y_t)$ grows to order n.

Proposition 24. *Let (X_t, Y_t) be the usual coupling started at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) \leq 1$. Let j be a number and let $\delta \in (0, 1]$ be a constant such that $\left|V_0(2^{j+1})\right| \geq \delta n$ and $s(\sigma, \tau) \geq 2^{j+1}$. If K and τ_K are defined as in Definition 23 and $\epsilon \in (0, 1/32)$, then*

$$\mathbb{P}\{s(X_{\tau_K}, Y_{\tau_K}) < \epsilon\delta n\} \leq O(1)\delta^{-1}\epsilon |\log(\epsilon\delta)| \tag{5.3}$$

where the constant implied in the $O(1)$ notation is universal.

Proof of Theorem 16. Propositions 21 and 24 can be used to prove Theorem 16: let $t_1 \geq 9n$, and condition on $(X_{t_1}, Y_{t_1}) \in Q_{t_1}$, where

$$Q_{t_1} = \left\{(X_{t_1}, Y_{t_1}) \text{ such that } s(X_{t_1}, Y_{t_1}) \geq n^{1/3}, \left|V_{t_1}^{\pi}\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \tag{5.4}$$

Letting $2^{j+1} = n^{1/3}$ and $\delta = \frac{1}{2}$, if $(X_{t_1}, Y_{t_1}) \in Q_{t_1}$, then

$$s(X_{t_1}, Y_{t_1}) \geq 2^{j+1} \text{ and } \left|V_{t_1}(2^{j+1})\right| \geq \delta n$$

Since $\rho(X_{t_1}, Y_{t_1}) \leq 1$, Proposition 24 applies to pairs (X_{t_1}, Y_{t_1}) in Q_{t_1}. Therefore, averaging over $(X_{t_1}, Y_{t_1}) \in Q_{t_1}$,

$$\mathbb{P}\left\{s(X_{t_1+\tau_K}, Y_{t_1+\tau_K}) < \epsilon\delta n \mid (X_{t_1}, Y_{t_1}) \in Q_{t_1}\right\} \leq O(1)\delta^{-1}\epsilon |\log(\epsilon\delta)|$$

for any $\epsilon \in (0, 1/32)$. Now, pick ϵ such that the right hand side of the above inequality is at most $1/2$. Then,

$$\mathbb{P}\left\{s(X_{t_1+\tau_K}, Y_{t_1+\tau_K}) \geq \epsilon\delta n \mid (X_{t_1}, Y_{t_1}) \in Q_{t_1}\right\} \geq \frac{1}{2}$$

and therefore, for sufficiently large n,

$$\mathbb{P}\left\{s(X_{t_1+\tau_K}, Y_{t_1+\tau_K}) \geq \epsilon\delta n\right\} \geq \frac{\mathbb{P}(Q_{t_1})}{2} \geq \frac{1}{4}$$

using Lemma 21. Therefore,

$$\mathbb{E}\left[s(X_{t_1+\tau_K}, Y_{t_1+\tau_K})\right] \geq \frac{\epsilon\delta n}{4} \tag{5.5}$$

17

It now just remains to show that is that $t_1 + \tau_K$ can be of order n. Since $\delta = \frac{1}{2}$ and $2^{j+1} = n^{1/3}$, by Equation (5.2)

$$\tau_K = \sum_{i=j}^{K-1} \lceil 2\delta^{-1} 2^{-i} n(\log_2 n - i) \rceil = O\left(n \log n \sum_{r=j}^{K-1} 2^{-i}\right)$$

$$= O\left(n \log n \cdot 2^{-j+1}\right) = O(n^{2/3} \log n)$$

Since $t_1 \geq 9n$ is arbitrary and τ_K is $o(n)$, Equation (5.5) implies that

$$\mathbb{E}\left[s(X_t, Y_t)\right] \geq \frac{\epsilon \delta n}{4}$$

for all $t \geq 10n$, which is precisely what is needed. $\qquad\square$

Before the next two sections, in which Propositions 21 and 24 are proved, some technical results are needed. These are proved in Section 8 below, and are instrumental for controlling the probabilities in the next two sections.

Lemma 25. *Let σ be in S_n, and let $(\bar{X}_t)_{t \geq 1}$ be the random transposition walk starting at σ. Then, the expected number of v such that $|C_1(v)| < |C_0(v)|$ and $|C_1(v)| < x$ is no greater than $\frac{x^2}{n}$.*

For the next four lemmas, let (X_t, Y_t) be the usual coupling starting at (σ, τ), where $\rho(\sigma, \tau) = 1$, $s(\sigma, \tau) = b$ and $m(\sigma, \tau) = c$.

Lemma 26. *If $x \leq c$, then*

$$\mathbb{P}\left\{m(X_1, Y_1) < x\right\} \leq \frac{2x^2}{n^2}.$$

Lemma 27. *If $x \leq c$, and if $|V_0(y)| \geq R$, then*

$$\mathbb{P}\{m(X_1, Y_1) \geq x + y\} \geq \frac{2c(R - 2c)}{n^2}$$

Lemma 28. *If $x \leq b$, then*

$$\mathbb{P}\{s(X_1, Y_1) < x\} \leq \frac{4x^2}{n^2}$$

Lemma 29. *If x and y satisfy $x \leq b < x + y \leq c$, and $|V_0(y)| \geq R$, then*

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b(R - 3x - 3y)}{n^2}$$

6 Growing to $\Theta\left(n^{1/3}\right)$

This section proves Proposition 21. It makes a lot of use of the results of Schramm in "Compositions of random transpositions" [19]. A number of definitions are needed to state his main result.

Definition 30. *If $(\bar{X}_t)_{t\geq 0}$ is the random transposition walk, define G_t to be the graph on $\{1, 2, \ldots, n\}$ such that $\{u, v\}$ is an edge in G_t if and only if the random transposition (u, v) has appeared in the first t steps of our walk. Furthermore, let W_t denote the set of vertices of the largest component of G_t.*

Note that the behavior of the W_t defined above is well-understood; indeed,by an Erdős-Rényi theorem (see for example [1]), if $t = cn$, then

$$\frac{|W_t|}{n} \to z(2c) \tag{6.1}$$

in probability as $n \to \infty$, where $z(s)$ is the positive solution of $1 - z = e^{-zs}$.

Definition 31. *The Poisson-Dirichlet $(PD(1))$ distribution is a probability measure on the infinite dimensional simplex $\Omega = \{(x_1, x_2, \ldots) \,|\, \sum_{i=0}^{\infty} x_i = 1\}$. Sample from this simplex as follows: let $U_1, U_2, \ldots$ be an i.i.d sequence of random variables uniform on $[0, 1]$. Then, set $x_1 = U_1$, and recursively,*

$$x_j = U_j \left(1 - \sum_{i=1}^{j-1} x_i\right)$$

Let (y_i) be the (x_i) sorted in nonincreasing order; then, the $PD(1)$ distribution is defined as the law of (y_i).

The main theorem (Theorem 1.1) of Schramm's paper [19] can now be stated. This remarkable result was proved using the tools of graph theory and coupling. A clever lemma showing that vertices that start in 'sufficiently large' cycles are likely to end up in cycles of order n also played a pivotal role (Lemma 35 below is an almost exact reproduction of the result.) The full strength of the result is not needed: while Schramm determines the law of the large parts of X_t, the only fact necessary here is that after a sufficiently long time, these cycles are of order n. For this theorem, treat X_t as an infinite vector by adding infinitely many 0s at the end of it.

Theorem 32 (Schramm). *Let $c > 1/2$, and take $t = cn$. As $n \to \infty$, the law of $\frac{X_t}{|W_t|}$ converges weakly to the $PD(1)$ distribution; that is, for every $\epsilon > 0$, if n is sufficiently large and $t \geq cn$, then there is a coupling of X_t and a $PD(1)$ sample Y such that*

$$P\left\{\left\|Y - \frac{X_t}{|W_t|}\right\|_{\infty} < \epsilon\right\} > 1 - \epsilon \tag{6.2}$$

where $\|\cdot\|_{\infty}$ is the standard l^{∞} distance.

The proof that follows uses Theorem 32 to show that at time $t = n$, more than half the vertices are in cycles of order n with high probability. This is used to 'grow' $m(X_t, Y_t)$ to order $n^{1/3}$, after which the same is done for $s(X_t, Y_t)$. The results for $m(X_t, Y_t)$ are needed before the results for $s(X_t, Y_t)$: since $s(X_t, Y_t) \leq m(X_t, Y_t)$, $m(X_t, Y_t)$ constrains the growth of $s(X_t, Y_t)$ from above. Good control on m is needed before tackling s.

Lemma 33. *Let k be a natural number not dependent on n. For sufficiently large n, that is, for $n > N = N(k)$,*

$$\mathbb{P}\left\{|V_n(n/k)| > n/2\right\} \geq 1 - \frac{6}{k}$$

Proof: For convenience of notation, let $X = (x_1, x_2, \dots)$ be X_n, let $Q = (q_1, q_2, \dots)$ be $\frac{X_n}{|W_n|}$, and let $Y = (y_1, y_2, \dots)$ be a $PD(1)$ sample which is coupled with Q to satisfy Theorem 32 above. With current notation,

$$|V_n(n/k)| = \sum_{x_i \geq \frac{n}{k}} x_i \tag{6.3}$$

For the rest of the proof, fix $\epsilon = \frac{1}{9k}$. First note that Equation (6.1) implies that

$$\frac{|W_n|}{n} \to z(2) \approx 0.797$$

in probability, which means that $\lim_{n\to\infty} \mathbb{P}\left\{|W_n|/n < 3/4\right\} = 0$. Since $Q = X_n/|W_n|$, for sufficiently large n,

$$\mathbb{P}\left\{x_i \geq \frac{3n}{4}q_i \text{ for all } i\right\} > 1 - \epsilon$$

Furthermore, Theorem 32 implies that for sufficiently large n,

$$\mathbb{P}\left\{q_i \geq y_i - \epsilon \text{ for all } i\right\} > 1 - \epsilon$$

Combining the above two equations,

$$\mathbb{P}\left\{x_i \geq \frac{3n}{4}(y_i - \epsilon) \text{ for all } i\right\} > 1 - 2\epsilon \tag{6.4}$$

for sufficiently large n.

Thus, to estimate $|V_n(n/k)|$ it suffices to consider the large parts of the $PD(1)$ sample Y. To that end, define the random variable

$$G_Y(x) = \sum_{y_i \geq x} y_i$$

It is easy to check that $\mathbb{E}[G_Y(x)] = 1 - x$, and therefore $\mathbb{E}[1 - G_Y(x)] = x$. Thus, Markov's inequality implies that

$$\mathbb{P}\{G_Y(x) \leq 3/4\} = \mathbb{P}\{1 - G_Y(x) \geq 1/4\} \leq 4x$$

20

Recall that $\epsilon = \frac{1}{9k}$. Then, combining the above with Equation (6.4),

$$\mathbb{P}\left\{x_i \geq \frac{3n}{4}(y_i - \epsilon) \text{ for all } i, G_Y\left(\frac{13}{9k}\right) \geq \frac{3}{4}\right\} \geq 1 - \frac{6}{k} \qquad (6.5)$$

Finally, assume that $x_i \geq \frac{3n}{4}(y_i - \epsilon)$ for each i, and that $G_Y\left(\frac{13}{9k}\right) \geq \frac{3}{4}$. Then, Equation (6.3) implies that

$$|V_n(n/k)| \geq \sum_{\frac{3n}{4}(y_i - \epsilon) \geq \frac{n}{k}} \frac{3n}{4}(y_i - \epsilon) = \frac{3n}{4}\left(\sum_{y_i \geq 13/9k} y_i - \sum_{y_i \geq 13/9k} \frac{1}{9k}\right)$$

$$\geq \frac{3n}{4}\left(G_Y\left(\frac{13}{9k}\right) - \frac{1}{13}\right) \geq \frac{n}{2} \qquad (6.6)$$

using the fact that there can be at most $\frac{9k}{13}$ values of y_i that are greater than $\frac{13}{9k}$, since the y_i are positive and sum to 1. Therefore, using Equation (6.5), for sufficiently large n

$$\mathbb{P}\left\{|V_n(n/k)| \geq \frac{n}{2}\right\} \geq 1 - \frac{6}{k}$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

The above lemma is now applied to find a t of order n such that the probability of having $m(X_t, Y_t) \geq n^{1/3}$ is sufficiently high. Lemmas 26 and 27 give control of $m(X_t, Y_t)$.

Lemma 34. *If n is sufficiently large and $t \geq 5n$, then*

$$\mathbb{P}\left\{m(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \geq \frac{4}{5}$$

Proof: From Lemma 33, at time $t = n$,

$$\mathbb{P}\left\{|V_t(n/k)| \geq \frac{n}{2}\right\} \geq 1 - \frac{6}{k} \qquad (6.7)$$

Average over the possible values of X_{t-n} to conclude that Equation (6.7) also holds for any time $t \geq n$. Now, for convenience of notation, define

$$S_t = \left\{(X_t, Y_t) \text{ s.t. } \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \qquad (6.8)$$

For sufficiently large n, $n^{1/3} \leq \frac{n}{k}$ for any fixed value of k. Fix $\epsilon > 0$. Then, for $t \geq n$ and sufficiently large n, Equation (6.7) implies that $\mathbb{P}(S_t) \geq 1 - \epsilon$. Furthermore, define

$$A_t = \left\{(X_t, Y_t) \mid m(X_t, Y_t) \geq n^{1/3}\right\} \qquad (6.9)$$

To find a lower bound for $\mathbb{P}(A_t \cap S_t)$ for $t \geq 10n$, note that

$$\mathbb{P}(A_t \cap S_t) \geq \mathbb{P}(A_t) - \mathbb{P}(S_t^c) \geq \mathbb{P}(A_t) - \epsilon \qquad (6.10)$$

21

and hence it suffices to bound $\mathbb{P}(A_t)$. This is done using a recursive argument: at each step t, calculate the probability that $m(X_t, Y_t)$ was too small, but $m(X_{t+1}, Y_{t+1})$ is large enough, and vice versa. The probability of A_t is shown to grow sufficiently quickly with t.

Start by bounding the probability that $m(X_{t+1}, Y_{t+1}) < n^{1/3}$ if $m(X_t, Y_t) \geq n^{1/3}$. By Lemma 26 with $x = n^{1/3}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin A_{t+1} \mid (X_t, Y_t) \in A_t\} \leq \frac{2x^2}{n^2} = \frac{2}{n^{4/3}}$$

and therefore

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin A_{t+1}, (X_t, Y_t) \in A_t\} \leq \frac{2}{n^{4/3}} \mathbb{P}(A_t) \qquad (6.11)$$

Now bound the probability that $m(X_t, Y_t) < n^{1/3}$, while $m(X_{t+1}, Y_{t+1}) \geq n^{1/3}$. In order to bound this in a satisfactory way, enough parts of size $n^{1/3}$ are needed; accordingly, work with $(X_t, Y_t) \in A_t^c \cap S_t$. If $m(X_t, Y_t) < n^{1/3}$ and $\left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}$, then using Lemma 27 with $x = 0, y = n^{1/3}$, and $R = \frac{n}{2}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in A_{t+1} \mid (X_t, Y_t) \in A_t^c \cap S_t\} \geq \frac{2\left(n/2 - 2n^{1/3}\right)}{n^2} \geq \frac{1-\epsilon}{n}$$

for sufficiently large n. Thus, for $t \geq n$, using the fact that $\mathbb{P}(S_t) \geq 1 - \epsilon$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in A_{t+1}, (X_t, Y_t) \notin A_t\} \geq \left(\frac{1-\epsilon}{n}\right) \mathbb{P}(A_t^c \cap S_t)$$
$$\geq \left(\frac{1-\epsilon}{n}\right)(1 - \mathbb{P}(A_t) - \epsilon) \qquad (6.12)$$

for sufficiently large n. Combining Equations (6.11) and (6.12),

$$\mathbb{P}(A_{t+1}) - \mathbb{P}(A_t) \geq -\frac{2}{n^{4/3}} \mathbb{P}(A_t) + \left(\frac{1-\epsilon}{n}\right)(1 - \mathbb{P}(A_t) - \epsilon)$$
$$\geq \frac{1 - \mathbb{P}(A_t) - 3\epsilon}{n}$$

for sufficiently large n and $t \geq n$. Rearranging the above,

$$(1 - 3\epsilon - \mathbb{P}(A_{t+1})) \leq \left(1 - \frac{1}{n}\right)(1 - 3\epsilon - \mathbb{P}(A_t)) \qquad (6.13)$$

and hence using recursion and the lower bound in Equation (6.10),

$$(1 - 3\epsilon - \mathbb{P}(A_t)) \leq \left(1 - \frac{1}{n}\right)^{t-n} \leq e^{-(t-n)/n}$$
$$\Rightarrow \mathbb{P}(A_t \cap S_t) \geq 1 - 4\epsilon - e^{-(t-n)/n}$$

Thus, for $t \geq 5n$, $\mathbb{P}(A_t \cap S_t) \geq 1 - 4\epsilon - e^{-4} \approx 1 - 4\epsilon - 0.018$, and picking ϵ appropriately completes the proof. $\square$

Proposition 21 (Restatement). *For sufficiently large n, and $t \geq 9n$,*

$$\mathbb{P}\left\{s(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\} \geq \frac{1}{2}$$

Proof: This proof is very similar to the one above. Let $t \geq 5n$, and define

$$R_t = \left\{m(X_t, Y_t) \geq n^{1/3}, \left|V_t\left(n^{1/3}\right)\right| \geq \frac{n}{2}\right\}$$

From the above lemma, $\mathbb{P}(R_t) \geq \frac{4}{5}$. Now, define

$$C_t = \left\{(X_t, Y_t) \mid s(X_t, Y_t) \geq n^{1/3}\right\}$$

It is shown below that $\mathbb{P}(C_t \cap R_t) \geq \frac{1}{2}$, which will clearly suffice. Note that for $t \geq 5n$,

$$\mathbb{P}(C_t \cap R_t) \geq \mathbb{P}(C_t) - \frac{1}{5} \tag{6.14}$$

and hence it suffices to find a lower bound on $\mathbb{P}(C_t)$. As above, this is done by finding recursive bounds on the probability of C_{t+1} given the probability of C_t. By Lemma 28 with $x = n^{1/3}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin C_{t+1} \mid (X_t, Y_t) \in C_t\} \leq \frac{4x^2}{n^2} = \frac{4}{n^{4/3}}$$

and therefore

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \notin C_{t+1}, (X_t, Y_t) \in C_t\} \leq \frac{4}{n^{4/3}}\mathbb{P}(C_t) \tag{6.15}$$

Now, assume that $(X_t, Y_t) \in C_t^c \cap R_t$. Then $m(X_t, Y_t) \geq n^{1/3} > s(X_t, Y_t)$ and $V_t\left(n^{1/3}\right) \geq \frac{n}{2}$. Therefore, using Lemma 29 with $x = 0, y = n^{1/3}$, and $R = \frac{n}{2}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in C_{t+1} \mid (X_t, Y_t) \in C_t^c \cap R_t\} \geq \frac{2\left(n/2 - 3n^{1/3}\right)}{n^2} = \frac{1}{n} - \frac{6}{n^{5/3}}$$

Thus, for $t \geq 5n$, using the fact that $\mathbb{P}(R_t) \geq \frac{4}{5}$,

$$\mathbb{P}\{(X_{t+1}, Y_{t+1}) \in C_{t+1}, (X_t, Y_t) \notin C_t\} \geq \left(\frac{1}{n} - \frac{6}{n^{5/3}}\right)\mathbb{P}(C_t^c \cap R_t)$$

$$\geq \left(\frac{1}{n} - \frac{6}{n^{5/3}}\right)\left(\frac{4}{5} - \mathbb{P}(C_t)\right) \tag{6.16}$$

for sufficiently large n. Therefore, combining Equations (6.15) and (6.16) and picking n sufficiently large,

$$\mathbb{P}(C_{t+1}) - \mathbb{P}(C_t) \geq -\frac{2}{n^{4/3}}\mathbb{P}(C_t) + \left(\frac{1}{n} - \frac{6}{n^{5/3}}\right)\left(\frac{4}{5} - \mathbb{P}(C_t)\right)$$

$$\geq \frac{3/4 - \mathbb{P}(C_t)}{n} \tag{6.17}$$

23

for $t \geq 5n$. Rearranging analogously to Equation (6.13),

$$\left(\frac{3}{4} - \mathbb{P}(C_{t+1})\right) \leq \left(1 - \frac{1}{n}\right)\left(\frac{3}{4} - \mathbb{P}(C_t)\right)$$

As before, for $t \geq 9n$, $\mathbb{P}(C_t) \geq \frac{7}{10}$. Combining this with Equation (6.14),

$$\mathbb{P}(C_t \cap R_t) \geq \frac{1}{2}$$

for $t \geq 9n$ and n sufficiently large, as required. $\qquad\square$

7   Growing to $\Theta(n)$

This section proves Proposition 24, which shows that $s(X_t, Y_t)$ can be grown to order n. This section is structured similarly to the previous one: proving a lemma about overall cycle sizes, then a lemma about $m(X_t, Y_t)$, and then finally Proposition 24. Again, use is made of the technical results in Lemmas 25 through 29.

The idea behind the proof is largely based on Lemma 2.3 from "Compositions of random transpositions" [19]. Let ϵ, δ and j be chosen as in Proposition 24. Recall that Definition 23 defines $K = \lceil \log_2(\epsilon\delta n) \rceil$ and

$$a_r = \lceil 2\delta^{-1} 2^{-r} n(\log_2 n - r) \rceil \ \text{ and } \ \tau_r = \sum_{i=j}^{r-1} a_i$$

for r between j and K, with $\tau_j = 0$. Then, define

$$I_r = [\tau_r, \tau_{r+1} - 1] \tag{7.1}$$

and for convenience of notation, define $I_K = \{\tau_K\}$.

As should be clear from the statement of Proposition 24, the argument starts with $s(\pi, \sigma) \geq 2^{j+1}$ and $V_0(2^{j+1}) \geq \delta n$, and shows that at time τ_K, the probability that $s(X_{\tau_K}, Y_{\tau_K})$ is less than $\epsilon\delta n$ is appropriately bounded above. In fact, something stronger is shown: for the intervals I_r as defined above, one 'expects' to have

$$V_t(2^{r+1}) \geq \frac{\delta n}{2}, s(X_t Y_t) \geq 2^r, \text{ and } m(X_t, Y_t) \geq 2^{r+1}$$

for all r between j and K. This would clearly suffice to prove the result.

The first lemma is almost identical to Lemma 2.3 from [19] – it is reproven here for completeness, and to illustrate the technique. This lemma starts with $\sigma \in S_n$, and $\left|V_0(2^{j+1})\right| \geq \delta n$. It gives an upper bound for the expected number of vertices that start in cycles of size at least 2^{j+1}, and that are not in cycles of size $\epsilon\delta n$ at time τ_K. This shows that 'most' vertices that start in cycles of size 2^{j+1} are in cycles of order n at time τ_K.

Lemma 35. *Let $\sigma \in S_n$. Let $\delta \in (0,1)$ be a constant such that $\left|V_0(2^{j+1})\right| \geq \delta n$, and let K and τ_K be defined as they are above and in Definition 23. Fix $\epsilon \in (0, 1/32)$. For the random transposition walk $\left(\bar{X}_t\right)_{t \geq 0}$,*

$$\mathbb{E}\left|V_0(2^{j+1}) \setminus V_{\tau_K}(2\epsilon\delta n)\right| \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| n \tag{7.2}$$

where the constant implied in the $O(1)$ notation is universal.

Proof: Before beginning the proof, consider what is being shown. Starting with a σ such that $\left|V_0(2^{j+1})\right| > \delta n$ means that at least δn of the vertices in σ are in cycles of size at least 2^{j+1}. An upper bound on the expected size of $V_0(2^{j+1}) \setminus V_{\tau_K}(2\epsilon\delta n)$ is needed: that is, an upper bound on the expected number of vertices that started off in cycles of size at least 2^{j+1} in σ, and ended up in cycles of size less than $2\epsilon\delta n$ at time τ_K.

Something stronger is shown: conditioned on $v \in V_0(2^{j+1})$,

$$\mathbb{E}\left|\left\{v \text{ s.t. } C_t(v) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } r \in [j, K]\right\}\right| \leq$$
$$O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| n \tag{7.3}$$

This requires an upper bound on the expected number of vertices that for any time $t \in I_r$ are 'too small' for I_r: they are of size less than 2^{r+1}. Note that the above set includes all vertices such that $C_{\tau_K}(v) < 2\epsilon\delta n \leq 2^{K+1}$, and hence the above bound would suffice.

Three different possibilities are considered. First of all, an upper bound is needed on the expected number of vertices v such that at any point, the cycle containing v is split, and becomes too small. Secondly, all vertices that appear in permutations with an insufficient number of large parts are rejected. And thirdly, it is necessary to bound the possibility that the cycle containing v does not grow sufficiently during I_r. Call the vertices that fall into any of these undesirable categories 'failed.'

In the next three sections, condition on $v \in V_0(2^{j+1})$: that is, assume that v is in a cycle of size 2^{j+1} in σ. This means that v has not failed at time 0.

The cycle containing v becomes too small Let $r \in [j, K-1]$, and let $t \in I_r + 1 = [\tau_r + 1, \tau_{r+1}]$. For $C_t(v)$ to be of size 2^{r+2} by time τ_{r+1}, calculate the probability that for any $t \in I_r + 1$, the cycle containing v is split, and v is then contained in a cycle of size less than 2^{r+2}. To be precise, define F_t to be the set of vertices at time t such that $|C_t(v)| < |C_{t-1}(v)|$ and $|C_t(v)| < 2^{r+2}$. Find the expected size of F_t: by definition, this is the expected number of vertices v, whose cycle is split from time $t-1$ to time t, and which are in cycles of size less than 2^{r+2} at time t. By Lemma 25,

$$\mathbb{E}\left|F_t\right| \leq \frac{2\left(2^{r+2}\right)^2}{n} = \frac{2^{2r+5}}{n}$$

Now, define the cumulative set $\tilde{F}_t = \bigcup_{x=1}^{t} F_x$. This is the set of all vertices up to time t, whose cycles have at any time $x \leq t$ been split into ones that are

'too small.' Clearly,

$$
\begin{aligned}
\mathbb{E}\left|\tilde{F}_{\tau_K}\right| &\leq \sum_{x=1}^{\tau_K} \mathbb{E}\left|F_x\right| \leq \sum_{r=j}^{K-1} a_r \frac{2^{2r+5}}{n} \\
&\leq \sum_{r=j}^{K-1} \lceil 2\delta^{-1}2^{-r}n(\log_2 n - r)\rceil \frac{2^{2r+5}}{n} \\
&\leq \sum_{r=j}^{K-1} \left(2\delta^{-1}2^{-r}n(\log_2 n - r) + 1\right) \frac{2^{2r+5}}{n} \\
&\leq \sum_{r=j}^{K-1} 2^6\delta^{-1}2^r (\log_2 n - r) + \sum_{r=j}^{K-1} \frac{2^{2r+5}}{n}
\end{aligned}
\tag{7.4}
$$

Now,

$$
\sum_{r=j}^{K-1} r2^r = (K-2)2^K - (j-2)2^j
$$

$$
\sum_{r=j}^{K-1} 2^r = 2^K - 2^j
$$

shows that

$$
\mathbb{E}\left|\tilde{F}_{\tau_K}\right| \leq 2^8 \left|\log_2(\epsilon\delta)\right| \epsilon n
\tag{7.5}
$$

Permutations with insufficiently many large parts It is also necessary to rule out vertices in permutations for which the union of the 'large parts' isn't sufficiently high. This will be useful for the next part of the proof. To be more precise, let $t \in I_r$: if $\left|V_t(2^{r+1})\right| < \delta n/2$, and this is the first t for which the inequality holds, then consider all vertices in X_t to have failed, and set $H_t = \{1, \ldots, n\}$. Otherwise, set $H_t = \emptyset$.

Again, define the cumulative set $\tilde{H}_t = \bigcup_{x=0}^{t} H_x$. This is the union of all vertices that up to time t have been in a permutation with insufficiently many large parts, by the above definition. It is clear that this set is either empty, or contains all the vertices. There is no current available upper bound on the expectation for $\tilde{H}_t$; one will be derived after the next section of the proof.

The cycle containing v doesn't grow sufficiently Next, consider how a vertex v might fail at time t, if it does not fall into $\tilde{F}_t$ or $\tilde{H}_{t-1}$. Assume t is the minimal time for which v fails: since failed vertices include all vertices contained in cycles that are 'too small', if $s < t$ and $s \in I_k$ then $|C_s(v)| \geq 2^{k+1}$. Now, assume that t, the first time at which v fails, is in I_r: thus, $t - 1$ is either in I_{r-1} or in I_r. Either way, since it was assumed that v is not in F_t^π, it can't be that $|C_t(v)| < |C_{t-1}(v)|$ and $|C_t(v)| < 2^{r+1}$. Since the vertex v fails at time t, $C_t(v)$ must contain fewer than 2^{r+1} vertices. Combine this with

26

the preceding statement to conclude that $C_{t-1}(v)$ also contains fewer than 2^{r+1} vertices. However, by definition the vertex v did not fail at time $t-1$. This implies $t-1$ must have been in I_{r-1}. Thus, the only remaining times at which vertices could fail are $t = \tau_r$, for $r \in \{j, j+1, \ldots, K\}$. Having conditioned on $v \in V_0(2^{j+1})$, it may be concluded that v can't fail at time $\tau_j = 0$.

Now, define B_r to be the set of vertices at time τ_r that are not in $\tilde{F}^{\tau_r} \cup \tilde{H}^{\tau_r-1}$, such that $|C_{\tau_r}(v)| < 2^{r+1}$, and that have not failed previously. As before, define $\tilde{B}_r = \bigcup_{x=j}^{r} B_x$ and estimate the expected size of B_r.

Condition on $v \notin \tilde{F}_{\tau_r} \cup \tilde{H}_{\tau_r-1}$ and calculate the probability that v fails at τ_r, given that it has not failed up to that time. First, for $t \in I_{r-1}$, $|C_t(v)| \geq 2^r$. Furthermore, since $v \notin \tilde{F}_{\tau_r}$, there was no time between τ_{r-1} and τ_r at which the cycle containing v was split to contain fewer than 2^{r+1} vertices. This means that if v failed at time τ_r, then $C_t(v)$ must have been of size less than 2^{r+1} for all $t \in [\tau_{r-1}, \tau_r - 1]$. Therefore, for $t \in I_{r-1}$,

$$2^r \leq |C_t(v)| < 2^{r+1} \tag{7.6}$$

Furthermore, since v is not in $\tilde{H}_t$ for any $t \in I_{r-1}$, for every $t \in I_{r-1}$, $|V_t(2^r)| \geq \delta n/2$. Consider the probability that from time t to time $t+1$, the cycle containing v is merged with a cycle of size at least 2^r. By (7.6) above, the size of $C_t(v)$ is at least 2^r, so such a merge would result in $C_{t+1}(v) \geq 2^{r+1}$. Using the above reasoning implies that $|C_{\tau_r}(v)| \geq 2^{r+1}$, and therefore v does not fail at time τ_r. Now, again by (7.6), the cycle containing v is of size at most 2^{r+1}. Since $|V_t(2^r)| \geq \delta n/2$, this means the union of the cycles disjoint from $C_t(v)$ of size at least 2^r contains at least $\delta n/2 - 2^{r+1}$ vertices. Now, since $r \leq K = \lceil \log_2(\epsilon \delta n) \rceil$, $2^{r+1} \leq 2^{K+1} \leq 4\epsilon \delta n$, and since $\epsilon < 1/32$,

$$\frac{\delta n}{2} - 2^{r+1} \geq \frac{\delta n}{4}$$

Thus, the union of the cycles of size at least 2^r disjoint from $C_t(v)$ is of size at least $\delta n/4$, and therefore

$$\mathbb{P}\{C_t(v) \text{ merges with a cycle of size} \geq 2^r\} \geq 2 \cdot 2^r \frac{\delta n/4}{n^2} = 2^{r-1}\delta n^{-1}$$

Clearly, for v to be in B_r, it cannot be that $C_t(v)$ merges with a cycle of size $\geq 2^r$ for any $t \in I_{r-1}$. Therefore,

$$\mathbb{P}\{v \in B_r\} \leq \left(1 - 2^{r-1}\delta n^{-1}\right)^{a_{r-1}} \tag{7.7}$$
$$\leq \exp(-2^{r-1}\delta n^{-1}a_{r-1})$$

and since $a_{r-1} \geq 2\delta^{-1}2^{-r+1}n(\log_2 n - r + 1)$,

$$\mathbb{P}\{v \in B_r\} \leq e^{2(r-1-\log_2 n)}$$

Now, $r - 1 \leq K - 1 \leq \log_2(\epsilon \delta n)$, and therefore, $\log_2 n - r + 1 \leq 0$. Thus,

$$\mathbb{P}\{v \in B_r\} \leq e^{2(\log_2 n - r + 1)} \leq 2^{\log_2 n - r + 1} = \frac{2^{r-1}}{n}$$

This yields

$$\mathbb{E}\,|B_r| \le n\mathbb{P}\{v \in B_r\} = 2^{r-1}$$

and therefore,

$$\mathbb{E}\left|\tilde{B}_K\right| \le \sum_{r=j}^{K} \mathbb{E}\,|B_r| \le \sum_{r=j}^{K} 2^{r-1}$$
$$\le 2^K \le \epsilon\delta n + 1 \tag{7.8}$$

Finally, bound the expected size of H_t, the set of vertices in permutations with insufficiently many large parts. Recall that for $t \in I_r$, if $\left|V_t(2^{r+1})\right| < \delta n/2$ and t was the first time this inequality held, H_t was defined to be the set of all vertices, and it was otherwise defined to be the empty set. If H_t is non-empty, then the set of vertices in X_t that are in cycles of size less than 2^{r+1} has size at least $n - \delta n/2 \ge \delta n/2$. Now, consider v in H_t such that $|C_t(v)| < 2^{r+1}$. By definition, v has failed by time t, and v is not in H_s for any $s < t$. Therefore, each such vertex is in $\tilde{F}_t \cup \tilde{B}_r$. Thus,

$$\mathbb{E}\left|\tilde{H}_{\tau_K}\right| \le n\mathbb{P}\left\{\left|\tilde{F}_{\tau_K} \cup \tilde{B}_K\right| \ge \delta n/2\right\} \le 2\delta^{-1}\mathbb{E}\left|\tilde{F}_{\tau_K} \cup \tilde{B}_K\right|$$

so using (7.5) and (7.8) above,

$$\mathbb{E}\left|\tilde{H}_{\tau_K}\right| \le 2\delta^{-1}(\epsilon\delta n + 1 + 2^8\,|\log(\epsilon\delta)|\,\epsilon n)$$
$$\le (2^9 + 1)\epsilon\delta^{-1}\,|\log_2(\epsilon\delta)|\,n \tag{7.9}$$

as desired. Finally, adding up the expectations for H_{τ_K}, B_K and F_{τ_K} in (7.9), (7.8) and (7.5) completes the proof. $\qquad\square$

The next lemma is similar. It shows that $m(X_t, Y_t)$ becomes sufficiently large at time τ_K. The proof is almost entirely analogous; the only substantial difference is in the bound for the probability of X_t having insufficiently many 'large parts.' For this bound, Equation (7.9) above has to be used. Lemmas 26 and 27 will also be used.

Lemma 36. *Assume $\rho(\sigma, \tau) = 1$. Let j be a natural number such that $m(\sigma, \tau) \ge 2^{j+1}$, and let $\delta \in (0, 1]$ be a constant such that $\left|V_0(2^{j+1})\right| \ge \delta n$ and $m(\pi, \sigma) \ge 2^{j+1}$. Let K and τ_K be defined as above, and let $\epsilon \in (0, 1/16)$. Then,*

$$\mathbb{P}\{m(X_{\tau_K}, Y_{\tau_K}) < 2\epsilon\delta n\} \le O(1)\delta^{-1}\epsilon\,|\log(\epsilon\delta)| \tag{7.10}$$

where the constant implied in the $O(1)$ notation is universal.

Proof: This proof is almost exactly analogous to the previous one, except that instead of keeping track of failed vertices, failed pairs of partitions will be tracked. Something stronger is shown:

$$\mathbb{P}\{m(X_t, Y_t) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } j \in [j, K]\} \le O(1)\delta^{-1}\epsilon\,|\log(\epsilon\delta)| \tag{7.11}$$

Again, the argument requires upper bounds on three different cases: the one where $m(X_t, Y_t)$ shrinks to become too small at time t, the one where X_t doesn't have sufficiently many large parts, and the one where $m(X_t, Y_t)$ fails to grow sufficiently during I_r. The only major difference in the proof is use of the bound from Lemma 35 to bound the probability of X_t having insufficiently many large parts.

Since the quantities specified are precisely analogous, use the names $\mathcal{F}_t, \mathcal{B}_t$ and $\mathcal{H}_t$.

Probability $m(X_t, Y_t)$ gets too small during I_r For $t \in I_r + 1$, define $\mathcal{F}_t$ to be the set of pairs (X_t, Y_t) such that $m(X_t, Y_t) < m(X_{t-1}, X_{t-1})$ and $m(X_t, Y_t) < 2^{r+2}$. Apply Lemma 26 above. Let $x = \min(2^{r+2}, m(X_{t-1}, Y_{t-1}))$. Then, $x \leq m(X_{t-1}, Y_{t-1})$, and therefore from Lemma 26, the probability that $m(X_t, Y_t)$ is less than x is bounded above by $\frac{2x^2}{n^2}$. By definition of $\mathcal{F}_t$, this means that

$$\mathbb{P}\{\mathcal{F}_t\} \leq \frac{2x^2}{n^2} \leq 2\frac{\left(2^{r+2}\right)^2}{n^2} = \frac{2^{2r+5}}{n^2}$$

Define the cumulative set $\tilde{\mathcal{F}}_t = \bigcup_{x=1}^t \mathcal{F}_x$. Therefore,

$$\mathbb{P}\{\tilde{\mathcal{F}}_{\tau_K}\} \leq \sum_{x=1}^{\tau_K} \mathbb{P}\{\mathcal{F}_x\} \leq \sum_{r=j}^{K-1} a_r \frac{2^{2r+5}}{n^2}$$

and doing a calculation almost identical to (7.5),

$$\mathbb{P}\{\tilde{\mathcal{F}}_{\tau_K}\} \leq 2^8 \epsilon \left|\log_2(\epsilon\delta)\right| \tag{7.12}$$

Note that the only difference in the calculation was an extra factor of n in the denominator.

Probability X_t doesn't have enough large parts Define $\mathcal{H}_t$ almost exactly as H_t in the last lemma, except that instead of making it a set of vertices, let it be a set of pairs (X_t, Y_t). (X_t, Y_t) is included in $\mathcal{H}_t$ precisely when X_t doesn't have enough large parts: that is, if $t \in I_r$, then (X_t, Y_t) is in $\mathcal{H}_t$ if $\left|V_t(2^{r+1})\right| < \delta n/2$, and t is the first time for which this inequality holds. Define $\tilde{\mathcal{H}}_t$ as usual to be the cumulative set.

Clearly, if $(X_t, Y_t) \in \mathcal{H}_t$, then H_t contains n vertices, and otherwise H_t is empty. Since $\left|V_0(2^{j+1})\right| \geq \delta n$, the results derived in Lemma 35 can be used. Therefore,

$$\mathbb{P}\{\mathcal{H}_t\} = \frac{1}{n}\mathbb{E}\left|H_t\right|$$

and thus, from (7.9) above,

$$\mathbb{P}\{\mathcal{H}_t\} \leq (2^9 + 1)\epsilon\delta^{-1}\left|\log_2(\epsilon\delta)\right| \tag{7.13}$$

29

Probability $m(X_t, Y_t)$ doesn't grow sufficiently during I_r As before, the only remaining times that (X_t, Y_t) can fail is at times τ_r. Accordingly, define $\mathcal{B}_r$ to be those pairs (X_{τ_r}, Y_{τ_r}) that are not in $\tilde{\mathcal{F}}_{\tau_r}$ or $\tilde{\mathcal{H}}_{\tau_r-1}$, such that $m(X_{\tau_r}, Y_{\tau_r}) < 2^{r+1}$ and that have not failed previously. As before, if (X_{τ_r}, Y_{τ_r}) is in $\mathcal{B}_r$, then it had not failed in I_{r-1}, and therefore, for $t \in I_{r-1}$, $m(X_t, Y_t) \geq 2^r$. Furthermore, since (X_{τ_r}, Y_{τ_r}) is not in $\tilde{\mathcal{F}}_{\tau_r}$, it must be that $m(X_t, Y_t)$ is less than 2^{r+1} for $t \in I_{r-1}$. Thus, for $t \in I_{r-1}$,

$$2^r \leq m(X_t, Y_t) < 2^{r+1} \tag{7.14}$$

Furthermore, since $\mathcal{B}_r$ is disjoint from $\mathcal{H}_{\tau_r-1}$, for every $t \in I_{r-1}$, $|V_t(2^r)| \geq \delta n/2$. Since $m(X_t, Y_t) \geq 2^r$, Lemma 27 holds with $R = \delta n/2$ and $x = y = 2^r$. Let $c = m(X_t, Y_t)$. Thus, for any $t \in I_r$,

$$\mathbb{P}\{m(X_{t+1}, Y_{t+1}) \geq 2^{r+1}\} \geq \frac{2c(\delta n/2 - 2c)}{n^2} \geq \frac{2^{r+1}(\delta n/2 - 2^{r+2})}{n^2}$$

and since $\epsilon < \frac{1}{32}$, and $r \leq K \leq \log_2(\epsilon\delta n) + 1$, $\delta n/2 - 2^{r+2} \geq \delta n/4$. Thus,

$$\mathbb{P}\{m(X_{t+1}, Y_{t+1}) \geq 2^{r+1}\} \geq 2^{r-1}\delta n^{-1}$$

Finally, the probability of $\mathcal{B}_r$ is the probability that $m(X_{t+1}, Y_{t+1})$ isn't at least 2^{r+1} for any $t \in I_r$, and therefore,

$$\mathbb{P}\{\mathcal{B}_r\} \leq \left(1 - 2^{r-1}\delta n^{-1}\right)^{a_{r-1}}$$

and since this is precisely the same inequality as in (7.7),

$$\mathbb{P}\{\mathcal{B}_r\} \leq \frac{2^{r-1}}{n}$$

and hence

$$\mathbb{P}\{\tilde{\mathcal{B}}_K\} \leq \sum_{r=j}^{K} \mathbb{P}\{\mathcal{B}_r\} \leq \sum_{r=j}^{K} \frac{2^{r-1}}{n} \leq \epsilon\delta + \frac{1}{n} \tag{7.15}$$

Thus, adding (7.12), (7.13), and (7.15),

$$\mathbb{P}\{m(X_t, Y_t) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \leq$$
$$(2^9 + 2^8 + 1)\epsilon\delta^{-1}\left|\log_2(\epsilon\delta)\right| \tag{7.16}$$

which is what is needed. $\qquad\square$

The stage is almost set to prove an analogous result for $s(X_t, Y_t)$. As above, the two technical Lemma 28 and 29 are used. As in the previous section, $m(X_t, Y_t)$ must be 'sufficiently large' to allow $s(X_t, Y_t)$ to grow. This is the reason for proving the lemma concerning $m(X_t, Y_t)$ first.

Proposition 24 (Restatement). *Let (X_t, Y_t) be the usual coupling started at $(X_0, Y_0) = (\sigma, \tau)$, where $\rho(\sigma, \tau) \leq 1$. Let j be a number and let $\delta \in (0, 1]$ be a constant such that $\left|V_0(2^{j+1})\right| \geq \delta n$ and $s(\sigma, \tau) \geq 2^{j+1}$. If K and τ_K are defined as in Definition 23 and $\epsilon \in (0, 1/32)$, then*

$$\mathbb{P}\{s(X_{\tau_K}, Y_{\tau_K}) < \epsilon \delta n\} \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right|$$

where the constant implied in the $O(1)$ notation is universal.

Proof of Lemma 24. This proof is analogous to the proof of Lemma 35 and 36, except that the previous two lemmas are used to bound the probability that $s(X_t, Y_t)$ shrinks or grows. As before, a stronger statement is proved:

$$\mathbb{P}\{s(X_t, Y_t) < 2^r \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \leq O(1)\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| \quad (7.17)$$

Again, bounds are needed for a number of different cases: for the probability that $s(X_t, Y_t)$ shrinks to become too small during I_r, the probability that X_t doesn't have enough large parts, and that the probability that $s(X_t, Y_t)$ doesn't grow sufficiently on I_r. Furthermore, note that Lemma 29 requires the assumption that $m(\sigma, \tau) \geq 2x$ to lower bound on the probability that $s(X_1, Y_1) \geq 2x$. Since $s(X_t, Y_t)$ must grow during I_r to be at least 2^{r+1} by τ_{r+1}, $m(X_t, Y_t)$ must be at least 2^{r+1} on I_r. Lemma 36 is used to bound the probability that $m(X_t, Y_t)$ is too small.

The quantities are precisely analogous to the ones in the two similar previous lemmas. Accordingly, name them $\mathscr{F}_t, \mathscr{H}_t$, and $\mathscr{B}_t$, using the same letters but yet another font. The new quantity $\mathscr{M}_t$ is added, as discussed above.

Probability $s(X_t, Y_t)$ gets too small during I_r For $t \in I_r + 1 = [\tau_r + 1, \tau_{r+1}]$, define $\mathscr{F}_t$ to be the set of pairs (X_t, Y_t) such that $s(X_t, Y_t) < s(X_{t-1}, Y_{t-1})$ and $s(X_t, Y_t) < 2^{r+1}$. Apply Lemma 28 above. Define $x = \min(2^{r+1}, s(X_{t-1}, Y_{t-1}))$. Then, $x \leq s(X_{t-1}, Y_{t-1})$, and therefore Lemma 28 applies. Plugging it in, the probability that $s(X_t, Y_t)$ is less than x is at most $\frac{4x^2}{n^2}$. Thus,

$$\mathbb{P}\{\mathscr{F}_t\} \leq \frac{4x^2}{n^2} \leq \frac{4(2^{r+1})^2}{n^2} = \frac{2^{2r+4}}{n^2}$$

Now, define the cumulative set $\tilde{\mathscr{F}}_t = \bigcup_{x=1}^t \mathscr{F}_x$. Then,

$$\mathbb{P}\{\tilde{\mathscr{F}}_{\tau_K}\} \leq \sum_{x=1}^{\tau_K} \mathbb{P}\{\mathscr{F}_x\} \leq \sum_{r=j}^{K-1} a_r \frac{2^{2r+4}}{n^2}$$

Doing a calculation identical to the one in (7.5) and (7.12),

$$\mathbb{P}\{\tilde{\mathscr{F}}_{\tau_K}\} \leq 2^7 \epsilon \left|\log_2(\epsilon\delta)\right| \quad (7.18)$$

31

Probability X_t doesn't have enough large parts For $t \in I_r$, define $\mathscr{H}_t$ very similarly to before, to be the set of (X_t, Y_t) such that $|V_t(2^r)| < \delta n/2$, whenever this is the first t for which this inequality holds. Define $\tilde{\mathscr{H}}_t$ to be the usual cumulative set. Now, from Lemma 36,

$$\mathcal{H}_t = \{(X_t, Y_t) \mid |V_t(2^{r+1})| < \delta n/2\}$$

Since $V_t(2^r) \supseteq V_t(2^{r+1})$, clearly $\mathscr{H}_t \supseteq \mathcal{H}_t$, and therefore, using (7.13)

$$\mathbb{P}\{\tilde{\mathscr{H}}_t\} \leq \mathbb{P}\{\tilde{\mathcal{H}}_t\} \leq (2^9 + 1)\epsilon\delta^{-1} |\log_2(\epsilon\delta)| \tag{7.19}$$

Probability $m(X_t, Y_t)$ is too small For $t \in I_r$, define $\mathscr{M}_t$ to be the set of all (X_t, Y_t) such that $m(X_t, Y_t) < 2^{r+1}$. As usual, define $\tilde{\mathscr{M}}_t$ to be the cumulative set. Since at the start $s(\sigma, \tau) \geq 2^{j+1}$, $m(\sigma, \tau) \geq 2^{j+1}$ is forced. By assumption, $V_0(2^{j+1}) \geq \delta n$, so any inequalities derived in Lemma 36 are in force. Thus, from Equation (7.16),

$$\mathbb{P}\{m(X_t, Y_t) < 2^{r+1} \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \leq (2^9 + 2^8 + 1)\epsilon\delta^{-1} |\log_2(\epsilon\delta)|$$

and clearly, from the definition of $\tilde{\mathscr{M}}^t$,

$$\mathbb{P}\{\tilde{\mathscr{M}}^t\} \leq (2^9 + 2^8 + 1)\epsilon\delta^{-1} |\log_2(\epsilon\delta)| \tag{7.20}$$

Probability $s(X_t, Y_t)$ doesn't grow sufficiently during I_r As before, the only remaining times that $s(X_t, Y_t)$ can fail is at time τ_r. Therefore, define $\mathscr{B}_r$ to be the set of (X_{τ_r}, Y_{τ_r}) that are not in $\tilde{\mathscr{F}}_{\tau_r}$, $\tilde{\mathscr{H}}_{\tau_r - 1}$ or $\tilde{\mathscr{M}}_{\tau_r}$, such that $s(X_{\tau_r}, Y_{\tau_r}) < 2^r$ and that have not failed previously. If (X_{τ_r}, Y_{τ_r}) is in $\mathscr{B}_r$, then it had not failed in I_{r-1}, and therefore for $t \in I_{r-1}$, $s(X_t, Y_t) \geq 2^{r-1}$. Furthermore, since (X_{τ_r}, Y_{τ_r}) is not in $\tilde{\mathscr{F}}_{\tau_r}$, for $t \in I_{r-1}$, $s(X_t, Y_t) < 2^r$. Thus, for $t \in I_{r-1}$,

$$2^{r-1} \leq s(X_t, Y_t) < 2^r \tag{7.21}$$

Furthermore, since $(X_{\tau_r}^\pi, X_{\tau_r}^\sigma)$ is not in $\tilde{\mathscr{M}}_{\tau_r}$, for $t \in I_{r-1}$

$$m(X_t, Y_t) \geq 2^r$$

Finally, since $\mathscr{B}_r$ is disjoint from $\mathscr{H}_{\tau_r - 1}$, for every $t \in I_{r-1}$, $|V_t(2^{r-1})| \geq \delta n/2$. Now apply Lemma 29 with $R = \delta n/2$ and $x = y = 2^{r-1}$. For any $t \in I_r$,

$$\mathbb{P}\{s(X_{t+1}, Y_{t+1}) \geq 2^r\} \geq \frac{2x(R - 3x - 3y)}{n^2} = \frac{2^r(\delta n/2 - 3 \cdot 2^r)}{n^2}$$

Since $r \leq K = \lceil \log_2(\epsilon\delta n) \rceil$, and since $\epsilon < \frac{1}{32}$, $3 \cdot 2^r \leq 6\epsilon\delta n \leq \frac{\delta n}{4}$. Thus,

$$\mathbb{P}\{s(X_{t+1}, Y_{t+1}) \geq 2^r\} \geq 2^{r-2}\delta n^{-1}$$

Finally, the probability of $\mathscr{B}_r$ is the probability that $s(X_{t+1}, Y_{t+1})$ isn't at least 2^{r+1} for any $t \in I_r$, and therefore,

$$\mathbb{P}\{\mathscr{B}_r\} \leq \left(1 - 2^{r-2}\delta n^{-1}\right)^{a_{r-1}} \leq \exp(-2^{r-2}\delta n^{-1}a_{r-1})$$

Now, since $a_{r-1} = \lceil 2\delta^{-1} 2^{-r+1} n (\log_2 n - r + 1) \rceil$,

$$\mathbb{P}\{\mathscr{B}_r\} \le e^{r-1-\log_2 n} \le 2^{r-1-\log_2 n} = \frac{2^{r-1}}{n}$$

using the fact that $r \le K \le \log_2 n + 1$, and hence $r - 1 - \log_2 n \le 0$. Therefore,

$$\mathbb{P}\{\tilde{\mathscr{B}}_r\} \le \sum_{r=j}^{K} \mathbb{P}\{\mathscr{B}_r\} \le \sum_{r=j}^{K} \frac{2^{r-1}}{n} \le \frac{2^r}{n} \le \epsilon\delta + \frac{1}{n} \qquad (7.22)$$

Now, adding (7.18), (7.19), (7.20) and (7.22),

$$\mathbb{P}\{s(X_t, Y_t) < 2^r \text{ for any } t \in I_r, \text{ for } r \in [j, K]\} \le 2^{11}\delta^{-1}\epsilon \left|\log(\epsilon\delta)\right| \qquad (7.23)$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

Remark 37. Assiduously tracking down all the constants in the above argument shows that the mixing time was bounded above by $2^{25} n \log n$ or so. This, of course, is very far from the correct answer of $\frac{1}{2} n \log n$. While this argument can almost certainly be mildly tweaked to give a less intimidating answer such as $10 n \log n$, it is unlikely that it could be manipulated to give the right constant.

8   Technical Lemmas

In this section, the technical results in Lemmas 25 through 29 are proved. For the convenience of the reader, the results are restated.

Lemma 25 (Restatement). *Let σ be in S_n, and let $(\bar{X}_t)_{t \ge 1}$ be the random transposition walk starting at σ. Then, the expected number of v such that $|C_1(v)| < |C_0(v)|$ and $|C_1(v)| < x$ is no greater than $\frac{x^2}{n}$.*

Proof: Let $\sigma = (a_1, \ldots, a_m)$. Clearly, the only way that $|C_1(v)| < |C_0(v)|$ is if the cycle containing v is split; furthermore, the only way that $|C_1(v)| < x$ is if v winds up in a piece of size less than x. The 'ordered' splitting formula shows that the probability of splitting a_i into $(r, a_i - r)$ is $\frac{a_i}{n^2}$. Consider the cases where either $r < x$ or $a_i - r < x$. Thus, summing over the possible a_i,

$$\mathbb{E}\left|\{v \text{ s.t. } |C_1(v)| < |C_0(v)|, |C_1(v)| < x\}\right| \le$$

$$\sum_{i=1}^{m} \left(\sum_{r=1}^{x-1} r \cdot \frac{a_i}{n^2} + \sum_{r=a_i-x+1}^{a_i-1} (a_i - r) \cdot \frac{a_i}{n^2} \right)$$

It's clear that

$$\sum_{r=a_i-x+1}^{a_i-1} (a_i - r) \frac{a_i}{n^2} = \sum_{r=1}^{x-1} r \cdot \frac{a_i}{n^2} = \frac{a_i}{n^2} \sum_{c=1}^{x-1} r \le \frac{a_i x^2}{2n^2}$$

Therefore,

$$\mathbb{E}\,|\{v \text{ s.t. } |C_1(v)| < |C_0(v)|, |C_1(v)| < x\}| \leq \sum_{i=1}^{m} \frac{a_i x^2}{n^2} = \frac{x^2}{n^2} \sum_{i=1}^{m} a_i$$

$$= \frac{x^2}{n^2} \cdot n = \frac{x^2}{n}$$

as required. $\qquad\qquad\Box$

For the next four lemmas, let (X_t, Y_t) be our usual coupling starting at (σ, τ), where $\rho(\sigma, \tau) = 1$, $s(\sigma, \tau) = b$ and $m(\sigma, \tau) = c$. For these proofs, it will be useful to reference the original definition of the coupling and the possible pairs (X_1, Y_1) in Definition 13.

Lemma 26 (Restatement). *If $x \leq c$, then*

$$\mathbb{P}\{m(X_1, Y_1) < x\} \leq \frac{2x^2}{n^2}.$$

Proof: Let us assume without loss of generality that

$$\sigma = (a_1, \ldots, a_n, b, c) \qquad\qquad\qquad (8.1)$$
$$\tau = (a_1, \ldots, a_n, b + c)$$

Consider how $m(X_1, Y_1)$ could be smaller than c. Note that performing an operation involving only the a_i on σ and τ, then X_1 and Y_1 will still differ in b, c and $b + c$, so $m(X_1, Y_1) = c$. Furthermore, merging a_i with b in σ and a_i with $b + c$ in τ, then X_1 and Y_1 will differ in the parts $(b + a_i, c, b + c + a_i)$, which are greater, respectively, than $(b, c, b+c)$. This means that $m(X_1, Y_1) \geq c$. Similar reasoning holds for merging a_i with c in σ, and hence these cases do not contribute to $\mathbb{P}\{m(X_1, Y_1) < x\}$.

Also, note that if b is split into $\{r, b - r\}$ for $r \leq \frac{b}{2}$, then

$$X_1 = (a_1, \ldots, a_m, r, b - r, c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b + c - r)$$

Clearly, $c \geq b \geq b - r$, and therefore $m(X_1, Y_1) = c$. Thus, m cannot decrease if b is split in σ. This gives cases: splitting c in σ, and merging b and c in σ. The cases in which the coupling meets can be ignored, since $m(\alpha, \alpha) = n \geq c$, and hence these cases do not contribute to $\mathbb{P}\{m(X_1, Y_1) < x\}$.

Splitting c in σ: If c is split into $\{r, c - r\}$ for $r \leq \frac{c}{2}$, then

$$X_1 = (a_1, \ldots, a_m, r, b, c - r)$$
$$Y_1 = (a_1, \ldots, a_m, r, b + c - r)$$

Clearly, $m(X_1, Y_1) \geq c - r$. Thus, to have $m(X_1, Y_1) < x$, it must be that $c - r < x$, and thus $r > c - x$. By definition, $r \leq \frac{c}{2}$, and hence

$$c - x < r \leq \frac{c}{2}$$

If $2x < c$, this set contains no elements, so assume for now that $2x \geq c$. Then the number of possible r is at most $\frac{c}{2} - (c - x) = \frac{2x - c}{2}$. Since the probability of splitting c into $\{r, c - r\}$ is at most $\frac{2c}{n^2}$ for each r,

$$\mathbb{P}\left(m(X_1, Y_1) < x, \ c \text{ split in } \sigma\right) \leq \frac{c(2x - c)}{n^2} \leq \frac{x^2}{n^2} \tag{8.2}$$

using the AM-GM inequality and the assumption that $2x - c \geq 0$. Furthermore, the above inequality also holds when $2x < c$, since in that case, the left-hand side is 0.

Merging b and c in σ: If b and c are merged in σ,

$$X_1 = (a_1, \ldots, a_m, b + c)$$
$$Y_1 = (a_1, \ldots, a_m, s, b + c - s)$$

for some $s \leq \frac{b+c}{2}$. Hence, $m(X_1, Y_1) = b + c - r$. Again, to have $b + c - r < x$, it must be that $s > b + c - x$, and the probability of each split is at most $\frac{2(b+c)}{n^2}$. Thus, analogously to above, consider

$$b + c - x < r \leq \frac{b + c}{2}$$

and hence the total number of such s is at most $\frac{2x - (b+c)}{2}$ if $2x \geq b + c$, and 0 otherwise. Therefore, if $2x \geq b + c$,

$$P\{m(X_1, Y_1) < x, \ b \text{ and } c \text{ merged in } \pi\} \leq \frac{(b + c)(2x - (b + c))}{n^2} \leq \frac{x^2}{n^2} \tag{8.3}$$

again using AM-GM. This clearly also holds for $2x < b + c$.

Finally, adding (8.2) and (8.3),

$$P\{m(X_1, Y_1) < x\} \leq \frac{2x^2}{n^2}$$

as required. $\qquad\square$

Lemma 27 (Restatement). *If $x \leq c$, and $|V_0(y)| \geq R$, then*

$$\mathbb{P}\{m(X_1, Y_1) \geq x + y\} \geq \frac{2c(R - 2c)}{n^2}$$

35

Proof: Consider both the possibilities that

$$\pi = (a_1, \ldots, a_m, b, c) \tag{8.4}$$
$$\sigma = (a_1, \ldots, a_m, b + c)$$

and that

$$\pi = (a_1, \ldots, a_m, b + c) \tag{8.5}$$
$$\sigma = (a_1, \ldots, a_m, b, c)$$

with $b \leq c$, since $V_t(y)$ is defined for X_t and not Y_t, and therefore the symmetry breaks down. Merging c with an $a_i \geq y$ will result in $m(X_1, Y_1) = c + a_i \geq x + y$. To calculate the probability of such a merge, the sum of these a_i is needed.

In both cases (8.4) and (8.5), since σ and τ agree on the a_i,

$$\sum_{a_i \geq y} a_i \geq |V_0(y)| - (b + c) \geq R - 2c \tag{8.6}$$

using Remark 20. For case (8.4), merging c and some $a_i \geq y$ in σ gives

$$X_1 = (a'_1, \ldots, a'_{m-1}, b, c + a_i)$$
$$Y_1 = (a'_1, \ldots, a'_{m-1}, b + c + a_i)$$

where $\{a'_1, \ldots, a'_{m-1}\} = \{a_1, \ldots, a_m\}/\{a_i\}$. Clearly, $c + a_i \geq b$, and therefore $m(X_1, Y_1) = c + a_i \geq x + y$. The probability of merging c with a_i in σ is $\frac{2ca_i}{n^2}$, and thus

$$\mathbb{P}\{m(X_1^\pi, X_1^\sigma) \geq x + y\} \geq \sum_{a_i \geq y} \frac{2ca_i}{n^2} = \frac{2c}{n^2} \sum_{a_i \geq y} a_i \geq \frac{2c(R - 2c)}{n^2}$$

using Equation (8.6) for the last inequality. Thus, in case (8.4) the proof is finished. Furthermore, since Equation (8.6) is symmetric for the cases (8.4) and (8.5), the second case is completely analogous. $\square$

Lemma 28 (Restatement). *If $x \leq b$, then*

$$\mathbb{P}\{s(X_1, Y_1) < x\} \leq \frac{4x^2}{n^2}$$

Proof: For simplicity, assume without loss of generality that π and σ satisfy (8.1) above. In the same way as in Lemma 26 above, any operations involving a_i cannot make $s(X_1, Y_1)$ smaller than b. Thus, the operations that might produce $s(X_1, Y_1) < x$ involve either splitting b in σ, splitting c in σ, or merging b and c in σ. Consider these cases separately. In the same way as before, the cases where the coupling meets can be ignored.

Splitting b in σ: Recall that if b is split into $\{r, b - r\}$ for $r \leq \frac{b}{2}$, then

$$X_1 = (a_1, \ldots, a_m, r, b - r, c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b + c - r)$$

Thus, $s(X_1, Y_1) = \min(b - r, c) = b - r$. To have $s(X_1, Y_1) < x$, $b - r < x$ is needed. Hence, consider r such that

$$b - x < r \leq \frac{b}{2}$$

If $2x < b$, this set contains no elements, so assume $2x \geq b$. Clearly, the above set is of size at most $x - \frac{b}{2}$. The probability of splitting b into $(s, b - s)$ is at most $\frac{2b}{n^2}$ for each $s \leq \frac{b}{2}$, and therefore

$$\mathbb{P}\{s(X_1, Y_1) < x, \ b \text{ split in } \sigma\} \leq \frac{2b}{n^2}\left(x - \frac{b}{2}\right) = \frac{b(2x - b)}{n^2} \leq \frac{x^2}{n^2} \qquad (8.7)$$

using AM-GM and the assumption that $2x \geq b$ for the last inequality. This clearly also holds if $2x < b$, since in that case the left-hand side is 0.

Splitting c in σ: This calculation is very similar to the above. The probability that c is split into $\{r, c - r\}$, where $r \leq \frac{c}{2}$ and $c - r < x$ is needed. Again, consider

$$c - x < r \leq \frac{c}{2}$$

and since the probability of a particular split is at most $\frac{2c}{n^2}$, assuming that $2x \geq c$, the total probability of all these cases is at most

$$\mathbb{P}\{s(X_1, Y_1) < x, \ c \text{ split in } \sigma\} \leq \frac{c(2x - c)}{n^2} \leq \frac{x^2}{n^2} \qquad (8.8)$$

which again holds trivially when $2x < c$.

Merging b and c in σ: Recall that merging b and c in σ is coupled with splitting $b + c$ into $\{r, b + c - r\}$ in τ, where each split in τ occurs with the probability that it has not already been coupled with a split of b or c in σ. Thus, in this case,

$$X_1 = (a_1, \ldots, a_m, b + c)$$
$$Y_1 = (a_1, \ldots, a_m, r, b + c - r)$$

Assuming as usual that $r \leq \frac{b+c}{2}$, $s(X_1, Y_1) = r$. Now calculate the probability that $r < x$. Define

$$P_r = \mathbb{P}\{b \text{ and } c \text{ merge in } \pi, \ b + c \text{ splits into } \{r, b + c - r\} \text{ in } \sigma\}$$

and bound P_r for various values of r. Consider three different cases:

37

- $r < \frac{b}{2}$: In this case, splitting $b + c$ into $\{r, b + c - r\}$ in τ is coupled with both spliting b into $\{r, b - r\}$ in σ and with splitting c into $\{r, c - r\}$ in σ. Thus,

$$P_r = \frac{2(b + c)}{n^2} - \frac{2b}{n^2} - \frac{2c}{n^2} = 0 \tag{8.9}$$

- $\frac{b}{2} \leq r < \frac{c}{2}$: In this case, splitting $b + c$ into $\{r, b + c - r\}$ in τ is coupled with splitting c into $\{r, c - r\}$ in σ. Thus,

$$P_r \leq \frac{2(b + c)}{n^2} - \frac{2c}{n^2} = \frac{2b}{n^2}$$

- $\frac{c}{2} \leq r$: In this case, splitting $b + c$ into $\{r, b + c - r\}$ in τ isn't coupled with any splits in σ. Hence,

$$P_r \leq \frac{2(b + c)}{n^2}$$

Therefore, the reasoning above shows

$$P_r \leq \begin{cases} 0 & r < \frac{b}{2} \\ \frac{2b}{n^2} & \frac{b}{2} \leq r < \frac{c}{2} \\ \frac{2b + 2c}{n^2} & \frac{c}{2} \leq r \leq \frac{b + c}{2} \end{cases} \implies P_r \leq \frac{4r}{n^2} \tag{8.10}$$

where the right-hand inequality uses the fact that $b \leq c$. Therefore,

$$\mathbb{P}\{s(X_1, Y_1) < x, \, b \text{ and } c \text{ merge in } \sigma\} = \sum_{r=1}^{x-1} P_r \leq \sum_{r=1}^{x-1} \frac{4r}{n^2} = 4 \frac{x(x-1)}{2n^2}$$

$$\leq \frac{2x^2}{n^2} \tag{8.11}$$

Adding Equations (8.7), (8.8) and (8.11) gives

$$\mathbb{P}\{s(X_1, Y_1) < x\} \leq \frac{4x^2}{n^2}$$

as required. $\qquad\square$

Lemma 29 (Restatement). *If x and y satisfy $x \leq b < x + y \leq c$, and $|V_0(y)| \geq R$,*

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b(R - 3x - 3y)}{n^2}$$

Proof: Just like in Lemma 27, consider the two possibilities that

$$\sigma = (a_1, \ldots, a_m, b, c) \tag{8.12}$$
$$\tau = (a_1, \ldots, a_m, b + c)$$

38

and that

$$\sigma = (a_1, \ldots, a_m, b + c) \tag{8.13}$$
$$\tau = (a_1, \ldots, a_m, b, c)$$

since $V_t(y)$ depends on X_t and not on Y_t. As in the previous lemma, in both cases (8.12) and (8.13),

$$\sum_{a_i \geq y} a_i \geq |V_0(y)| - (b + c) \geq R - (b + c) \tag{8.14}$$

so case (8.12) may be assumed. Identical arguments will apply for (8.13).

There are two possible ways to have $s(X_1, Y_1) \geq x + y$: either b can merge with an $a_i \geq y$ in σ, or b and c can merge in σ, while $b + c$ can be split into $\{r, b + c - r\}$ in τ, where $r \geq x + y$. Consider those cases separately.

Merging b and $a_i \geq y$ in σ: Note that if b and a_i are merged in σ, then

$$X_1 = (a'_1, \ldots, a'_{m-1}, b + a_i, c)$$
$$Y_1 = (a'_1, \ldots, a'_{m-1}, b + a_i + c)$$

where $\{a'_1, \ldots, a'_{m-1}\} = \{a_1, \ldots, a_m\}/\{a_i\}$. Therefore, $s(X_1, Y_1) = \min(b + a_i, c)$. Since $b \geq x$ and $a_i \geq y$, $b + a_i \geq x + y$. By assumption, $c \geq x + y$, and so $s(X_1, Y_1) \geq x + y$.

The probability of b merging with a particular a_i is $\frac{2ba_i}{n^2}$, and using the bound in Equation (8.14),

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y, b \text{ merges with some } a_i \text{ in } \sigma\} = \sum_{a_i \geq y} \frac{2ba_i}{n^2} = \frac{2b}{n^2} \sum_{a_i \geq y} a_i$$
$$\geq \frac{2b(R - (b + c))}{n^2} \tag{8.15}$$

Merging b and c in σ: If $c < 2x + 2y$, it will later show that the above bound in Equation (8.15) suffices. Therefore, for this case, assume that $c \geq 2x + 2y$. Consider the probability of merging b and c in σ, while splitting $b + c$ in τ into $\{r, b + c - r\}$, where $r \geq x + y$.

Let P_r be defined as in Equation (8.9). Now a lower bound on

$$\sum_{r \geq x + y} P_r = \mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma\}$$

$$- \mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\} - \sum_{r < x + y} P_r$$

is needed. The above equality follows because merging b and c in σ is always either coupled with splitting $b + c$ into $\{r, b + c - r\}$ in τ, or staying at τ. Here is a lower bound for the right-hand side.

To start, $c \geq 2x + 2y > 2b$. By Equation (3.5),

$$\mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\} = \min\left(p, \frac{1}{n}\right)$$

where $p = \mathbb{P}(b + c \text{ split into } \{b, c\} \text{ in } \tau, \text{ staying at } \sigma)$. Now, since $c > 2b$,

$$p = \mathbb{P}(b + c \text{ split into } \{b, c\} \text{ in } \tau) - \mathbb{P}(c \text{ split into } \{b, c - b\} \text{ in } \sigma)$$
$$= \frac{2(b + c)}{n^2} - \frac{2c}{n^2} = \frac{2b}{n^2}$$

and hence, since $b \leq \frac{n}{2}$,

$$\mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\} = \min\left(\frac{2b}{n^2}, \frac{1}{n}\right) = \frac{2b}{n^2}$$

Furthermore, since $x + y \leq \frac{c}{2}$, Equation (8.10) above implies that if $r < x + y$ then $P_r \leq \frac{2b}{n^2}$, and therefore

$$\mathbb{P}\{\text{merge } b \text{ and } c \text{ in } \sigma, \text{ stay at } \tau\} + \sum_{r < x+y} P_r \leq \frac{2b(x + y)}{n^2}$$

Thus, since the probability of merging b and c is $\frac{2bc}{n^2}$,

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y, b \text{ and } c \text{ merge in } \sigma\} = \sum_{r \geq x+y} P_r \geq \frac{2bc}{n^2} - \frac{2b(x + y)}{n^2}$$
$$= \frac{2b(c - x - y)}{n^2} \qquad (8.16)$$

Combining all this information, if $c < 2x + 2y$, then Equation (8.15) shows that

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b(R - (b + c))}{n^2} \geq \frac{2b(R - 3x - 3y)}{n^2}$$

using the fact that $b \leq x + y$. Furthermore, if $c \geq 2x + 2y$, then combining Equation (8.15) and (8.16),

$$\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b\left(R - (b + c)\right)}{n^2} + \frac{2b\left(c - x - y\right)}{n^2}$$
$$\geq \frac{2b(R - 2x - 2y)}{n^2}$$

Hence, in either case $\mathbb{P}\{s(X_1, Y_1) \geq x + y\} \geq \frac{2b(R - 2x - 2y)}{n^2}$, completing the proof. $\qquad \square$

References

[1] Noga Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1992. With an appendix by Paul Erdős, A Wiley-Interscience Publication.

[2] Nathanael Berestycki, Oded Schramm, and Ofer Zeitouni. Mixing times for random k-cycles and coalescence-fragmentation chains. `http://arxiv.org/abs/1001.1894`. [Online; accessed 4-July-2011].

[3] Andrei Z. Broder. Unpublished manuscript. 1985.

[4] R. Bubley and M. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *Proceedings of the 38th Annual Symposium on Foundations of Computer Science*, pages 223–, Washington, DC, USA, 1997. IEEE Computer Society.

[5] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.

[6] Persi Diaconis, Eddy Mayer-Wolf, Ofer Zeitouni, and Martin P. W. Zerner. The Poisson-Dirichlet law is the unique invariant distribution for uniform split-merge transformations. *Ann. Probab.*, 32(1B):915–938, 2004.

[7] Persi Diaconis and Laurent Saloff-Coste. Comparison techniques for random walk on finite groups. *Ann. Probab.*, 21(4):2131–2156, 1993.

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

[9] W. Doeblin. Exposé de la théorie des chaînes simple constantes de Markov à un nombre fini d'états. *Rev. Math. Union Interbalkan*, 2:77–105, 1938.

[10] David Griffeath. A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:95–106, 1974/75.

[11] T.P. Hayes and E. Vigoda. A non-markovian coupling for randomly sampling colorings. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 618–627, oct. 2003.

[12] J. Jonasson. Mixing time bounds for overlapping cycles shuffles. *Electronic Journal of Probability*, 16:1281–1295, 2011.

[13] Yevgeniy Kovchegov and Robert Burton. Mixing times via super-fast coupling. `arXiv:0912.2759v1`. [Online; accessed 4-July-2011].

[14] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[15] Torgny Lindvall. *Lectures on the coupling method*. Dover Publications Inc., Mineola, NY, 2002. Corrected reprint of the 1992 original.

[16] Peter Matthews. A strong uniform time for random transpositions. *J. Theoret. Probab.*, 1(4):411–423, 1988.

[17] Yuval Peres. Mixing for markov chains and spin systems. `http://www.stat.berkeley.edu/users/peres/ubc.pdf`. [Online; accessed 4-July-2011].

[18] J. W. Pitman. On coupling of Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 35(4):315–322, 1976.

[19] Oded Schramm. Compositions of random transpositions. *Israel Journal of Mathematics*, vol. 147:pp.221–243, 2005.

[20] Hermann Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.