Probability Theory and Related Fields
https://doi.org/10.1007/s00440-026-01479-9

RESEARCH

Hitting time mixing for the random transposition walk

Vishesh Jain[1] · **Mehtaab Sawhney**[2]

Received: 24 February 2025 / Accepted: 7 March 2026

Abstract

Consider shuffling a deck of n cards, labeled 1 through n, as follows: at each time step, pick one card uniformly with your right hand and another card, independently and uniformly with your left hand; then swap the cards. How long does it take until the deck is close to random? Diaconis and Shahshahani showed that this process undergoes cutoff in total variation distance at time $t = \lfloor (n \log n)/2 \rfloor$. Confirming a conjecture of N. Berestycki, we prove the definitive "hitting time" version of this result: let τ denote the first time at which all cards have been touched. The total variation distance between the stopped distribution at τ and the uniform distribution on permutations is $o(1)$; this is best possible, since at time $\tau - 1$, the total variation distance is at least $(1 + o(1))e^{-1}$.

Mathematics Subject Classification 60J10 · 60B15

1 Introduction

The object of study in this paper is the random transposition walk on the symmetric group $\mathfrak{S}_n$. Informally, viewing elements of $\mathfrak{S}_n$ as possible permutations of n cards, this is the Markov chain with the following transitions: at each time step, pick one card uniformly with your right hand and another card, independently and uniformly with your left hand (note that with probability $1/n$, the cards will be the same); then swap the cards. From this point of view, perhaps the most natural question to ask is the following: what is the "first time" at which the cards are (approximately) shuffled?

✉ Vishesh Jain
visheshj@uic.edu

Mehtaab Sawhney
m.sawhney@columbia.edu

[1] Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, Chicago, IL 60607, USA

[2] Department of Mathematics, Columbia University, New York, NY 10027, USA

Published online: 16 March 2026

Formally, let P_n be the probability measure on $\mathfrak{S}_n$ defined by

$$P_n = \begin{cases} \mathrm{Id} & \text{with probability } 1/n \\ (ij) & \text{with probability } 2/n^2 \text{ for } i < j. \end{cases}$$

Then, the process described above can be modeled by the discrete time Markov chain given by $X_{t+1} = X_t \cdot \pi_t$ where $\pi_t \sim P_n$ (with $X_0 = \mathrm{id}_n$, the identity permutation). Equivalently we may denote $\pi_t = (i_t\, j_t)$ where i_t and j_t are drawn iid from $[n]$; note that if $i_t = j_t$ then this permutation is trivial. In particular, $X_t \sim P_n^{*t}$, the t-fold convolution of P_n.

Let $U_{\mathfrak{S}_n}$ denote the uniform measure on $\mathfrak{S}_n$. Since the random transposition walk is ergodic with respect to $U_{\mathfrak{S}_n}$, the distribution of X_t converges in total variation distance to $U_{\mathfrak{S}_n}$ as $t \to \infty$. Recall that for two (probability) measures μ and ν on a common ground set E, the total variation distance (or TV distance) and L^1-distance are defined by

$$d_{\mathrm{TV}}(\mu, \nu) = \frac{d_1(\mu, \nu)}{2} = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|.$$

If $\mu = \mathrm{Law}(X)$ for a random variable X, we will often write $d_{\mathrm{TV}}(\mu, \cdot)$ as $d_{\mathrm{TV}}(X, \cdot)$.

Returning to the question at the end of the first paragraph, a natural lower bound is the first time at which the random transposition walk has "touched" all the cards. formally, using the notation above, consider the random (stopping) time

$$\tau := \min_{t \geq 1} \mathbb{1}\Bigg[\Big(\min_{k \in [n]} \sum_{\ell=1}^{t} \mathbb{1}[i_\ell = k \vee j_\ell = k]\Big) \geq 1\Bigg].$$

Here, $(i_\ell\, j_\ell)$ denotes the transposition applied at the ℓ-th step of the Markov chain. Observe that at time $\tau - 1$, the cards are certainly not well-shuffled, even in an approximate sense – indeed, $X_{\tau-1}$ is supported entirely on permutations with at least one fixed point. However it is classical that the number of fixed points of a random permutation drawn from $\mathfrak{S}_n$ converges in distribution to Pois(1); hence, the probability of no fixed points if $(1 + o(1))e^{-1}$. Moreover, since X_τ is not distributed exactly uniformly on $\mathfrak{S}_n$ (already for $n = 2$, a direct computation shows that $X_\tau(\mathrm{id}_2) = 1/6$), the best scenario one can hope for is that X_τ is distributed asymptotically uniformly (in the limit $n \to \infty$); in fact, this has been conjectured by Nathanaël Berestycki (see [15, Conjecture 1.2]). Our main result confirms this conjecture.

Theorem 1.1 *With notation as above*

$$d_{\mathrm{TV}}(X_\tau, U_{\mathfrak{S}_n}) \leq \exp\big(-(\log n)^{1/2+o(1)}\big).$$

1.1 Background and additional results

There is a rich and extensive literature surrounding the random transposition walk. For a historical account of the study of this walk, as well as examples of (sometimes surprising) appearances in other areas such as geometry, physics, and biology, we refer the reader to [4]. Moreover, using comparison theory, understanding the convergence of the random transposition walk leads to an understanding of the convergence of very general shuffling procedures; see the work of Diaconis and Saloff-Coste [5].

In a pioneering work [6] from 1981, Diaconis and Shahshahani used techniques from nonabelian Fourier analysis (or equivalently representation theory) to show that the random transposition walk undergoes cutoff at $(n \log n)/2$ with cutoff window $O(n)$, i.e.

$$d_{\mathrm{TV}}(P_n^{*\lfloor (n\log n)/2+cn\rfloor}, U_{\mathfrak{S}_n}) \xrightarrow[c\to\infty]{n\to\infty} 0 \quad \text{and} \quad d_{\mathrm{TV}}(P_n^{*\lfloor (n\log n)/2-cn\rfloor}, U_{\mathfrak{S}_n}) \xrightarrow[c\to\infty]{n\to\infty} 1.$$

To use terminology common in probabilistic combinatorics, whereas this result determines the "sharp threshold" for TV-mixing, Theorem 1.1 establishes the "hitting time" for TV-mixing; this is typically a much more challenging task since it involves a detailed understanding of the process in the threshold window. We note that the result of [6] has been generalized from transpositions to k-cycles for fixed k by Berestycki, Schramm and Zeitouni [1], and even further to conjugacy classes with support $o(n)$ by Berestycki and Şengül [2], both using probabilistic arguments. See also work of Saloff-Coste and Zúñiga [13] for a refinement of [6] for the L^2-cutoff as well as [2, Section 1.3] for more references and details about early work on the mixing of the random transposition walk.

Relevant to the present paper is a recent work of Teyssier [15] which established the cutoff profile for the random transposition walk: for fixed $c \in \mathbb{R}$,

$$d_{\mathrm{TV}}(P_n^{*\lfloor (n\log n)/2+cn\rfloor}, U_{\mathfrak{S}_n}) \xrightarrow{n\to\infty} d_{\mathrm{TV}}(\mathrm{Pois}(1+e^{-2c}), \mathrm{Pois}(1)),$$

where $\mathrm{Pois}(a)$ denotes the Poisson distribution with mean a; the same cutoff profile appears for several other random walks (see, e.g., [9, 10]). As a key intermediate step in our work, we establish the following strengthening of Teyssier's result, which provides a distributional approximation for X_t near cutoff (instead of only finding the distance to uniformity). In order to state this result, we need some notation. Let

$$t_n = \lfloor (n \log n)/2 \rfloor.$$

Given a non-negative integer t, we define $c_n(t)$ by

$$t = t_n + nc_n(t).$$

We also define

$$\gamma_{n,t} = e^{-2(t-t_n)/n} = e^{-2c_n(t)}.$$

We will approximate the distribution of X_t by the following distribution.

Definition 1.2 Let ν_t denote the measure on $\mathfrak{S}_n$ defined by the following sampling process: first, sample $M_t \in \{0, 1, \dots, n\}$ according to the distribution $\mathbb{P}[M_t = x] = \mathbb{P}[\mathrm{Pois}(\gamma_{n,t}) = x]/\mathbb{P}[\mathrm{Pois}(\gamma_{n,t}) \le n]$, then sample a uniformly random subset S_t of size M_t in $[n]$ (the truncation $M_t \le n$ ensures that S_t is well-defined), and finally, sample a uniformly random element of $\mathfrak{S}_{[n]\setminus S_t}$ and view it as an element of $\mathfrak{S}_n$ by fixing all of the elements in S_t.

Theorem 1.3 *With notation as above, if* $|c_n(t)| \le (\log\log n)/4 - \log\log\log n$*, then*

$$d_{\mathrm{TV}}(X_t, \nu_t) \le n^{-1+o(1)}.$$

Remark From the above theorem, we have that

$$d_{\mathrm{TV}}(X_t, U_{\mathfrak{S}_n}) = d_{\mathrm{TV}}(\nu_t, U_{\mathfrak{S}_n}) + n^{-1+o(1)} = d_{\mathrm{TV}}(\mathrm{Pois}(1+\gamma_{n,t}), \mathrm{Pois}(1)) + o(1). \tag{1.1}$$

Here, the first equality follows from the triangle inequality for d_{TV} and Theorem 1.3. The second equality follows by observing that for any $S \subseteq [n]$, the distribution of a permutation drawn from ν_t, conditioned on the set of its fixed points being S, is identical to a uniformly random permutation, conditioned on the set of its fixed points being S.

Note that (1.1) is exactly the main result of Teyssier [15, Theorem 1.1]. Additionally, identifying the stronger statement in Theorem 1.3 leads to a significantly simpler proof of the main result of [15]; in particular, our proof of Theorem 1.3 uses only the L^2-estimates from the work of Diaconis-Shahshahani [6] along with an application of Young's rule, thereby completely bypassing the L^1-theory and intricate combinatorial analysis in [15]; see Section 1.2 for further discussion. Finally, we have not carefully optimized the upper bound $|c_n(t)| \le (\log\log n)/4 - \log\log\log n$; however, for very large negative $c_n(t)$ we expect that the description in Theorem 1.3 does not hold.

We discuss two additional lines of work which are related to our result. First, recall that for an ergodic Markov chain with stationary distribution π, a stopping time κ is said to be a stationary time if $X_\kappa \sim \pi$; κ is said to be a strong stationary time if it is stationary and additionally, X_κ is independent of κ. In 1985, Broder (see [8]) devised a strong stationary time for the random transposition walk concentrated around time $2n\log n$; an improved strong stationary time concentrated around $n\log n$ was given by Matthews [8] who also claimed a further improvement to $(n\log n)/2$; however, his proof contained a subtle error which was only noticed nearly 30 years later by White. Subsequently, White [16] constructed a rather intricate strong stationary time concentrated in a window of size $O(n)$ around time $(n\log n)/2$; in particular, this establishes the analogue of [6] for the separation distance. The relationship of the present work to [16] is as follows: White showed that there exists some stopping time κ, concentrated around $(n\log n)/2$, for which X_κ is exactly uniformly distributed; in contrast, Theorem 1.1 shows that at time τ, which is a lower bound for mixing as discussed above, X_τ is (TV-)approximately uniform. We remark that extending White's result to the random k-cycle walk is an open problem, even for $k = 3$; on the other hand, our analysis can be adapted to prove an analogous result for any fixed k (see, for instance, works of Hough [7], Nestoridi and Olesker-Taylor [10] or Olesker-Taylor, Teyssier and Thévenin [11] for cutoff results in this direction).

Second, note that the random transposition walk is naturally associated to a random graph process, where we add the edge $\{i, j\}$ to the graph whenever the transposition

(ij) is chosen. Let G_t denote the corresponding (random) graph at time t and let C_t denote the largest connected component of G_t; by standard results in random graphs, C_t is macroscopic if $c > 1$ is fixed and $t \geq cn/2$ for sufficiently large n. A well-known result of Schramm [14] shows that the distribution of the lengths of the largest cycles of $X_t|_{C_t}$ is close to the distribution of the lengths of the largest cycles for the uniform measure on $\mathfrak{S}_{C_t}$; the natural conclusion of this line of work would be to establish the following conjecture due to Nathanaël Berestycki (see [15, Conjecture 1.3]).

Conjecture 1.4 *Fix $c > 1$. With notation as above, suppose $t \geq cn/2$. Then,*

$$d_{\mathrm{TV}}(X_t|_{C_t}, U_{\mathfrak{S}_{C_t}}) = o(1).$$

Here, the implicit constant in the notation $o(1)$ is allowed to depend on c.

Our final main result confirms this conjecture for t near cutoff.

Theorem 1.5 *With notation as above, if $|c_n(t)| \leq (\log\log n)/4 - \log\log\log n$, then*

$$d_{\mathrm{TV}}(X_t|_{C_t}, U_{\mathfrak{S}_{C_t}}) \leq n^{-1+o(1)}.$$

1.2 Proof outline

Our proof of Theorem 1.1 can be roughly decomposed into three steps, which we briefly discuss.

The first step is to prove Theorem 1.3, which characterizes, up to a vanishingly small error, the distribution of the random transposition walk at time t. As mentioned in the introduction, this is a strengthening of work of Teyssier [15] on the cutoff profile for the random transposition walk. Interestingly, our proof is considerably simpler, essentially due to the reason that once the correct statement has been identified, the task amounts to establishing an upper bound of the form $o(1)$ (as opposed to [15], where the (macroscopic) total variation distance to the uniform distribution is computed asymptotically exactly). Our proof of Theorem 1.3 proceeds along the same lines as that of Diaconis-Shahshahani [6] in that we pass from the L^1 distance to the L^2 distance using Cauchy-Schwarz, and then use Plancherel's formula along with the same character estimates as [6]. The only point of departure is that the uniform distribution on $\mathfrak{S}_n$ is replaced by the distribution ν_t; however, the Fourier coefficients of ν_t can be calculated exactly using a simple application of Young's rule. In contrast, the proof in [15] is based on approximating the L^1 distance using the Fourier inversion formula, as well as fairly intricate combinatorial analysis involving the Murnaghan—Nakayama rule. We believe that our method of distributional approximation can be applied to obtain cutoff profiles in many settings.

The second step is to bootstrap Theorem 1.3 to show that at times sufficiently close to $(n \log n)/2$, the number of untouched points is an "approximate sufficient statistic" for the random transposition walk (Proposition 2.1) in the sense that, conditioned on the untouched set being $M \subseteq [n]$ (which is assumed to be not too large), the distribution of the random transposition walk is close to the uniform distribution on permutations which leave each element of M fixed. The proof is carried out entirely in physical space. The key idea is to exploit the following "self-reducibility" of the random transposition walk: the distribution of the walk at time t, conditioned on the untouched set *containing* M, coincides exactly with the distribution of the random transposition walk at time t run on the ground set $[n] \setminus M$. To see why this is useful, observe that the indicator of the event that the untouched set at time t is exactly M can be written as a signed combination of events of the form "the untouched set at time t contains T" using the principle of inclusion-exclusion. Applying Theorem 1.3 to each of these summands and performing some explicit calculations using the form of the distribution ν_t completes the proof.

Finally, given Proposition 2.1, Theorem 1.5 follows immediately by using the law of total probability and classical results from the theory of random graphs. The deduction of Theorem 1.5 from Theorem 1.1 is also relatively standard, and can be viewed as a blend of the classical hitting-time result for connectivity in random graph theory along with Broder's strong stationary time for the random transposition walk.

1.3 Notation

We use standard asymptotic notation throughout, as follows. For functions $f = f(n)$ and $g = g(n)$, we write $f = O(g)$ or $f \lesssim g$ to mean that there is a constant C such that $|f(n)| \le C|g(n)|$ for sufficiently large n. Similarly, we write $f = \Omega(g)$ or $f \gtrsim g$ to mean that there is a constant $c > 0$ such that $f(n) \ge c|g(n)|$ for sufficiently large n. Finally, we write $f \asymp g$ or $f = \Theta(g)$ to mean that $f \lesssim g$ and $g \lesssim f$, and we write $f = o(g)$ or $g = \omega(f)$ to mean that $f(n)/g(n) \to 0$ as $n \to \infty$.

1.4 Organization

The remainder of this paper is organized as follows. In Section 2, we show how to deduce Theorems 1.1 and 1.5 from Theorem 1.3. The main work in this section is to prove Proposition 2.1, which we undertake in Section 2.1, before completing the proofs in Section 2.2. Then we prove Theorem 1.3 in Section 3. Finally in Appendix A we give the proof of Lemma 3.4.

2 Reduction to distributional approximation at a fixed time

2.1 An approximate sufficient statistic

Throughout this section, we will make use of the following notation: denote the (random) sequence of transpositions chosen by the random transposition walk up to time t by $(i_1 j_1), \ldots, (i_t j_t)$. For a subset $T \subseteq [n]$, let $\mathcal{G}(T) = \mathcal{G}^t(T)$ denote the event that $\{i_1, j_1, \ldots, i_t, j_t\} \cap T = \emptyset$; in words, the transposition walk does not "touch" T up to (and including) time t. Let $\mathcal{F}(T) = \mathcal{F}^t(T)$ denote the event $\mathcal{G}(T) \cap \bigcap_{j \in T^c} \mathcal{G}^c(\{j\})$; in words, the transposition walk touches everything outside T and does not touch T

up to (and including) time t. For convenience of notation, we will denote $\mathcal{F}([M])$ and $\mathcal{G}([M])$ by $\mathcal{F}(M)$ and $\mathcal{G}(M)$ for an integer $1 \le M \le n$.

We begin with the following proposition, which is central to proving our main results. Roughly, it asserts that for times close to $(n \log n)/2$, the "untouched set" is an approximate sufficient statistic in the sense that the distribution of the random transposition walk, conditioned on the untouched set being S, is close to the uniform distribution on $\mathfrak{S}_{[n]\setminus S}$ (provided that S is not too large). More precisely, we have the following.

Proposition 2.1 *Recall that $t_n = \lfloor (n \log n)/2 \rfloor$. Let $t = t_n + nc_n(t)$ and M_n be such that $|c_n(t)| \le (\log\log n)/4 - 2\log\log\log n$ and $M_n \le (\log n)^2$. Let $X_t^{M_n}$ denote the random variable X_t conditioned on the event $\mathcal{F}(M_n)$. Let μ^{M_n} denote the measure on $\mathfrak{S}_n$ induced by the uniform distribution on $\mathfrak{S}_{n\setminus[M_n]}$ via the natural inclusion $\mathfrak{S}_{n\setminus[M_n]} \hookrightarrow \mathfrak{S}_n$. Then*

$$d_{\mathrm{TV}}(X_t^{M_n}, \mu^{M_n}) \le n^{-1+o(1)}.$$

The proof of this proposition requires a few intermediate lemmas. The first lemma provides a relationship between the probabilities of the events $\mathcal{F}(T)$ and $\mathcal{G}(T)$.

Lemma 2.2 *Recall that $t_n = \lfloor (n \log n)/2 \rfloor$, $t = t_n + nc_n(t)$, and $\gamma_{n,t} = \exp(-2c_n(t))$. Let $M_n \in \mathbb{N}$ and $T_n \subseteq [n]$ such that $M_n \le |T_n| = n^{o(1)}$. Suppose that $|c_n(t)| \le (\log\log n)/4 - 2\log\log\log n$. Then,*

(1) $\mathbb{P}[\mathcal{G}(T_n)] = (1 + n^{-1+o(1)})\mathbb{P}[\mathcal{G}(M_n)]\gamma_{n,t}^{|T_n|-M_n} n^{M_n-|T_n|}$,
(2) $\mathbb{P}[\mathcal{F}(M_n)] = (1 + n^{-1+o(1)})\mathbb{P}[\mathcal{G}(M_n)]\exp(-\gamma_{n,t})$,
(3) $\sum_{S \supseteq [M_n], S \le n^{o(1)}} \mathbb{P}[\mathcal{G}(S)] \le n^{o(1)}\mathbb{P}[\mathcal{F}(M_n)]$.

Proof For the first item, note that by our bound on t,

$$\begin{aligned}
\mathbb{P}[\mathcal{G}(T_n)] = \left(\frac{n - |T_n|}{n}\right)^{2t} &= (1 + n^{-1+o(1)}) \exp\left(-\frac{2t|T_n|}{n}\right) \\
&= (1 + n^{-1+o(1)}) \exp\left(-\frac{2tM_n}{n}\right) \exp\left(-\frac{2t(|T_n| - M_n)}{n}\right) \\
&= (1 + n^{-1+o(1)}) \frac{\mathbb{P}[\mathcal{G}(M_n)]\gamma_{n,t}^{|T_n|-M_n}}{n^{|T_n|-M_n}}. \qquad (2.1)
\end{aligned}$$

For the second item, observe that

$$\mathcal{F}(M_n) = \mathcal{G}([M_n]) \cap \bigcap_{j > M_n} \mathcal{G}^c(\{j\}) = \mathcal{G}([M_n]) \setminus \bigcup_{j > M_n} \mathcal{G}([M_n] \cup \{j\})$$

so that

$$\mathbb{1}_{\mathcal{F}(M_n)} = \mathbb{1}_{\mathcal{G}([M_n])} - \mathbb{1}_{\bigcup_{j > M_n} \mathcal{G}([M_n]\cup\{j\})}.$$

By the Bonferroni inequality, for any non-negative integer k,

$$\sum_{\substack{J \subseteq [M_n+1, n] \\ 1 \le |J| \le 2k}} (-1)^{|J|+1} \mathbb{1}_{\mathcal{G}([M_n]\cup J)} \le \mathbb{1}_{\bigcup_{j > M_n} \mathcal{G}([M_n]\cup\{j\})} \le \sum_{\substack{J \subseteq [M_n+1, n] \\ 1 \le |J| \le 2k+1}} (-1)^{|J|+1} \mathbb{1}_{\mathcal{G}([M_n]\cup J)}.$$

Substituting this in the above expression for the indicator of $\mathcal{F}(M_n)$, we have for any non-negative integer k that

$$\sum_{\substack{T\supseteq[M_n]\\|T|\le M_n+2k+1}}(-1)^{|T|-M_n}\mathbb{1}_{\mathcal{G}(T)}\le\mathbb{1}_{\mathcal{F}(M_n)}\le\sum_{\substack{T\supseteq[M_n]\\|T|\le M_n+2k}}(-1)^{|T|-M_n}\mathbb{1}_{\mathcal{G}(T)}.$$

Taking expectations (with, say, $k=\lfloor(\log n)^2\rfloor$), we have that

$$\mathbb{P}[\mathcal{F}(M_n)]=\sum_{\substack{T\supseteq[M_n]\\|T|\le M_n+2k}}(-1)^{|T|-M_n}\mathbb{P}[\mathcal{G}(T)]\pm\sum_{\substack{T\supseteq[M_n]\\|T|=M_n+2k+1}}\mathbb{P}[\mathcal{G}(T)].$$

Therefore using (2.1) and $\gamma_{n,t}=o(\log n)$, we have

$$\begin{aligned}\frac{\mathbb{P}[\mathcal{F}(M_n)]}{\mathbb{P}[\mathcal{G}(M_n)]}&=(1+n^{-1+o(1)})\cdot\Big(\sum_{j=0}^{2k}\binom{n-M_n}{j}\frac{(-\gamma_{n,t})^j}{n^j}\pm\binom{n-M_n}{2k+1}\frac{(-\gamma_{n,t})^{2k+1}}{n^{2k+1}}\Big)\\&=(1+n^{-1+o(1)})\left(\sum_{j=0}^{2k}\frac{(-\gamma_{n,t})^j}{j!}\pm\frac{(-\gamma_{n,t})^{2k+1}}{(2k+1)!}\right)\\&=(1+n^{-1+o(1)})\exp(-\gamma_{n,t}),\end{aligned}$$

where the second line uses $M_n=O(n^{o(1)})$ and $k\le(\log n)^2$. Finally, for the third item, we have by combining the first two items that

$$\begin{aligned}\sum_{S\supseteq[M_n],S\le n^{o(1)}}\mathbb{P}[\mathcal{G}(S)]&\lesssim\sum_{s=0}^{n^{o(1)}}\binom{n-M_n}{s}\mathbb{P}[\mathcal{F}(M_n)]e^{\gamma_{n,t}}\Big(\frac{\gamma_{n,t}}{n}\Big)^s\\&\lesssim\mathbb{P}[\mathcal{F}(M_n)]e^{\gamma_{n,t}}\sum_{s=0}^{n^{o(1)}}\Big(\frac{3n}{s}\Big)^s\Big(\frac{\gamma_{n,t}}{n}\Big)^s\\&=\mathbb{P}[\mathcal{F}(M_n)]e^{\gamma_{n,t}}\sum_{s=0}^{n^{o(1)}}\Big(\frac{3\gamma_{n,t}}{s}\Big)^s\le\mathbb{P}[\mathcal{F}(M_n)]\cdot n^{o(1)},\end{aligned}$$

where in the second inequality, we have used that $\binom{n-M_n}{s}\le\binom{n}{s}\le(3n/s)^s$ and in final inequality, we have used that $\gamma_{n,t}\le\sqrt{\log n}$. ◻

In order to state the next lemma, we need some notation. As in the statement of Proposition 2.1, $t_n=\lfloor(n\log n)/2\rfloor$ and let $t=t_n+nc_n(t)$.

For any $T\subseteq[n]$, we define a new time which depends on $|T|$ and also on t:

$$t_T=t_{n-|T|}+(n-|T|)c_n(t).$$

Analogous to our definition of $\gamma_{n,t}$, we define

$$\gamma_{n-|T|,t_T} = \exp(-(t_T - t_{n-|T|})/(n-|T|)) = \exp(-2c_n(t));$$

therefore, $\gamma_{n-|T|,t_T} = \gamma_{n,t}$. We also define the analog of the distribution in Definition 1.2.

Definition 2.3 Let $\nu^T = \nu^T_{t_T}$ denote the distribution on $\mathfrak{S}_{[n]\setminus T}$ which is defined in the same way as the distribution ν_t in Definition 1.2 with two modifications: the time t is replaced by t_T and the ground set $[n]$ is replaced by $[n]\setminus T$. Using the natural inclusion $\mathfrak{S}_{[n]\setminus T} \hookrightarrow \mathfrak{S}_n$, we can (and will) view ν^T as a distribution on $\mathfrak{S}_n$.

Lemma 2.4 *For any $\sigma_n \in \mathfrak{S}_n$ and $M_n \subseteq [n]$ such that $\mathrm{Fix}(\sigma_n) \supseteq [M_n]$ and $|\mathrm{Fix}(\sigma_n)| \le n^{o(1)}$, we have that*

$$\sum_{[M_n]\subseteq T\subseteq \mathrm{Fix}(\sigma_n)} (-1)^{|T|-M_n}\mathbb{P}[\mathcal{G}(T)]\cdot \nu^T(\sigma_n) = (1+n^{-1+o(1)})\frac{\mathbb{P}[\mathcal{F}(M_n)]}{(n-M_n)!}.$$

Proof For convenience of notation, we will omit the subscript n from σ_n and M_n. Consider $\sigma \in \mathfrak{S}_n$ satisfying the assumptions of the lemma and let $T \subseteq \mathrm{Fix}(\sigma)$. By definition,

$$\nu^T(\sigma) = \sum_{r=0}^{|\mathrm{Fix}(\sigma)|-|T|} \frac{\mathbb{P}[\mathrm{Pois}(\gamma_{n-|T|,t_T}) = r]}{\mathbb{P}[\mathrm{Pois}(\gamma_{n-|T|,t_T}) \le n-|T|]} \cdot \frac{\binom{|\mathrm{Fix}(\sigma)|-|T|}{r}}{\binom{n-|T|}{r}(n-|T|-r)!}.$$

Let us simplify the individual terms in the sum. Since $\gamma_{n-|T|,t_T} = \gamma_{n,t} = \exp(-2c_n(t))$ and $|T| \le n^{o(1)}$, it follows that

$$1-\mathbb{P}[\mathrm{Pois}(\gamma_{n-|T|,t_T}) \le n-|T|] = O(n^{-1+o(1)}).$$

Moreover, for any $0 \le m \le s \le n$ with $s = n^{o(1)}$ and $r = n^{o(1)}$, we have

$$\begin{aligned}\frac{\mathbb{P}[\mathrm{Pois}(\gamma_{n-|T|,t_T}) = r]}{n^{s-m}\cdot\binom{n-s}{r}(n-r-s)!} &= \frac{e^{-\gamma_{n,t}}(\gamma_{n,t})^r}{r!}\cdot\frac{1}{n^{s-m}\cdot\binom{n-s}{r}(n-r-s)!}\\ &= \frac{e^{-\gamma_{n,t}}(\gamma_{n,t})^r}{(n-m)!}(1+n^{-1+o(1)}).\end{aligned}$$

Substituting $m = M$ and $s = |T|$ in the above, it follows that

$$\nu^T(\sigma) = (1+n^{-1+o(1)})\frac{e^{-\gamma_{n,t}}n^{|T|-M}}{(n-M)!}\sum_{r=0}^{|\mathrm{Fix}(\sigma)|-|T|}\gamma^r_{n,t}\binom{|\mathrm{Fix}(\sigma)|-|T|}{r}.$$

Therefore, by Lemma 2.2, we have

$$(1+n^{-1+o(1)})(n-M)!\sum_{[M]\subseteq T\subseteq \mathrm{Fix}(\sigma)}(-1)^{|T|-M}\mathbb{P}[\mathcal{G}(T)]\nu^T(\sigma)$$

$$
\begin{aligned}
&= \mathbb{P}[\mathcal{F}(M)] \sum_{[M]\subseteq T\subseteq \mathrm{Fix}(\sigma)} (-\gamma_{n,t})^{|T|-M} \sum_{r=0}^{|\mathrm{Fix}(\sigma)|-|T|} \gamma_{n,t}^{r} \binom{|\mathrm{Fix}(\sigma)|-|T|}{r} \\
&= \mathbb{P}[\mathcal{F}(M)] \sum_{[M]\subseteq T\subseteq \mathrm{Fix}(\sigma)} (-\gamma_{n,t})^{|T|-M} (1+\gamma_{n,t})^{|\mathrm{Fix}(\sigma)|-|T|} \\
&= \mathbb{P}[\mathcal{F}(M)] \sum_{s\geq 0} \binom{\mathrm{Fix}(\sigma)-M}{s} (1+\gamma_{n,t})^{s} (-\gamma_{n,t})^{\mathrm{Fix}(\sigma)-M-s} \\
&= \mathbb{P}[\mathcal{F}(M)].
\end{aligned}
$$

□

The next lemma is the crucial juncture where we exploit the "self-reducibility" of the random transposition walk. In its proof, we will use Theorem 1.3; the proof of Theorem 1.3 is presented in Section 3.

Lemma 2.5 *Recall that $t_n = \lfloor (n\log n)/2 \rfloor$. Let $t = t_n + nc_n(t)$ be such that $|c_n(t)| \leq (\log\log n)/4 - 2\log\log\log n$. For $T_n \subseteq [n]$, let $\widetilde{X}_t^T$ denote the random variable X_t conditioned on the event $\mathcal{G}(T)$. Let ν^T denote the distribution on $\mathfrak{S}_n$ in Definition 2.3. If $|T_n| \leq n^{o(1)}$, then*

$$d_{\mathrm{TV}}(\widetilde{X}_t^T, \nu^T) \leq n^{-1+o(1)}.$$

Proof Recall that

$$t_T = t_{n-|T|} + (n-|T|)c_n(t)$$

and

$$\gamma_{n-|T|,t_T} = \exp(-2c_n(t)) = \gamma_{n,t}.$$

We define a new hybrid time

$$\widetilde{t}_T = t_{n-|T|} + nc_n(t),$$

and correspondingly, let

$$\gamma_{n-|T|,\widetilde{t}_T} = \exp(-2nc_n(t)/(n-|T|)).$$

Let $\widetilde{\nu}^T = \widetilde{\nu}^T_{\widetilde{t}_T}$ denote the distribution on $\mathfrak{S}_{[n]\setminus T}$ which is defined in the same way as the distribution ν_t in Definition 1.2 with two modifications: the time t is replaced by $\widetilde{t}_T$ and the ground set $[n]$ is replaced by $[n]\setminus T$. Using the natural inclusion $\mathfrak{S}_{[n]\setminus T} \hookrightarrow \mathfrak{S}_n$, we can (and will) view $\widetilde{\nu}^T$ as a distribution on $\mathfrak{S}_n$.

Since $|\gamma_{n-|T|,t_T} - \gamma_{n-|T|,\widetilde{t}_T}| = n^{-1+o(1)}$ and $\mathrm{Pois}(a+b)$ has the same distribution as the independent sum $\mathrm{Pois}(a) + \mathrm{Pois}(b)$ for $a, b \geq 0$, it readily follows that $d_{\mathrm{TV}}(\nu^T, \widetilde{\nu}^T) \leq n^{-1+o(1)}$. Therefore, it suffices to show that $d_{\mathrm{TV}}(\widetilde{X}_t^T, \widetilde{\nu}^T) \leq n^{-1+o(1)}$.

But this follows immediately from Theorem 1.3 upon noting that:

(i) The distribution of $\widetilde{X}_t^T$, which is defined to be the distribution of $X_t \mid \mathcal{G}(T)$ coincides with the distribution of the random transposition walk on $[n] \setminus T$ at time t, and

(ii) the assumed bound on $|c_n(t)|$ and $|T|$ implies that

$$|nc_n(t)| + T\log n \leq (n - |T|)((\log\log(n - |T|))/4 - \log\log\log(n - |T|))$$

so that Theorem 1.3 is indeed applicable for $[n] \setminus T$ at time t. □

We are now in position to prove Proposition 2.1.

Proof of Proposition 2.1 For notational convenience, we will omit the subscript n in M_n. Let $\mathfrak{S}_n^M \subseteq \mathfrak{S}_n$ denote the subset of permutations which fix each element of $[M]$. The statement of Proposition 2.1 can be rewritten as

$$\sum_{\sigma\in\mathfrak{S}_n^M} \left|\mathbb{P}[\{X_t = \sigma\} \cap \mathcal{F}(M)] - \mathbb{P}[\mathcal{F}(M)] \cdot (n - M)!^{-1}\right| \leq n^{-1+o(1)} \cdot \mathbb{P}[\mathcal{F}(M)]. \quad (2.2)$$

We break up the sum on the left hand side into two parts depending on the size of $\mathrm{Fix}(\sigma)$. Namely, let $\mathcal{S} = \mathfrak{S}_n^M \cap \{\sigma \in \mathfrak{S}_n : |\operatorname{Fix}(\sigma)| \leq 3(\log n)^2\}$ and let $\mathcal{L} = \mathfrak{S}_n^M \setminus \mathcal{S}$.

Recall that for any $T \subseteq [n]$,

$$t_T = t_{n-|T|} + (n - |T|)c_n(t),$$

and $\widetilde{X}_t^T$ denotes the random variable X_t conditioned on the event $\mathcal{G}(T)$.

We define

$$t_M := t_{[M]}, \qquad \widetilde{X}_t^M := \widetilde{X}_t^{[M]}.$$

Contribution from $\mathcal{L}$. Recall that μ^M is the uniform distribution on $\mathfrak{S}_{[n]\setminus M}$ viewed as a distribution on $\mathfrak{S}_n$ via the natural inclusion. Additionally let $\nu^M := \nu_{t_M}^{[M]}$ (see Definition 2.3, but to recap, this corresponds to sampling $\mathrm{Pois}[\gamma_{n-M,t_M}]$ fixed points on $[n] \setminus [M]$, sampling a uniform permutation on $[n] \setminus [M]$ modulo these forced fixed points and then viewing this permutation of $[n] \setminus [M]$ as a permutation in $\mathfrak{S}_n$).

We have

$$\begin{aligned}
\sum_{\sigma\in\mathcal{L}} |\mathbb{P}[\widetilde{X}_t^M = \sigma]| &\leq n^{-1+o(1)} + \nu^M(\mathcal{L}) \\
&\leq n^{-1+o(1)} + \mathbb{P}[\mathrm{Pois}(\gamma_{n-M,t_M}) \geq (\log n)^2] + \sup_{\substack{Q\subseteq[n]\setminus M\\ |Q|\leq(\log n)^2}} \mu^{M\cup Q}(\mathcal{L}) \\
&\leq n^{-1+o(1)}.
\end{aligned} \quad (2.3)$$

Here, the first inequality uses Lemma 2.5 and Theorem 1.3. For the second inequality, consider the distribution ν^M and the set of "forced" fixed points sampled. The second term corresponds to the event that we sample $\geq (\log n)^2$ of these "forced" fixed points on $[n] \setminus M$. Otherwise, the set of "forced" fixed points is given by $M \cup Q$ with $Q \subseteq [n] \setminus M$ and $|Q| \leq (\log n)^2$; this gives us the third term.

For the final inequality, by the basic combinatorial fact that the probability of a random permutation having exactly k fixed points is at most $1/k!$ (see, e.g. [15, Eq 3.3]), the term $\mathbb{P}[\mathrm{Pois}(\gamma_{n-M,t_M}) \geq (\log n)^2] = n^{-1+o(1)}$. This same fact also shows that

$$\sup_{\substack{Q \subseteq [n]\setminus M \\ |Q| \leq (\log n)^2}} \mu^{M\cup Q}(\mathcal{L}) = n^{-1+o(1)},$$

since $\mathcal{L}$ consists of permutations with at least $3(\log n)^2$ fixed points, while for any such Q, $|M \cup Q| \leq 2(\log n)^2$, so that $\mu^{M\cup Q}(\mathcal{L})$ is bounded above by the probability that a random permutation, drawn according to $\mu^{M\cup Q}$, has at least $(\log n)^2$ fixed points.

Using this, we have

$$\begin{aligned}\sum_{\sigma\in\mathcal{L}} \mathbb{P}[\{X_t = \sigma\} \cap \mathcal{F}(M)] &\leq \sum_{\sigma\in\mathcal{L}} \mathbb{P}[\{X_t = \sigma\} \cap \mathcal{G}(M)] \\ &= \sum_{\sigma\in\mathcal{L}} \mathbb{P}[\widetilde{X}_t^M = \sigma]\mathbb{P}[\mathcal{G}[M]] \\ &\leq n^{-1+o(1)}\mathbb{P}[\mathcal{G}(M)] \leq n^{-1+o(1)}\mathbb{P}[\mathcal{F}(M)],\end{aligned}$$

where the first line uses $\mathcal{F}(M) \subseteq \mathcal{G}(M)$, the second line uses the definition of $\widetilde{X}_t^M$, the third line uses (2.3), and the final line uses the third item of Lemma 2.2. Moreover, by the above combinatorial fact on fixed points of a random permutation, we have

$$\sum_{\sigma\in\mathcal{L}} |\mathbb{P}[\mathcal{F}(M)](n-M)!^{-1}| = \mathbb{P}[\mathcal{F}(M)]\mu^M(\mathcal{L}) \leq n^{-1+o(1)}\mathbb{P}[\mathcal{F}(M)].$$

Combining the above two display equations with the triangle inequality shows that

$$\sum_{\sigma\in\mathcal{L}} \left|\mathbb{P}[\{X_t = \sigma\} \cap \mathcal{F}(M)] - \mathbb{P}[\mathcal{F}(M)] \cdot (n-M)!^{-1}\right| \leq n^{-1+o(1)} \cdot \mathbb{P}[\mathcal{F}(M)].$$

Contribution from $\mathcal{S}$. By the principle of inclusion-exclusion, we have that

$$\mathbb{P}[\{X_t = \sigma\} \cap \mathcal{F}(M)] = \sum_{[M]\subseteq T\subseteq \mathrm{Fix}(\sigma)} (-1)^{|T|-M}\mathbb{P}[\{X_t = \sigma\} \cap \mathcal{G}(T)]$$

Recall that by Lemma 2.4, for any $\sigma \in \mathcal{S}$,

$$\sum_{[M]\subseteq T\subseteq \mathrm{Fix}(\sigma)} (-1)^{|T|-M}\mathbb{P}[\mathcal{G}(T)] \cdot \nu^T(\sigma) = (1 + n^{-1+o(1)})\mathbb{P}[\mathcal{F}(M)](n-M)!^{-1}.$$

It follows that for any $\sigma \in \mathcal{S}$,

$$
\begin{aligned}
&\Big|\mathbb{P}[\{X_t = \sigma\} \cap \mathcal{F}(M)] - \mathbb{P}[\mathcal{F}(M)] \cdot (n-M)!^{-1}\Big| \\
&\quad \le n^{-1+o(1)}(n-M)!^{-1}\mathbb{P}[\mathcal{F}(M)] \\
&\qquad + \left|\sum_{[M]\subseteq T \subseteq \mathrm{Fix}(\sigma)} (-1)^{|T|-M}\Big(\mathbb{P}[\{X_t=\sigma\}\cap \mathcal{G}(T)] - \mathbb{P}[\mathcal{G}(T)]\cdot \nu^T(\sigma)\Big)\right| \\
&\quad \le n^{-1+o(1)}(n-M)!^{-1}\mathbb{P}[\mathcal{F}(M)] + \sum_{[M]\subseteq T \subseteq \mathrm{Fix}(\sigma)} \Big|\mathbb{P}[\{X_t=\sigma\}\cap \mathcal{G}(T)] - \mathbb{P}[\mathcal{G}(T)]\nu^T(\sigma)\Big|.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\sum_{\sigma\in\mathcal{S}}\Big|\mathbb{P}[\{X_t = \sigma\} \cap \mathcal{F}(M)] - \mathbb{P}[\mathcal{F}(M)] \cdot (n-M)!^{-1}\Big| \\
&\qquad \le n^{-1+o(1)}\mathbb{P}[\mathcal{F}(M)] + \sum_{\sigma\in\mathcal{S}}\sum_{[M]\subseteq T \subseteq \mathrm{Fix}(\sigma)} \Big|\mathbb{P}[\{X_t=\sigma\}\cap \mathcal{G}(T)] - \mathbb{P}[\mathcal{G}(T)]\nu^T(\sigma)\Big| \\
&\qquad \le n^{-1+o(1)}\left(\mathbb{P}[\mathcal{F}(M)] + \sum_{[M]\subseteq T, |T|\le n^{o(1)}} \mathbb{P}[\mathcal{G}(T)]\right) \\
&\qquad \le n^{-1+o(1)}\mathbb{P}[\mathcal{F}(M)],
\end{aligned}
$$

where the first inequality uses our estimate for a fixed $\sigma \in \mathcal{S}$ along with $|\mathcal{S}| \le (n-M)!$ and the final inequality uses the third item of Lemma 2.2. Let us explain the second inequality. Recalling the notation from Lemma 2.5 that $\widetilde{X}_t^T$ denotes the random variable X_t conditioned on the event $\mathcal{G}(T)$, we have

$$
\begin{aligned}
&\sum_{\sigma\in\mathcal{S}}\sum_{[M]\subseteq T \subseteq \mathrm{Fix}(\sigma)} \Big|\mathbb{P}[\{X_t=\sigma\}\cap \mathcal{G}(T)] - \mathbb{P}[\mathcal{G}(T)]\nu^T(\sigma)\Big| \\
&\le \sum_{\sigma\in\mathcal{S}}\sum_{[M]\subseteq T, |T|\le n^{o(1)}} \Big|\mathbb{P}[\{X_t=\sigma\}\cap \mathcal{G}(T)] - \mathbb{P}[\mathcal{G}(T)]\nu^T(\sigma)\Big| \\
&= \sum_{[M]\subseteq T, |T|\le n^{o(1)}}\sum_{\sigma\in\mathcal{S}} \Big|\mathbb{P}[\{X_t=\sigma\}\cap \mathcal{G}(T)] - \mathbb{P}[\mathcal{G}(T)]\nu^T(\sigma)\Big| \\
&= \sum_{[M]\subseteq T, |T|\le n^{o(1)}}\sum_{\sigma\in\mathcal{S}} \Big|\mathbb{P}[\widetilde{X}_t^T=\sigma]\mathbb{P}[\mathcal{G}(T)] - \mathbb{P}[\mathcal{G}(T)]\nu^T(\sigma)\Big| \\
&\le \sum_{[M]\subseteq T, |T|\le n^{o(1)}} \mathbb{P}[\mathcal{G}(T)] d_{\mathrm{TV}}(\widetilde{X}_t^T, \nu^T) \\
&\le n^{-1+o(1)} \sum_{[M]\subseteq T, |T|\le n^{o(1)}} \mathbb{P}[\mathcal{G}(T)],
\end{aligned}
$$

where the final inequality uses Lemma 2.5.

Together with the above estimate for $\sigma \in \mathcal{L}$, this completes the proof of (2.2). $\square$

2.2 Finishing the proof

Given the key Proposition 2.1, the deduction of Theorems 1.1 and 1.5 is relatively straightforward.

Proof of Theorem 1.5 Let t and C_t be as in the statement of the theorem and let $S = S_t \subseteq [n]$ denote the set of indices which are not touched by time t. Let $\mathcal{C}_t$ denote the event that $C_t = [n] \setminus S_t$ and $|S_t| \leq (\log n)^2$; by standard results in the theory of random graphs, it follows that $\mathbb{P}[\mathcal{C}_t^c] \leq n^{-1+o(1)}$. We sketch the argument for the reader's convenience.

First, note that it suffices to prove $\mathbb{P}[\mathcal{C}_{t^*}^c] \leq n^{-1+o(1)}$ for $t^* = t_n + nc_n(t^*)$ and $c_n(t^*) = -(\log\log n)/4 + \log\log\log n$, since the statement for $t \geq t^*$ easily follows from the statement for t^* by the following observation: conditioned on $\mathcal{C}_{t^*}$, the probability that a vertex in S_{t^*} gets connected to another vertex in S_{t^*} before it gets connected to a vertex in C_{t^*} is $n^{-1+o(1)}$. We now prove that $\mathbb{P}[\mathcal{C}_{t^*}^c] \leq n^{-1+o(1)}$. The bound $\mathbb{P}[|S_{t^*}| \geq (\log n)^2] \leq n^{-\omega(1)}$ follows using the Chernoff bound for negatively dependent random variables and our choice of t^*. The more interesting bound $\mathbb{P}[C_{t^*} = [n] \setminus S_{t^*}] \geq 1 - n^{-1+o(1)}$ follows from [3, Theorem 7.2] with $k_0 = 1$. In the statement of this theorem, $E_{t^*}(T_k) = (1+o_n(1))\binom{n}{k}k^{k-2}p^{k-1}(1-p)^{k(n-k)}$, where $p = 2t^*/n^2$, T_k denotes the number of components which are trees of order k, and $E_t(T_k)$ denotes the expected value of T_k for the random graph process at time t. In particular, for our choice of t^*, we have that $E_{t^*}(T_1) = (1 + o(1))e^{-2c_n(t^*)} = (\log n)^{1/2+o(1)}$ and $E_t(T_2) = (1 + o(1))e^{-4c_n(t^*))} \log n/2n = n^{-1+o(1)}$, so that the assumptions of the theorem are satisfied for $k_0 = 1$. Note that, while the statement of the theorem is asymptotic, the proof gives the probability bound $O(t^* k_0/n^2) = n^{-1+o(1)}$, as claimed.

Recall that our goal is to show that $d_{\mathrm{TV}}(X_t|_{C_t}, U_{\mathfrak{S}_{C_t}}) \leq n^{-1+o(1)}$. By the triangle inequality and the definition of total variation distance, we have:

$$d_{\mathrm{TV}}(X_t|_{C_t}, U_{\mathfrak{S}_{C_t}}) \leq \mathbb{P}[\mathcal{C}_t^c] + d_{\mathrm{TV}}\big(\mathrm{Law}(X_t|_{C_t} \mid \mathcal{C}_t), \mathrm{Law}(U_{\mathfrak{S}_{C_t}} \mid \mathcal{C}_t)\big).$$

Since the first term is $n^{-1+o(1)}$, it suffices to show that the second term is also $n^{-1+o(1)}$.

Let $\mathcal{M} = \{M \subseteq [n] : |M| \leq (\log n)^2\}$. The event $\mathcal{C}_t$ is the disjoint union of events E_M indexed by $M \in \mathcal{M}$, where

$$E_M := \{S_t = M\} \cap \{C_t = [n] \setminus M\}.$$

By the convexity of total variation distance, we have:

$$\begin{aligned} &d_{\mathrm{TV}}\big(\mathrm{Law}(X_t|_{C_t} \mid \mathcal{C}_t), \mathrm{Law}(U_{\mathfrak{S}_{C_t}} \mid \mathcal{C}_t)\big) \\ &\quad \leq \sum_{M \in \mathcal{M}} \mathbb{P}[E_M \mid \mathcal{C}_t] \cdot d_{\mathrm{TV}}\big(\mathrm{Law}(X_t|_{C_t} \mid E_M), U_{\mathfrak{S}_{[n]\setminus M}}\big). \end{aligned} \tag{2.4}$$

Fix $M \in \mathcal{M}$. Note that on the event E_M, we have $C_t = [n] \setminus M$, so $X_t|_{C_t}$ is identical to the restriction $X_t|_{[n]\setminus M}$. Let μ^M denote the uniform distribution on $\mathfrak{S}_{[n]\setminus M}$. Let

$\mathcal{F}(M) := \{S_t = M\}$ By the triangle inequality, we have:

$$\begin{aligned} d_{\mathrm{TV}}\Big(\mathrm{Law}(X_t|_{C_t} \mid E_M), \mu^M\Big) &= d_{\mathrm{TV}}\Big(\mathrm{Law}(X_t|_{[n]\setminus M} \mid E_M), \mu^M\Big) \\ &\le d_{\mathrm{TV}}\big(\mathrm{Law}(X_t|_{[n]\setminus M} \mid E_M), \mathrm{Law}(X_t|_{[n]\setminus M} \mid \mathcal{F}(M))\big) \\ &\quad + d_{\mathrm{TV}}\Big(\mathrm{Law}(X_t|_{[n]\setminus M} \mid \mathcal{F}(M)), \mu^M\Big). \end{aligned}$$

By Proposition 2.1, the second term is bounded by $n^{-1+o(1)}$. For the first term, observe that $E_M = \mathcal{F}(M) \cap \mathcal{K}_M$, where $\mathcal{K}_M := \{C_t = [n] \setminus M\}$. Using the identity $d_{\mathrm{TV}}(\nu, \nu(\cdot \mid A)) = \nu(A^c)$ for $\nu = \mathrm{Law}(X_t|_{[n]\setminus M} \mid \mathcal{F}(M))$ and $A = \mathcal{K}_M$, we obtain

$$d_{\mathrm{TV}}\big(\mathrm{Law}(X_t|_{[n]\setminus M} \mid E_M), \mathrm{Law}(X_t|_{[n]\setminus M} \mid \mathcal{F}(M))\big) = \mathbb{P}[\mathcal{K}_M^c \mid \mathcal{F}(M)].$$

Substituting this into (2.4), we have:

$$\begin{aligned} d_{\mathrm{TV}}\big(\mathrm{Law}(X_t|_{C_t} \mid \mathcal{C}_t), \mathrm{Law}(U_{\mathfrak{S}_{C_t}} \mid \mathcal{C}_t)\big) &\le n^{-1+o(1)} + \sum_{M\in\mathcal{M}} \mathbb{P}[E_M \mid \mathcal{C}_t] \cdot \mathbb{P}[\mathcal{K}_M^c \mid \mathcal{F}(M)] \\ &\le n^{-1+o(1)} + \sum_{M\in\mathcal{M}} \frac{\mathbb{P}[\mathcal{F}(M)]}{\mathbb{P}[\mathcal{C}_t]} \cdot \frac{\mathbb{P}[\mathcal{K}_M^c \cap \mathcal{F}(M)]}{\mathbb{P}[\mathcal{F}(M)]} \\ &\le n^{-1+o(1)} + \mathbb{P}[\mathcal{C}_t]^{-1} \sum_{M\in\mathcal{M}} \mathbb{P}[K_M^c \cap \mathcal{F}(M)] \\ &\le n^{-1+o(1)} + \mathbb{P}[\mathcal{C}_t]^{-1}\mathbb{P}[\mathcal{C}_t^c] \\ &\le n^{-1+o(1)}. \end{aligned}$$

In the second line, we have used that $E_M \subseteq \mathcal{F}(M)$; in the fourth line, we have used that the events $K_M^c \cap \mathcal{F}(M)$ are disjoint for different choices of M and $\cup_M (K_M^c \cap \mathcal{F}(M)) \subseteq \mathcal{C}_t^c$ by definition; in the last line, we have used that $\mathbb{P}[\mathcal{C}_t^c] \le n^{-1+o(1)}$. This completes the proof. □

Proof of Theorem 1.1 Let $t^* = \left\lfloor \frac{n(\log n)}{2} \right\rfloor - \frac{n \log\log n}{4} + 2n \log\log\log n$. Let S_{t^*} denote the set of indices which are not touched by time t^*.

Consider the following two "marking schemes".

- At time t^*, mark all of the elements in $[n] \setminus S_{t^*}$. At each subsequent step s, choose a random pair $(i_s j_s)$ according to the distribution Q_n on the set of ordered pairs $\mathcal{O}_n = \{(ij) : i, j \in [n]\}$ defined as follows:

$$Q_n = \begin{cases} (ii) & \text{with probability } 2/n^2 \text{ for } i \in S_{t^*} \\ (jj) & \text{with probability } n^{-2} \cdot (n - 2|S_{t^*}|)/(n - |S_{t^*}|) \text{ for } j \notin S_{t^*} \\ (ij) & \text{with probability } 2/n^2 \text{ for } i < j. \end{cases}$$

  Update the set of marked elements as follows: if the transposition given by $(i_s j_s)$ consists of an unmarked element and a marked element, or if it consists of the same unmarked element twice, then mark the unmarked element. Let κ_m denote the first time when all of the elements are marked. We will discuss the reason for the introduction of Q_n and κ_m after introducing some further notation.

- Same as above, except the random pairs are chosen according to the distribution P_n on $\mathcal{O}_n$ as in the random transposition walk: [1]

$$P_n = \begin{cases} (ii) & \text{with probability } 1/n^2 \text{ for } i \in [n] \\ (ij) & \text{with probability } 2/n^2 \text{ for } i < j. \end{cases}$$

Let τ_m denote the first time when all of the elements are marked.

Corresponding to the two marking schemes above, we consider the following two Markov processes:

- $(Z_t)_{t \geq t^*}$, where Z_{t^*} is sampled from the uniform distribution on $\mathfrak{S}_{[n] \setminus S_{t^*}}$ (viewed as a distribution on $\mathfrak{S}_n$) and the process evolves according to the random transposition walk with transitions sampled from Q_n.
- $(Y_t)_{t \geq t^*}$, defined as above, except with transitions sampled from P_n.

The reason behind the introduction of Q_n and κ_m is the following: essentially due to Broder's argument (see [8])[2], $Z_{\kappa_m} \sim U_{\mathfrak{S}_n}$. To see this, let M_t denote the set of marked elements at time t. We will prove by induction on $t \geq t^*$ that the distribution of $Z_t|_{M_t}$ is uniform over all of the $|M_t|!$ permutations. This is true at time t^* by construction. Suppose it is true at time $t \geq t^*$. At time $t + 1$, there are two possibilities. The first possibility is that $M_t = M_{t+1}$, so the transposition chosen was either contained entirely inside the marked set or is between two distinct unmarked elements; in either case, $X_{t+1}|_{M_{t+1}}$ is uniform due to the inductive hypothesis. The second possibility is that $M_{t+1} = M_t \cup \{v\}$ for some $v \notin M_t$. Since $X_t|_{M_t}$ is uniform, and since, by definition of Q_n, all of the transpositions (vj), $j \in [n]$ are equally likely, it follows that $X_{t+1}|_{M_{t+1}}$ is also uniformly distributed.

Since the distributions Q_n and P_n on $\mathcal{O}_n$ agree outside "diagonal" elements (jj), it follows by direct computation that $d_{\mathrm{TV}}(P_n, Q_n) \leq |S_{t^*}|/n^2$. Moreover, it follows by direct computations as in Lemma 2.2 that

$$\mathbb{P}[|S_{t^*}| \geq (\log n)^2] \leq n^{-\omega(1)}, \text{ and } \mathbb{P}[\max\{\tau_m, \kappa_m\} - t^* \geq n \log n] \leq n^{-\omega(1)}.$$

In particular, the total variation distance between P_n and Q_n is at most $(\log n)^2/n^2 + n^{-\omega(1)} = n^{-2+o(1)}$. Additionally, since we need to run the process for only $n^{1+o(1)}$ steps (except with probability $n^{-\omega(1)}$), it follows that $\mathbb{P}[\tau_m \neq \kappa_m] \leq n^{-\omega(1)} + n^{-2+o(1)} \cdot n^{1+o(1)} = n^{-1+o(1)}$. Since $Z_{\kappa_m} \sim U_{\mathfrak{S}_n}$, it therefore follows that

$$d_{\mathrm{TV}}(Y_{\tau_m}, U_{\mathfrak{S}_n}) \leq n^{-1+o(1)}.$$

[1] We have slightly abused notation here: in the introduction, P_n was defined to be a probability measure on $\mathfrak{S}_n$, whereas here, P_n is a probability measure on $\mathcal{O}_n$. By interpreting an element of $\mathcal{O}_n$ as a transposition, the measure P_n on $\mathcal{O}_n$ define here induces the measure P_n on $\mathfrak{S}_n$ corresponding to the random transposition walk.

[2] Broder's argument produces a strong stationary time at roughly $n \log n$ instead of at the cutoff time of $n(\log n)/2$. A key point in our analysis is that we are only simulating Broder's argument for a small number of steps.

By the definition of t^*, it follows from direct computation that $\mathbb{P}[\tau \le t^*] \le \exp(-(\log n)^{1/2+o(1)})$. Moreover, on the event that $|S_{t^*}| = n^{o(1)}$, $\tau \le n^{1+o(1)}$, and $\tau > t^*$, we have that $\tau = \tau_m$, except with probability $n^{-1+o(1)}$; this is because a transposition involving two elements of S_{t^*} is chosen throughout the process with probability at most $|S_{t^*}|^2 n^{-1+o(1)} = n^{-1+o(1)}$. Therefore, as above, it follows that $\mathbb{P}[\tau = \tau_m] \ge 1 - \exp(-(\log n)^{1/2+o(1)})$, so that it suffices to prove Theorem 1.1 with X_τ replaced by X_{τ_m}.

By the above discussion, in order to prove the statement of the theorem, it suffices to show that $d_{\mathrm{TV}}(X_{\tau_m}, Y_{\tau_m}) \le n^{-1+o(1)}$. By the proof of Theorem 1.5, we have that $d_{\mathrm{TV}}(X_{t^*}|_{[n]\setminus S_t^*}, Y_{t^*}) \le n^{-1+o(1)}$. In particular, by the coupling characterization of total variation distance, there exists a coupling (W, V) of these distributions such that $\mathbb{P}[W \ne V] \le n^{-1+o(1)}$. Evolving this coupling using the same updates for the random transposition walk gives a coupling (W', V') of (X_{τ_m}, Y_{τ_m}) such that $\mathbb{P}[W' \ne V'] \le \mathbb{P}[W \ne V] \le n^{-1+o(1)}$. Therefore, $d_{\mathrm{TV}}(X_{\tau_m}, Y_{\tau_m}) \le n^{-1+o(1)}$, as required. $\square$

3 Approximating the distribution at a fixed time

3.1 Nonabelian Fourier Transform

In this subsection, we recall various standard notions concerning the (nonabelian) Fourier transform over finite groups. Given a finite group G, we let $\widehat{G}$ denote the set of irreducible representations. We define the convolution of $f_1, f_2 : G \to \mathbb{C}$ to be

$$(f_1 * f_2)(z) := \sum_{x\in G} f_1(x) f_2(x^{-1}z), \qquad z \in G.$$

For $\rho \in \widehat{G}$, let d_ρ denote the dimension of the representation. The nonabelian Fourier transform for a function $f : G \to \mathbb{C}$ is the map $\widehat{f} : \widehat{G} \to \oplus_{\rho\in\widehat{G}}(\mathbb{C}^{d_\rho \times d_\rho})$ given by

$$\widehat{f}(\rho) = \sum_{g\in G} f(g)\rho(g), \qquad \rho \in \widehat{G}.$$

Recall that for any $f_1, f_2 : G \to \mathbb{C}$ and $\rho \in \widehat{G}$,

$$\widehat{f_1 * f_2}(\rho) = \widehat{f_1}(\rho) \cdot \widehat{f_2}(\rho).$$

In our applications, we will restrict attention to functions f which are class functions (i.e. are constant on conjugacy classes). For such functions, it follows from Schur's lemma (see e.g. [6, Lemma 5]) that for $\rho \in \widehat{G}$,

$$\widehat{f}(\rho) = \left(\frac{\sum_{g\in G} f(g)\,\mathrm{Tr}(\rho(g))}{d_\rho}\right) \cdot \mathrm{Id}_{d_\rho}. \tag{3.1}$$

Recall that the character of $\rho \in \widehat{G}$ is defined by

$$\chi_\rho(g) = \mathrm{Tr}(\rho(g)), \qquad g \in G.$$

Observe that $\chi_\rho(g)$ is a class function.

3.2 Character estimates for P_n^{*t}

The following is immediate from the notions introduced above and the definition of the random transposition walk.

Lemma 3.1 *Let $\rho \in \widehat{\mathfrak{S}_n}$ and $\tau \in \mathfrak{S}_n$ be an (arbitrary) transposition. Then*

$$\widehat{P_n}(\rho) := \left(\frac{1}{n} + \frac{n-1}{n} \cdot \frac{\chi_\rho(\tau)}{d_\rho}\right) \cdot \mathrm{Id}_{d_\rho}.$$

We need the following bound on the character ratio $\chi_\rho(\tau)/d_\tau$ appearing in the above expression. As is standard, we will identify elements of $\widehat{\mathfrak{S}_n}$ with (positive integer) partitions λ of n. We use $\lambda \vdash n$ to denote that λ is a partition of n. Given a partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 \geq \cdots \geq \lambda_k$, we let λ' denote the conjugate partition and $\lambda^* = (\lambda_2, \ldots, \lambda_k)$ to be the partition of $n - \lambda_1$ obtained by truncating λ.

Lemma 3.2 *Let $\lambda \vdash n$ and $\tau \in \mathfrak{S}_n$ be a transposition. Then*

$$\chi_\lambda(\tau) = -\chi_{\lambda'}(\tau) = d_\lambda \binom{n}{2}^{-1} \left(\sum_{i=1}^{n} \binom{\lambda_i}{2} - \binom{\lambda_i'}{2}\right). \tag{3.2}$$

In particular, if $\lambda_1 \geq n - n^{o(1)}$, then

$$\frac{\chi_\lambda(\tau)}{d_\lambda} = \frac{\binom{\lambda_1}{2}}{\binom{n}{2}} + O(n^{-2+o(1)}) = 1 - \frac{2(n-\lambda_1)}{n} + O(n^{-2+o(1)}). \tag{3.3}$$

Proof The evaluation (3.2) for $\chi_\lambda(\tau)$ when τ is a transposition appears as [6, Lemma 7]. The bound for $\lambda_1 \geq n - n^{o(1)}$ follows by direct computation using the evaluation (3.2). In this computation we are using that $\lambda_1 \geq n - n^{o(1)}$ implies that $\lambda_2^2 + \cdots + \lambda_n^2 \leq \lambda_2(\lambda_2 + \cdots + \lambda_n) \leq n^{o(1)} \cdot n^{o(1)} = n^{o(1)}$. Moreover, $\lambda_1 \geq n - n^{o(1)}$ also implies that the number of terms among $\lambda_1', \ldots, \lambda_n'$ which are strictly greater than 1 is at most $n^{o(1)}$, and each such term is at most $n^{o(1)}$. □

We also need the following bounds on the dimension of irreducible representations which follow from the hook-length formula. The first part of the proposition is [6, Corollary 2]; the second is a mild extension of [15, Proposition 3.2] and follows from exactly the same proof there.

Lemma 3.3 *Let $\lambda \vdash n$. Then*

$$d_\lambda \leq \binom{n}{\lambda_1} \cdot \sqrt{(n-\lambda_1)!}.$$

Furthermore if $\lambda_1 = n - x$ *with* $x \leq (\log n)^2$,

$$d_\lambda = \binom{n}{\lambda_1} \cdot d_{\lambda^*}\Big(1 - \frac{x}{n} + O(n^{-2+o(1)})\Big).$$

Given the character estimates in Lemma 3.2, it is natural to decompose the irreducible representations based on the size of the largest part. Accordingly, let

$$\mathcal{L} = \{\lambda_1 \vdash n : \lambda_1 \geq n - (\log n)^2\}.$$

The next lemma allows us to control the mass of the Fourier coefficients of P_n^{*t} on representations outside $\mathcal{L}$. The proof essentially appears in [6, pg. 168-174] (in [6], the corresponding result is stated only for $c_n(t) \geq 0$; an extension to negative $c_n(t)$ for the L^1 case is provided in [15, Section 4.1]; for our purpose, owing to the bound on $c_n(t)$ and the definition of $\mathcal{L}$, we are able to directly use the estimates in [6]). For completeness, we record the details in Appendix A.

Lemma 3.4 *Recall that* $t_n = \lfloor (n \log n)/2 \rfloor$. *Let* $t = t_n + nc_n(t)$ *be such that* $|c_n(t)| \leq (\log\log n)/4 - \log\log\log n$. *We have that*

$$\sum_{\substack{\lambda \vdash n \\ \lambda \notin \mathcal{L}}} d_\lambda^2 \cdot \left|\frac{1}{n} + \frac{n-1}{n} \cdot \frac{\chi_\lambda(\tau)}{d_\lambda}\right|^{2t} \leq n^{-2+o(1)}.$$

3.3 Character estimates for ν_t

In order to estimate the Fourier coefficients of ν_t, we define ξ_M to be the distribution on $\mathfrak{S}_n$ obtained as follows: sample a uniform subset $S \subseteq [n]$ of size M, then sample a uniformly random permutation of $\mathfrak{S}_{[n]\setminus S}$, and extend this to an element of $\mathfrak{S}_n$ by fixing each element in S. It turns out that $\widehat{\xi_M}$ vanishes on all irreducible representations λ with $\lambda_1 < n - M$; on the remaining irreducible representations, we have an explicit evaluation of $\widehat{\xi_M}$ (provided that M is not too large). More precisely, we have the following.

Lemma 3.5 *Let* $\lambda \vdash n$ *and recall that* $\lambda^* = (\lambda_2, \lambda_3, \ldots) \vdash n - \lambda_1$.

- *If* $\lambda_1 < n - M$, *then* $\widehat{\xi_M}(\lambda) = 0$.
- *If* $\lambda_1 \geq n - M$ *and* $M \leq n/3$, *then*

$$\widehat{\xi_M}(\lambda) = \frac{d_{\lambda^*}\binom{M}{n-\lambda_1}}{d_\lambda} \operatorname{Id}_{d_\lambda}.$$

The proof of this lemma utilizes several basic facts about representations of the symmetric group; we refer the reader to [12] for a very readable account.

Proof Since $\xi = \xi_M$ is a class function (i.e. conjugation invariant), it follows from Schur's lemma that for any irreducible representation λ, $\widehat{\xi}(\lambda) = a_\lambda \operatorname{Id}_{d_\lambda}$, where

$$a_\lambda = \frac{\sum_{\sigma \in \mathfrak{S}_n} \xi(\sigma)\chi^\lambda(\sigma)}{d_\lambda}.$$

Consider the partition $\mu = \mu_M = (n - M, 1, \ldots, 1)$ of n; the corresponding Young subgroup of $\mathfrak{S}_n$ [12, Definition 2.1.2] is given by

$$\mathfrak{S}_\mu = \mathfrak{S}_{\{1\}} \times \cdots \times \mathfrak{S}_{\{M\}} \times \mathfrak{S}_{\{M+1,\ldots,n\}}.$$

Let $\psi = \psi^{\text{triv}}$ denote the trivial character on $\mathfrak{S}_\mu$ and let $\psi^{\uparrow \mathfrak{S}_n}$ denote the character of the induced representation [12, Definition 1.12.2] on $\mathfrak{S}_n$. Furthermore, let $\chi^\lambda_{\downarrow \mathfrak{S}_\mu}$ denote the character of the restricted representation on $\mathfrak{S}_\mu$. The crucial fact for our analysis is that for $\sigma \in \mathfrak{S}_\mu$, we have

$$\chi^\lambda(\sigma) = \binom{n}{M} \chi^\lambda_{\downarrow \mathfrak{S}_\mu}(\sigma).$$

Then, by definition of ξ, we have that

$$\begin{aligned}
\sum_{\sigma \in \mathfrak{S}_n} \xi(\sigma)\chi^\lambda(\sigma) &= \binom{n}{M} \sum_{\sigma \in \mathfrak{S}_\mu} \frac{\chi^\lambda_{\downarrow \mathfrak{S}_\mu}(\sigma)}{\binom{n}{M}(n-M)!} \\
&= \frac{1}{(n-M)!} \sum_{\sigma \in \mathfrak{S}_\mu} \psi(\sigma)\chi^\lambda_{\downarrow \mathfrak{S}_\mu}(\sigma) \\
&= \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \psi^{\uparrow \mathfrak{S}_n}(\sigma)\chi^\lambda(\sigma),
\end{aligned}$$

where the last equality is Frobenius reciprocity [12, Theorem 1.12.6]. Since $\psi^{\uparrow \mathfrak{S}_n}$ is the character for the permutation representation associated to μ [12, Definition 2.1.5], it follows from Young's rule [12, Theorem 2.11.2] that

$$\psi^{\uparrow \mathfrak{S}_n} = \sum_{\alpha \vdash n} K_{\alpha\mu} \chi^\alpha(\sigma),$$

where $K_{\alpha\mu}$ are the Kostka numbers. Hence, by the orthonormality of characters, it follows that

$$\sum_{\sigma \in \mathfrak{S}_n} \xi(\sigma)\chi^\lambda(\sigma) = K_{\lambda\mu}.$$

Finally, let us evaluate the Kostka numbers $K_{\lambda\mu}$. By definition [12, Definition 2.11.1], these are equal to the number of semistandard Young tableaux [12, Definition 2.9.5] with $n - M$ occurrences of 1 and a single occurrence each of $2, \ldots, M + 1$.

- If $\lambda_1 < n - M$, then $K_{\lambda\mu} = 0$, since all the occurences of 1 must necessarily be in the first row. This proves the first item.
- Now, suppose $\lambda_1 \geq n - M$ and $M \leq n/3$. Notice that any valid semistandard Young tableaux can be obtained as follows: (i) first, we put 1 in the leftmost $n - M$ entries of the top row (ii) next, we choose $n - \lambda_1$ elements of $\{2, \dots, M+1\}$ and populate rows 2 and onwards with a standard Young tableaux of shape λ^* with these entries [12, Definition 2.5.1] (iii) finally, we place the remaining entries in increasing order in the top row. Therefore,

$$\begin{aligned} K_{\lambda\mu} &= \binom{M}{n-\lambda_1} \cdot \#\text{ of standard Young tableaux of shape } \lambda^* \\ &= \binom{M}{n-\lambda_1} \cdot d_{\lambda^*}, \end{aligned}$$

where the last equality follows from [12, Theorem 2.6.5]. This completes the proof of the second item. □

3.4 Proof of Theorem 1.3

We have the necessary preparation to prove Theorem 1.3.

Proof of Theorem 1.3 Recall that $t_n = \lfloor (n \log n)/2 \rfloor$, $t = t_n + nc_n(t)$, $|c_n(t)| \leq (\log\log n)/4 - 2\log\log\log n$ and $\gamma_{n,t} = e^{-2c_n(t)}$. In order to simplify computations, we consider the following truncated version μ_t of ν_t (see Definition 1.2), which is defined in an identical fashion, except that the size M_t of the random subset S_t has the distribution $\mathbb{P}[M_t = x] = \mathbb{P}[\text{Pois}(\gamma_{n,t}) = x]/\mathbb{P}[\text{Pois}(\gamma_{n,t}) \leq (\log n)^2]$, for $x \leq (\log n)^2$.

Since $\gamma_{n,t} = o(\log n)$ by our assumption on $|c_n(t)|$, it follows via standard bounds on Poisson random variables that

$$\mathbb{P}[\text{Pois}(\gamma_{n,t}) \geq (\log n)^2] \leq (\log n)^{-\Omega((\log n)^2)} = n^{-\omega(1)},$$

from which, it readily follows using the natural coupling that $d_{\text{TV}}(\nu_t, \mu_t) = n^{-\omega(1)}$. Thus, in order to prove Theorem 1.3, it suffices to prove that

$$d_{\text{TV}}(X_t, \mu_t) \leq n^{-1+o(1)}.$$

Let f_1 be the probability density function of X_t and f_2 be the probability density function of μ_t. Recall that $\mathcal{L} = \{\lambda \vdash n : \lambda_1 \geq n - (\log n)^2\}$.

Using Cauchy-Schwarz followed by Plancherel, as in [6], we have that

$$
\begin{aligned}
4\cdot d_{\mathrm{TV}}(X_t,\mu_t)^2 &\le n!\sum_{\sigma\in\mathfrak{S}_n}(f_1(\sigma)-f_2(\sigma))^2\\
&=\sum_{\lambda\in\widehat{\mathfrak{S}_n}} d_\lambda \operatorname{Tr}\Big((\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))(\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))^\dagger\Big)\\
&=\sum_{\lambda\notin\mathcal{L}} d_\lambda^2\left|\frac{1}{n}+\frac{n-1}{n}\cdot\frac{\chi_\lambda(\tau)}{d_\lambda}\right|^{2t}\\
&\quad+\sum_{\lambda\in\mathcal{L}} d_\lambda \operatorname{Tr}\Big((\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))(\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))^\dagger\Big)\\
&\le n^{-2+o(1)}+\sum_{\lambda\in\mathcal{L}} d_\lambda \operatorname{Tr}\Big((\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))(\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))^\dagger\Big);
\end{aligned}
$$

here, the third line follows from Lemmas 3.1 and 3.5 and the final line follows from Lemma 3.4.

We are now left with bounding

$$
\sum_{\lambda\in\mathcal{L}} d_\lambda \operatorname{Tr}\Big((\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))(\widehat{f_1}(\lambda)-\widehat{f_2}(\lambda))^\dagger\Big). \tag{3.4}
$$

For $\lambda\in\mathcal{L}$, $\lambda_1=n-x$ with $x\le(\log n)^2$. Therefore, by (3.3), we have that for $\lambda\in\mathcal{L}$

$$
\begin{aligned}
\widehat{f_1}(\lambda) &= \left(\frac{1}{n}+\frac{(n-1)}{n}\cdot\frac{\chi_\lambda(\tau)}{d_\lambda}\right)^t\cdot \mathrm{Id}_{d_\lambda}\\
&=\left(1-\frac{2x}{n}\cdot\frac{n-1}{n}+O(n^{-2+o(1)})\right)^t\cdot \mathrm{Id}_{d_\lambda}\\
&=\exp\left(\frac{-2xt}{n}\right)\cdot(1+O(tn^{-2+o(1)}))\cdot \mathrm{Id}_{d_\lambda}.
\end{aligned}
$$

Now note that by the definition of μ_t we have that

$$
\mu_t \overset{d}{=} \sum_{\ell=0}^{(\log n)^2}\xi_\ell\cdot\frac{\mathbb{P}[\mathrm{Pois}(\gamma_{n,t})=\ell]}{\mathbb{P}[\mathrm{Pois}(\gamma_{n,t})\le(\log n)^2]};
$$

here ξ_ℓ is the distribution in the statement of Lemma 3.5. (In other words, we are representing the distribution μ_t as a convex combination of the distributions ξ_ℓ.)

It then follows from Lemma 3.5

$$
\widehat{f_2}(\lambda)=\mathbb{P}[\mathrm{Pois}(\gamma_{n,t})\le(\log n)^2]^{-1}\cdot\sum_{\ell=0}^{(\log n)^2}\frac{d_{\lambda^*}\binom{\ell}{x}}{d_\lambda}\cdot\mathbb{P}[\mathrm{Pois}(\gamma_{n,t})=\ell]\,\mathrm{Id}_{d_\lambda}.
$$

By the second item of Lemma 3.3, we have that

$$\frac{d_{\lambda^*}}{d_\lambda} = \binom{n}{\lambda_1} \cdot (1 + n^{-1+o(1)})$$

for $\lambda \in \mathcal{L}$. This, combined with $\mathbb{P}[\mathrm{Pois}(\gamma_{n,t}) \leq (\log n)^2]^{-1} = 1 - n^{-\omega(1)}$, gives that

$$\begin{aligned}
\widehat{f_2}(\lambda) &= (1 + n^{-1+o(1)}) \sum_{\ell=0}^{(\log n)^2} \binom{\ell}{x} \cdot x! n^{-x} \cdot \frac{\gamma_{n,t}^{\ell} e^{-\gamma_{n,t}}}{\ell!} \mathrm{Id}_{d_\lambda} \\
&= (1 + n^{-1+o(1)}) \sum_{\ell=x}^{(\log n)^2} n^{-x} \cdot \frac{\gamma_{n,t}^{\ell} e^{-\gamma_{n,t}}}{(\ell - x)!} \mathrm{Id}_{d_\lambda} \\
&= (1 + n^{-1+o(1)}) n^{-x} \cdot \gamma_{n,t}^{x} \cdot \mathbb{P}[\mathrm{Pois}(\gamma_{n,t}) \leq (\log n)^2 - x] \, \mathrm{Id}_{d_\lambda} \\
&= (1 + n^{-1+o(1)}) \exp\left(\frac{-2xt}{n}\right) \cdot \mathbb{P}[\mathrm{Pois}(\gamma_{n,t}) \leq (\log n)^2 - x] \cdot \mathrm{Id}_{d_\lambda} .
\end{aligned}$$

At this point, we are almost done. We break the sum in (3.4) into two parts, depending on whether $x \geq (\log n)^2/2$ or not.

Case 1: $x \geq (\log n)^2/2$ In this case, we use the crude bound $\widehat{f_1}(\lambda) - \widehat{f_2}(\lambda) = \alpha_\lambda \, \mathrm{Id}_{d_\lambda}$ for $|\alpha_\lambda| \leq 3 \exp(-2tx/n)$, along with the dimension bound $d_\lambda \leq n^x/\sqrt{x!}$ (Lemma 3.3) to see that

$$\begin{aligned}
\sum_{\substack{\lambda \in \mathcal{L} \\ x \geq (\log n)^2/2}} d_\lambda \operatorname{Tr}\Big((\widehat{f_1}(\lambda) - \widehat{f_2}(\lambda))^2 \Big) &= \sum_{x \in \mathcal{L}, x \geq (\log n)^2/2} d_\lambda^2 |\alpha_\lambda|^2 \\
&\leq 9 \sum_{x=(\log n)^2/2}^{(\log n)^2} \sum_{\lambda \in \mathcal{L}, \lambda_1 = n-x} \exp(-4tx/n) \frac{n^{2x}}{x!} \\
&\leq n^{o(1)} \max_{x \in [(\log n)^2/2, (\log n)^2]} \frac{\exp(-4c_n(t)x)}{x!} \\
&\leq n^{-\omega(1)},
\end{aligned}$$

where in the last inequality, we used that $|c_n(t)| \leq (\log \log n)/4$.

Case 2: $x \leq (\log n)^2/2$ In this case, we have that $\mathbb{P}[\mathrm{Pois}(\gamma_{n,t}) \leq (\log n)^2 - x] = 1 - n^{-\omega(1)}$. Therefore, $\widehat{f_1}(\lambda) - \widehat{f_2}(\lambda) = \alpha_\lambda \, \mathrm{Id}_{d_\lambda}$ for $|\alpha_\lambda| \leq \exp(-2tx/n) \cdot n^{-1+o(1)}$. Using the same bounds on the dimension as above, we see that

$$\begin{aligned}
\sum_{\substack{\lambda \in \mathcal{L} \\ x < (\log n)^2/2}} d_\lambda \operatorname{Tr}\Big((\widehat{f_1}(\lambda) - \widehat{f_2}(\lambda))^2 \Big) &= \sum_{x \in \mathcal{L}, x < (\log n)^2/2} d_\lambda^2 |\alpha_\lambda|^2 \\
&\leq n^{-2+o(1)} \sum_{x=0}^{(\log n)^2/2} \sum_{\lambda \in \mathcal{L}, \lambda_1 = n-x} \exp(-4tx/n) \frac{n^{2x}}{x!}
\end{aligned}$$

$$\leq n^{-2+o(1)} \max_{x\in[0,(\log n)^2/2]} \frac{\exp(-4c_n(t)x)}{x!}$$
$$\leq n^{-2+o(1)},$$

For the last inequality, we have used the following computation. Let $0 \leq x \leq (\log n)^2/2$. Since $|c_n(t)| \leq (\log\log n)/4 - 2\log\log\log n$ by assumption, we have

$$\begin{aligned}\exp(-4c_n(t)x)/x! &\leq \exp(x\log\log n - 8x\log\log\log n)/x!\\ &\leq \frac{((\log n)/(\log\log n)^8)^x}{(x/e)^x}\\ &\leq \left(\frac{e(\log n)/(\log\log n)^8}{x}\right)^x\\ &\leq \exp(e\log n/(\log\log n)^8)\\ &= n^{o(1)}.\end{aligned}$$

□

Appendix A. Proof of Lemma 3.4

For completeness, we deduce Lemma 3.4 from the estimates in [6].

Proof of Lemma 3.4 Throughout, let $p(n)$ denote the partition number of n; we have via the Hardy–Ramanujan formula for partitions that $p(n) \lesssim e^{\pi\sqrt{2n/3}}$. Following [6], we break the sum appearing on the left hand side in the statement of Lemma 3.4 into 5 parts: S_1 is the sum over $\lambda \vdash n$ satisfying $\lambda_1 \leq n/3$ and $\lambda_1' \leq n/3$; S_2 is the sum over $\lambda \vdash n$ satisfying ($n/3 < \lambda_1 \leq n/2$ and $\lambda_1' \leq n/2$) or ($n/3 < \lambda_1' \leq n/2$ and $\lambda_1 \leq n/2$); S_3 is the sum over $\lambda \vdash n$ satisfying $n/2 < \lambda_1 \leq 0.7n$ or $n/2 < \lambda_1' \leq 0.7n$; S_4 is the sum over $\lambda \vdash n$ satisfying $0.7n < \lambda_1 \leq n - (\log n)^2$, and finally, S_5 is the sum over $\lambda \vdash n$ satisfying $0.7n < \lambda_1' \leq n$.

From [6, Equation 3.2], we have $S_1 \leq (1/3)^{2t} n! \leq n^{-\omega(1)}$; from [6, Equation 3.3], we have $S_2 \lesssim 8^n (1/2)^{2t} n^{2n/3} \leq n^{-\omega(1)}$, and from [6, Equation 3.5], we have $S_3 \lesssim 8^n (n/2)!(0.6)^{2t} \leq n^{-\omega(1)}$, where the final inequality uses the numerical inequality $\sqrt{e}\cdot 0.6 \leq 0.99$.

For the remaining summands, we have

$$\begin{aligned}S_4 &\lesssim \sum_{j=(\log n)^2}^{0.3n} \frac{p(j)}{j!}\exp\Big(2j(j-1)\log n/n - 4t'(j/n - j^2/n^2 + j/n^2)\Big)\\ &\lesssim \sum_{j=(\log n)^2}^{0.3n} \frac{p(j)}{j!}\exp(2j^2\log n/n + 4|t'|j/n)\\ &\lesssim \sum_{j=(\log n)^2}^{0.3n} \exp\Big(j + 2j^2\log n/n + j\log\log n - j\log j\Big) \lesssim n^{-\omega(1)};\end{aligned}$$

here, the first line follows from the proof of [6, Equation 3.13] (see the bottom of [6, Page 172]), and for the final estimate, we have used the easily verified numerical inequality: for $(\log n)^2 \leq j \leq 0.3n$ and all sufficiently large n

$$j + 2j^2 \log n/n + j \log\log n \leq j \log j - j.$$

Finally, we have

$$\begin{aligned}
S_5 &\lesssim e^{-4t/n} \sum_{j=0}^{0.3n} \frac{p(j)}{j!} \exp\Big(2j \log n - 4t(j/n - j^2/n^2)\Big) \\
&\lesssim e^{-4t/n} \sum_{j=0}^{0.3n} \frac{p(j)}{j!} \exp\Big(2j^2 \log n/n\Big) \exp\big(4|t'|j/n\big) \\
&\lesssim e^{-4t/n} \sum_{j=0}^{0.3n} \exp(2j + 2j^2 \log n/n - j \log j) \exp(j \log\log n) \exp(-4j \log\log\log n) \\
&\lesssim n^{-2+o(1)};
\end{aligned}$$

here, the first line follows from [6, Equation 3.6] and the first part of Lemma 3.3, and in the final line, we have used the estimate $\exp(2j + 2j^2 \log n/n + j \log\log n) = n^{o(1)}$ in the range $j \leq \log n/(\log\log n)^2$ and the estimate

$$2j + 2j^2 \log n/n + j \log\log n - j \log j - 4j \log\log\log n \leq -j$$

in the complementary range. □

Acknowledgements MS thanks Evita Nestoridi and Dominik Schmid for useful conversations. We thank Nathanaël Berestycki, Persi Diaconis, Mark Sellke and Lucas Teyssier for comments on the manuscript. VJ is supported by NSF CAREER award DMS-2237646. This research was conducted during the period MS served as a Clay Research Fellow. We thank the referees for extensive comments which have improved the manuscript considerably.

Author Contributions V.J and M.S carried out all aspects of the research and prepared the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Berestycki, N., Schramm, O., Zeitouni, O.: Mixing times for random k-cycles and coalescence-fragmentation chains. Ann. Probab. 1815–1843 (2011)
2. Berestycki, N., Şengül, B.: Cutoff for conjugacy-invariant random walks on the permutation group. Probab. Theory Relat. Fields **173**, 1197–1241 (2019)
3. Bollobás, B.: Random Graphs. Cambridge University Press, Cambridge, UK (2001)
4. Diaconis, P: Random walks on groups: characters and geometry (2003)
5. Diaconis, P., Saloff-Coste, L.: Comparison techniques for random walk on finite groups. Ann. Probab. 2131–2156 (1993)
6. Diaconis, P., Shahshahani, M.: Generating a random permutation with random transpositions. Z. Wahrsch. Verw. Gebiete **57**, 159–179 (1981)
7. Hough, B.: The random k cycle walk on the symmetric group. Probab. Theory Related Fields **165**, 447–482 (2016)
8. Matthews, P.: A strong uniform time for random transpositions. J. Theor. Probab. **1**, 411–423 (1988)
9. Nestoridi, E.: Comparing limit profiles of reversible markov chains. Electron. J. Probab. **29**, 1–14 (2024)
10. Nestoridi, E., Olesker-Taylor, S.: Limit profiles for reversible Markov chains. Probab. Theory Relat. Fields **182**, 157–188 (2022)
11. Olesker-Taylor, S., Teyssier, L., Thévenin, P.: Sharp character bounds and cutoff for symmetric groups arXiv:2503.12735v1
12. Sagan, B.E.: The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions, vol. 203. Springer Science & Business Media (2013)
13. Saloff-Coste, L., Zúñiga, J.: Refined estimates for some basic random walks on the symmetric and alternating groups. ALEA **4**, 359–392 (2008)
14. Schramm, O.: Compositions of random transpositions. Israel J. Math. **147**, 221–243 (2005)
15. Teyssier, L.: Limit profile for random transpositions. Ann. Probab. **48**, 2323–2343 (2020)
16. White, G.: A strong stationary time for random transpositions. arXiv preprint arXiv:1910.00770 (2019)