Near-independence of permutations and an almost sure polynomial bound on the diameter of the symmetric group

László Babai* and Thomas P. Hayes†

Abstract

We address the long-standing conjecture that all permutations have polynomially bounded word length in terms of any set of generators of the symmetric group S_n. This is equivalent to polynomial-time ($O(n^c)$) mixing of the (lazy) random walk on S_n where one step is multiplication by a generator or its inverse.

We prove that the conjecture is true for almost all pairs of generators. Specifically, our bound is $\widetilde{O}(n^7)$. For almost all pairs of generators, words of this length representing any given permutation can be constructed in Las Vegas polynomial time. The best previous bound on the word length for a random pair of generators was $n^{\ln n(1/2+o(1))}$ (Babai–Hetyei, 1992).

We build on recent major progress by Babai–Beals–Seress (SODA, 2004), confirming the conjecture under the assumption that at least one of the generators has degree $< 0.33n$.

The main technical contribution of the present paper is the following **near-independence** result for permutations. The *first cycle* of a permutation is the trajectory of the first element of the permutation domain. For a random permutation, the distribution of the length of the first cycle is uniform. We show that if $\tau \in S_n$ is a given permutation of degree $\geq n^{3/4}$ and $\sigma \in S_n$ is chosen at random, then the distributions of the length of the first cycle of σ and the length of the first cycle in $\sigma\tau$ are nearly independent. The ability of an essentially arbitrarily fixed permutation (τ) to "scramble" another permutation in this technical sense may be of independent interest and suggests new directions in the statistical theory of permutations pioneered by Goncharov and Erdős–Turán.

*Department of Computer Science, University of Chicago. Email: laci@cs.uchicago.edu.

†Division of Computer Science, University of California Berkeley. Email: hayest@cs.uchicago.edu. Much of this work was done while the second author was visiting Toyota Technological Institute at Chicago.

1 Pairwise near-independence of permutations

By a *random* element of a nonempty finite set S we mean an element chosen uniformly from S.

Given two independent random residue classes mod p, X and Y, one can easily generate $p+1$ *pairwise independent* random residue classes: Y and $X_i := X + iY$ ($i = 0, 1, \ldots, p-1$). This fact gives rise to a small universal family of hash functions and has been used in myriad ways in computer science in contexts where full independence would be too costly to achieve.

Given two independent random permutations $\tau, \sigma \in S_n$ (where S_n denotes the symmetric group of degree n, i. e., the set of $n!$ permutations of an n-set), can we construct a large family of pairwise independent random permutations from them, using only multiplications and inversions? Clearly, σ, τ, and $\sigma\tau$ are pairwise independent, but it is easy to show that no four words in τ and σ can be pairwise independent.

Yet one would expect that τ and $\sigma\tau^i$ ($i = 1, \ldots, r$) should be pairwise "nearly independent" for rather large values of r (like $r = \Omega(n)$; but r might even grow at a slightly superpolynomial rate). The question is, how to measure "near-independence" of permutations.

Definition 1.1. We say that a sequence of pairs of (real-valued) random variables X_n and Y_n is *nearly independent* if for all $x, y \in \mathbb{R}$,

$$\mathbf{Pr}(X_n \leq x \text{ and } Y_n \leq y) = \mathbf{Pr}(X_n \leq x)\mathbf{Pr}(Y_n \leq y) + o(1),$$

where the $o(1)$ term approaches zero (uniformly in x and y) as $n \to \infty$.

One of the most important parameters of a permutation π is the length of its **first cycle**, i. e., the length of the trajectory of the first point of the permutation domain: $1, 1^\pi, 1^{\pi^2}, \ldots$. For a random permutation $\pi \in S_n$, the length of the first cycle of π is uniformly distributed over $\{1, 2, \ldots, n\}$ (cf. [20, Ex.3.3]).

In this paper we show:

Theorem 1.2. *Given a pair of independent random permutations $\tau, \sigma \in S_n$, the lengths of the first cycles* *of the $s+1$ permutations τ and $\sigma\tau^i$ $(i=1,\dots,s)$ are pairwise nearly independent up to $s = n^{(7/32-o(1))\ln n}$.*

In fact, to generate this degree of pairwise near-independence, we do not need the amount of randomness afforded by a pair of random permutations; one random permutation suffices. This is formalized in the next result which is the **main technical contribution** of this paper. The *degree* $\deg(\tau)$ of a permutation τ is the number of elements moved by τ.

THEOREM 1.3. *Let us fix $\tau \in S_n$ and choose $\sigma \in S_n$ at random. Assume $\deg(\tau) \geq n^{3/4}$. Let Y denote the length of the first cycle of σ and X the length of the first cycle of $\sigma\tau$. Then for all k, ℓ $(1 \leq k, \ell \leq n)$,*

$$\left| \mathbf{Pr}\left(X \leq k \text{ and } Y \leq \ell\right) - \frac{k\ell}{n^2} \right| \leq n^{-1/8+o(1)}.$$

Here, the $o(1)$ term approaches zero as $n \to \infty$ uniformly in k and ℓ. The proof of Theorem 1.3 is presented in Sections 7 to 10.

The immediate motivation for these results is described in Sections 2, 3, and 6. We believe, however, that Theorem 1.3 is of wider interest and suggests a new direction in the statistical theory of permutations pioneered by Goncharov [15] and, in a series of papers, by Erdős and Turán [11, 12, 13] (see the Open Problems section).

2 Diameter of Cayley graphs

Let G be a finite group and S a set of generators of G. We consider the undirected Cayley graph $\Gamma(G,S)$ which has G as its vertex set; the pairs $\{\{g, gs\} : g \in G, s \in S\}$ are the edges. Let $\operatorname{diam}(G,S)$ denote the diameter of $\Gamma(G,S)$.

In this paper we consider the cases $G = S_n$ and $G = A_n$ (the symmetric group consisting of the $n!$ permutations of a set of n elements, and the alternating group consisting of the $n!/2$ even permutations). We address the following long-standing conjecture.

CONJECTURE 2.1. ([7]) *For $G = S_n$ and $G = A_n$, the diameter $\operatorname{diam}(G,S)$ is polynomially bounded in terms of n for all sets S of generators.*

The best upper bound known to hold for the diameter of all Cayley graphs of A_n and S_n is $\exp(\sqrt{n \ln n}(1+o(1)))$ [7].

Regarding random pairs of generators, Dixon's classical result states that almost all pairs of permutations in S_n generate either S_n or A_n [9] (cf. [8, 2]).

The **main result** of the present paper shows that the generation a.a. leads to polynomial diameter:

THEOREM 2.2. *For almost all pairs of permutations $S = \{\sigma, \tau\}$ of $G = S_n$ or A_n, the diameter of the Cayley graph $\Gamma(G,S)$ is bounded by $O(n^C)$ where C is a constant. (Our current estimate of C is $7+o(1)$.)*

The best previously known bound for almost all pairs of generators was $n^{\ln n(1/2+o(1))}$ [4].

3 Previous work

The diameter of Cayley graphs has been studied in a number of contexts, including interconnection networks, expanders, puzzles such as Rubik's cube and Rubik's rings, card shuffling and rapid mixing, random generation of group elements, and combinatorial group theory.

Even and Goldreich [14] proved that finding the diameter of a Cayley graph of a permutation group is NP-hard even for the basic case when the group is an elementary abelian 2-group (every element has order 2). Jerrum [18] proved that for directed Cayley graphs of permutation groups, to find the directed distance between two permutations is PSPACE-hard. No approximation algorithm is known for distance in and the diameter of Cayley graphs of permutation groups. Strikingly, the question of the diameter of the Rubik's cube Cayley graph appears to be wide open (cf. [19]). We refer to [5] for more information on the history of the diameter problem and related results and to the survey [16] for applications of Cayley graphs to interconnection networks.

Prior to [3], Conjecture 2.1 had only been verified for very special classes of generating sets. Driscoll and Furst [10] proved an $O(n^2)$ bound for the case when all generators are cycles of bounded lengths and McKenzie [21] gave a polynomial bound for the case when the generators have bounded degree. Major progress on Conjecture 2.1 was made by Babai-Beals-Seress in [3].

THEOREM 3.1. ([3]) *Let S be a set of generators for $G = S_n$ or $G = A_n$. Assume that S contains a permutation of degree $\leq 0.33n$. Then $\operatorname{diam}(G,S) = O(n^C)$ where C is a constant.*

The bound on C stated in [3] was $7+o(1)$; and $6+o(1)$ for bounded $|S|$.

For random pairs of generators, it was pointed out in [3] that Theorem 3.1 implies a polynomial bound on the diameter of the Cayley graphs of S_n and A_n with probability $1-\varepsilon$; but the exponent depended on $\varepsilon > 0$. The move from probability $(1-\varepsilon)$ to "almost all" (Theorem 2.2) was remarkably difficult and required new insights into the "near-independence" of deterministically related permutations. We believe that these insights bring us closer to settling Conjecture 2.1;

and the new type of questions in the statistical theory of permutations initiated by Theorem 1.3 may have further applications.

4 Some statistical group theory

Henceforth, we will assume n is sufficiently large. By an **"almost certain" event** we shall mean a sequence of events depending on the parameter n such that the probability of the events approaches 1 as $n \to \infty$. The abbreviation "a.a." ("almost always") or equivalently, "almost surely," etc., refers to such a sequence of events.

In this section we study the typical behavior of permutations. We shall use the following terminology.

Definition 4.1. Let $\sigma \in S_n$ be a permutation acting on the set $[n] = \{1, \ldots, n\}$. The "first cycle" of σ is the trajectory of 1; the second cycle is the trajectory of the smallest i that does not belong to the first cycle, etc. Let ℓ_i denote the length of the i-th cycle; set $\ell_i = 0$ if the number of cycles is less than i. Let T_i denote the set of elements in $[n]$ not covered by the first $i-1$ cycles.

Observation 4.2. *Let $\sigma \in S_n$ be a random permutation. Given T_k, the restriction $\sigma_k := (\sigma \,|\, T_k)$ is a random permutation of T_k. Consequently, ℓ_k is uniformly distributed in $\{1, \ldots, |T_k|\}$.*

First we give an explicit upper bound on the probability that a random permutation has very low order. This is required for the proof of Theorem 2.2 (Section 6).

Proposition 4.3. *The probability that the order of a random permutation $\sigma \in S_n$ is $\le n$ is $O(n^{-1/4})$.*

Proof. (Sketch.) We prove that the probability that l.c.m.$[\ell_1, \ell_2] \le n$ is $1 - O(n^{-1/4})$. It is easy to prove that with this probability, both ℓ_1 and ℓ_2 are greater than $n^{3/4}$, so it suffices to prove that the probability that g.c.d.$(\ell_1, \ell_2) \ge n^{1/4}$ is $O(n^{-1/4})$. To this end, we observe that for any d and i the probability that $d \,|\, \ell_i$, conditioned under any sequence $\ell_1, \ldots, \ell_{i-1}$ such that $\sum_{j=1}^{i-1} \ell_j < n$, is $\le 1/d$ since ℓ_i is uniformly distributed in a prefix of the positive integers. Therefore, assuming $\ell_1 < n$, the probability that $d \,|\, \ell_1$ and $d \,|\, \ell_2$ is at most $1/d^2$. Hence the probability that g.c.d.$(\ell_1, \ell_2) \ge r$ is less than $1/n$ plus $\sum_{d=r}^{\infty} 1/d^2 < \sum_{d=r}^{\infty} 1/d(d-1) = 1/(r-1)$. Apply this bound with $r = n^{1/4}$. $\square$

The following result, the main result of this section, is is an ingredient in the proof of Theorem 1.2 (Section 11).

Theorem 4.4. *For a permutation σ and a number K, let $e(\sigma, K)$ denote the smallest $j \ge 1$ such that $\deg(\sigma^j) \le K$. Let us fix p, $0 < p < 1$. Then, for a.a. $\sigma \in S_n$, we have $e(\sigma, n^p) \ge \exp\left(((1-p^2)/2 - o(1)) \ln^2 n\right)$.*

We shall use the following classical result of Erdős and Turán on the typical order of a permutation.

Theorem 4.5. (Erdős – Turán [11]) *A.a., the order of a random permutation $\sigma \in S_n$ is*

$$\mathrm{ord}(\sigma) = \exp\left((1/2 + o(1)) \ln^2 n\right).$$

We shall also make use of the following result concerning the typical distribution of cycle lengths in a random permutation. For the notation see Def. 4.1.

Lemma 4.6. (Babai – Hetyei [4], Lemma 3.1) *Let us fix r, $0 < r < 1$, and let $k = \lfloor (1-r) \ln n \rfloor$. Let $\sigma \in S_n$ be a random permutation. Then a.a., $|T_k| = n^{r(1+o(1))}$.*

Corollary 4.7. *Using the notation of Lemma 4.6, a.a., the length of each of the first k cycles is at least $n^{r(1+o(1))}$.*

Proof. Let us fix $\varepsilon > 0$. Recall Def. 4.1 and Obs. 4.2. Let us consider the following events.

- A : $|T_k| \le n^{r-\varepsilon}$. We have $\mathbf{Pr}\,(A) = o(1)$ by Lemma 4.6.
- $A_i : |T_i| \le n^{r-\varepsilon}$. Note that $\bigcup_{i=1}^{k} A_i = A_k = A$.
- $B_i : \ell_i \le n^{r-2\varepsilon}$.

  Observe that $\mathbf{Pr}\,(B_i \,|\, \neg A_i) \le n^{-\varepsilon}$ because ℓ_i is uniform in $\{1, ..., |T_i|\}$.
- $B : (\exists i \le k)(\ell_i \le n^{r-2\varepsilon})$.

Now $\mathbf{Pr}\,(B) \le \mathbf{Pr}\,(A) + \mathbf{Pr}\,(B \wedge \neg A) \le \mathbf{Pr}\,(A) + \sum_{i=1}^{k} \mathbf{Pr}\,(B_i \wedge \neg A) \le \mathbf{Pr}\,(A) + \sum_{i=1}^{k} \mathbf{Pr}\,(B_i \wedge \neg A_i) \le \mathbf{Pr}\,(A) + \sum_{i=1}^{k} \mathbf{Pr}\,(B_i \,|\, \neg A_i) \le \mathbf{Pr}\,(A) + kn^{-\varepsilon} = o(1)$, proving the claim. $\square$

Proof of Theorem 4.4. Let us fix r such that $1 > r > p$ and let $k = \lfloor (1-r) \ln n \rfloor$. For convenience we shall omit rounding. Let us also fix $\varepsilon > 0$. As before, ℓ_i denotes the length of the i-th cycle in σ.

Let us consider the following events, all of which will be shown to have probability $o(1)$. Fix $\varepsilon > 0$.

- $B : (\exists i < k)(\ell_i \le n^{r-\varepsilon})$.

  We have $\mathbf{Pr}\,(B) = o(1)$ by Corollary 4.7.
- C: $|T_k| > n^{r+\varepsilon}$. We have $\mathbf{Pr}\,(D) = o(1)$ by Lemma 4.6.
- D : the order of $\sigma_k := (\sigma \,|\, T_k)$ (the restriction of σ to T_k) is $\mathrm{ord}(\sigma) \le \exp((\ln^2 |T_k|/2)(1+\varepsilon))$.

  We have $\mathbf{Pr}\,(D) \le \mathbf{Pr}\,(C) + \mathbf{Pr}\,(D \,|\, \neg C) = o(1)$.

  This follows from Theorem 4.5, noting that under condition $\neg C$, the restriction σ_k is a uniform random permutation of T_k.

- $E : \mathrm{ord}(\sigma_k) \ge \exp((r^2/2)\ln^2 n(1+2\varepsilon))$.

  $E \subseteq D \cup C$ so $\mathbf{Pr}(E) = o(1)$.

- $F_j : \deg(\sigma^j) \le n^{r-2\varepsilon}$.

- $F(s) : (\exists j)(1 \le j \le s)(\deg(\sigma^j) \le n^{r-2\varepsilon})$.

  Note that $F(s) := \bigcup_{j=1}^{s} F_j$. Our goal is to prove that for appropriate choice of s, $F(s)$ has vanishing probability.

- $G_j : \mathrm{ord}(\sigma)$ divides $j\,\mathrm{ord}(\sigma_k)$.

  Note that $F_j \subseteq B \cup G_j$ since if $\ell_i > n^{r-2\varepsilon}$ for all $i < k$ then the only way for σ^j to have degree $\le |T_k|$ is to fix all points outside T_k.

- $G(s) : \mathrm{ord}(\sigma) \le s\,\mathrm{ord}(\sigma_k)$.

  So $\bigcup_{j=1}^{s} G_j \subseteq G(s)$, hence $F(s) \subseteq B \cup G(s)$.

- $H(s) : \mathrm{ord}(\sigma) \le s\exp((r^2/2)(\ln^2 n)(1+2\varepsilon))$.

- $K : \mathrm{ord}(\sigma) \le \exp((\ln^2 n/2)(1-\varepsilon))$.

  Note that $\mathbf{Pr}(K) = o(1)$ by Theorem 4.5.

Let now $s = \exp((1-r^2-\varepsilon)/2)\ln^2 n$. Note that $s < \exp((\ln^2 n/2)(1-\varepsilon)) - (r^2/2)\ln^2 n(1+2\varepsilon))$. Therefore $H(s) \subseteq K \cup E$; consequently, $\mathbf{Pr}(H(s)) = o(1)$. Note further that $G(s) \subseteq H(s) \cup E$; therefore $\mathbf{Pr}(G(s)) = o(1)$. The relation $F(s) \subseteq B \cup G(s)$ then implies that $\mathbf{Pr}(F(s)) = o(1)$.

So for $s = \exp(((1-r^2-\varepsilon)/2)\ln^2 n$, the statement $F(s)$ a.a. fails to hold. Noting that this is true for all $r > p$ and all $\varepsilon > 0$, Theorem 4.4 follows. $\square$

Finally we include an observation on cycle lengths, to be used in the proof of Theorem 1.3 (Section 10).

PROPOSITION 4.8. *For a random $\sigma \in S_n$, let X_k denote the total length of cycles of lengths $\le k$. Then, for every $\ell > 0$, $\mathbf{Pr}(X_k \ge \ell) \le k/\ell$.*

Proof. It is easy to see that $\mathbf{E}(X_k) = k$; therefore the stated bound follows by Markov's inequality. $\square$

5 Tail Estimates

In this section, we will present several general tail estimates which will be needed later. In particular, Theorem 5.6 will be one of the key ingredients in the proof of Theorem 1.3 (Section 10).

The following version of Chernoff's bound is especially useful for variables with a "biased" distribution. This version is an immediate consequence of [22, Theorem 4.1, p. 68].

THEOREM 5.1. (CHERNOFF) *Let $X_1, \dots, X_N$ be independent random variables, bounded by $0 \le X_i \le 1$. Let $X = \sum_{i=1}^{N} X_i$ have expectation μ. Then for every $r > \mu$,*

$$\mathbf{Pr}(X \ge r) \le \left(\frac{\mathrm{e}\mu}{r}\right)^r. \tag{5.1}$$

For every $r < \mu$,

$$\mathbf{Pr}(X \le r) \le \exp\left(\frac{-(\mu-r)^2}{2\mu}\right). \tag{5.2}$$

The following corollary will be used in the proof of Lemma 8.2.

COROLLARY 5.2. *Let $b, \mu_1, \dots, \mu_N$ be positive reals. Let $\mu = \sum_{i=1}^{N} \mu_i$. Suppose $Y_1, \dots, Y_N$ is a sequence of random variables such that $0 \le Y_i \le b$. Assume that, for every $i \ge 0$, for every history $Y_1, \dots, Y_{i-1}$, the conditional expectation of the indicator of the event $\{Y_i \ne 0\}$ satisfies*

$$\mathbf{Pr}(Y_i \ne 0 \mid Y_1, \dots, Y_{i-1}) \le \mu_i.$$

Then, letting $Y = \sum_{i=1}^{N} Y_i$, for every $s > 0$,

$$\mathbf{Pr}(Y \ge s) \le \left(\frac{\mathrm{e}b\mu}{s}\right)^{s/b}.$$

Proof. It is easy to construct independent random variables $X_1, \dots, X_N$ with values in $\{0,1\}$, satisfying the conditions of Theorem 5.1 such that $\mathbf{E}(X_i) = \mu_i$ and $bX_i \ge Y_i$ holds with probability one. Substituting $r = s/b$ in (5.1) completes the proof. $\square$

Our next bound, on the size of the intersection of a given set with a random set, will be used in the proof of Lemma 8.1.

LEMMA 5.3. *Let $T \subseteq V$ and $\alpha := |T|/n$. Sample Π uniformly at random from among subsets of V of size t. Then*

$$\mathbf{Pr}(|\Pi \cap T| \le \alpha t/2) \le \exp(-\alpha t/8).$$

Proof. Generate Π by sampling t elements of V sequentially, without replacement. Note that, if this sampling were performed with replacement, the probability of the event $|\Pi \cap T| \le \alpha t/2$ would only increase, since

$$\mathbf{Pr}(|\Pi \cap T| \le \alpha t/2 \mid |\Pi| = s)$$

would then clearly be a decreasing function of s. That the desired upper bound holds even when Π is sampled "with replacement" follows from (5.2). $\square$

We will need the following extension of the (unbiased) Chernoff's bound to real-valued martingales with bounded differences, due to W. Hoeffding [17], and often referred to as the Hoeffding-Azuma inequality (cf. [1, Thm.7.2.1]).

Theorem 5.4. (Hoeffding-Azuma) *Let $X_0, \ldots, X_n$ be a real martingale, i. e., a random process such that $\mathbf{E}(X_i \mid X_0, \ldots, X_{i-1}) = X_{i-1}$ for every i. Suppose further that $|X_i - X_{i-1}| \le 1$. Then, for every $a \ge 0$,*

$$\mathbf{Pr}(X_n < X_0 - a) \le \exp\left(-a^2/2n\right).$$

Finally, we will prove two concentration inequalities on the distributions of random subset sums. In the first case, we condition on the size of the subset.

Lemma 5.5. *Let $a_1, \ldots, a_m \in [0,1]$, let $b = \sum_{i=1}^m a_i$, and let $k \in \{1, \ldots, m\}$. Let S be a random subset of $\{1, \ldots, m\}$ of size k, and let $Y_k = \sum_{i \in S} a_i$. Then, for every $s > 0$,*

$$\mathbf{Pr}(Y_k \le kb/m - s) < \exp(-s^2/2k). \tag{5.3}$$

Proof. Define random variables $Y_0, \ldots, Y_k$ by first selecting $\psi \in S_m$ uniformly at random, then setting $Y_i = \sum_{j=1}^i a_{i^\psi}$, for $i \in \{0, \ldots, k\}$. Note that Y_k has the same distribution as in the lemma.

For $0 \le i \le k$, set $Z_i = \mathbf{E}(Y_k \mid Y_0, \ldots, Y_i)$, so that $Z_0, \ldots, Z_k$ is a (Doob) martingale. Note that $Z_0 = \mathbf{E}(Y_k) = kb/m$, and $Z_k = Y_k$. It is clear that, for all i, $|Z_i - Z_{i-1}| \le 1$. (In fact, an easy shuffling argument shows that $|Z_i - Z_{i-1}| \le (m-k)/(m+1-i)$.)

Inequality (5.3) now follows from Theorem 5.4 applied to the deviation $Z_k - Z_0 = Y_k - kb/m$. □

Next, we consider the case when k is chosen uniformly at random from $\{1, \ldots, m\}$, and the random subset has size k. Theorem 5.6, which will be applied in the proof of Theorem 1.3 (Section 10), says that, as long as the set of values is not too skewed, the distribution of the sum is nearly uniform.

Theorem 5.6. *Let $m > 1$, let $a_1, \ldots, a_m \in [0,1]$, and let $b = \sum_{i=1}^m a_i$. Let $k \in \{1, \ldots, m\}$ be uniformly random, and let S be a randomly chosen subset of $\{1, \ldots, m\}$ of size k. Let $Y = \sum_{i \in S} a_i$. Then, for every $x \in [0, b]$,*

$$\left|\mathbf{Pr}(Y \le x) - \frac{x}{b}\right| \le \frac{2\sqrt{m \ln m}}{b}. \tag{5.4}$$

Proof. Assume $b > 2\sqrt{m \ln m}$; otherwise the conclusion holds vacuously.

First, we estimate the probability that $Y \le x$. Define $Y_1, \ldots, Y_m$ as in the proof of Lemma 5.5. Then we can think of Y as Y_k, where k is uniformly random in $\{1, \ldots, m\}$.

For every ℓ, we have

$$\begin{aligned} \mathbf{Pr}(Y \le x) &= \frac{1}{m} \sum_{k=1}^m \mathbf{Pr}(Y_k \le x) \\ &\le \frac{\ell}{m} + \frac{1}{m} \sum_{k=\lceil \ell \rceil}^m \mathbf{Pr}(Y_k \le x). \end{aligned}$$

Let $\ell = (x+s)m/b$, where $s > 0$ will be specified later. Then, by Lemma 5.5, for each $k \ge \ell$,

$$\mathbf{Pr}(Y_k \le x) \le \mathbf{Pr}(Y_k \le kb/m - s) \le \exp(-s^2/2m),$$

and therefore,

$$\mathbf{Pr}(Y \le x) \le \frac{x+s}{b} + \exp(-s^2/2m). \tag{5.5}$$

Setting $s = \sqrt{2m \ln(b/\sqrt{m \ln m})} \in \left[0, \sqrt{m \ln m}\right]$, equation (5.5) implies

$$\mathbf{Pr}(Y \le x) \le \frac{x}{b} + \frac{2\sqrt{m \ln m}}{b} \tag{5.6}$$

Observing that the distribution of $b - Y$ is identical to the distribution of Y, we infer that

$$\mathbf{Pr}(b - Y \le b - x) \le \frac{b-x}{b} + \frac{2\sqrt{m \ln m}}{b}. \tag{5.7}$$

Combining (5.6) and (5.7) yields (5.4). □

6 Almost sure diameter bound via pairwise near-independence: a Chebyshev argument

In this section we deduce our main result, Theorem 2.2, from Theorem 3.1 combined with our "main technical contribution," Theorem 1.3.

First we state a corollary to Theorem 1.3. For $\pi \in S_n$, we use $c(\pi, i)$ to denote the π-cycle containing point i, so $|c(\pi, 1)|$ is the length of the *first cycle* of π.

Corollary 6.1. *Suppose $\tau \in S_n$ has degree at least $n/4$. Sample σ uniformly at random from S_n. Then the events $A = \{|c(\sigma, 1)| > 3n/4\}$ and $B = \{|c(\sigma\tau, 1)| > 3n/4\}$ are nearly independent. More precisely, $\mathbf{Pr}(A) = 1/4 + o(1)$, $\mathbf{Pr}(B) = 1/4 + o(1)$, and $\mathbf{Pr}(A \text{ and } B) = 1/16 + o(1)$.*

The Chebyshev argument follows.

Lemma 6.2. *Suppose $\tau \in S_n$ is such that for $1 \le i \le 10 \log n$, τ^i has degree at least $n/4$. Sample σ uniformly at random from S_n. Then, a.a., for at least one of the permutations $\sigma\tau^i$, for $1 \le i \le 10 \log n$, the length of the first cycle is greater than $3n/4$.*

Proof. Let $\tau \in S_n$, and suppose that each of $\tau, \tau^2, \ldots, \tau^{10 \log n}$ has degree at least $n/4$. Sample $\sigma \in S_n$ uniformly at random.

For $1 \le i \le 10 \log n$, let A_i denote the event that $|c(\sigma\tau^i, 1)| > 3n/4$. Since $\sigma\tau^i$ is distributed uniformly over S_n, $|c(\sigma\tau^i, 1)|$ is uniformly distributed over $1, \ldots, n$, and so $\mathbf{Pr}(A_i) = 1/4 + o(1)$.

By Corollary 6.1 applied to τ and σ, it follows that A_0 and A_1 are nearly independent. More generally, for $0 \le i < j \le 10 \log n$, applying Corollary 6.1 to τ^{j-i} (in place of τ) and $\sigma\tau^i$ (in place of σ), it follows that A_i and A_j are nearly independent.

Let X denote the number of events A_i which occur. Then $X = \sum_{i=1}^{10 \log n} X_i$, where X_i denotes the indicator variable for event A_i. It follows that $\mathbf{E}(X) = \Theta(\log n)$ and $\mathbf{Var}(X) = o(\log^2 n)$. Hence, by Chebyshev's inequality, $\mathbf{Pr}(X = 0) \le \mathbf{Var}(X)/\mathbf{E}(X)^2 = o(1)$. Hence a.a., at least one of the events A_i occurs. □

Proof of Theorem 2.2, assuming Theorem 1.3. According to Theorem 3.1, it suffices to show that almost all pairs of permutations generate a nonidentity permutation of degree $\le n/4$ as a short word. This, in turn, follows from Lemma 6.2.

Indeed, either some τ^i has degree $< n/4$, or there is almost always some $\sigma\tau^i$ with a cycle of length $j > 3n/4$, where in both cases $i \le 10 \log n$. In the latter case, $(\sigma\tau^i)^j$ has degree $< n/4$.

Finally, we need to show that none of these permutations is the identity. Note that τ as well as every $\sigma\tau^i$ is uniformly distributed in S_n. Therefore the probability that either τ or any of the $\sigma\tau^i$ $(1 \le i \le 10 \log n)$ has order $\le n$ is $O(\log n/n^{1/4})$ by Proposition 4.3. □

7 Partial Permutation Graphs

In this section we develop notation and terminology for a theory of partial permutation graphs which will provide the language for the asymptotic structure theory to follow in the subsequent sections.

Let V be a set of size n. For $\pi \in \mathrm{Sym}(V)$, let $E_\pi = \{(v, v^\pi) \mid v \in V\} \subseteq V \times V$. Then the map $\pi \mapsto (V, E_\pi)$ defines a bijection between $\mathrm{Sym}(V)$ and the set of directed graphs on vertex set V such that every vertex has in-degree 1 and out-degree 1. We will refer to such graphs as *permutation graphs*. More generally, if G is a graph on vertex set V such that every in-degree and out-degree is *at most* 1, then we call G a *partial permutation graph*.

Notation 7.1. Let $E \subseteq V^2$. We will implicitly identify the edge set E with the graph (V, E). Except where otherwise specified, by "component" of E, we shall always mean "weakly connected component containing at least one edge."

Definition 7.2. Let $\pi \in \mathrm{Sym}(V)$, let $t \ge 0$, and let $v \in V$. The *t-trajectory of v under* π, denoted $\mathrm{traj}(v, \pi, t)$, is the partial permutation graph with edge set

$$\{(v^{\pi^{j-1}}, v^{\pi^j}) \mid 1 \le j \le t\}.$$

Note that (for $t > 0$) these edges form a connected subgraph of E_π, which is therefore either a path or a cycle, according to whether the π-cycle containing v has size greater than t or not. $|\mathrm{traj}(v, \pi, t)|$ is always the lesser of t and the size of the π-cycle containing v.

Definition 7.3. Let $\tau \in \mathrm{Sym}(V)$. The *τ-shear* function is the map $\mathrm{shear}_\tau : V^2 \to V^2$ defined by $\mathrm{shear}_\tau(x, y) = (x, y^\tau)$. For $E \subseteq V^2$, we define $\mathrm{shear}_\tau(E) := \{\mathrm{shear}_\tau(e) \mid e \in E\}$.

The shear operator provides a way of viewing composition of permutations as an edgewise operation on partial permutation graphs, which respects trajectories, as we now state formally.

Observation 7.4. *Let $\sigma, \tau \in \mathrm{Sym}(V)$, and set $\pi = \sigma\tau$. Then*

$$E_\pi = \mathrm{shear}_\tau(E_\sigma),$$

where E_π and E_σ denote the edge sets of the permutation graphs of π and σ, respectively.

Consequently, for every partial permutation $\rho \subseteq \sigma$, we have $\mathrm{shear}_\tau(E_\rho) \subseteq E_\pi$.

Definition 7.5. Let $\tau \in \mathrm{Sym}(V)$. We say τ is *fixed-point-free* if τ has degree n, i. e., for every $v \in V$, $v^\tau \ne v$.

Next we define a projection operator, which will be the key tool allowing us to apply results about fixed-point-free τ to the general case.

Definition 7.6. For $T \subseteq V$, we define the *projection* $\pi \mapsto \pi_T$, mapping $\mathrm{Sym}(V) \to \mathrm{Sym}(T)$, as follows. Let $\pi \in \mathrm{Sym}(V)$. For $i \in T$, let k denote the smallest positive integer such that $i^{\pi^k} \in T$. Set $i^{\pi_T} = i^{\pi^k}$.

We extend this notion to partial permutations π in the natural way: if there exists k such that $i^{\pi^k} \in T$, then take k minimal and set $i^{\pi_T} = i^{\pi^k}$.

We now state some basic facts about projections. Observations 7.8 and 7.10 will be used in the proof of Lemma 8.1.

Observation 7.7. *Let π be any partial permutation on V, and let $T \subseteq V$. Projection onto T defines a one-to-one correspondence between those cycles of π which have non-empty intersection with T, and the cycles of π_T. It also defines a one-to-one correspondence between those paths of π which intersect T in at least 2 points, with the paths of π_T.*

Observation 7.8. *Let $v \in V$, $\ell \ge 1$, $T \subseteq V$, and let π be uniformly random in* $\mathrm{Sym}(V)$. *Let $E = \mathrm{traj}(v, \pi, \ell)$.*

(A) Suppose $v \notin T$. Then, for every $s \ge 0$,

(A1) *Conditioned on the event that E is a cycle containing exactly s points of T, the distribution of E_T is uniformly random over cycles of length s on T.*

(A2) *Conditioned on E a path containing exactly s points of T, the E_T is uniformly random over paths of length $s-1$ on T.*

(B) *Suppose $v \in T$. Then, for every $s \geq 0$,*

(B1) *Conditioned on the event that E is a cycle containing exactly s points of T, the distribution of E_T is uniformly random over cycles of length s on T, containing v.*

(B2) *Conditioned on E a path containing exactly s points of T, the E_T is uniformly random over paths of length $s-1$ on T, starting from v.*

Observation 7.9. *Let $\tau \in \mathrm{Sym}(V)$. Then projection onto* $\mathrm{supp}(\tau)$ *commutes with* shear_τ, *considered as operators on the set of all partial permutations of V.*

Observation 7.10. *Let $\tau \in \mathrm{Sym}(V)$ and $T = \mathrm{supp}(\tau)$. Let C be a cycle containing at least one vertex of T. Then* $\mathrm{shear}_\tau(C)$ *has the same number of connected components as* $\mathrm{shear}_\tau(C_T)$. *Let P be a path containing at least one vertex of T. Then* $\mathrm{shear}_\tau(P)$ *has either* 0, 1 *or* 2 *more components than* $\mathrm{shear}_\tau(P_T)$. *Hence for any trajectory E (hitting T or not)* $\mathrm{shear}_\tau(E)$ *has* 0, 1, *or* 2 *more components than than* $\mathrm{shear}_\tau(E_T)$.

Finally, we introduce a notion of "contraction" by a partial permutation.

Definition 7.11. (Contraction of a set by a partial permutation) Let ρ be a partial permutation on vertex set R. Note that each component of ρ is either a path (this includes isolated points) or a cycle (this includes cycles of length one). We define the contracted set R/ρ to be the set of those components of ρ which are paths (including isolated points). (Note that we throw out all cycles and contract all paths.) If $v \in V$ is contained in a path of ρ, and hence has a corresponding point in V/ρ, we say v *survives* contraction by ρ; if not, we say v is *killed* under contraction by ρ.

Definition 7.12. (Contraction of a permutation by a partial permutation) Let ρ be a partial permutation on R, and $\pi \in \mathrm{Sym}(R)$. Assume $\rho \subseteq \pi$. We define the contraction $\pi^* = \pi/\rho \in \mathrm{Sym}(R/\rho)$, as follows. For $p \in R/\rho$, let p be a path from $\mathrm{tail}(p)$ to $\mathrm{head}(p)$. For $p, q \in R/\rho$, we set $p^{\pi^*} = q$ if $\mathrm{head}(p)^\pi = \mathrm{tail}(q)$.

Observation 7.13. *Fix a partial permutation ρ of R. Sample $\pi \in \mathrm{Sym}(R)$ uniformly at random, conditioned on $\rho \subseteq \pi$. Then π/ρ is distributed uniformly over* $\mathrm{Sym}(R/\rho)$.

Observation 7.13 will be used in the proofs of Lemma 9.1 and Theorem 1.3.

8 Shears usually have many components

Suppose $\tau \in \mathrm{Sym}(V)$ has degree αn. For a random path of length t, we would intuitively expect the operation of shearing by τ to cut the path into about αt pieces. We now present a formal justification for this intuition, as long as t is not too large. Indeed, if $\alpha t = \Omega(\log n)$ and $t = o(n)$, then it is almost certain that *all* trajectories of length t are cut into $\Theta(\alpha t)$ pieces (not including trajectories in cycles of length less than t, of course).

Lemma 8.1. *Let $v \in V$, $t \geq 1$, and $\tau \in \mathrm{Sym}(V)$. Let $\alpha := \deg(\tau)/n$. Sample $\pi \in \mathrm{Sym}(V)$ uniformly at random, and let p denote the number of components of* $\mathrm{shear}_\tau(\mathrm{traj}(v,\pi,t))$. *Then*

$$\mathbf{Pr}\left(p \leq \frac{\alpha t}{4} \;\middle|\; |\,\mathrm{traj}(v,\pi,t)|=t\right) \leq 2\mathrm{e}^{-\alpha t/8} + \left(\frac{54\mathrm{e}t}{n}\right)^{\alpha t/12}.$$

We will first prove that, when τ is fixed-point-free, small random sets are very far from being fixed by τ. The proof of Lemma 8.1 is at the end of the section.

Lemma 8.2. *Let $v \in V$, $t \geq 1$, and let $\tau \in Sym(V)$ be fixed-point-free. Sample Π uniformly at random from among subsets of V of size $t+1$ containing v. Then, for all $s \geq 1$,*

$$\mathbf{Pr}\left(|\Pi \cap \Pi^\tau| \geq s\right) \leq \left(\frac{6\mathrm{e}t^2}{sn}\right)^{s/2}.$$

Proof. The theorem holds vacuously if $t > n/4$, so we assume $t \leq n/4$.

Generate $\Pi = \{v_1, \ldots, v_{t+1}\}$ by setting $v_1 = v$, and sequentially sampling $v_2, \ldots, v_{t+1}$ from $V \setminus \{v\}$, without replacement. For $0 \leq i \leq t+1$, let $\Pi_i = \{v_j \mid j \leq i\}$, and let $Z_i = |\Pi_i \cap \Pi_i^\tau|$. For $1 \leq i \leq t+1$, set $Y_i = Z_i - Z_{i-1}$. Note that Y_i is a random variable taking values in $\{0,1,2\}$. We investigate $Z_{t+1} = \sum_{i=1}^{t+1} Y_i = |\Pi \cap \Pi^\tau|$.

Suppose we are given Π_i. Conditioned on this information, v_{i+1} is uniformly distributed over $V \setminus \Pi_i$, a set of size $n-i$. Hence

$$\mathbf{Pr}\left(Y_{i+1} \neq 0\right) \leq \frac{|\Pi_i^\tau \cup \Pi_i^{\tau^{-1}}|}{n-i} \leq \frac{2i}{n-i}.$$

Now we apply Corollary 5.2 to the random variables Y_i with $\mu_i = \frac{2(i-1)}{n-(i-1)}$, $b = 2$ and

$$\mu = \sum_{i=1}^{t+1} \mu_i \le \frac{2}{n-t} \sum_{i=0}^{t} i = \frac{t(t+1)}{n-t} \le \frac{3t^2}{n},$$

obtaining, for every $s > 0$,

$$\mathbf{Pr}\left(Z_{t+1} \ge s\right) \le \left(\frac{6\mathrm{e}t^2}{sn}\right)^{s/2}. \quad \square$$

We are now ready to prove the main result of this section.

Proof of Lemma 8.1. Let $T = \operatorname{supp}(\tau)$, let $E = \operatorname{traj}(v, \pi, t)$, and let S denote the set of vertices of T hit by E.

By Observation 7.10, the number of components of $\operatorname{shear}_\tau(E)$ is at least that of $\operatorname{shear}_\tau(E_T)$. Since all the tails of $\operatorname{shear}_\tau(E_T)$ are in S and all the heads of $\operatorname{shear}_\tau(E_T)$ are in S^τ, it follows that the number of components of $\operatorname{shear}_\tau(E_T)$ is at least $|S| - 2\,|S \cap S^\tau|$. Hence, if p denotes the number of components of $\operatorname{shear}_\tau(E)$, then

$$\mathbf{Pr}\left(p \le \frac{\alpha t}{6} \,\middle|\, |E| = t\right) \le \mathbf{Pr}\left(|S| \notin \left[\frac{\alpha t}{2}, 3\alpha t\right] \,\middle|\, |E| = t\right) + \mathbf{Pr}\left(|S \cap S^\tau| > \frac{|S|}{3} \,\middle|\, |S| \in \left[\frac{\alpha t}{2}, 3\alpha t\right], |E| = t\right).$$

By Lemma 5.3,

$$\mathbf{Pr}\left(|S| \notin \left[\frac{\alpha t}{2}, 3\alpha t\right] \,\middle|\, |E| = t\right) \le 2\mathrm{e}^{-\alpha t/8}.$$

By Observation 7.8, if we condition on the value of $|S|$, and on the vertex $w \in S$ that occurs first along trajectory E starting from v, then S is a uniformly random subset of T of size $|S|$ containing w. Now, since τ acts on T without fixed points, we can apply Lemma 8.2 to S, obtaining

$$\mathbf{Pr}\left(|S \cap S^\tau| \ge \frac{|S|}{3} \,\middle|\, |S| \in \left[\frac{\alpha t}{2}, 3\alpha t\right]\right) \le \max_s \left(\frac{18\mathrm{e}s}{\alpha n}\right)^{s/6} \le \left(\frac{54\mathrm{e}t}{n}\right)^{\alpha t/12},$$

where the maximum is over $s \in \left[\frac{\alpha t}{2}, 3\alpha t\right]$. $\square$

9 Shears rarely have large components

Our main result in this section is in some sense dual to the main result of the previous section. As before, suppose $\tau \in \operatorname{Sym}(V)$ has support of size αn. For any $v \in V$, we would expect the trajectory of v under a random permutation π, unless a very short cycle, to go at most $O(1/\alpha)$ steps before hitting the support of τ. This suggests the main result of this section (Lemma 9.2) which says that, even if ℓ is quite large, the operation of shearing by τ splits trajectories of length ℓ into only small components (as long as $n - \ell$ is sufficiently bigger than n/α). The full strength of this result will be needed in the proof of Theorem 1.3 (Section 10).

To prove this result, we approach it from another perspective, which strengthens the connection to the previous section. If shearing by τ^{-1} cuts even fairly short cycles of $\pi = \sigma\tau$ into enough pieces, then they are unlikely to all fall into a single ℓ-step trajectory of σ.

LEMMA 9.1. *Let $\tau \in \operatorname{Sym}(V)$, $t, \ell > 0$ and $v, w \in V$. Sample $\sigma \in \operatorname{Sym}(V)$ uniformly at random, and let $\pi = \sigma\tau$. Let $C = \operatorname{traj}(v, \pi, t)$. Let p denote the number of components of $\operatorname{shear}_{\tau^{-1}}(C)$. Then*

$$\mathbf{Pr}\left(\operatorname{shear}_{\tau^{-1}}(C) \subseteq \operatorname{traj}(w, \sigma, \ell) \mid C\right) \le \left(\frac{\ell}{n-t-1}\right)^{p-1}.$$

Proof. Let $\rho = \operatorname{shear}_{\tau^{-1}}(C)$.

If ρ is a single cycle, then $p = 1$ and the result is trivial. If ρ properly contains a cycle, then since $\operatorname{traj}(w, \sigma, \ell)$ is a subset of a single cycle, the probability is zero.

Finally, if ρ contains no cycle, then it consists of p paths of length ≥ 1, with total length t. Let $V(\rho)$ denote the set of points hit by ρ, so that $V(\rho)/\rho$ is a set of p points in V/ρ. Now, if $\rho \subseteq \operatorname{traj}(w, \sigma, \ell)$, then

$$V(\rho)/\rho \subseteq V(\operatorname{traj}(w/\rho, \sigma/\rho, \ell)). \tag{9.8}$$

But by Observation 7.13, the conditional distribution of σ/ρ, given ρ, is uniform over $\operatorname{Sym}(V/\rho)$. Even if we assume pessimistically that $w/\rho \in V(\rho)/\rho$, and that $|\operatorname{traj}(w/\rho, \sigma/\rho, \ell)| = \ell$, the probability of the event in (9.8) is at most $\frac{\binom{\ell}{p-1}}{\binom{n-t-1}{p-1}} \le \left(\frac{\ell}{n-t-1}\right)^{p-1}$. $\square$

Our next Lemma says that, as long as $\operatorname{supp}(\tau)$ is not too small, revealing a (not too long) trajectory of σ probably reveals only very short paths in $\pi = \sigma\tau$.

LEMMA 9.2. *Let $\tau \in \operatorname{Sym}(V)$ and set $\alpha = \deg(\tau)/n$. Let $\ell \in \{1, \ldots, n\}$, and let $w \in V$. Sample $\sigma \in \operatorname{Sym}(V)$ uniformly at random, and let $\pi = \sigma\tau$. Let X be the number of edges in the largest component of $\operatorname{shear}_\tau(\operatorname{traj}(w, \sigma, \ell))$. Then, for all $t \ge 0$,*

$$\mathbf{Pr}(X \ge t) \le n\left(\left(\frac{\ell}{n-t-1}\right)^{\alpha t/4-1} + 2\mathrm{e}^{-\alpha t/8} + \left(\frac{54\mathrm{e}t}{n}\right)^{\alpha t/12}\right).$$

Proof. For every $v \in V$, define events $A_v = \{\mathrm{traj}(v, \pi, t) \subseteq \mathrm{shear}_\tau(\mathrm{traj}(w, \sigma, \ell))\}$ and $B_v = \{|\mathrm{traj}(v, \pi, t)| = t\}$. Note that $X \geq t$ if and only if there exists v such that both A_v and B_v occur.

Hence, by the union bound,

$$\mathbf{Pr}\left(X \geq t\right) \leq \sum_v \mathbf{Pr}\left(A_v \wedge B_v\right),$$

and so it suffices to prove, for every $v \in V$, that

$$\mathbf{Pr}(A_v \wedge B_v) \leq \left(\frac{\ell}{n-t-1}\right)^{\alpha t/4-1} + 2\mathrm{e}^{-\alpha t/8} + \left(\frac{12\mathrm{e}t}{n}\right)^{\alpha t/32}.$$

Let $v \in V$, let p_v denote the number of connected components of $\mathrm{shear}_{\tau^{-1}}(\mathrm{traj}(v, \pi, t))$, and define the event $C_v = \{p_v \leq \alpha t/4\}$. Then

$$A_v \wedge B_v \subseteq (A_v \wedge \neg C_v) \cup (B_v \wedge C_v),$$

and hence

$$\mathbf{Pr}\left(A_v \wedge B_v\right) \leq \mathbf{Pr}\left(A_v \mid \neg C_v\right) + \mathbf{Pr}\left(C_v \mid B_v\right).$$

Now Lemma 8.1 says exactly that $\mathbf{Pr}\left(C_v \mid B_v\right) \leq 2\mathrm{e}^{-\alpha t/8} + \left(\frac{54\mathrm{e}t}{n}\right)^{\alpha t/12}$, while Lemma 9.1 implies that

$$\begin{aligned}\mathbf{Pr}(A_v \mid \neg C_v) &\leq \mathbf{E}\left(\left(\frac{\ell}{n-t-1}\right)^{p_v-1} \middle|\, p_v > \alpha t/4\right) \\ &\leq \left(\frac{\ell}{n-t-1}\right)^{\alpha t/4-1},\end{aligned}$$

which completes the proof. □

10 Proof of Theorem 1.3

OBSERVATION 10.1. *Let $v \in V$, $\tau \in \mathrm{Sym}(V)$, $\ell \geq 1$, and let $\alpha = \deg(\tau)/n$. Choose $\sigma \in \mathrm{Sym}(V)$ uniformly at random. Let $\rho = \mathrm{shear}_\tau(\mathrm{traj}(v, \sigma, \ell))$. Then, for every $t \in \{1, \ldots, n\}$, the probability that the point v survives contraction by ρ is at least*

$$1 - \frac{t}{n} - n\left(\left(\frac{\ell}{n-t-1}\right)^{\alpha t/4-1} + 2\mathrm{e}^{-\alpha t/8} + \left(\frac{54\mathrm{e}t}{n}\right)^{\alpha t/12}\right).$$

Proof. If v is killed under contraction by ρ, then either v is in a cycle of $\pi = \sigma\tau$ of length $< t$, which has probability $(t-1)/n$, or ρ has a component of size $\geq t$, which has probability bounded in Lemma 9.2. □

Proof of Theorem 1.3. The conclusion of the theorem is equivalent to this: for every $1 \leq k, \ell \leq n$,

(10.9)
$$\left|\mathbf{Pr}\left(X \leq k \text{ and } Y > \ell\right) - \frac{k(n-\ell)}{n^2}\right| \leq n^{-1/8+o(1)}.$$

Since $\mathbf{Pr}\left(Y > n - n^{7/8-o(1)}\right) \leq n^{-1/8+o(1)}$, it suffices to prove (10.9) in the case $\ell < n - n^{7/8-o(1)}$. We will prove the stronger claim that, for $1 \leq k \leq n$, $1 \leq \ell \leq n - n^{7/8}\log^2 n$,

$$\left|\mathbf{Pr}\left(X \leq k \mid Y > \ell\right) - \frac{k}{n}\right| \leq n^{-1/8+o(1)}.$$

Henceforth, let $\ell \leq n - n^{7/8}\log^2 n$ be fixed.

Let v be the first vertex in V, so $Y = |c(\sigma, v)|$ and $X = |c(\pi, v)|$. Consider the partial permutation $\rho := \mathrm{shear}_\tau(\mathrm{traj}(v, \sigma, \ell)) \subseteq \pi$. Since $\pi = \sigma\tau$ is uniformly distributed *a priori*, it follows that the conditional distribution of π, given $\mathrm{traj}(v, \sigma, \ell)$, is uniform over the set of permutations containing ρ. We will analyze the conditional distribution of X, given ρ, by studying the contraction of π by ρ.

Let C denote the cycle of π containing v. We say ρ is *bad* if at least one of the following holds:

(a) v is in a cycle of ρ, i. e., $C \subseteq \rho$,
(b) ρ has a component of size $\geq n^{3/8}$, or
(c) $\geq \sqrt{n}$ edges of ρ are contained in cycles.

By Observation 10.1, Lemma 9.2, and Proposition 4.8, the probability that ρ is bad is $\leq n^{-1/8+o(1)}$, even conditioned on the event $Y \geq \ell$.

From now on, fix ρ, and assume ρ is good. When we talk about probabilities below, we will always mean conditional probabilities, conditioned on this value of ρ.

Label each point $w \in V/\rho$ by the number p_w of points in the maximal path of ρ which contracts to it. Now, by Observation 7.13, since π is uniformly distributed over extensions of ρ, π/ρ is uniformly distributed over $\mathrm{Sym}(V/\rho)$. Since ρ is good, the contraction C/ρ is non-empty. This implies the length of C/ρ is uniformly distributed over $1, \ldots, n-\ell$, and that the set of vertices hit by C/ρ consists of v/ρ, together with a random $(|C/\rho|-1)$-subset S of the remaining $n-\ell-1$ vertices. If Z denotes the sum of the labels on S, then $X = Z + p_v$.

Condition further on the values of all the labels, so that the only randomness remaining is the set S. Since ρ is good, these $n-\ell$ labels are all in $\{1, \ldots, n^{3/8}\}$ and sum to $n - c = n - O(\sqrt{n})$, where c is the number of edges in cycles of ρ. Normalizing these labels to $[0, 1]$, and applying Theorem 5.6, we find, for every $z \in [0, n-c]$,

$$\left|\mathbf{Pr}\left(Z \leq z\right) - \frac{z}{n-c}\right| \leq \frac{4\sqrt{n\log n}}{n^{5/8}} \leq n^{-1/8+o(1)}.$$

Since $\mathbf{Pr}\left(X \leq x\right) = \mathbf{Pr}\left(Z \leq x - p_v\right)$ and since $(x - p_v)/(n-c) = x/n + O(n^{-1/2})$ the result follows by the triangle inequality. □

11 Proof of Theorem 1.2

Proof. Given a pair of independent random permutations $\tau, \sigma \in S_n$, we need to show that the lengths of the first cycles of $\sigma\tau^i$ and $\sigma\tau^j$ are nearly independent for $1 \le i < j < s$ where $s = n^{(7/32-\varepsilon)\ln n}$.

Let $\sigma_1 = \sigma\tau^i$. Note that σ_1 is uniformly distributed over S_n. Let $k = j - i$. Then $\sigma\tau^j = \sigma_1\tau^k$. According to Theorem 1.3, we are done if we can show that a.a., for all $k \le s$, $\deg(\tau^k) \ge n^{3/4}$. But this is exactly the assertion of Theorem 4.4 for $p = 3/4$.

We also need to show that the lengths of the first cycles of τ and $\sigma\tau^i$ are nearly independent. In fact, these two permutations are independent. □

12 Open Problems

Theorem 1.3 introduces a new type of problem in the statistical theory of permutations: the near-independence of parameters of the permutations σ and $\tau\sigma$ where τ is given and satisfies a lower bound on its degree (such as $\deg(\tau) = \Omega(n)$) and σ is selected uniformly at random.

Theorem 1.3 shows that under these conditions, the lengths of the first cycle of σ and of $\tau\sigma$ are nearly independent. Does the same hold for the second, third, etc. cycles; for the number of cycles; for the logarithm of the order? We remark that the number of cycles of a random permutation, as well as the logarithm of its order, are asymptotically normally distributed ([15, 12]).

We believe that all these parameters will be nearly independent for σ and $\tau\sigma$. In fact we expect that for any constant k, if $\tau_1, \ldots, \tau_k$ are given permutations such that the degree of each quotient $\tau_i^{-1}\tau_j$ $(i \ne j)$ is $\Omega(n)$ and σ is random then the lengths of the first cycle of $\tau_i\sigma$ $(i = 1, \ldots, k)$ are nearly independent; and the same holds for the other parameters mentioned.

References

[1] N. Alon, J. Spencer: *The Probabilistic Method.* Wiley-Interscience, 1992.

[2] L. Babai: The probability of generating the symmetric group. *J. Comb. Th.–A* **52** (1989), 148–153.

[3] L. Babai, R. Beals, Á. Seress: On the diameter of the symmetric group: polynomial bounds. *15th ACM-SIAM SODA* (2004), 1101–1105.

[4] L. Babai, G. L. Hetyei: On the diameter of random Cayley graphs of the symmetric group. *Combinatorics, Probability, and Computing* **1** (1992), 201–208.

[5] L. Babai, G. Hetyei, W.M. Kantor, A. Lubotsky, Á. Seress: On the diameter of finite groups. *31st IEEE FOCS,* 1990, pp. 857–865.

[6] L. Babai, Á. Seress: On the Degree of Transitivity of Permutation Groups: A Short Proof. *J. Combinatorial Theory–A* **45** (1987), 310–314.

[7] L. Babai, Á. Seress: On the diameter of Cayley graphs of the symmetric group. *J. Combinatorial Theory–A* **49** (1988), 175–179.

[8] J. D. Bovey: The probability that some power of a permutation has small degree. *Bull. London Math. Soc.* **12** (1980), 47–51.

[9] J. D. Dixon: The probability of generating the symmetric group. *Math. Z.* **110** (1969), 199–205.

[10] J. R. Driscoll, M. L. Furst: Computing short generator sequences. *Info. & Comp.* **72** (1987), 117–132.

[11] P. Erdős, P. Turán: On some problems of a statistical group theory I., *Z. Wahrscheinlichkeitstheorie verw. Geb.* **4** (1965), 175–186.

[12] P. Erdős and P. Turán: On some problems of a statistical group theory III, *Acta Math. Acad. Sci. Hung.* **18** (1967), 309–320.

[13] P. Erdős and P. Turán: On some problems of a statistical group theory VII, *Period. Math. Hung.* **2** (1972), 149–163.

[14] S. Even, O. Goldreich: The minimum length generator sequence is NP-hard. *J. Algor.* **2** (1981), 311–313.

[15] V. L. Goncharov: On the Field of Combinatory Analysis, *Isvestija Akad. Nauk. SSSR*, Ser. mat. **8** (1944), 3-48 (Russian. English translation: *Translations of the AMS*, Ser. 2, **19** (1962), 1-46.)

[16] M-C. Heydemann: Cayley graphs and interconnection networks. In: *Graph Symmetry (Montreal, 1996),* NATO Adv. Sci. Inst. Ser. C, Math. Phys. Sci. 497, Kluwer Acad. Publ., Dordrecht 1997, pp. 167–224.

[17] W. Hoeffding: Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58** (1963), 13–30.

[18] M. R. Jerrum: The complexity of finding minimum length generator sequences. *Theoretical Computer Science* **36** (1985), 265–289.

[19] R. E. Korf: Finding optimal solutions to Rubik's Cube using pattern databases. In: *Proc. 14th Nat. Conf. on Artificial Intelligence (AAAI-97)*, Amer. Assoc. for Artificial Intelligence, 1997, pp. 700-705.

[20] L. Lovász: *Combinatorial Problems and Exercises.* Akadémiai Kiadó, Budapest, and North-Holland, Amsterdam, 1979.

[21] P. McKenzie: Permutations of bounded degree generate groups of polynomial diameter. *Info. Proc. Lett.* **19** (1984), 253–254.

[22] R. Motwani, P. Raghavan: *Randomized Algorithms.* Cambridge University Press, Cambridge, 1995.

[23] Á. Seress: *Permutation Group Algorithms.* Cambridge Univ. Press, 2003.