

On the LSH Distortion of Ulam and Cayley Similarities

Flavio Chierichetti* Mirko Giacchini† Ravi Kumar‡ Erasmo Tani†

Abstract

Locality-sensitive hashing (LSH) has found widespread use as a fundamental primitive, particularly to accelerate nearest neighbor search. An LSH scheme for a similarity function $S : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ is a distribution over hash functions on \mathcal{X} with the property that the probability of collision of any two elements $x, y \in \mathcal{X}$ is exactly equal to $S(x, y)$. However, not all similarity functions admit exact LSH schemes. The notion of LSH distortion measures how multiplicatively close a similarity function is to having an LSH scheme.

In this work, we study the LSH distortion of the Ulam and Cayley similarities, which are popular similarity measures on permutations of n elements. We show that the Ulam similarity admits a sublinear LSH distortion of $O(n/\sqrt{\log n})$; we also prove a lower bound of $\Omega(n^{0.12})$ on the best LSH distortion achievable. On the other hand, we show that the LSH distortion of the Cayley similarity is $\Theta(n)$.

*Reddit, San Francisco. Work done in part while at Sapienza University of Rome.
flavio.chierichetti@reddit.com

†Sapienza University of Rome. {giacchini, tani}@di.uniroma1.it

‡Google, Mountain View. ravi.k53@gmail.com

1 Introduction

Locality-sensitive hashing (LSH) is an important tool for processing high dimensional data (Broder, 1997; Broder et al., 2000; Indyk et al., 1997; Indyk and Motwani, 1998). Classic examples of LSH include minHash for the Jaccard similarity (Broder, 1997; Broder et al., 2000), simHash for the cosine similarity (Charikar, 2002), and sampling hash for the Hamming similarity (Indyk and Motwani, 1998). LSH has been extensively used for solving (approximate) nearest-neighbor search (Andoni and Indyk, 2008) and similarity search (Gionis et al., 1999; Lv et al., 2007; Satuluri and Parthasarathy, 2012); see <https://www.mit.edu/~andoni/LSH/> for more pointers on classical results.

LSH is a way to represent a similarity S —which in the context of this paper is a bi-variate, reflexive, symmetric function—by a distribution over partitions. These partitions correspond to the pre-images of hash functions, so that two entities are considered similar by a hash function if they both hash to the same value. In other words, an LSH family \mathcal{H} induces a *collision kernel* of the form $S(x, y) = \Pr_{h \sim \mathcal{H}}[h(x) = h(y)]$, for entities x, y in the universe.

Such collision kernels have to satisfy two constraints: they have unit diagonals (capturing reflexivity) and are a convex combination of equivalence-relation matrices (capturing hashing to the same value). Hence, it turns out that many natural and useful similarity measures do not admit an LSH. An interesting line of work studies which similarities have an LSH. Charikar (2002) identified a necessary condition for a similarity to admit an LSH: the corresponding distance function (which is often one minus the similarity) must be a metric and must be isometrically embeddable into ℓ_1 .

To circumvent this, Chierichetti et al. (2019) introduced the notion of LSH distortion, which answers the question: how much is a non-LSHable similarity S multiplicatively close to an LSHable similarity?

Formally, given a similarity S that does not admit an LSH, is there another similarity S' that has an LSH and satisfies:

$$\forall x, y : \quad \frac{1}{\Delta} \cdot S(x, y) \leq S'(x, y) \leq S(x, y) \quad (1)$$

We call the optimal Δ in (1) the *LSH distortion* of S .

In their paper, Chierichetti et al. showed tight LSH distortion bounds for several set similarities. Note that, since we are considering multiplicative approximations over **similarities**, and we wish to achieve them using collision kernels, the LSH distortion of a similarity can differ significantly from its metric counterpart, and known results for metric distortion do not transfer to this setting.

Our contributions. In this work, we study the LSH distortion of two natural similarities derived from popular metrics on permutations. The first is the normalized Ulam similarity, which measures the length of the longest common subsequence between two permutations. The second is the Cayley similarity, which is the normalized number of cycles in the relative permutation (its distance counterpart is number of transpositions needed to turn one permutation into another).

Our main results are the following:

- We show that the Ulam similarity has sublinear LSH distortion. In particular, we show that it can be approximated by a similarity that has an LSH, with distortion $O(n/\sqrt{\log n})$ (Theorem 5),
- We also show a lower bound of $\Omega(n^{0.12})$ on the LSH distortion of the Ulam similarity (Theorem 10),

- In contrast, we show that the Cayley similarity has an $\Omega(n)$ lower bound on its LSH distortion (Theorem 19), matching a simple $O(n)$ upper bound (Theorem 18).

Techniques. The proof of the upper bound for the Ulam similarity is based on computing a suitable measure of correlation between the given permutation and a random permutation. The collision probability of our scheme depends on how well the input permutations are correlated with a random one. Showing that the collision probability is upper bounded by $\text{LCS}(\cdot, \cdot)/n$ is relatively easy. On the other hand, showing it is also lower bounded by the LCS (up to sublinear distortion) is more challenging.

For the lower bounds on the distortion, we crucially utilize the PSDness of the collision kernel (Lemma 4). The lower bound for the Ulam similarity is based on first constructing a small set of permutations on an 8-element universe that have an LSH distortion of at least $9/7$, and then amplifying its cardinality to work for a universe of arbitrarily large size, by using the wreath product on permutations.

The lower bound for the Cayley similarity is based on representation theory, and uses both the fact that every bi-invariant PSD kernel on \mathbb{S}_n can be written as a convex combination of normalized characters, as well as Roichman’s character-ratio estimates applied to derangements. The distortion lower bound is obtained by constructing two derangements that share the high-order terms of this decomposition and hence must have close collision probabilities with the identity permutation, but have vastly different Cayley similarities (linear vs constant) to it.

Related work. For permutation metrics, the Ulam metric has been extensively studied through its connection to edit distance and longest common subsequences. Charikar and Krauthgamer (2006) proved that the Ulam metric embeds into ℓ_1 with $O(\log n)$ distortion; Andoni and Krauthgamer (2010) later proved that embeddings of the Ulam metric into ℓ_1 requires a near-tight distortion of $\Omega(\log n / \log \log n)$. Andoni et al. (2009) bypassed the ℓ_1 barrier by embedding Ulam distance with constant distortion into an iterated product metric. Charikar et al. (2018) obtained an $O(\log n)$ distortion randomized embedding of Ulam into Hamming space that also carries alignment information. Note that all these previous results are for the Ulam metric; to the best of our knowledge, the corresponding question of Ulam similarity has not been considered in the literature.

Unlike Ulam, the Cayley metric is not hard to embed into ℓ_1 . Our work can be viewed as a hardness result for LSHability of the Cayley similarity.

Outline. We begin, in Section 2, by introducing some basic notation. In Section 3, we review the definition of LSH distortion and introduce in detail the central problems of the paper. In Section 4, we provide our $O(n/\sqrt{\log n})$ upper bound on the LSH distortion for the Ulam similarity. We then give a $\Omega(n^{0.12})$ lower bound in Section 5. In Section 6, we give a simple $O(n)$ upper bound on the distortion of the Cayley similarity, that we complement with a matching lower bound in Section 7.

In Appendix A we also review some background on the representation theory of \mathbb{S}_n , which we rely on for our proofs in Section 7.

2 Notation

Permutations. For a positive integer n , let $[n] := \{1, \dots, n\}$. Let \mathbb{S}_n be the set of all permutations of $[n]$. We denote by id the identity permutation. For a permutation $\pi \in \mathbb{S}_n$ we will sometimes use

the one-line notation and sometimes the cycle decomposition.¹ For a permutation $\pi \in \mathbb{S}_n$ we denote with $c(\pi)$ the number of cycles in its cycle decomposition. Note that $c(\pi) = c(\pi^{-1}) = c(\sigma\pi\sigma^{-1})$ for any $\sigma \in \mathbb{S}_n$. For $\pi \in \mathbb{S}_n$, $a, b \in [n]$ we write $a <_\pi b$ to indicate that a comes before b in the one-line representation of π (alternatively that $\pi^{-1}(a) < \pi^{-1}(b)$). We write $a \leq_\pi b$ to indicate that either $a = b$ or $a <_\pi b$. For $\pi, \sigma \in \mathbb{S}_n$ we denote with $\text{LCS}(\pi, \sigma)$ the length of the longest common subsequence between the one-line representations of π and σ . For $\pi \in \mathbb{S}_n$, we denote with $\text{fix}(\pi)$ the number of fixed points of π and with $\text{cycle}_2(\pi)$ the number of cycles of length 2 in the cycle decomposition of π . We also let $\text{supp}(\pi) = n - \text{fix}(\pi)$.

Partitions and Compositions. Recall that, for any positive integer n , a partition $\lambda = (\lambda_1, \dots, \lambda_r)$ of n is a finite sequence of non-increasing positive integers that sum to n . We write $\lambda \vdash n$ to indicate that λ is a partition of n . We denote by (1^n) the partition $(1, \dots, 1)$ containing n 1s. Given a partition $\lambda = (\lambda_1, \dots, \lambda_r) \vdash n$ its *transpose partition* (sometimes also called the *conjugate partition*) is the partition $\lambda' \vdash n$ defined by $\lambda' = (\lambda'_1, \dots, \lambda'_{\lambda_1})$, where:

$$\forall i \in [\lambda_1] : \lambda'_i = \max\{k \mid \lambda_k \geq i\}.$$

A *composition* of n is an ordered partition of n , i.e. an integer vector $\mu = (\mu_1, \dots, \mu_\ell)$ such that $\mu_i \geq 0$ for every $i \in [\ell]$ and $\sum_{i \in [\ell]} \mu_i = n$.

Matrices. A symmetric matrix $M \in \mathbb{R}^{n \times n}$ is positive semidefinite (PSD, write $M \succeq 0$) if for all $x \in \mathbb{R}^n$, we have:

$$x^T M x \geq 0.$$

The trace operator is the function that takes in a square matrix $M \in \mathbb{R}^{n \times n}$ and returns $\text{tr}(M) = \sum_{i \in [n]} M(i, i)$. Given two square matrices X, Y , we denote by $\text{diag}(X, Y)$ the block matrix:

$$\text{diag}(X, Y) = \begin{pmatrix} X & \mathbf{0} \\ \mathbf{0} & Y \end{pmatrix}$$

where $\mathbf{0}$ is the all-zeros matrix.

3 Similarities, LSH, and Distortion

Definition 1 (Similarity). Given a set \mathcal{X} , a *similarity* (or *similarity function*) on \mathcal{X} is a function $S : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$, satisfying:

- (a) **Reflexivity:** $S(x, x) = 1$ for all $x \in \mathcal{X}$,
- (b) **Symmetry:** $S(x, y) = S(y, x)$ for all $x, y \in \mathcal{X}$.

A *similarity space* is simply a pair (\mathcal{X}, S) , where \mathcal{X} is a set, and S is a similarity function on \mathcal{X} .

Given a similarity space, it is natural to ask whether one can find a small representation for elements of \mathcal{X} that preserve the values of $S(\cdot, \cdot)$. A natural approach to obtaining this is locality-sensitive hashing, which aims to construct randomized embeddings of \mathcal{X} into discrete (0-1) similarity spaces that preserve $S(\cdot, \cdot)$ in expectation. We recall the key definition here.

¹For example, for the permutation $\pi : [5] \rightarrow [5]$ such that $\pi(1) = 3$, $\pi(2) = 5$, $\pi(3) = 1$, $\pi(4) = 2$, $\pi(5) = 4$; the one-line representation is $(3, 5, 1, 2, 4)$ and the cycle decomposition is $(1\ 3)(2\ 5\ 4)$.

Definition 2 (Locality-Sensitive Hashing (LSH)). A locality-sensitive hashing (LSH) scheme for a similarity space (\mathcal{X}, S) is a distribution \mathcal{H} over functions from \mathcal{X} into some finite set \mathcal{Y} with the property that:

$$\Pr_{h \sim \mathcal{H}} [h(x) = h(y)] = S(x, y),$$

for all $x, y \in \mathcal{X}$.

We say that a similarity function S is *LSHable* if it admits an LSH scheme.

Chierichetti et al. (2019) give a notion of distortion for similarity functions, which quantifies the degree to which non-LSHable similarities are far from being LSHable. We recall the key definitions.

Definition 3 (Chierichetti et al. (2019)). We say a distribution \mathcal{H} over functions from \mathcal{X} to some finite set \mathcal{Y} Δ -approximates the space (\mathcal{X}, S) , if:

$$\frac{1}{\Delta} \cdot S(x, y) \leq \Pr_{h \sim \mathcal{H}} [h(x) = h(y)] \leq S(x, y),$$

for all $x, y \in \mathcal{X}$. The LSH *distortion* (or simply distortion) of a similarity space (\mathcal{X}, S) is the smallest $\Delta \geq 1$ such that there exists a distribution \mathcal{H} that Δ -approximates (\mathcal{X}, S) .

A simple argument shows that the distortion of a similarity space (\mathcal{X}, S) could be characterized as the smallest $\Delta \geq 1$ such that there exists an LSHable similarity function $S' : \mathcal{X} \times \mathcal{X} \rightarrow [0, 1]$ with the property that:

$$\frac{1}{\Delta} \cdot S(x, y) \leq S'(x, y) \leq S(x, y), \tag{2}$$

for all $x, y \in \mathcal{X}$.

The focus of this work is the distortion of two similarity spaces in which \mathcal{X} is the symmetric group \mathbb{S}_n : the Ulam similarity and the Cayley similarity.

Ulam Similarity. The *Ulam distance* between two permutations $\pi, \sigma \in \mathbb{S}_n$ is defined as

$$d^U(\pi, \sigma) := n - \text{LCS}(\pi, \sigma),$$

We define the (normalized) *Ulam similarity* in a natural way:

$$S^U(\pi, \sigma) = 1 - \frac{n - \text{LCS}(\pi, \sigma)}{n} = \frac{\text{LCS}(\pi, \sigma)}{n}.$$

Cayley Similarity. The *Cayley distance* between two permutations $\pi, \sigma \in \mathbb{S}_n$ is defined as the minimum number of (not necessarily adjacent) transpositions required to transform one permutation into the other, and it is also equivalent to $d^C(\pi, \sigma) := n - c(\pi^{-1}\sigma) = n - c(\pi\sigma^{-1})$ (see, e.g., (Diaconis, 1988, page 118)). We define a notion of (normalized) *Cayley similarity* in a natural way:

$$S^C(\pi, \sigma) = 1 - \frac{d^C(\pi, \sigma)}{n} = \frac{c(\pi^{-1}\sigma)}{n}.$$

Positive Semi-Definiteness. In order to prove lower bounds on the achievable distortion, we will make use of the following result. Let \mathcal{H} be a probability distribution over functions from a finite domain \mathcal{X} into a finite range \mathcal{Y} . Fixing any ordering of the elements of \mathcal{X} , we define matrix $P_{\mathcal{H}} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$ such that:

$$P_{\mathcal{H}}(x_1, x_2) := \Pr_{h \sim \mathcal{H}} [h(x_1) = h(x_2)],$$

for $x_1, x_2 \in \mathcal{X}$.

We now show that this matrix is always PSD.

Lemma 4. *For any distribution \mathcal{H} over functions from \mathcal{X} to \mathcal{Y} , where \mathcal{X} and \mathcal{Y} are finite sets, the matrix $P_{\mathcal{H}}$ is PSD. That is, for all $\{w_x\}_{x \in \mathcal{X}} \in \mathbb{R}^{|\mathcal{X}|}$, we have:*

$$\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} w_{x_1} w_{x_2} P_{\mathcal{H}}(x_1, x_2) \geq 0.$$

Proof. Note that for any $x_1, x_2 \in \mathcal{X}$,

$$P(x_1, x_2) = \Pr_{h \sim \mathcal{H}} [h(x_1) = h(x_2)] = \mathbb{E}_{h \sim \mathcal{H}} [\mathbf{1}_{[h(x_1)=h(x_2)]}],$$

where $\mathbf{1}_{[\cdot]}$ is the indicator function, and hence:

$$\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} w_{x_1} w_{x_2} P(x_1, x_2) = \sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} w_{x_1} w_{x_2} \mathbb{E}_{h \sim \mathcal{H}} [\mathbf{1}_{[h(x_1)=h(x_2)]}] = \mathbb{E}_{h \sim \mathcal{H}} \left[\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} w_{x_1} w_{x_2} \mathbf{1}_{[h(x_1)=h(x_2)]} \right].$$

Now, for a fixed $h : \mathcal{X} \rightarrow \mathcal{Y}$, we show that the inner sum is non-negative, which in turns completes the proof. We have:

$$\sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} w_{x_1} w_{x_2} \mathbf{1}_{[h(x_1)=h(x_2)]} = \sum_{y \in \mathcal{Y}} \sum_{\substack{x_1 \in \mathcal{X} \\ h(x_1)=y}} \sum_{\substack{x_2 \in \mathcal{X} \\ h(x_2)=y}} w_{x_1} w_{x_2} = \sum_{y \in \mathcal{Y}} \left(\sum_{\substack{x \in \mathcal{X} \\ h(x)=y}} w_x \right)^2 \geq 0. \quad \square$$

4 Upper Bound for Ulam Similarity

In this section, we prove the following result:

Theorem 5. *The LSH distortion of $(\mathbb{S}_n, \mathbb{S}^U)$ is $O(n/\sqrt{\log n})$.*

For two permutations $\tau, \pi \in \mathbb{S}_n$ and $z \in [n]$, define:

$$\text{Rec}_{\tau, z}(\pi) := \{a \in [n] \mid z \leq_{\pi} a \text{ and } b <_{\tau} a \text{ for all } b \text{ such that } z \leq_{\pi} b <_{\pi} a\}.$$

Intuitively, $\text{Rec}_{\tau, z}(\pi)$ is the set of new record maxima according to τ obtained while scanning the one-line representation of π starting from z .

For $\tau, \pi \in \mathbb{S}_n$, $a, z \in [n]$ define:

$$h_{\tau, z, a}(\pi) := \begin{cases} 1 & \text{if } a \in \text{Rec}_{\tau, z}(\pi), \\ \pi & \text{otherwise.} \end{cases}$$

Let \mathcal{H} be the uniform distribution over the set of functions: $\{h_{\tau, z, a} \mid \tau \in \mathbb{S}_n, z \in [n], a \in [n]\}$. Recall that for $\pi, \sigma \in \mathbb{S}_n$:

$$P_{\mathcal{H}}(\pi, \sigma) := \Pr_{h_{\tau, z, a} \sim \mathcal{H}} [h_{\tau, z, a}(\pi) = h_{\tau, z, a}(\sigma)].$$

We will show that: $O(\sqrt{\log n}/n) \cdot \mathbb{S}^U(\pi, \sigma) \leq P_{\mathcal{H}}(\pi, \sigma) \leq \mathbb{S}^U(\pi, \sigma)$.

Lemma 6. For $\pi, \sigma \in \mathbb{S}_n$, $\pi \neq \sigma$, it holds that:

$$P_{\mathcal{H}}(\pi, \sigma) = \frac{1}{n} \cdot \mathbb{E}_{\tau, z} [|\text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)|].$$

Proof. Note that for $\pi \neq \sigma$, we have that $h_{\tau, z, a}(\pi) = h_{\tau, z, a}(\sigma)$ if and only if $a \in \text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)$. Conditioned on the choice of τ and z , there are n possible values for a and only $|\text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)|$ of them make the hashes of π and σ equal. This concludes the proof. \square

Lemma 7. For any $\tau, \pi, \sigma \in \mathbb{S}_n$ and $z \in [n]$, it holds that:

$$1 \leq |\text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)| \leq \text{LCS}(\pi, \sigma).$$

Proof. Since $z \in \text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)$, the set $\text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)$ is always non-empty, proving the first inequality.

We now show the second inequality. If there is only one element in the intersection, the result is trivially true, since $\text{LCS}(\pi, \sigma) \geq 1$ for any $\pi, \sigma \in \mathbb{S}_n$. Otherwise, consider any distinct $x, y \in \text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)$. Suppose without loss of generality that $x <_{\pi} y$. By definition of $\text{Rec}_{\tau, z}(\pi)$, since y is observed after x in the one-line representation of π , it must hold that $x <_{\tau} y$. Moreover, we have $x <_{\sigma} y$, for otherwise we would have $x >_{\tau} y$, leading to a contradiction.

Therefore, if we let $\text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma) = r_1 <_{\pi} \dots <_{\pi} r_t$, we also have $r_1 <_{\sigma} \dots <_{\sigma} r_t$, meaning that r_1, \dots, r_t is a common subsequence of π and σ , and therefore $t \leq \text{LCS}(\pi, \sigma)$. \square

Fixed π, σ such that $\text{LCS}(\pi, \sigma) \geq 2$, define $L := \text{LCS}(\pi, \sigma)$ and let $C = (c_1, \dots, c_L)$ be a longest common subsequence. For $1 \leq i < j \leq L$, define the intervals:

$$I_{i,j}^{\pi} := \{x \in [n] \mid c_i \leq_{\pi} x \leq_{\pi} c_j\}, \quad \text{and} \quad I_{i,j}^{\sigma} := \{x \in [n] \mid c_i \leq_{\sigma} x \leq_{\sigma} c_j\}.$$

Define also $W_{i,j} := |I_{i,j}^{\pi} \cup I_{i,j}^{\sigma}|$.

Lemma 8. For any $\pi, \sigma \in \mathbb{S}_n$ such that $L := \text{LCS}(\pi, \sigma) \geq 2$, it holds that:

$$P_{\mathcal{H}}(\pi, \sigma) \geq \frac{1}{n^2} \cdot \sum_{1 \leq i < j \leq L} \frac{1}{W_{i,j}}.$$

Proof. Consider the case $\pi \neq \sigma$, otherwise the result is trivial, since the right-hand side is always less than 1. For $\tau \in \mathbb{S}_n$, and $1 \leq i < j \leq L$, suppose that:

$$c_j \geq_{\tau} x \quad \text{for all } x \in I_{i,j}^{\pi} \cup I_{i,j}^{\sigma}.$$

Under this assumption, it holds that: $c_j \in \text{Rec}_{\tau, c_i}(\pi) \cap \text{Rec}_{\tau, c_i}(\sigma)$. Indeed, we have that c_j is the maximum (according to τ) of all indices $I_{i,j}^{\pi}$ between c_i and c_j in the one-line representation of π and therefore it will be added to $\text{Rec}_{\tau, z}(\pi)$ when starting from $z = c_i$ (and the same argument holds for σ).

By Lemma 6 and the remark above, we have:

$$\begin{aligned} P_{\mathcal{H}}(\pi, \sigma) &= \frac{1}{n} \cdot \mathbb{E}_{\tau, z} [|\text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)|] \\ &\geq \frac{1}{n} \cdot \sum_{i=1}^{L-1} \Pr[z = c_i] \cdot \mathbb{E}_{\tau} [|\text{Rec}_{\tau, z}(\pi) \cap \text{Rec}_{\tau, z}(\sigma)| \mid z = c_i] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \cdot \sum_{i=1}^{L-1} \frac{1}{n} \cdot \sum_{j=1}^n \Pr_{\tau} [j \in \text{Rec}_{\tau, c_i}(\pi) \cap \text{Rec}_{\tau, c_i}(\sigma)] \\
&\geq \frac{1}{n^2} \sum_{i=1}^{L-1} \sum_{j=i+1}^L \Pr_{\tau} [c_j \in \text{Rec}_{\tau, c_i}(\pi) \cap \text{Rec}_{\tau, c_i}(\sigma)] \\
&\geq \frac{1}{n^2} \sum_{i=1}^{L-1} \sum_{j=i+1}^L \Pr_{\tau} [c_j \geq_{\tau} x \text{ for all } x \in I_{i,j}^{\pi} \cup I_{i,j}^{\sigma}] \\
&= \frac{1}{n^2} \sum_{i=1}^{L-1} \sum_{j=i+1}^L \frac{1}{W_{i,j}},
\end{aligned}$$

where the last inequality follows from the remark above and the last equality uses that for a uniform at random $\tau \in \mathbb{S}_n$, the element ranked highest among $I_{i,j}^{\pi} \cup I_{i,j}^{\sigma}$ is chosen uniformly at random among its elements. \square

Lemma 9. *For any $\pi, \sigma \in \mathbb{S}_n$ such that $L := \text{LCS}(\pi, \sigma) \geq 4$, it holds that:*

$$\sum_{1 \leq i < j \leq L} \frac{1}{W_{i,j}} \geq \frac{L^2 \ln L}{32n}.$$

Proof. For $1 < t \leq L$, define:

$$w_t := |I_{t-1,t}^{\pi}| + |I_{t-1,t}^{\sigma}|.$$

For $1 \leq i < L$ and $1 \leq d \leq L - i$, define $S_{i,d} := \sum_{t=i+1}^{i+d} w_t$. Note that, for $1 \leq i < j \leq L$, we have:

$$W_{i,j} \leq |I_{i,j}^{\pi}| + |I_{i,j}^{\sigma}| \leq \sum_{t=i+1}^j w_t = S_{i,(j-i)}. \quad (3)$$

Therefore, we have:

$$\sum_{1 \leq i < j \leq L} \frac{1}{W_{i,j}} = \sum_{i=1}^{L-1} \sum_{j=i+1}^L \frac{1}{W_{i,j}} \stackrel{(3)}{\geq} \sum_{i=1}^{L-1} \sum_{j=i+1}^L \frac{1}{S_{i,(j-i)}} = \sum_{i=1}^{L-1} \sum_{d=1}^{L-i} \frac{1}{S_{i,d}} = \sum_{d=1}^{L-1} \sum_{i=1}^{L-d} \frac{1}{S_{i,d}}. \quad (4)$$

For a fixed $1 \leq d < L$, by the Cauchy–Schwarz inequality, we have:

$$\left(\sum_{i=1}^{L-d} \frac{1}{S_{i,d}} \right) \cdot \left(\sum_{i=1}^{L-d} S_{i,d} \right) \geq \left(\sum_{i=1}^{L-d} \sqrt{\frac{1}{S_{i,d}}} \cdot \sqrt{S_{i,d}} \right)^2 = (L-d)^2. \quad (5)$$

Note that each element of $[n]$ can be counted at most twice in w_t and it can appear in at most two w_t 's, therefore:

$$\sum_{t=2}^L w_t \leq 4n.$$

Moreover, each w_t can appear in at most d distinct $S_{i,d}$ for $i \in \{1, \dots, L-d\}$, giving:

$$\sum_{i=1}^{L-d} S_{i,d} \leq d \sum_{t=2}^L w_t \leq 4 \cdot d \cdot n.$$

Combining this inequality with (5) gives:

$$\sum_{i=1}^{L-d} \frac{1}{S_{i,d}} \geq \frac{(L-d)^2}{\sum_{i=1}^{L-d} S_{i,d}} \geq \frac{(L-d)^2}{4 \cdot d \cdot n}.$$

Plugging this inequality into (4) we obtain:

$$\sum_{1 \leq i < j \leq L} \frac{1}{W_{i,j}} \geq \sum_{d=1}^{L-1} \frac{(L-d)^2}{4 \cdot d \cdot n} \geq \sum_{d=1}^{\lfloor L/2 \rfloor} \frac{(L-d)^2}{4 \cdot d \cdot n} \geq \frac{L^2}{16n} \sum_{d=1}^{\lfloor L/2 \rfloor} \frac{1}{d} \geq \frac{L^2 \ln L}{32n},$$

where the last inequality uses that for $L \geq 4$:

$$\sum_{d=1}^{\lfloor L/2 \rfloor} \frac{1}{d} \geq \ln(\lfloor L/2 \rfloor + 1) \geq \ln(\sqrt{L}) = \frac{\ln L}{2}. \quad \square$$

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

Proof of Theorem 5. It is sufficient to show that for each $\pi, \sigma \in \mathbb{S}_n$ it holds that:

$$\Theta(\sqrt{\log n/n}) \cdot S^U(\pi, \sigma) \leq P_{\mathcal{H}}(\pi, \sigma) \leq S^U(\pi, \sigma).$$

Note that if $\pi = \sigma$ then $P_{\mathcal{H}}(\pi, \sigma) = S^U(\pi, \sigma) = 1$. Therefore, consider the case where $\pi \neq \sigma$. By combining Lemma 6 and Lemma 7 we have:

$$\frac{1}{n} \leq P_{\mathcal{H}}(\pi, \sigma) \leq \frac{\text{LCS}(\pi, \sigma)}{n} = S^U(\pi, \sigma). \quad (6)$$

Therefore, we only have to prove $P_{\mathcal{H}}(\pi, \sigma) \geq \Omega(\sqrt{\log n/n}) \cdot S^U(\pi, \sigma)$ to conclude the proof. Consider first the case where $\text{LCS}(\pi, \sigma) \leq 3$. In this case we have:

$$P_{\mathcal{H}}(\pi, \sigma) \geq \frac{1}{n} \geq \frac{1}{3} \cdot \frac{\text{LCS}(\pi, \sigma)}{n} = \frac{1}{3} \cdot S^U(\pi, \sigma).$$

Consider now the case $\text{LCS}(\pi, \sigma) \geq 4$ and denote $L := \text{LCS}(\pi, \sigma)$. By combining Lemma 8 and Lemma 9 we obtain:

$$P_{\mathcal{H}}(\pi, \sigma) \geq \frac{L^2 \ln L}{32 \cdot n^3} = \frac{L \ln L}{32 \cdot n^2} \cdot S^U(\pi, \sigma). \quad (7)$$

By (6) we also have:

$$P_{\mathcal{H}}(\pi, \sigma) \geq \frac{1}{L} \cdot S^U(\pi, \sigma). \quad (8)$$

We split the case $\text{LCS}(\pi, \sigma) \geq 4$ in two sub-cases. Suppose first that $\text{LCS}(\pi, \sigma) \leq \frac{n}{\sqrt{\ln n}}$. Then, (8) gives: $P_{\mathcal{H}}(\pi, \sigma) \geq \sqrt{\ln n}/n \cdot S^U(\pi, \sigma)$. Suppose instead that $\text{LCS}(\pi, \sigma) > \frac{n}{\sqrt{\ln n}}$. Then, by (7) we have:

$$P_{\mathcal{H}}(\pi, \sigma) \geq \frac{\frac{n}{\sqrt{\ln n}} \ln \left(\frac{n}{\sqrt{\ln n}} \right)}{32n^2} \cdot S^U(\pi, \sigma) = \left(\frac{\ln n - \frac{1}{2} \ln \ln n}{32 \cdot \sqrt{\ln n} \cdot n} \right) \cdot S^U(\pi, \sigma) \geq \frac{\sqrt{\ln n}}{64 \cdot n} \cdot S^U(\pi, \sigma),$$

where we used that $\frac{\ln n}{2} \geq \frac{\ln \ln n}{2}$ for $n \geq 3$. □

5 Lower Bound for Ulam Similarity

In this section, we prove the following result.

Theorem 10. *Let n be any integer of the form 8^{2^k} , for integer $k \geq 0$. The LSH distortion of $(\mathbb{S}_n, \mathbb{S}^U)$ is $\Omega(n^\delta)$, where $\delta \geq 0.12$.*

At a high level, we obtain a lower bound on the LSH distortion by starting with a small instance that exhibits constant distortion, and then iteratively amplifying it to handle larger values of n . In each step, both the distortion bound and the permutation domain grow quadratically. Our base case has $n = 8$, therefore our construction produces instances for any n of the form 8^{2^k} , where k is a non-negative integer. The main tool enabling this amplification is the wreath product.

For any two subsets $A, B \subseteq \mathbb{S}_n$, we define the similarity matrix $M_{A,B} \in \mathbb{R}^{|A| \times |B|}$ by $M_{A,B}(\pi, \sigma) = \mathbb{S}^U(\pi, \sigma)$ for any $\pi \in A$ and $\sigma \in B$. For simplicity, when A and B are the same, we will adopt a lighter notation, and denote by M_A the matrix $M_{A,A}$. We note that the matrix $M_{(A \cup B)}$ has a simple block structure:

$$M_{(A \cup B)} = \begin{pmatrix} M_A & M_{A,B} \\ M_{B,A} & M_B \end{pmatrix},$$

where we fixed an ordering in which the permutations in A come before those in B .

Definition 11 (Witness). For $n > 0$, let $A, B \subseteq \mathbb{S}_n$ be disjoint subsets such that $|A| = |B|$ and $M_A = M_B$. Let $W \in \mathbb{R}^{(|A|+|B|) \times (|A|+|B|)}$ be a PSD block matrix such that:

$$W = \begin{pmatrix} U & -V \\ -V^T & U \end{pmatrix} \succeq 0,$$

where $U, V \in \mathbb{R}^{|A| \times |A|}$ are entry-wise non-negative symmetric matrices such that $\text{tr}(U \cdot M_A) \neq 0$. We call the triple (A, B, W) a *witness* of size n and we define its value to be:

$$\nu(A, B, W) := \frac{\text{tr}(V \cdot M_{A,B})}{\text{tr}(U \cdot M_A)}.$$

We now show that the existence of a witness implies a lower bound to the LSH distortion for Ulam similarity on \mathbb{S}_n .

Lemma 12. *Let (A, B, W) be a witness of size n (Definition 11). Then, the LSH distortion for Ulam similarity space $(\mathbb{S}_n, \mathbb{S}^U)$ is at least the value of the witness, $\nu(A, B, W)$, even when restricted only to permutations in $A \cup B \subseteq \mathbb{S}_n$.*

Proof. Consider any distribution \mathcal{H} over functions on \mathbb{S}_n which Δ -approximates $(A \cup B, \mathbb{S}^U)$ for $\Delta \geq 1$. For every $\pi, \sigma \in A \cup B$ we have:

$$\frac{M_{A \cup B}(\pi, \sigma)}{\Delta} = \frac{\mathbb{S}^U(\pi, \sigma)}{\Delta} \leq P_{\mathcal{H}}(\pi, \sigma) \leq \mathbb{S}^U(\pi, \sigma) = M_{A \cup B}(\pi, \sigma). \quad (9)$$

Note that, having fixed an ordering of the elements of $A \cup B$ that puts the elements of A before the elements of B , $P_{\mathcal{H}}$ gets naturally divided into four square blocks:

$$P_{\mathcal{H}} = \begin{pmatrix} P_A & P_{A,B} \\ P_{B,A} & P_B \end{pmatrix}.$$

By Lemma 4, $P_{\mathcal{H}} \succeq 0$ and $W \succeq 0$, we have $\text{tr}(WP_{\mathcal{H}}) \geq 0$. Expanding blockwise, and using $V^T = V$ and $P_{B,A}^T = P_{A,B}$, we have:

$$\text{tr}(UP_A) + \text{tr}(UP_B) - 2\text{tr}(VP_{A,B}) \geq 0.$$

Note that, by Equation (9), $P_{\mathcal{H}}(\pi, \sigma) \leq M_{A \cup B}(\pi, \sigma)$ and $P_{\mathcal{H}}(\pi, \sigma) \geq (1/\Delta) \cdot M_{A \cup B}(\pi, \sigma)$ for all π, σ . Therefore, we have:

$$\text{tr}(U \cdot M_A) + \text{tr}(U \cdot M_B) \geq \frac{2}{\Delta} \cdot \text{tr}(V \cdot M_{A,B}).$$

Since $M_A = M_B$, we have $\text{tr}(U \cdot M_A) = \text{tr}(U \cdot M_B)$. Hence,

$$\Delta \geq \frac{\text{tr}(V \cdot M_{A,B})}{\text{tr}(U \cdot M_A)} = \nu(A, B, W). \quad \square$$

For permutations $\pi, \sigma \in \mathbb{S}_n$ the *wreath product* $\pi \wr \sigma \in \mathbb{S}_{n^2}$ is defined as follows. The elements of $[n^2]$ are partitioned into n blocks where the i th block, $i \in [n]$, contains elements $\Sigma_i := \{(i-1) \cdot n + 1, \dots, i \cdot n\}$. The resulting permutation $\pi \wr \sigma$ is obtained by permuting the elements in each block with the inner permutation σ and then permuting the blocks with the outer permutation π . For two sets $A, B \subseteq \mathbb{S}_n$, we define $A \wr B := \{\pi \wr \sigma \mid \pi \in A, \sigma \in B\}$. We use the following key property of the wreath product for amplification:

Lemma 13. *For any $\pi, \sigma, \tau_1, \tau_2 \in \mathbb{S}_n$ we have:*

$$\text{LCS}(\pi \wr \sigma, \tau_1 \wr \tau_2) = \text{LCS}(\pi, \tau_1) \cdot \text{LCS}(\sigma, \tau_2).$$

Proof. In the permutation $\pi \wr \sigma$, the blocks appear in the order given by π , and within every block the elements appear in the order given by σ ; similarly for $\tau_1 \wr \tau_2$.

We first prove that $\text{LCS}(\pi \wr \sigma, \tau_1 \wr \tau_2) \leq \text{LCS}(\pi, \tau_1) \cdot \text{LCS}(\sigma, \tau_2)$. Let Z be any common subsequence of $\pi \wr \sigma$ and $\tau_1 \wr \tau_2$. Since the alphabets Σ_i are pairwise disjoint, an element from a block Σ_i can only be matched with the same element from the same block Σ_i . Therefore, if we track the sequence of block indices that contribute at least one matched element to Z , then this is a common subsequence of the outer permutations π and τ_1 ; thus, the number of contributing blocks is at most $\text{LCS}(\pi, \tau_1)$.

Let Σ_i be any contributing block. The matched elements of Z in this block is a common subsequence of the ordering of Σ_i inside $\pi \wr \sigma$ and the ordering of Σ_i inside $\tau_1 \wr \tau_2$. These two internal orderings are σ and τ_2 , relabeled with the alphabet Σ_i ; therefore, this block contributes at most $\text{LCS}(\sigma, \tau_2)$ matched elements. Summing over all contributing blocks, we obtain $\text{LCS}(\pi \wr \sigma, \tau_1 \wr \tau_2) \leq \text{LCS}(\pi, \tau_1) \cdot \text{LCS}(\sigma, \tau_2)$.

We now show $\text{LCS}(\pi \wr \sigma, \tau_1 \wr \tau_2) \geq \text{LCS}(\pi, \tau_1) \cdot \text{LCS}(\sigma, \tau_2)$. Consider an optimal common subsequence of the outer permutations π and τ_1 , consisting of $\text{LCS}(\pi, \tau_1)$ block indices. For each such matched block Σ_i , consider an optimal common subsequence of the inner permutations σ and τ_2 , relabeled using Σ_i . Since the block indices appear in the same relative order in both π and τ_1 , concatenating these common subsequences over the chosen blocks yields a valid common subsequence of $\pi \wr \sigma$ and $\tau_1 \wr \tau_2$ of length $\text{LCS}(\pi, \tau_1) \cdot \text{LCS}(\sigma, \tau_2)$. \square

We now show how to use the wreath product to amplify the size of the witness. We will make use of the *Kronecker product*. We recall that for two matrices $X \in \mathbb{R}^{n \times m}$ and $Y \in \mathbb{R}^{s \times t}$, the Kronecker product $X \otimes Y \in \mathbb{R}^{ns \times mt}$ is the following block matrix:

$$X \otimes Y = \begin{pmatrix} x_{11} \cdot Y & \dots & x_{1m} \cdot Y \\ \vdots & \ddots & \vdots \\ x_{n1} \cdot Y & \dots & x_{nm} \cdot Y \end{pmatrix}.$$

Lemma 14. Consider any disjoint sets $A, B \subseteq \mathbb{S}_n$ with $|A| = |B|$. Let $A' = A \wr A$ and $B' = B \wr B$. Then, we have:

$$M_{(A' \cup B')} = \begin{pmatrix} M_A \otimes M_A & M_{A,B} \otimes M_{A,B} \\ M_{B,A} \otimes M_{B,A} & M_B \otimes M_B \end{pmatrix}.$$

Proof. Let $\pi, \sigma \in A' \cup B'$ and suppose that $\pi = \pi_1 \wr \pi_2$, $\sigma = \sigma_1 \wr \sigma_2$. By Lemma 13,

$$M_{(A' \cup B')}(\pi, \sigma) = \frac{\text{LCS}(\pi, \sigma)}{n^2} = \frac{\text{LCS}(\pi_1, \sigma_1)}{n} \cdot \frac{\text{LCS}(\pi_2, \sigma_2)}{n} = M_{(A \cup B)}(\pi_1, \sigma_1) \cdot M_{(A \cup B)}(\pi_2, \sigma_2).$$

Then, by considering the ordering over A' obtained by the ordering of A when we consider first the outer permutation and then the inner permutation (and similarly for B'), we have that each entry of the matrix is equivalent to the Kronecker product. \square

Lemma 15. Let (A, B, W) be a witness of size n . Let $A' = A \wr A$, $B' = B \wr B$, and:

$$W' = \begin{pmatrix} U \otimes U & -(V \otimes V) \\ -(V^T \otimes V^T) & U \otimes U \end{pmatrix}.$$

Then, (A', B', W') is a witness of size n^2 and $\nu(A', B', W') \geq \nu(A, B, W)^2$.

Proof. Note that $U \otimes U$ and $V \otimes V$ are non-negative and symmetric. Moreover, since $W \succeq 0$, we have that $W \otimes W \succeq 0$. Define:

$$W'' = \begin{pmatrix} U \otimes U & V \otimes V \\ V^T \otimes V^T & U \otimes U \end{pmatrix}.$$

Since W'' is a principal sub-matrix of $W \otimes W$, we have that $W'' \succeq 0$. Observe now that $W' = \text{diag}(I, -I) \cdot W'' \cdot \text{diag}(I, -I)$, and therefore $W' \succeq 0$ since for any matrix $X \succeq 0$ and any matrix Y , $YXY^T \succeq 0$. Moreover, by Lemma 14 and since $M_A = M_B$, we also have $M_{A'} = M_{B'}$. Then, (A', B', W') is a valid witness of size n^2 . We now compute its value.

For matrices A, B, C, D it holds that $\text{tr}((A \otimes B)(C \otimes D)) = \text{tr}(AC) \cdot \text{tr}(BD)$. By using this property, and by Lemma 14, we have:

$$\begin{aligned} \text{tr}((V \otimes V) \cdot M_{A', B'}) &= \text{tr}((V \otimes V) \cdot (M_{A,B} \otimes M_{A,B})) = \text{tr}(V \cdot M_{A,B})^2, \text{ and} \\ \text{tr}((U \otimes U) \cdot M_{A'}) &= \text{tr}((U \otimes U) \cdot (M_A \otimes M_A)) = \text{tr}(U \cdot M_A)^2. \end{aligned}$$

Note that since $\text{tr}(U \cdot M_A) \neq 0$, $\text{tr}((U \otimes U)M_{A'}) \neq 0$. Then:

$$\nu(A', B', W') = \frac{\text{tr}((V \otimes V) \cdot M_{A', B'})}{\text{tr}((U \otimes U) \cdot M_{A'})} = \left(\frac{\text{tr}(V \cdot M_{A,B})}{\text{tr}(U \cdot M_A)} \right)^2 = \nu(A, B, W)^2. \quad \square$$

We now provide our starting instance.

Proposition 16. There exists a witness (A_0, B_0, W_0) of size 8 and value $\geq 9/7$.

Proof. Let $A_0 = \{A_1, \dots, A_4\}$ and $B_0 = \{B_1, \dots, B_4\}$, where $\tau = (1278)(3654) = (2, 7, 6, 3, 4, 5, 8, 1)$ and $B_i = \tau A_i$:

$$\begin{aligned} A_1 &= (1, 2, 3, 4, 5, 6, 7, 8), & B_1 &= (2, 7, 6, 3, 4, 5, 8, 1), \\ A_2 &= (4, 3, 2, 1, 8, 7, 6, 5), & B_2 &= (3, 6, 7, 2, 1, 8, 5, 4), \\ A_3 &= (6, 5, 8, 7, 2, 1, 4, 3), & B_3 &= (5, 4, 1, 8, 7, 2, 3, 6), \\ A_4 &= (7, 8, 5, 6, 3, 4, 1, 2), & B_4 &= (8, 1, 4, 5, 6, 3, 2, 7). \end{aligned}$$

Since B_0 is obtained from A_0 by relabeling, we have $M_{A_0} = M_{B_0}$. A direct computation gives

$$M_{A_0} = \frac{1}{8} \begin{pmatrix} 8 & 2 & 2 & 2 \\ 2 & 8 & 2 & 2 \\ 2 & 2 & 8 & 2 \\ 2 & 2 & 2 & 8 \end{pmatrix}, \quad M_{A_0, B_0} = \frac{1}{8} \begin{pmatrix} 5 & 4 & 4 & 5 \\ 4 & 5 & 5 & 4 \\ 4 & 5 & 5 & 4 \\ 5 & 4 & 4 & 5 \end{pmatrix}.$$

The matrix W_0 for this instance is given by $U = V = J_4$ (where J_4 is the all-ones 4×4 matrix). First,

$$W = \begin{pmatrix} U & -V \\ -V^T & U \end{pmatrix} = \begin{pmatrix} J_4 & -J_4 \\ -J_4 & J_4 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes J_4 \succeq 0.$$

Next, it is easy to calculate

$$\text{tr}(V \cdot M_{A_0, B_0}) = \frac{1}{8}(18 \cdot 4) = 9, \quad \text{tr}(U \cdot M_{A_0}) = \frac{1}{8}(14 \cdot 4) = 7.$$

Hence, $\nu(A_0, B_0, W_0) = 9/7$. □

Lemma 17. *There exists a constant $\delta > 0$ such that, for any $n_0 > 0$, there exists $n \geq n_0$ such that the LSH distortion of the Ulam similarity S^U on \mathbb{S}_n is at least $\Omega(n^\delta)$.*

Proof. Let $\Delta_0 = 9/7$ and $m = 8$. Let k be the smallest integer such that $m^{2^k} \geq n_0$, we set $n = m^{2^k}$. Note that $k = \log_2 \log_m(n)$.

Starting from the witness of Proposition 16 and applying Lemma 15 k times we obtain a witness of size $m^{2^k} = n$ and value $\geq \Delta_0^{2^k}$. By Lemma 12, this implies that the LSH distortion on \mathbb{S}_n is at least:

$$\Delta_0^{2^k} = \Delta_0^{\log_m(n)} = n^{\log_m \Delta_0} \geq n^{0.120856} = \Omega(n^\delta),$$

for $\delta = 0.12$. □

Note that Lemma 17 directly implies Theorem 10. The proof is elementary.

6 Upper Bound for Cayley Similarity

In this section, we show that the LSH distortion for Cayley similarity is no more than n .

Theorem 18. *The LSH distortion of (\mathbb{S}_n, S^C) is at most n .*

Proof. Let \mathcal{H} be the uniform at random distribution over the functions $h : \mathbb{S}_n \rightarrow [n]$. Clearly, for each $\pi, \sigma \in \mathbb{S}_n$, $\pi \neq \sigma$:

$$\Pr_{h \sim \mathcal{H}}[h(\pi) = h(\sigma)] = \frac{1}{n}.$$

Since $\frac{1}{n} \leq S^C(\pi, \sigma) \leq 1$ for all $\pi, \sigma \in \mathbb{S}_n$, we have:

$$\frac{1}{n} \cdot S^C(\pi, \sigma) \leq \Pr_{h \sim \mathcal{H}}[h(\pi) = h(\sigma)] \leq S^C(\pi, \sigma). \quad \square$$

7 Lower Bound for Cayley Similarity

In this section, we show the following result:

Theorem 19. *The LSH distortion of $(\mathbb{S}_n, \mathbb{S}^C)$ is $\Omega(n)$.*

The proof will rely on some standard results in the representation theory of \mathbb{S}_n . For convenience, we review these standard results in Appendix A.

We begin with the following definition:

Definition 20 (Bi-invariance). We say that a function $f : \mathbb{S}_n \times \mathbb{S}_n \rightarrow \mathbb{R}$ is *bi-invariant* if:

$$f(\pi, \sigma) = f(\tau_1 \pi \tau_2, \tau_1 \sigma \tau_2),$$

for any $\pi, \sigma, \tau_1, \tau_2 \in \mathbb{S}_n$. We say a bi-invariant function is a *bi-invariant PSD kernel* if, in addition, it satisfies:

$$\sum_{\pi, \sigma \in \mathbb{S}_n} a_\sigma a_\pi f(\pi, \sigma) > 0,$$

for every $a \in \mathbb{R}^{\mathbb{S}_n} \setminus \{0\}$. We also require that $f(\text{id}, \text{id}) = 1$ for f to be a PSD kernel.

Recall that a *class function* on \mathbb{S}_n is a function that is constant on each conjugacy class of \mathbb{S}_n . We recall the following property of bi-invariant functions. (Note that equivalent results are known in the literature, and hold for more general groups—see, e.g. (Kondor, 2008, Chapter 4.5)—but we include a proof anyway for completeness).

Lemma 21. *A function $f : \mathbb{S}_n \times \mathbb{S}_n \rightarrow \mathbb{R}$ is bi-invariant if and only if there exists a class function $g : \mathbb{S}_n \rightarrow \mathbb{R}$ such that:*

$$f(\sigma, \pi) = g(\sigma \pi^{-1}),$$

for every $\sigma, \pi \in \mathbb{S}_n$.

Proof. Let $f : \mathbb{S}_n \times \mathbb{S}_n \rightarrow \mathbb{R}$ be a bi-invariant function. Define $g : \mathbb{S}_n \rightarrow \mathbb{R}$ as:

$$g(\pi) := f(\pi, \text{id}).$$

Then we have:

$$f(\pi, \sigma) = f(\pi \sigma^{-1}, \sigma \sigma^{-1}) = f(\pi \sigma^{-1}, \text{id}) = g(\pi \sigma^{-1})$$

for any $\pi, \sigma \in \mathbb{S}_n$. Moreover, g is a class function, since, for any $\pi, \sigma \in \mathbb{S}_n$:

$$g(\sigma \pi \sigma^{-1}) = f(\sigma \pi \sigma^{-1}, \text{id}) = f(\pi, \sigma^{-1} \sigma) = f(\pi, \text{id}) = g(\pi),$$

where the second equality follows from the bi-invariance of f .

On the other hand, suppose that $f : \mathbb{S}_n \times \mathbb{S}_n \rightarrow \mathbb{R}$ is a function such that $f(\pi, \sigma) = g(\pi \sigma^{-1})$ for some class function g , then, for any $\pi, \sigma, \tau \in \mathbb{S}_n$:

$$f(\pi \tau, \sigma \tau) = g(\pi \tau \tau^{-1} \sigma^{-1}) = g(\pi \sigma^{-1}) = f(\pi, \sigma),$$

and:

$$f(\tau \pi, \tau \sigma) = g(\tau \pi \sigma^{-1} \tau^{-1}) = g(\pi \sigma^{-1}) = f(\pi, \sigma),$$

where the second equality uses the fact that g is a class function. Hence f is bi-invariant. \square

A crucial property of the Cayley similarity is that it is bi-invariant, as we now show.

Lemma 22. $S^C(\cdot, \cdot)$ is bi-invariant.

Proof. We have that, for any choice of π and σ , $S^C(\pi, \sigma) = \frac{c(\pi\sigma^{-1})}{n}$. Since any two conjugate permutations have the same cycle type, the function $\pi \mapsto \frac{c(\pi)}{n}$ is a class function, and the result follows from Lemma 21. \square

Definition 23 (Bi-invariant Distribution over Functions). We will say that a distribution \mathcal{H} over functions from \mathbb{S}_n into some set \mathcal{Y} is *bi-invariant* if the function $P_{\mathcal{H}}(\cdot, \cdot)$ is bi-invariant.

We now show that when studying the minimum achievable distortion for (\mathbb{S}_n, S^C) , any LSH family can be assumed to be bi-invariant.

Lemma 24. Suppose there exists a distribution \mathcal{H} over functions from \mathbb{S}_n into a finite set \mathcal{Y} that Δ -approximates (\mathbb{S}_n, S^C) , then there exists a bi-invariant distribution \mathcal{H}' that Δ -approximates (\mathbb{S}_n, S^C) .

Proof. We construct \mathcal{H}' by pre-composing the elements of \mathcal{H} with random left and right permutations: define a function $h' \sim \mathcal{H}'$ as $h'(\pi) = h(\alpha\pi\beta)$, where h is drawn from \mathcal{H} and α, β are drawn uniformly and independently from \mathbb{S}_n . We first show that \mathcal{H}' Δ -approximates (\mathbb{S}_n, S^C) . For every choice of $\pi, \sigma \in \mathbb{S}_n$, we have:

$$\begin{aligned} \Pr_{h' \sim \mathcal{H}'} [h'(\pi) = h'(\sigma)] &= \mathbb{E}_{\alpha, \beta \sim \mathbb{S}_n} \left[\Pr_{h \sim \mathcal{H}} [h(\alpha\pi\beta) = h(\alpha\sigma\beta)] \right] \\ &\leq \mathbb{E}_{\alpha, \beta \sim \mathbb{S}_n} [S^C(\alpha\pi\beta, \alpha\sigma\beta)] \\ &\stackrel{\text{Lemma 22}}{=} \mathbb{E}_{\alpha, \beta \sim \mathbb{S}_n} [S^C(\pi, \sigma)] = S^C(\pi, \sigma), \end{aligned}$$

where the first inequality uses that \mathcal{H} Δ -approximates (\mathbb{S}_n, S^C) . Similarly:

$$\begin{aligned} \Pr_{h' \sim \mathcal{H}'} [h'(\pi) = h'(\sigma)] &= \mathbb{E}_{\alpha, \beta \sim \mathbb{S}_n} \left[\Pr_{h \sim \mathcal{H}} [h(\alpha\pi\beta) = h(\alpha\sigma\beta)] \right] \\ &\geq \mathbb{E}_{\alpha, \beta \sim \mathbb{S}_n} \left[\frac{1}{\Delta} \cdot S^C(\alpha\pi\beta, \alpha\sigma\beta) \right] \\ &\stackrel{\text{Lemma 22}}{=} \mathbb{E}_{\alpha, \beta \sim \mathbb{S}_n} \left[\frac{1}{\Delta} \cdot S^C(\pi, \sigma) \right] = \frac{1}{\Delta} S^C(\pi, \sigma), \end{aligned}$$

Showing that indeed \mathcal{H}' Δ -approximates (\mathbb{S}_n, S^C) . Finally, we verify that $P_{\mathcal{H}'}$ is bi-invariant. For any $\pi, \sigma \in \mathbb{S}_n$ and any left multiplier τ_1 and right multiplier τ_2 :

$$P_{\mathcal{H}'}(\tau_1\pi\tau_2, \tau_1\sigma\tau_2) = \mathbb{E}_{\alpha, \beta \sim \mathbb{S}_n} [P_{\mathcal{H}}(\alpha\tau_1\pi\tau_2\beta, \alpha\tau_1\sigma\tau_2\beta)].$$

Because α and β are drawn uniformly from the group, the products $\alpha' = \alpha\tau_1$ and $\beta' = \tau_2\beta$ are also uniformly distributed. Thus, the expectation is identical to $\mathbb{E}_{\alpha', \beta'} [P_{\mathcal{H}}(\alpha'\pi\beta', \alpha'\sigma\beta')] = P_{\mathcal{H}'}(\pi, \sigma)$, completing the proof. \square

Note also that, by Lemma 4, $P_{\mathcal{H}'}$ is PSD and it is therefore a bi-invariant PSD kernel.

Recall that each partition $\lambda \vdash n$ is associated with a distinct irreducible representation S^λ of \mathbb{S}_n , we denote by $\chi_\lambda : \mathbb{S}_n \rightarrow \mathbb{C}$ its character, and by $d_\lambda \in \mathbb{N}$ its dimension.

Lemma 25. Any bi-invariant PSD kernel P on \mathbb{S}_n can be written as a convex combination of normalized characters of irreducible representations of \mathbb{S}_n as follows:

$$\forall \sigma, \pi \in \mathbb{S}_n : \quad P(\sigma, \pi) = \sum_{\lambda \vdash n} w_\lambda \frac{\chi_\lambda(\sigma\pi^{-1})}{d_\lambda},$$

where $w_\lambda \geq 0$ and $\sum_\lambda w_\lambda = 1$.

Proof. Let $\phi : \mathbb{S}_n \rightarrow \mathbb{R}$ be the class function satisfying:

$$\phi(\sigma\pi^{-1}) = P(\sigma, \pi),$$

the existence of which is guaranteed by Lemma 21. By a result of Kondor (2008, Corollary 4.5.5²), since ϕ is PSD, we have:

$$\phi = \sum_{\lambda \vdash n} a_\lambda \cdot \chi_\lambda, \tag{10}$$

for some choice of $\{a_\lambda\}_{\lambda \vdash n}$, where each $a_\lambda \in \mathbb{R}_{\geq 0}$.

In particular:

$$1 = P(\text{id}, \text{id}) = \phi(\text{id}) = \sum_{\lambda \vdash n} a_\lambda \cdot \chi_\lambda(\text{id}) = \sum_{\lambda \vdash n} a_\lambda \cdot d_\lambda,$$

where we use that $\chi_\lambda(\text{id}) = d_\lambda$. Letting $w_\lambda := a_\lambda \cdot d_\lambda$, together with Equation (10) completes the proof. \square

For our asymptotic bounds, it is necessary to isolate the trivial representation (n) , where $\chi_{(n)}(\cdot) = 1$ and $d_{(n)} = 1$ and the sign representation (1^n) , where $\chi_{(1^n)}(\sigma) = \text{sgn}(\sigma)$ and $d_{(1^n)} = 1$. Thus, it is helpful to write

$$P(\text{id}, \sigma) = w_0 + w_{\text{sgn}} \cdot \text{sgn}(\sigma) + \sum_{\lambda \neq (n), (1^n)} w_\lambda \frac{\chi_\lambda(\sigma)}{d_\lambda}. \tag{11}$$

For a permutation $\sigma \in \mathbb{S}_n$, the *fixed-point* count is defined as $\text{fix}(\sigma) = |\{i \in [n] \mid \sigma(i) = i\}|$, i.e., the number of length-1 cycles. A *derangement* is a permutation σ that moves all elements, i.e., $\text{fix}(\sigma) = 0$. For any $\sigma \in \mathbb{S}_n$, we also define the value $\text{supp}(\sigma)$ as $n - \text{fix}(\sigma)$. We recall the following result.

Theorem 26 (Roichman (1996)). *There exist absolute constants $\varphi > 0$ and $0 < q < 1$ such that for any partition $\lambda = (\lambda_1, \dots, \lambda_k) \vdash n$, any $n \geq 4$ and any permutation $\sigma \in \mathbb{S}_n$:*

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq \left(\max \left\{ \frac{\lambda_1}{n}, \frac{k}{n}, q \right\} \right)^{\varphi \cdot \text{supp}(\sigma)}.$$

Theorem 26 allows us to prove the following.

Lemma 27. *There exist absolute constants $\varphi > 0$ and $0 < q < 1$ such that for any $n \geq 4$, any partition $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k) \vdash n$, and any derangement $\sigma \in \mathbb{S}_n$,*

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq e^{-\varphi \cdot \alpha},$$

where $\alpha = \min\{n - \lambda_1, n - k, (1 - q)n\}$.

²We remark that Kondor's result is stated in terms of what he refers to as "Positive Definite" kernels, which are, in reality, positive semidefinite kernels, as he himself explains earlier in the thesis.

Proof. Since σ is a derangement, $\text{fix}(\sigma) = 0$; therefore $\text{supp}(\sigma) = n - \text{fix}(\sigma) = n$. Furthermore, from the definition of α , we have:

$$\max \left\{ \frac{\lambda_1}{n}, \frac{k}{n}, q \right\} = 1 - \frac{\alpha}{n}.$$

Substituting these into the bound from Theorem 26 yields:

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq \left(1 - \frac{\alpha}{n} \right)^{\varphi n} \leq e^{-\varphi \cdot \alpha}. \quad \square$$

Lemma 28 (Theorem 1.2 (i) from Larsen and Shalev (2008), paraphrased). *For any derangement $\sigma \in \mathbb{S}_n$ and any partition $\lambda \vdash n$:*

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq \frac{1}{d_\lambda^{1/2-o(1)}}.$$

Lemma 29. *For any partition $\lambda = (\lambda_1, \dots, \lambda_k) \vdash n$ with $\lambda_1 \geq \frac{11}{12} \cdot n$, it holds that: $d_\lambda \geq \left(\frac{n}{2(n-\lambda_1)} \right)^{n-\lambda_1}$*

Proof. Let $\gamma = n - \lambda_1$. By the hook length formula (see Appendix A), we have:

$$\begin{aligned} d_\lambda &= \frac{n!}{\prod h(i, j)} \geq \frac{n!}{(n-2\gamma)! \cdot n^\gamma \cdot \gamma^\gamma} \\ &= \frac{n \cdots (n-2\gamma+1)}{n^\gamma \gamma^\gamma} \\ &= \frac{n \cdots n-\gamma+1}{n} \cdot \frac{(n-\gamma) \cdots (n-2\gamma+1)}{\gamma^\gamma} \\ &\geq \left(\frac{n-\gamma+1}{n} \right)^\gamma \cdot \left(\frac{n-2\gamma+1}{\gamma} \right)^\gamma \\ &\geq \left(\frac{n^2 - 6 \cdot n \cdot \gamma}{\gamma \cdot n} \right)^\gamma \\ &= \left(\frac{n}{\gamma} - 6 \right)^\gamma \\ &= \left(\frac{n}{\gamma} \left(1 - \frac{6\gamma}{n} \right) \right)^\gamma \\ &\geq \left(\frac{n}{2\gamma} \right)^\gamma, \end{aligned}$$

where the last inequality used the fact that $\gamma \leq \frac{n}{12}$. □

Lemma 30. *Consider any sufficiently large n , any derangement $\sigma \in \mathbb{S}_n$, and let P be a bi-invariant PSD kernel. Let:*

$$P(\tau, \pi) = \sum_{\lambda \vdash n} w_\lambda \frac{\chi_\lambda(\tau \pi^{-1})}{d_\lambda},$$

be the decomposition guaranteed by Lemma 25, then:

$$\left| \sum_{\substack{\lambda \vdash n \\ \lambda \neq (n), (1^n)}} w_\lambda \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq \frac{C}{n},$$

for some constant C .

Proof. We have:

$$\left| \sum_{\substack{\lambda \vdash n \\ \lambda \neq (n), (1^n)}} w_\lambda \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq \max_{\substack{\lambda \vdash n \\ \lambda \neq (n), (1^n)}} \left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \sum_{\substack{\lambda \vdash n \\ \lambda \neq (n), (1^n)}} w_\lambda \leq \max_{\substack{\lambda \vdash n \\ \lambda \neq (n), (1^n)}} \left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right|,$$

where we first apply Hölder's inequality on ℓ_∞ and ℓ_1 and then we use the fact that $\sum_{\lambda \vdash n} w_\lambda = 1$, as well as the non-negativity of the w_λ s. We now give an upper bound for the right-hand side. Let φ and q be the constants given by Lemma 27. Let $\lambda = (\lambda_1, \dots, \lambda_k) \vdash n$, and let $\alpha = \min\{n - \lambda_1, n - k, (1 - q)n\}$. Since $\lambda \neq (n), (1^n)$, we have $\min\{n - \lambda_1, n - k\} \geq 1$. We obtain an upper bound by considering five possible cases for the value of α .

Case 1: $\alpha = (1 - q)n$. Using Lemma 27, we have:

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq e^{-\varphi\alpha} = e^{-\Theta(n)} = o\left(\frac{1}{n}\right).$$

For the rest of the proof we will assume that $\alpha = n - \lambda_1 \geq 1$. This is without loss of generality, since, for any partition that does not satisfy this condition (i.e., for which $\alpha = n - k$), its transpose partition does satisfy it, and the values of $\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right|$ are the same for the two partitions.

Case 2: $\alpha \geq \frac{1}{\varphi} \ln n$. Using Lemma 27 we have:

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq e^{-\varphi\alpha} \leq \frac{1}{n}.$$

Case 3: $3 \leq \alpha < \frac{1}{\varphi} \ln n$. Using Lemma 28 and Lemma 29 and recalling that $\alpha = n - \lambda_1$, and that therefore $\lambda_1 = n - \alpha > n - \frac{1}{\varphi} \ln n \geq \frac{11}{12}n$ for sufficiently large n , we obtain:

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| \leq \left(\frac{1}{d_\lambda}\right)^{1/2-o(1)} \leq \left(\frac{2\alpha}{n}\right)^{\alpha(\frac{1}{2}-o(1))} \leq \left(\frac{(2/\varphi) \cdot \ln(n)}{n}\right)^{3/2-o(1)} = O\left(\frac{1}{n}\right),$$

where we use that $\alpha \geq 3$ and $\alpha < O(\log n)$.

Case 4: $\alpha = 1$. In this case, the only partition is $\lambda = (n - 1, 1)$, corresponding to the standard representation, which has dimension $n - 1$. Its character is known to satisfy $\chi_\lambda(\pi) = \text{fix}(\pi) - 1$ (See (Sagan, 2001, Example 2.3.8)). Since σ is a derangement, $\text{fix}(\sigma) = 0$, and the character $\chi_\lambda(\sigma) = -1$. We then have:

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| = \left| \frac{-1}{n-1} \right| = O\left(\frac{1}{n}\right).$$

Case 5: $\alpha = 2$. In this case, the only two partitions are $\lambda = (n - 2, 2)$ and $\mu = (n - 2, 1, 1)$.

By (19) in Example 32 in Appendix A, we can write the character of the permutation module $M^{(n-2,2)}$ as the sum of irreducible characters as follows:

$$\chi_{M^{(n-2,2)}} = \chi_{(n-2,2)} + \chi_{(n-1,1)} + \chi_{(n)}.$$

By (17), this equation becomes:

$$\text{cycle}_2(\pi) + \binom{\text{fix}(\pi)}{2} = \chi_{(n-2,2)}(\pi) + \text{fix}(\pi) - 1 + 1$$

for all $\pi \in \mathbb{S}_n$, where $\text{cycle}_2(\pi)$ is the number of length-two cycles in the cycle decomposition of π . And hence:

$$\chi_{(n-2,2)}(\pi) = \text{cycle}_2(\pi) + \binom{\text{fix}(\pi)}{2} - \text{fix}(\pi)$$

for all $\pi \in \mathbb{S}_n$.

Similarly, by (20) in Example 32 in Appendix A we can write the character of the permutation module $M^{(n-2,1,1)}$ as:

$$\chi_{M^{(n-2,1,1)}} = \chi_{(n-2,1,1)} + \chi_{(n-2,2)} + 2\chi_{(n-1,1)} + \chi_{(n)}.$$

By (18), this equation becomes:

$$\text{fix}(\pi)(\text{fix}(\pi) - 1) = \chi_{(n-2,1,1)} + \chi_{(n-2,2)} + 2(\text{fix}(\pi) - 1) + 1$$

for all $\pi \in \mathbb{S}_n$, which gives:

$$\chi_{(n-2,1,1)} = 1 - \text{cycle}_2(\pi) + \frac{\text{fix}(\pi)(\text{fix}(\pi) - 3)}{2}$$

for all $\pi \in \mathbb{S}_n$. Since σ is a derangement and $\text{fix}(\sigma) = 0$, we have:

$$\chi_{(n-2,2)}(\sigma) = \text{cycle}_2(\sigma),$$

and:

$$\chi_{(n-2,1,1)}(\sigma) = 1 - \text{cycle}_2(\sigma).$$

By Example 31 in Appendix A: $d_\lambda = n(n-3)/2$ and $d_\mu = (n-1)(n-2)/2$. Putting things together, we have that:

$$\left| \frac{\chi_\lambda(\sigma)}{d_\lambda} \right| = \frac{2 \cdot \text{cycle}_2(\sigma)}{n(n-3)} \leq \frac{n}{n(n-3)} = \frac{1}{n-3} = O\left(\frac{1}{n}\right),$$

and:

$$\left| \frac{\chi_\mu(\sigma)}{d_\mu} \right| = \frac{2 \cdot (\text{cycle}_2(\sigma) - 1)}{(n-1)(n-2)} \leq \frac{n}{(n-1)(n-2)} = O\left(\frac{1}{n}\right).$$

This completes the proof of Lemma 30. □

We can now complete the proof of Theorem 19.

Proof of Theorem 19. Let $n \geq 6$ be a large enough integer so that Lemma 30 holds. Let \mathcal{H} be a distribution that Δ -approximates $(\mathbb{S}_n, \mathbb{S}^C)$. Without loss of generality, by Lemma 24, we can assume that $P_{\mathcal{H}}$ is a bi-invariant PSD kernel.

We let σ^* be a derangement with $\Theta(n)$ cycles and π^* be a derangement with $O(1)$ cycles and having the same sign as σ^* . We explicitly define these permutations expressing them in their cycle decomposition. If n is even, we let:

$$\sigma^* := \prod_{i=1}^{n/2} (2i-1, 2i),$$

and observe that $c(\sigma^*) = n/2$. If instead n is odd, we define:

$$\sigma^* := (1, 2, 3) \prod_{i=2}^{\frac{n-1}{2}} (2i, 2i+1),$$

and observe that $c(\sigma^*) = (n-1)/2$ and σ^* can be decomposed in $(n+1)/2$ transpositions. Now, if n is even and $\text{sgn}(\sigma^*) = 1$ or if n is odd and $\text{sgn}(\sigma^*) = -1$, we define:

$$\pi^* = (1, 2, \dots, n-3)(n-2, n-1, n),$$

and note that $\text{sgn}(\pi^*) = \text{sgn}(\sigma^*)$ since it can be decomposed into $n-2$ transpositions. If instead n is even and $\text{sgn}(\sigma^*) = -1$ or if n is odd and $\text{sgn}(\sigma^*) = 1$, we define:

$$\pi^* = (1, 2, \dots, n),$$

and note that $\text{sgn}(\pi^*) = \text{sgn}(\sigma^*)$ since it can be decomposed into $n-1$ transpositions.

Observe that, since $n \geq 6$:

$$\text{S}^{\text{C}}(\text{id}, \sigma^*) \geq \frac{(n-1)/2}{n} \geq \frac{1}{3} \quad \text{and} \quad \text{S}^{\text{C}}(\text{id}, \pi^*) \leq \frac{2}{n}.$$

Since \mathcal{H} Δ -approximates $(\mathbb{S}_n, \text{S}^{\text{C}})$, we have:

$$P_{\mathcal{H}}(\text{id}, \pi^*) \leq \text{S}^{\text{C}}(\text{id}, \pi^*) \leq \frac{2}{n}.$$

Plugging this into (11), and letting C be the constant in the statement of Lemma 30:

$$w_0 + w_{\text{sgn}} - \frac{C}{n} \stackrel{\text{Lemma 30}}{\leq} w_0 + w_{\text{sgn}} \cdot \text{sgn}(\pi^*) + \sum_{\substack{\lambda \vdash n \\ \lambda \neq (n), (1^n)}} w_{\lambda} \frac{\chi_{\lambda}(\pi^*)}{d_{\lambda}} \stackrel{(11)}{=} P_{\mathcal{H}}(\text{id}, \pi^*) \leq \frac{2}{n},$$

giving:

$$w_0 + w_{\text{sgn}} \leq \frac{2+C}{n}. \tag{12}$$

Again applying Equation (11) and Lemma 30, we have:

$$P_{\mathcal{H}}(\text{id}, \sigma^*) \stackrel{(11), \text{Lemma 30}}{\leq} w_0 + w_{\text{sgn}} + \frac{C}{n} \stackrel{(12)}{\leq} \frac{2+2C}{n}.$$

On the other hand:

$$P_{\mathcal{H}}(\text{id}, \sigma^*) \geq \frac{1}{\Delta} \text{S}^{\text{C}}(\text{id}, \sigma^*) = \frac{1}{3\Delta}.$$

Combining the last two equations yields:

$$\Delta \geq \frac{n}{6(1+C)},$$

completing the proof. □

8 Conclusions and Open Questions

Our work shows that the Ulam similarity enjoys sublinear LSH distortion whereas the Cayley similarity does not. An obvious open problem is to close the glaring yet tantalizing gap between the upper bound $(n/\sqrt{\log n})$ and the lower bound $(\Omega(n^{0.12}))$ for the Ulam similarity; our current belief is that the correct LSH distortion bound is $\tilde{\Theta}(\sqrt{n})$. Improving the base case (Proposition 16) is a natural direction, though there are computational limitations to finding a larger and better base instance. While it seems possible to improve our upper bound to $\tilde{O}(n/\log n)$, going beyond this to $\tilde{O}(n^{1-\Omega(1)})$ might require an entirely new approach. Extending our results to other permutation similarities such as reversal is another interesting direction; our current lower bound techniques seem inadequate for the block reversal similarity, for which we conjecture an $\Omega(n)$ lower bound.

Acknowledgments

The authors wish to thank Alessandro Panconesi for providing useful ideas for the project.

The Ulam upper bound was obtained after weeks of interaction with and help from an LLM; the LLM-generated proof was hand-verified and somewhat simplified. The solver for the base case of the Ulam lower bound was written by an LLM; the instance itself was hand-verified. The high-level approach of using representation theory, in particular, Roichman’s Theorem and derangements, to obtain the Cayley lower bound was suggested by an LLM (after the initial attempts to extend Ulam lower bound proof approach to Cayley failed). All the final proofs were written and verified by the authors.

References

- Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *CACM*, 51(1):117–122, 2008.
- Alexandr Andoni and Robert Krauthgamer. The computational hardness of estimating edit distance. *SICOMP*, 39(6):2398–2429, 2010.
- Alexandr Andoni, Piotr Indyk, and Robert Krauthgamer. Overcoming the l_1 non-embeddability barrier: algorithms for product metrics. In *SODA*, pages 865–874, 2009.
- Andrei Z Broder. On the resemblance and containment of documents. In *SEQUENCES*, pages 21–29, 1997.
- Andrei Z. Broder, Moses Charikar, Alan M. Frieze, and Michael Mitzenmacher. Min-wise independent permutations. *JCSS*, 60(3):630–659, 2000.
- Moses Charikar. Similarity estimation techniques from rounding algorithms. In *STOC*, pages 380–388, 2002.
- Moses Charikar and Robert Krauthgamer. Embedding the Ulam metric into l_1 . *Theory Comput.*, 2(11):207–224, 2006.
- Moses Charikar, Ofir Geri, Michael P. Kim, and William Kuszmaul. On estimating edit distance: Alignment, dimension reduction, and embeddings. In *ICALP*, pages 34:1–34:14, 2018.
- Flavio Chierichetti, Ravi Kumar, Alessandro Panconesi, and Erisa Terolli. On the distortion of locality sensitive hashing. *SICOMP*, 48(2):350–372, 2019.

- Persi Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, 1988.
- Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Similarity search in high dimensions via hashing. In *VLDB*, pages 518–529, 1999.
- Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC*, pages 604–613, 1998.
- Piotr Indyk, Rajeev Motwani, Prabhakar Raghavan, and Santosh S. Vempala. Locality-preserving hashing in multidimensional spaces. In *STOC*, pages 618–625, 1997.
- Imre Risi Kondor. *Group Theoretical Methods in Machine Learning*. Columbia University, 2008.
- Michael Larsen and Aner Shalev. Characters of symmetric groups: sharp bounds and applications. *Inventiones mathematicae*, 174(3):645–687, 2008.
- Qin Lv, William Josephson, Zhe Wang, Moses Charikar, and Kai Li. Multi-probe LSH: efficient indexing for high-dimensional similarity search. In *VLDB*, pages 950–961, 2007.
- Paul Renteln. A combinatorial proof of a symmetric group character involution. *The American Mathematical Monthly*, 130(9):855–858, 2023.
- Yuval Roichman. Upper bound on the characters of the symmetric groups. *Invent. Math.*, 125:451–485, 1996.
- Bruce Sagan. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*. Springer Science & Business Media, 2001.
- Venu Satuluri and Srinivasan Parthasarathy. Bayesian locality sensitive hashing for fast similarity search. *Proc. VLDB Endow.*, 5(5):430–441, 2012.

A Background: The Symmetric Group and its Representations

In this section, we review some basic results in the representation theory of symmetric groups. This is not intended to be a comprehensive introduction to the subject, and we refer the reader to some references (e.g., Sagan (2001); Kondor (2008)) for a more detailed treatment. In the discussion that follows every group is assumed to be finite.

Group Theory of \mathbb{S}_n . Two elements x, y of a group G are conjugate to each other if there exists some element $z \in G$ such that $zxz^{-1} = y$. This defines an equivalence relation and the corresponding classes are called *conjugacy classes* of G . A *class function* is a function that is constant on each equivalence class. \mathbb{S}_n is the group of permutations of $[n]$ under composition. Any permutation in \mathbb{S}_n can be written as the composition of disjoint cycles in a way that is unique up to changing their order. This representation is known as the *cycle decomposition* of the permutation. The action of an element σ by conjugation on an element π simply re-labels the elements of $[n]$ according to σ . A key implication of this fact, is that the conjugacy classes of \mathbb{S}_n are labeled by the distinct shapes of cycle decompositions (the *cycle types*), and are in one-to-one correspondence with the partitions of n . It is also known that any permutation can be written as the product of transpositions: permutations that simply swap two elements i and j . While this representation is not unique, the number of transpositions in each of these factorizations always has the same parity. This allows us to define the sign $\text{sgn}(\pi)$ of a permutation π , as 1 if the permutation can be written as the product of an even number of transpositions, and -1 otherwise. Given a permutation $\pi \in \mathbb{S}_n$, we denote by $\text{fix}(\pi)$ the number of fixed points of π , and by $\text{cycle}_2(\pi)$ the number of cycles of length 2 in the cycle decomposition of π .

Representations of Finite Groups. A matrix representation of a finite group G is a homomorphism $\rho : G \rightarrow \text{GL}_d$ where GL_d is the group of $d \times d$ invertible matrices in $\mathbb{C}^{d \times d}$. Given a complex vector space V of dimension d , we let $GL(V)$ be the set of invertible linear transformation on V . Clearly, $GL(V)$ and GL_d are isomorphic, and a representation can be then thought of as a homomorphism of G into $GL(V)$. Under this equivalence, a representation is equivalent to a G -Module: a vector space V on which G acts linearly, in a way that preserves the group operation. A subrepresentation is a subspace (/submodule) $W \subseteq V$ with the property that $\rho(G) \cdot W \subseteq W$, i.e., W is closed under the linear action of elements of G . A representation V is irreducible if it contains no non-trivial subrepresentation, i.e. no subrepresentation other than $\{0\}$ and V itself.

Two representations $\rho : G \rightarrow \text{GL}_d$ and $\theta : G \rightarrow \text{GL}_d$ are equivalent (or isomorphic) if there exists some invertible matrix T with the property that $T \cdot \theta(g) = \rho(g) \cdot T$ for every $g \in G$. The character χ of a representation ρ is the operator $\chi : G \rightarrow \mathbb{C}$ given by $\chi(g) = \text{tr}(\rho(g))$. A consequence of these definitions, is that for any representation $\rho : G \rightarrow \text{GL}_d$, its character χ satisfies $\chi(e_G) = d$, where e_G is the identity element in G . A simple argument shows that characters are class functions, and, in fact, it is known that the characters of the non-equivalent irreducible representations of a finite group G form a basis for the space of class functions. A consequence of this fact is that the number of such characters is the same as the number of conjugacy classes of G .

Young Diagrams, Tableaux and Tabloids. A *Young diagram* is a visual, geometric way to represent an integer partition. Given a partition $\lambda = (\lambda_1, \dots, \lambda_r)$, a Young diagram of shape λ is a shape containing r left-aligned rows of congruent squares, where the i^{th} row contains λ_i squares. The reader might find the following example, showing the Young diagram for $\lambda = (5, 5, 2, 2, 1)$, more illuminating than the definition:

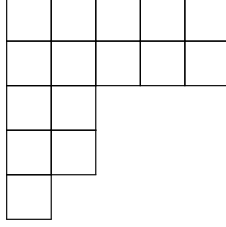


Figure 1: The Young diagram for the partition $\lambda = (5, 5, 2, 2, 1)$.

A *Young tableau* for a partition $\lambda \vdash n$, sometimes called a λ -tableau or a tableau of shape λ , is a Young diagram for λ in which every square has been filled by a distinct number in $[n]$.

11	8	3	10	14
13	7	2	9	
6	4			
12	1			
5				

Figure 2: A Young tableau for the partition $\lambda = (5, 4, 2, 2, 1)$.

If $\lambda \vdash n$, then \mathbb{S}_n has a natural action on the set of λ -tableaux: given a tableau T the tableau $\pi \cdot T$ is obtained by replacing each number i with $\pi(i)$.

A *Young tabloid* of shape λ is an equivalence class of λ -tableaux under the equivalence relation in which two tableaux are equivalent if you can obtain one from the other by permuting the content of each row. These are typically depicted by removing the vertical lines from one of the tableaux in the class. For example, a tabloid of shape $\lambda = (2, 2, 1) \vdash n$ is:

$$\overline{\begin{array}{|c|c|} \hline 1 & 3 \\ \hline 5 & 4 \\ \hline 2 & \\ \hline \end{array}} = \left\{ \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 4 & 5 \\ \hline 2 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 4 & 5 \\ \hline 2 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 1 & 3 \\ \hline 5 & 4 \\ \hline 2 & \\ \hline \end{array}, \begin{array}{|c|c|} \hline 3 & 1 \\ \hline 5 & 4 \\ \hline 2 & \\ \hline \end{array} \right\}$$

The group \mathbb{S}_n acts naturally on tabloids too. The action of a permutation π on a tabloid is found by choosing an arbitrary tableau t in the tabloid, taking the action of the permutation of that tableau to find a new tableau $t' = \pi \cdot t$, and finally returning the tabloid containing t' . The reader may want to verify that this operation is indeed well-defined, and does not depend on the choice of the representative t . Given a partition $\lambda \vdash n$, the permutation module M^λ corresponding to λ is the vector space $\mathbb{C}\{\{t_1\}, \dots, \{t_r\}\}$ of formal complex linear combinations of the complete set of distinct tabloids $\{t_1\} \dots \{t_r\}$ of shape λ . The action of \mathbb{S}_n on tabloids of shape $\lambda \vdash n$ extends linearly to M^λ , giving a representation of \mathbb{S}_n .

A *generalized semistandard Young tableau* of shape $\lambda = (\lambda_1, \dots, \lambda_r) \vdash n$ and content $\mu = (\mu_1, \dots, \mu_\ell)$ where μ is a composition of n , is a Young diagram of shape λ in which each square is filled with a number, satisfying the following properties:

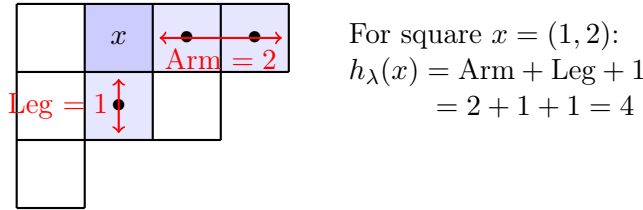
- (a) For each $i \in [\ell]$, the number i appears μ_i times in the diagram,
- (b) The numbers in each row are weakly increasing,

(c) The numbers in each columns are strictly increasing.

The number of generalized semistandard Young tableaux of shape λ and content μ is the so-called *Kostka number* $K_{\lambda\mu}$.

Representations of \mathbb{S}_n . Non-equivalent, irreducible representations of \mathbb{S}_n are in one-to-one correspondence with partitions $\lambda \vdash n$. Each of these λ corresponds to the representation S^λ , sometimes called the λ -Specht module, and its character is denoted by χ_λ . (The construction of these modules is beyond the scope of this paper, and the curious reader is advised to refer to Sagan (2001).) Under this correspondence, $\lambda = (n)$ corresponds to the trivial representation, in which every group element gets mapped to the identity matrix, and $\chi_{(n)}(\pi) = 1$ for every $\pi \in \mathbb{S}_n$. On the other hand, the partition (1^n) corresponds to the sign representation, and satisfies $\chi_{(1^n)}(\pi) = \text{sgn}(\pi)$ for every $\pi \in \mathbb{S}_n$. It is known that (see e.g. Renteln (2023)), for any permutation $\sigma \in \mathbb{S}_n$ and any partition λ : $|\chi_\lambda(\sigma)| = |\chi_{\lambda'}(\sigma)|$ (in fact, the values of $\chi_\lambda(\sigma)$ and $\chi_{\lambda'}(\sigma)$ are related by a factor of $\text{sgn}(\sigma)$, but we will not need this stronger fact). Moreover $d_\lambda = d_{\lambda'}$.

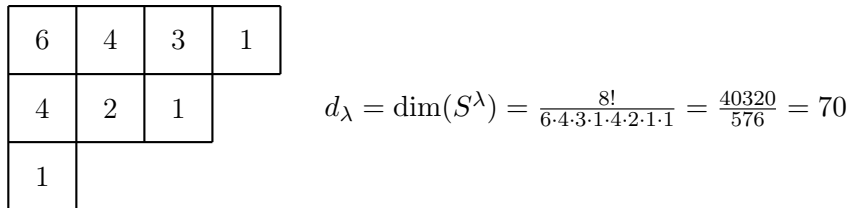
Given the Young diagram for a $\lambda = (\lambda_1, \dots, \lambda_k)$, the *arm* of a square (i, j) in the diagram, is the collection of squares on the same row that are to the right of (i, j) , while the *leg* of (i, j) is the collection of squares on the same column that are below it. The hook length of (i, j) is the value $h_\lambda(i, j)$ given by the sum of the cardinalities of the arm and the leg, plus one. Consider the following example for the partition $\lambda = (4, 3, 1)$:



The dimension d_λ of the irreducible representation S^λ is given by the so-called *hook length* formula:

$$d_\lambda := \frac{n!}{\prod_{i,j} h_\lambda(i, j)}.$$

In the following Young diagram for the partition $\lambda = (4, 3, 1) \vdash 8$, we have placed the value of $h_\lambda(i, j)$ in each square (i, j) . We can then compute the dimension d_λ by dividing $8!$ by the product of all these numbers:



We apply the same procedure in the following examples, to calculate the dimension of some irreducible representations that are used in the body of the paper.

Example 31. For any $n \geq 4$ we can compute the dimension of various irreducible representation. For $\lambda = (n)$

n	$n-1$	$n-2$	\cdots	1
-----	-------	-------	----------	-----

$$d_{(n)} = \frac{n!}{n!} = 1 \tag{13}$$

For $\lambda = (n-1, 1)$:

n	$n-2$	$n-3$	\cdots	1
1				

$$d_{(n-1,1)} = \frac{n!}{(n-2)! \cdot n} = n-1 \tag{14}$$

For $\lambda = (n-2, 1, 1)$:

n	$n-3$	$n-4$	\cdots	1
2				
1				

$$d_{(n-2,1,1)} = \frac{n!}{2 \cdot (n-3)! \cdot n} = \frac{(n-1)(n-2)}{2}, \tag{15}$$

For $\lambda = (n-2, 2)$:

$n-1$	$n-2$	$n-4$	\cdots	1
2	1			

$$d_{(n-2,2)} = \frac{n!}{2 \cdot (n-4)! \cdot (n-2) \cdot (n-1)} = \frac{n(n-3)}{2}, \tag{16}$$

For any partition $\lambda = (\lambda_1, \dots, \lambda_r) \vdash n$, the dimension of the permutation module M^λ is the number of tabloids of shape λ , i.e.:

$$\dim M^\lambda = \frac{n!}{\prod_{i \in [r]} \lambda_i!}.$$

The group \mathbb{S}_n acts on M^λ by permuting its basis vectors, and hence, in the standard basis, the actions of the elements of \mathbb{S}_n are all represented by permutation matrices. As a result, for every $\pi \in \mathbb{S}_n$ the character $\chi_{M^\lambda}(\pi)$ equals the number of tabloids of shape λ fixed by the action of π . From this, we can infer, for example, that for every $\pi \in \mathbb{S}_n$:

$$\chi_{M^{(n-2,2)}}(\pi) = \text{cycle}_2(\pi) + \binom{\text{fix}(\pi)}{2}, \tag{17}$$

$$\text{and } \chi_{M^{(n-2,1,1)}}(\pi) = \text{fix}(\pi)(\text{fix}(\pi) - 1). \quad (18)$$

By Maschke's Theorem, each representation M^μ , decomposes as the direct sum of irreducible representations. The number of times an irreducible representation S^λ appears in this sum is given by $K_{\lambda\mu}$. We illustrate the usefulness of this fact in the following example.

Example 32. A simple combinatorial argument shows that:

$$K_{(n)(n-2,2)} = K_{(n-1,1)(n-2,2)} = K_{(n-2,2)(n-2,2)} = 1.$$

As a consequence of this fact, we can see that the representations $S^{(n-2,2)}$, $S^{(n-1,1)}$ and $S^{(n)}$ each appear exactly once inside of $M^{(n-2,2)}$. Since:

$$\dim M^{(n-2,2)} = \binom{n}{2} = d_{(n-2,2)} + d_{(n-1,1)} + d_{(n)},$$

we find that:

$$M^{(n-2,2)} \cong S^{(n-2,2)} \oplus S^{(n-1,1)} \oplus S^{(n)}$$

and hence:

$$\chi_{M^{(n-2,2)}} = \chi_{(n-2,2)} + \chi_{(n-1,1)} + \chi_{(n)}. \quad (19)$$

Similarly, we have:

$$K_{(n)(n-2,1,1)} = K_{(n-2,1,1)(n-2,1,1)} = K_{(n-2,2)(n-2,1,1)} = 1,$$

and:

$$K_{(n-1,1)(n-2,1,1)} = 2.$$

This shows that each of the representations $S^{(n)}$, $S^{(n-2,1,1)}$ and $S^{(n-2,2)}$ appears exactly once within $M^{(n-2,1,1)}$ while $S^{(n-1,1)}$ appears twice. Since:

$$\dim M^{(n-2,1,1)} = n(n-1) = d_{(n)} + d_{(n-2,1,1)} + d_{(n-2,2)} + 2d_{(n-1,1)}$$

we get that:

$$M^{(n-2,1,1)} \cong S^{(n)} \oplus S^{(n-2,1,1)} \oplus S^{(n-2,2)} \oplus S^{(n-1,1)} \oplus S^{(n-1,1)},$$

giving:

$$\chi_{M^{(n-2,1,1)}} = \chi_{(n)} + \chi_{(n-2,1,1)} + \chi_{(n-2,2)} + 2\chi_{(n-1,1)}. \quad (20)$$