ISIT'2013 1569719047

Sequence Reconstruction for Grassmann Graphs and Permutations

Eitan Yaakobi
Electrical Engineering
Caltech
yaakobi@caltech.edu

Moshe Schwartz
Electrical and Computer Engineering
Ben-Gurion University of the Negev
schwartz@ee.bgu.ac.il

Michael Langberg
Mathematics and Computer Science
Open University of Israel
mikel@openu.ac.il

Jehoshua Bruck
Electrical Engineering
Caltech
bruck@caltech.edu

***Abstract*—The sequence-reconstruction problem was first proposed by Levenshtein in 2001. This problem studies the model where the same word is transmitted over multiple channels. If the transmitted word belongs to some code of minimum distance d and there are at most r errors in every channel, then the minimum number of channels that guarantees a successful decoder (under the assumption that all channel outputs are distinct) has to be greater than the largest intersection of two balls of radius r and with distance at least d between their centers.**

This paper studies the combinatorial problem of computing the largest intersection of two balls for two cases. In the first part we solve this problem in the Grassmann graph for all values of d and r. In the second part we derive similar results for permutations under Kendall's τ-metric for some special cases of d and r.

I. Introduction

The *sequence-reconstruction problem* was first proposed by Levenshtein in [10], [11]. In this setup, a codeword is transmitted through multiple channels and the decoder's task is to decode the transmitted codeword once it receives all the output transmissions. Levenshtein studied the minimum number of transmission channels which guarantees the existence of a successful decoder, under the assumption that all channel outputs are distinct. More specifically, assume that all words belong to some space V. Let ρ be a distance metric over the words in V. Assume the transmitted word $\boldsymbol{c}$ belongs to some code of minimum distance d and on every channel there are at most r errors. Then the number of channels has to be greater than the largest intersection of two balls where the distance between their centers is at least d,

$$N(d,r) = \max_{\substack{\boldsymbol{x}_1,\boldsymbol{x}_2 \in V \\ \rho(\boldsymbol{x}_1,\boldsymbol{x}_2) \geqslant d}} \left\{ |B_r(\boldsymbol{x}_1) \cap B_r(\boldsymbol{x}_2)| \right\}, \tag{1}$$

where $B_r(\boldsymbol{x})$ is the ball of radius r surrounding $\boldsymbol{x}$, $B_r(\boldsymbol{x}) = \{\boldsymbol{y} \in V \mid \rho(\boldsymbol{x},\boldsymbol{y}) \leqslant r\}$. We refer to the problem of finding $N(d,r)$ as the *reconstruction problem*.

Levenshtein studied the reconstruction problem for the Hamming graph, the Johnson graph, and the case of deletions and insertions. More results on the latter case were given in [12] and reconstruction algorithms for this model were presented in [1], [4], [18]. The deletion-only case was explored in the context of trace reconstruction in [2]. The information-theoretic study of a special model of deletions, applied for DNA sequences, was studied in [15], [16]. In [6]–[8], this problem was analyzed over permutations, and in [13], [14] for error graphs. Recently, this problem was extended in the context of associative memories [19]. In this setup, the largest intersection of multiple balls was studied, where the distance between all the centers is at least some prescribed parameter. In the reconstruction model, this problem is equivalent to the required number of sequences in order to output a list of some L words which contains the transmitted word.

The reconstruction problem over permutations has received considerable attention. In particular, in [6], [7] the case of permutations with reversal errors was studied and in [8], [14] permutations with transpositions errors were investigated. In [8], Kendall's τ-distance for permutations was briefly studied for some very special cases of $d = 1$ and $r = 1, 2$.

In this work, we study the reconstruction problem for two special cases. In the first part we find the value of $N(d,r)$ in the Grassmann graph for all values of d and r. The second part is dedicated to derive similar results for permutations with Kendall's τ-distance. We give lower bounds on the value of $N(d,r)$ for $d = 1, 2$ and find the exact value when $d = 2r$ and $r \leqslant n/4$.

The rest of the paper is organized as follows. In Section II, we formally define the problem and state some special properties which we use later on in the paper. In Section III, we solve the problem for the Grassmann graph. Section IV studies the reconstruction problem for permutations with Kendall's τ-distance.

II. Definitions and Basic Properties

We follow the definitions as presented in [11] and are summarized as follows. Let $\mathcal{G} = \{V, E\}$ be an undirected graph with a finite set V of vertices and E is its set of edges. The path metric of $\mathcal{G}$ is defined as a function $\rho : V \times V \to \mathbb{Z}^+ \cup \{\infty\}$ such that for every $u, v \in V$, $\rho(u,v)$ is the minimum number of edges in a path connecting u and v, and $\rho(u,v) = \infty$ in case such a path does not exist. For every $v \in V$ the ball of radius r centered in v is defined as

$$B_r(v) = \{u \in V \mid \rho(u,v) \leqslant r\},$$

and the sphere of radius r centered in v is similarly defined to be

$$S_r(v) = \{u \in V \mid \rho(u,v) = r\}.$$

Given two integers d and r, let $I(\mathcal{G}; d, r)$ be the size of the largest intersection of two balls of radius r and distance d

This work was supported in part by NSF grant ECCS-0801795, BSF grant 2010075, and ISF grant 480/08. This work was done while Michael Langberg was at the California Institute of Technology.

between their centers. That is,

$$I(\mathcal{G};d,r) = \max_{\substack{u,v\in V\\ \rho(u,v)=d}} |B_r(u)\cap B_r(v)|.$$

Let $N(\mathcal{G};d,r)$ be the size of the maximum intersection of two balls of radius r and distance at least d between their centers,

$$N(\mathcal{G};d,r) = \max_{\ell\geqslant d} I(\mathcal{G};\ell,r).$$

A code $\mathcal{C}$ with minimum distance d is a subset $\mathcal{C}\subseteq V$ such that $u,v\in\mathcal{C}$ and $u\neq v$ imply $\rho(u,v)\geqslant d$. The elements of $\mathcal{C}$ are called *codewords*. As is usually the case, we may transmit a codeword $\boldsymbol{c}\in\mathcal{C}$ over a channel and receive a corrupted version of it $\boldsymbol{y}\in\mathcal{G}$. We say r errors occurred during the process if $\rho(\boldsymbol{c},\boldsymbol{y})=r$.

Assume $\mathcal{C}$ is a code in $\mathcal{G}$ with minimum distance d and a codeword $\boldsymbol{c}\in\mathcal{C}$ is transmitted over N channels, where on each channel there are at most r errors and all channel outputs are different from each other. Then, it was shown by Levenshtein [11] that the minimum number of channels that guarantees the existence of a decoder that will successfully decode any transmitted codeword from $\mathcal{C}$ is given by $N(\mathcal{G};d,r)+1$.

Note that in general the value of $N(\mathcal{G};d,r)$ is not necessarily achieved for $\ell=d$, that is, it may happen that $I(\mathcal{G};d,r)<I(\mathcal{G};d',r)$ when $d<d'$. (See for example the case of permutations with transpositions errors for $d=1$ and $d=2$ [14].) Levenshtein gave sufficient conditions, stated in Lemma 6 in [11], for such a scenario to not happen. A graph $\mathcal{G}$ is called *monotone on intersections* if for any d and r, $I(\mathcal{G};d,r)\geqslant I(\mathcal{G};d+1,r)$. In this case, the following holds

$$N(\mathcal{G};d,r) = I(\mathcal{G};d,r) = \max_{\substack{u,v\in V\\ \rho(u,v)=d}} |B_r(u)\cap B_r(v)|.$$

Before we move on to the next sections, where we solve the reconstruction problem, let us start with a useful lemma for the case $d=1$.

Lemma 1. *Assume $u,v\in V$ and $\rho(u,v)=1$, then*

$$B_r(u)\cap B_r(v) = (B_{r-1}(u)\cup B_{r-1}(v))\cup(S_r(u)\cap S_r(v)).$$

Proof:

$\subseteq$: let $w\in B_r(u)\cap B_r(v)$, then $\rho(w,u),\rho(w,v)\leqslant r$. If $\rho(w,u)\leqslant r-1$ or $\rho(w,v)\leqslant r-1$ then $w\in B_{r-1}(u)\cup B_{r-1}(v)$. Otherwise $\rho(w,u)=\rho(w,v)=r$ so we get $w\in S_r(u)\cap S_r(v)$.

$\supseteq$: if $w\in B_{r-1}(u)$ then $\rho(w,u)\leqslant r-1$ and $\rho(w,v)\leqslant\rho(w,u)+\rho(u,v)\leqslant r$, so $B_{r-1}(u)\subseteq B_r(u)\cap B_r(v)$. Similarly we get $B_{r-1}(v)\subseteq B_r(u)\cap B_r(v)$. It is also clear that $S_r(u)\cap S_r(v)\subseteq B_r(u)\cap B_r(v)$. ∎

III. The Reconstruction Problem over the Grassmann Graph

In this section we study the reconstruction problem over the Grassmann graph. Let V be a vector space of dimension n over $\mathrm{GF}(q)$. For any integer $0\leqslant k\leqslant n$, we denote by $\begin{bmatrix}V\\k\end{bmatrix}$ the set of all k-dimensional subspaces of V. The q-number of k is defined as

$$[k]_q = 1+q+q^2+\cdots+q^{k-1} = \frac{q^k-1}{q-1}.$$

By abuse of notation we denote

$$[k]_q! = [k]_q[k-1]_q\ldots[1]_q.$$

The *Gaussian coefficient* is defined for n, k, and q as

$$\begin{bmatrix}n\\k\end{bmatrix}_q = \frac{[n]_q!}{[k]_q![n-k]_q!} = \frac{(q^n-1)(q^{n-1}-1)\ldots(q^{n-k+1}-1)}{(q^k-1)(q^{k-1}-1)\ldots(q-1)}.$$

It is well known that the number of k-dimensional subspaces of an n-dimensional space over $\mathrm{GF}(q)$ is given by $\begin{bmatrix}n\\k\end{bmatrix}_q$. In a more general form, the number of k'-dimensional subspaces of V which intersect a given k-dimensional subspace of V in an i-dimensional subspace is given by

$$q^{(k'-i)(k-i)}\begin{bmatrix}n-k\\k'-i\end{bmatrix}_q\begin{bmatrix}k\\i\end{bmatrix}_q.$$

The *Grassmann graph*, $\mathcal{G}_q(n,k)$ is defined by the vertex set $\begin{bmatrix}V\\k\end{bmatrix}$, and two vertices are connected by an undirected edge if their corresponding k-dimensional subspaces intersect in a $(k-1)$-dimensional subspace. This graph induces a distance measure, denoted $d(U,U')$, which is simply the distance between U and U' in $\mathcal{G}_q(n,k)$. The distance $d(U,U')$ can also be expressed as $k-\dim(U\cap U')$.

Given a vertex $U\in\mathcal{G}_q(n,k)$, the *ball of radius r* around U is defined as

$$\mathcal{B}_r(U) = \left\{U'\in\mathcal{G}_q(n,k)\mid d(U,U')\leqslant r\right\},$$

and the *sphere of radius r* around U is defined as

$$\mathcal{S}_r(U) = \left\{U'\in\mathcal{G}_q(n,k)\mid d(U,U')=r\right\}.$$

The size of a ball and a sphere does not depend on the choice of center, and we denote them, respectively, as

$$b_r = \sum_{i=0}^{r} q^{i^2}\begin{bmatrix}n-k\\i\end{bmatrix}_q\begin{bmatrix}k\\i\end{bmatrix}_q,$$
$$s_r = q^{r^2}\begin{bmatrix}n-k\\r\end{bmatrix}_q\begin{bmatrix}k\\r\end{bmatrix}_q.$$

We now move towards calculating the value of $I(\mathcal{G}_q(n,k);d,r)$ for all values of d and r. We start in the next lemma.

Lemma 2. *Let $n\geqslant k\geqslant d\geqslant 1$ be integers, and let $U_1,U_2\in\mathcal{G}_q(n,k)$ be two k-dimensional subspaces such that $d(U_1,U_2)=d$. Let i^*, i_*, i_1, and i_2, be non-negative integers. Then the number of subspaces W of dimension k that satisfy:*

1) $\dim(W\cap U_1)=i_1$
2) $\dim(W\cap U_2)=i_2$
3) $\dim(W\cap(U_1+U_2))=i^*$
4) $\dim(W\cap(U_1\cap U_2))=i_*$

is given by

$$\begin{aligned}\xi(i_*,i_1,i_2,i^*) &= q^{(k-d-i_*)(i^*-i_*)+(k+d-i^*)(k-i^*)+\binom{i^*-i_1-i_2+i_*}{2}} \\ &\cdot (q-1)^{i^*-i_1-i_2+i_*}[i^*-i_1-i_2+i_*]_q! \\ &\cdot \begin{bmatrix} k-d \\ i_* \end{bmatrix}_q \begin{bmatrix} d \\ i_1-i_* \end{bmatrix}_q \begin{bmatrix} d \\ i_2-i_* \end{bmatrix}_q \begin{bmatrix} n-k-d \\ k-i^* \end{bmatrix}_q \\ &\cdot \begin{bmatrix} d-i_1+i_* \\ i^*-i_1-i_2+i_* \end{bmatrix}_q \begin{bmatrix} d-i_2+i_* \\ i^*-i_1-i_2+i_* \end{bmatrix}_q\end{aligned}$$

Proof: For convenience, let us denote

$$A = U_1 \cap U_2 \qquad B_1 = U_1 \setminus U_2 \qquad B_2 = U_2 \setminus U_1$$

We note that A is a $(k-d)$-dimensional subspace, U_1+U_2 is a $(k+d)$-dimensional subspace, and $|B_1| = |B_2| = q^k - q^{k-d}$. We shall now count the number of subspaces W satisfying the above constraints.

We start choosing a basis for W by picking i_* linearly-independent vectors from A. The number of ways of doing so is

$$P_* = \prod_{j=0}^{i_*-1}(q^{k-d}-q^j).$$

We now need to extend this basis by picking $i_1 - i_*$ basis vectors from B_1 so that $\dim(W \cap U_1) = i_1$. This may be done in

$$P_1 = \prod_{j=0}^{i_1-i_*-1}(q^k - q^{k-d+j})$$

ways. Similarly, we extend the basis by choosing vectors from B_2 in

$$P_2 = \prod_{j=0}^{i_2-i_*-1}(q^k - q^{k-d+j})$$

ways.

We continue by choosing $i^* - i_1 - i_2 + i_*$ basis vectors from U_1+U_2, while being careful not to change the previous intersections. An analysis we omit due to space limitation gives the number of ways of doing so as

$$P^* = q^{(k-d)(i^*-i_1-i_2+i_*)} \cdot \prod_{j=0}^{i^*-i_1-i_2+i_*-1}(q^d - q^{i_1-i_*+j})(q^d - q^{i_2-i_*+j}).$$

Finally, we need to choose $k-i^*$ basis vectors from outside of U_1+U_2. This may be easily done in

$$P' = \prod_{j=0}^{k-i^*-1}(q^n - q^{k+d+j})$$

ways. Thus, the total number of bases is given by $P = P_*P_1P_2P^*P'$.

We are interested in counting subspaces, and the counting done so far is obviously an over-counting since the order of elements in a basis is irrelevant, and several bases may produce the same subspace. We shall correct by dividing by the over-counting factor. The over-counting factor for the part of the basis that is in A is given by

$$Q_* = \prod_{j=0}^{i_*-1}(q^{i_*} - q^j),$$

since every i_* dimensional subspace of A contains that many ordered bases. In a similar manner we have:

$$\begin{aligned} Q_1 &= \prod_{j=0}^{i_1-i_*-1}(q^{i_1} - q^{i_*+j}) \\ Q_2 &= \prod_{j=0}^{i_2-i_*-1}(q^{i_2} - q^{i_*+j}) \\ Q^* &= \prod_{j=0}^{i^*-i_1-i_2+i_*-1}(q^{i^*} - q^{i_1+i_2-i_*+j}) \\ Q' &= \prod_{j=0}^{k-i^*-1}(q^k - q^{i^*+j}) \end{aligned}$$

and we define $Q = Q_*Q_1Q_2Q^*Q'$. A tedious rearrangement of P/Q gives the desired result. ∎

According to the last lemma, we are now ready to find the value of $I(\mathcal{G}_q(n,k);d,r)$ for all d and r.

Theorem 3. *Let $n \geqslant k \geqslant d \geqslant 1$ be integers, then*

$$I(\mathcal{G}_q(n,k);d,r) = \sum_{\substack{i_*,i^* \\ 0\leqslant r_1,r_2\leqslant r}} \xi(i_*,k-r_1,k-r_2,i^*).$$

We note that it is possible to show using the conditions in Lemma 6 from [12] that the graph $\mathcal{G}_q(n,k)$ is monotone on intersections. Therefore, we conclude that $N(\mathcal{G}_q(n,k);d,r) = I(\mathcal{G}_q(n,k);d,r)$ for all d and r.

IV. The Reconstruction Problem for Permutations with Kendall's τ-Metric

Let $\mathbb{S}_n$ denote the set of all $n!$ permutations of n elements, conveniently chosen to be $\{1,2,\ldots,n\}$. We shall use the vector notation to denote a permutation $\pi = [\pi_1,\pi_2,\ldots,\pi_n]$ mapping $i \mapsto \pi_i$. An *adjacent transposition* of a permutation $\sigma \in \mathbb{S}_n$ is the local exchange of two adjacent elements in σ. For every two permutations $\sigma,\pi \in \mathbb{S}_n$, $d_\tau(\sigma,\pi)$ denotes the minimal number of adjacent transpositions needed to change σ into π. This distance measure induces a metric over $\mathbb{S}_n$ and is called Kendall's τ-metric in statistics [5] or the bubble-sort distance. The graph $\mathcal{G}_\tau(\mathbb{S}_n)$ with the vertex set $\mathbb{S}_n$ is constructed such that its set of edges are all pairs of permutations with Kendall's τ-distance one. Kendall's τ-metric is graphic, i.e., the path metric ρ in $\mathcal{G}_\tau(\mathbb{S}_n)$ equals d_τ.

It is well known that Kendall's τ-distance between every two permutations $\sigma,\pi \in \mathbb{S}_n$ can be expressed as $d_\tau(\sigma,\pi) = |\{(i,j) \mid i<j, (\sigma^{-1}(i)-\sigma^{-1}(j))(\pi^{-1}(i)-\pi^{-1}(j)) < 0\}|$. We let id be the identity permutation, $\mathrm{id} = [1,2,\ldots,n]$. For a permutation σ, we let $W_\tau(\sigma) = \{(i,j) \mid i<j, \sigma^{-1}(i) > \sigma^{-1}(j)\}$, and its weight is defined as $w_\tau(\sigma) = d_\tau(\mathrm{id},\sigma) = |W_\tau(\sigma)|$. It is also known that for any two permutations σ,π, $d_\tau(\sigma,\pi) = |W_\tau(\sigma) \triangle W_\tau(\pi)|$, where for any two sets A,B, $A \triangle B$ is the symmetric difference between A and B, that is, $A \triangle B = (A \setminus B) \cup (B \setminus A)$ and for finite sets $|A \triangle B| = |A| + |B| - 2|A \cap B|$.

For two permutations $\sigma,\pi \in \mathbb{S}_n$, the product $\alpha = \sigma\pi$ is defined according to the identity $\alpha(i) = \sigma(\pi(i))$ for $1 \leqslant i \leqslant n$. Under this definition of product, Kendall's τ-distance is

a left-invariant metric, that is, for all $\sigma, \pi, \beta \in \mathrm{S}_n$, we have $d_\tau(\sigma, \pi) = d_\tau(\beta\sigma, \beta\pi)$. In particular,

$$d_\tau(\sigma, \pi) = d_\tau(\sigma^{-1}\sigma, \sigma^{-1}\pi) = d_\tau(\mathrm{id}, \sigma^{-1}\pi) = w_\tau(\sigma^{-1}\pi).$$

For $1 \leqslant i \leqslant n-1$ the ith adjacent transposition, denoted by ϵ_i, swaps the elements in positions i and $i+1$ and keeps all other elements fixed. That is, $\epsilon_i(i) = i+1, \epsilon_i(i+1) = i$, and for all other values j, $\epsilon_i(j) = j$. The permutation $\sigma\epsilon_i$ is the permutation σ where the elements in positions i and $i+1$ are swapped, while in the permutation $\epsilon_i\sigma$ the elements labeled i and $i+1$ interchange their locations.

The ball of radius r, centered in $\sigma \in \mathrm{S}_n$, is the set $B_r(\sigma) = \{\pi \in \mathrm{S}_n \mid d_\tau(\sigma, \pi) \leqslant r\}$, and its size is denoted by b_r. Note that this value does not depend on the choice of the center σ. Similarly the sphere of radius r is $S_r(\sigma) = \{\pi \in \mathrm{S}_n \mid d_\tau(\sigma, \pi) = r\}$, and its size is s_r.

We seek to study the values of $I(\mathcal{G}_\tau(\mathrm{S}_n); d, r)$ and $N(\mathcal{G}_\tau(\mathrm{S}_n); d, r)$. In contrast to our previous analysis on the Grassmann graph, we were not able to prove (or disprove) for the Kendalls τ-distance that $N(\mathcal{G}_\tau(\mathrm{S}_n); d, r) = I(\mathcal{G}_\tau(\mathrm{S}_n); d, r)$ (i.e., we were not able to prove monotonicity of $\mathcal{G}_\tau(\mathrm{S}_n)$ on intersections). We thus focus our results on the analysis of $I(\mathcal{G}_\tau(\mathrm{S}_n); d, r)$ and in such obtain lower bounds on $N(\mathcal{G}_\tau(\mathrm{S}_n); d, r)$. It was shown in [8] that $N(\mathcal{G}_\tau(\mathrm{S}_n); 1, 1) = 2, \quad N(\mathcal{G}_\tau(\mathrm{S}_n); 1, 2) = 2(n-1)$.

A. The case $d = 1$

In this section we study the case $d = 1$. In order to apply Lemma 1, we first prove the following property.

Lemma 4. *For any σ, π such that $d_\tau(\sigma, \pi) = 1$, we have $S_r(\sigma) \cap S_r(\pi) = \emptyset$ for all $r \geqslant 0$.*

Proof: Due to the left invariance of the metric, without loss of generality, we can assume that $\sigma = \mathrm{id}$. Hence, we can assume that $\pi = \epsilon_\ell$ for some $1 \leqslant \ell \leqslant n-1$, and so $W_\tau(\pi) = \{(\ell, \ell+1)\}$. Let $\alpha \in S_r(\mathrm{id})$ so $w_\tau(\alpha) = r$. If $(\ell, \ell+1) \in W_\tau(\alpha)$ then $d_\tau(\alpha, \pi) = r-1$ and otherwise $d_\tau(\alpha, \pi) = r+1$. In any event $\alpha \notin S_r(\pi)$ and thus $S_r(\sigma) \cap S_r(\pi) = \emptyset$. ∎

We are now ready to show a recursive formula for the value of $I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r)$.

Theorem 5. *For $r \geqslant 2$, the values of $I(\mathrm{S}_n; 1, r)$ satisfy the following recursive formula*

$$I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r) = 2b_{r-1} - I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r-1),$$

where $I(\mathcal{G}_\tau(\mathrm{S}_n); 1, 1) = 2$.

Proof: Let σ, π be two permutations such that $d_\tau(\sigma, \pi) = 1$. According to Lemma 1, we have that

$$B_r(\sigma) \cap B_r(\pi) = (B_{r-1}(\sigma) \cup B_{r-1}(\pi)) \cup (S_r(\sigma) \cap S_r(\pi)),$$

and together with Lemma 4 we get

$$B_r(\sigma) \cap B_r(\pi) = B_{r-1}(\sigma) \cup B_{r-1}(\pi).$$

Therefore we conclude the following recursive formula

$$\begin{aligned} |B_r(\sigma) \cap B_r(\pi)| &= |B_{r-1}(\sigma) \cup B_{r-1}(\pi)| \\ &= |B_{r-1}(\sigma)| + |B_{r-1}(\pi)| - |B_{r-1}(\sigma) \cap B_{r-1}(\pi)|. \end{aligned}$$

If we apply it recursively $r-1$ times we will get

$$\begin{aligned} |B_r(\sigma) \cap B_r(\pi)| &= \sum_{i=1}^{r-2} (-1)^{i-1} (|B_{r-i}(\sigma)| + |B_{r-i}(\pi)|) \\ &\quad + (-1)^{r-1} |B_1(\sigma) \cap B_1(\pi)| \\ &= 2 \sum_{i=1}^{r-2} (-1)^{i-1} b_{r-i} + 2(-1)^{r-1} \end{aligned}$$

This also proves that for any other two permutations $\sigma', \pi' \in \mathrm{S}_n$ such that $d_\tau(\sigma', \pi') = 1$, we have

$$|B_r(\sigma) \cap B_r(\pi)| = |B_r(\sigma') \cap B_r(\pi')| = I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r),$$

and in particular,

$$I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r) = 2b_{r-1} - I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r-1),$$

where $I(\mathcal{G}_\tau(\mathrm{S}_n); 1, 1) = 2$. ∎

Alternatively, it follows that

$$\begin{aligned} I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r) &= 2b_{r-1} - 2b_{r-2} + 2b_{r-3} \cdots \\ &= 2 \sum_{i=1}^{r-1} (-1)^{i-1} b_{r-i} \\ &= \begin{cases} 2 \sum_{i=1}^{\lfloor \frac{r}{2} \rfloor} s_{r+1-2i} & r \text{ is odd,} \\ 2 + 2 \sum_{i=1}^{\lfloor \frac{r}{2} \rfloor} s_{r+1-2i} & r \text{ is even.} \end{cases} \end{aligned}$$

If $f(x)$ is a polynomial in x, the coefficient of x^r in $f(x)$ is denoted by $[x^r]f(x)$. Muir [17] showed that s_r, the size of a sphere or radius r, is given by

$$s_r = [x^r] \prod_{i=1}^{n} (x^{i-1} + x^{i-2} + \cdots + 1) = [x^r] \prod_{i=1}^{n} \frac{1 - x^i}{1 - x}.$$

Thus, the size of a ball of radius r, b_r, is given by $b_r = [x^r] \frac{1}{1-x} \prod_{i=1}^{n} \frac{1-x^i}{1-x}$. Therefore, $I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r)$ is given by

$$I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r) = [x^{r-1}] \frac{2}{1 - x^2} \prod_{i=1}^{n} \frac{1 - x^i}{1 - x}.$$

B. The case $d = 2$

The case of $d = 2$ can be solved using the same methods used for the case of $d = 1$. Specifically, we show that

Theorem 6. $I(\mathcal{G}_\tau(\mathrm{S}_n); 2, r) \geqslant I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r)$.

Assuming monotonicity of $\mathcal{G}_\tau(\mathrm{S}_n)$ on intersections would imply that $N(\mathcal{G}_\tau(\mathrm{S}_n); 2, r) = I(\mathcal{G}_\tau(\mathrm{S}_n); 2, r) = I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r) = N(\mathcal{G}_\tau(\mathrm{S}_n); 1, r)$. However, as stated previously, monotonicity is left open in this work.

Given two permutations σ and π such that $d_\tau(\sigma, \pi) = 2$, there are two options:

1) There is only a single permutation α such that $d_\tau(\sigma, \alpha) = d_\tau(\alpha, \pi) = 1$.
2) There are two distinct permutations α, β such that $d_\tau(\sigma, \alpha) = d_\tau(\alpha, \pi) = d_\tau(\sigma, \beta) = d_\tau(\beta, \pi) = 1$.

If the first option holds then we say that σ and π are of *type I*, and otherwise, we say they are of *type II*.

Lemma 7. *Assume σ, π satisfy $d_\tau(\sigma,\pi) = 2$ and they are of type II. Let α, β be such that $d_\tau(\sigma,\alpha) = d_\tau(\alpha,\pi) = d_\tau(\sigma,\beta) = d_\tau(\beta,\pi) = 1$. Then the following holds*

$$B_r(\sigma) \cap B_r(\pi) = B_{r-1}(\alpha) \cup B_{r-1}(\beta).$$

Proof: $\subseteq$: Using the left invariance of the metric, we can, without loss of generality, assume that $\sigma = \mathrm{id}$ and $\pi = (1 \cdots i-1, i+1, i, \cdots, j-1, j+1, j, \cdots, n) = \epsilon_i\epsilon_j$, where $|i-j| \geqslant 2$. In this case, we have that $\alpha = \epsilon_i$ and $\beta = \epsilon_j$. Assume that $\gamma \in B_r(\mathrm{id}) \cap B_r(\pi)$. Then, the weight of the permutation γ is at most r. Now, consider the following cases:

1) If $(i, i+1) \in W_\tau(\gamma)$, then $d_\tau(\gamma,\alpha) \leqslant r-1$.
2) If $(j, j+1) \in W_\tau(\gamma)$, then $d_\tau(\gamma,\beta) \leqslant r-1$.
3) If $(i, i+1) \notin W_\tau(\gamma)$ and $(j, j+1) \notin W_\tau(\gamma)$, then $w_\tau(\gamma) \leqslant r-2$ (since $d_\tau(\gamma,\pi) \leqslant r$), and thus $d_\tau(\gamma,\alpha), d_\tau(\gamma,\beta) \leqslant r-1$.

Combining all the cases we get $\gamma \in B_{r-1}(\alpha) \cup B_{r-1}(\beta)$.

$\supseteq$: if $\gamma \in B_{r-1}(\alpha)$ then since $d_\tau(\sigma,\alpha) = 1$, we have that $d_\tau(\gamma,\sigma) \leqslant d_\tau(\gamma,\alpha) + d_\tau(\alpha,\sigma) \leqslant r$. Similarly, $d_\tau(\gamma,\pi) \leqslant r$, thus $\gamma \in B_r(\sigma) \cap B_r(\pi)$ and $B_{r-1}(\alpha) \subseteq B_r(\sigma) \cap B_r(\pi)$. In the same way $B_{r-1}(\beta) \subseteq B_r(\sigma) \cap B_r(\pi)$. ∎

Lemma 8. *Assume σ, π satisfy $d_\tau(\sigma,\pi) = 2$ and they are of type II. Then, $|B_r(\sigma) \cap B_r(\pi)| = I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r)$.*

Proof: Let us choose two permutations σ and π of distance two and type II, i.e., $\pi = \sigma\epsilon_i\epsilon_j$ with $|i-j| \geqslant 2$. Let $\alpha = \sigma\epsilon_i$ and $\beta = \sigma\epsilon_2$ be the two permutations at distance one from σ and π. Note that α and β are also at distance two and of type II. Since $\epsilon_1^{-1} = \epsilon_1$ we have

$$\begin{aligned} B_r(\sigma) \cap B_r(\pi) &= B_r(\mathrm{id}) \cap B_r(\epsilon_1\epsilon_2) = B_r(\epsilon_1) \cap B_r(\epsilon_2) \\ &= B_r(\sigma\epsilon_1) \cap B_r(\sigma\epsilon_2) = B_r(\alpha) \cap B_r(\beta). \end{aligned}$$

If we denote $M_r = |B_r(\sigma) \cap B_r(\pi)|$, then we get from Lemma 7 the following recursive formula:

$$\begin{aligned} M_r &= |B_r(\sigma) \cap B_r(\pi)| = |B_{r-1}(\alpha) \cup B_{r-1}(\beta)| \\ &= 2b_{r-1} - |B_{r-1}(\alpha) \cap B_{r-1}(\beta)| = 2b_{r-1} - M_{r-1}, \end{aligned}$$

where $M_1 = 2$.

It follows that the value of M_r satisfies the same recursive formula with the same initial value as the value of $I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r)$ from Theorem 5. Therefore, $M_r = I(\mathcal{G}_\tau(\mathrm{S}_n); 1, r)$.

∎

Theorem 6 now follows from Lemma 8.

C. The case $d = 2r$

In this section we discuss the other extreme case where now the distance between the centers is relatively large. In particular, we know that if $d \geqslant 2r+1$ then $N(\mathcal{G}_\tau(\mathrm{S}_n); d, r) = 0$. We study the case $d = 2r$.

Lemma 9. *Let $\sigma, \pi \in \mathrm{S}_n$ be such that $d_\tau(\sigma,\pi) = d = 2r$. Then, $|B_r(\sigma) \cap B_r(\pi)| \leqslant \binom{2r}{r}$.*

Proof: Without loss of generality, assume that $\sigma = \mathrm{id}$, so π satisfies $w_\tau(\pi) = 2r$. For every $\alpha \in \mathrm{S}_n$, if $w_\tau(\alpha) \neq r$ then $\alpha \notin B_r(\mathrm{id}) \cap B_r(\pi)$. Now assume that $w_\tau(\alpha) = r$. If $|W_\tau(\alpha) \cap W_\tau(\pi)| < r$ then $d_\tau(\alpha,\pi) > r$. Hence, necessarily $W_\tau(\alpha) \subseteq W_\tau(\pi)$. Since the size of $W_\tau(\pi)$ is $2r$ and the size of $W_\tau(\alpha)$ is r there cannot be more than $\binom{2r}{r}$ such permutations α, and thus $|B_r(\mathrm{id}) \cap B_r(\pi)| \leqslant \binom{2r}{r}$. ∎

Lemma 10. *For $r \leqslant n/4$, $N(\mathcal{G}_\tau(\mathrm{S}_n); 2r, r) = \binom{2r}{r}$.*

Proof: According to Lemma 9, we already have $N(\mathcal{G}_\tau(\mathrm{S}_n); 2r, r) \leqslant \binom{2r}{2}$. In order to prove equality, we show an example of σ, π such $d_\tau(\sigma,\pi) = 2r$ and $|B_r(\sigma) \cap B_r(\pi)| = \binom{2r}{r}$. Let $\sigma = \mathrm{id}$ and $\pi = \epsilon_1\epsilon_3 \ldots \epsilon_{2r-1}$, so π has $2r$ independent adjacent transpositions. Every permutation α which has exactly r out of these $2r$ adjacent transpositions satisfies $d_\tau(\mathrm{id},\alpha) = d_\tau(\sigma,\alpha) = r$ and thus we get $|B_r(\mathrm{id}) \cap B_r(\pi)| = \binom{2r}{r}$. Therefore, we conclude that $N(\mathcal{G}_\tau(\mathrm{S}_n); 2r, r) = \binom{2r}{r}$. ∎

REFERENCES

[1] T. Batu, S. Kannan, S. Khanna, and A. McGregor, "Reconstructing strings from random traces," *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 903–911, 2004.

[2] T. Holenstein, M. Mitzenmacher, R. Panigrahy, and U. Wieder, "Trace reconstruction with constant deletion probability and related results," *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 389-398, 2008.

[3] D.E. Knuth, *The Art of Computer Programmoing Volume 3: Sorting and Searching*, 2nd ed. Reading, MA: Addison-Wesley, 1998.

[4] S. Kannan and A. McGregor, "More on reconstructing strings from random traces: insertions and deletions," *Proc. IEEE International Symposium on Information Theory*, pp. 297–301, Australia, Sep. 2005.

[5] M. Kendall and J.D. Gibbons, *Rank Correlation Methods*. New York: Oxford Univ. Press, 1990.

[6] E. Konstantinova, "On reconstruction of signed permutations distorted by reversal errors," *Discrete Mathematics*, vol. 308, pp. 974–984, 2008.

[7] E. Konstantinova, "Reconstruction of permutations distorted by single reversal errors," *Discrete Applied Math.*, vol. 155, pp. 2426–2434, 2007.

[8] E. Konstantinova, V. Levenshtein, and J. Siemons, "Reconstruction of permutations distorted by single transposition errors," *http://arxiv.org/abs/math/0702191v1*, February 2007.

[9] A. Jiang, M. Schwartz, and J. Bruck, "Correcting charge-constrained errors in the rank-modulation scheme," *IEEE Trans. on Information Theory*, vol. 56, no. 5, pp. 2112–2120, May 2010.

[10] V.I. Levenshtein, "Reconstructing objects from a minimal number of distorted patterns", (in Russian), *Dokl. Acad. Nauk 354* pp. 593-596; English translation, *Doklady Mathematics*, vol. 55 pp. 417–420, 1997.

[11] V.I. Levenshtein, "Efficient reconstruction of sequences," *IEEE Trans. on Information Theory*, vol. 47, no. 1, pp. 2–22, January 2001.

[12] V.I. Levenshtein, "Efficient reconstruction of sequences from their subsequences or supersequences," *Journal of Combin. Theory, Ser. A*, vol. 93, no. 2, pp. 310–332, 2001.

[13] V.I. Levenshtein, E. Konstantinova, E. Konstantinov, and S. Molodtsov, "Reconstruction of a graph from 2-vicinities of its vertices," *Discrete Applied Mathematics*, vol. 156, pp. 1399–1406, 2008.

[14] V.I. Levenshtein and J. Siemons, "Error graphs and the reconstruction of elements in groups," *Journal of Combin. Theory, Ser. A*, vol. 116, pp. 795–815, 2009.

[15] S. Motahari, G. Bresler, and D. Tse, "Information theory of DNA sequencing," http://arxiv.org/abs/1203.6233, 2012.

[16] S. Motahari, G. Bresler, and D. Tse, "Information Theory for DNA Sequencing: Part 1: A Basic Model," *Proc. IEEE International Symposium on Information Theory*, pp. 2741–2745, Cambridge, MA, July 2012.

[17] T. Muir, "On a simple term of a determinant," *Proc. Royal Soc. Edinburg*, vol. 21, pp. 441–477, 1898.

[18] K. Viswanathan and R. Swaminathan, "Improved string reconstruction over insertion-deletion channels," *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 399–408, 2008.

[19] E. Yaakobi and J. Bruck, "On the uncertainty of information retrieval in associative memories," *Proc. IEEE International Symposium on Information Theory*, pp. 106–110, Cambridge, MA, July 2012.