

Codes and mutations in between groups and graphs

Elena Konstantinova

Sobolev Institute of Mathematics
Mathematical Center in Akademgorodok
Novosibirsk State University

e_konsta@math.nsc.ru

Groups, Graphs and Everything in between
on the occasion of
Dragan Marušič's 70th birthday

Lake Bled, Slovenia, May 2023

Important remark

All graphs under consideration in this talk are **hamiltonian** with giving respect to one of the favourite Dragan's topic.

Levenshtein, 2001: Efficient reconstruction of sequences, IEEE Transactions on Information Theory

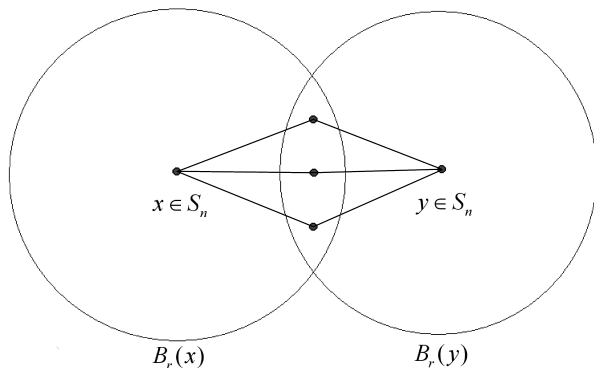
In coding theory, sequences (or any other information) are considered as vertices of a graph and there is an edge if there exist single errors of the type under consideration (substitutions, transpositions, deletions, insertions of symbols) transforming one vertex into another.

The problem of reconstructing an unknown vertex (sequence) $x \in V$ from the minimal number of vertices (erroneous patterns) of $B_r(x)$:

Given a graph $\Gamma = (V, E)$ and an integer r , is there the minimum number $N = N(\Gamma, r)$ of vertices in the metric balls $B_r(x)$ of radius r centered at a vertex $x \in V(\Gamma)$ such that an arbitrary $x \in V$ can be identified from any $N + 1$ distinct vertices of $B_r(x)$?

Related to: an [error correcting code](#) that is an encoding scheme that transmits messages as binary numbers, in such a way that the message can be recovered even if some bits are erroneously flipped.

Levenshtein, 2001: Efficient reconstruction of sequences



$$N = N(\Gamma, r) = \max_{x, y \in V, x \neq y} |B_r(x) \cap B_r(y)|$$

$N + 1$ vertices are sufficient to reconstruct x or y

Intersections in error graphs: Hamming graph

Hamming graph $L_n(q)$: is presented by vectors from F_q^n , $n \geq 2$, $q \geq 2$, with edges $\{x, y\}$ if and only if the Hamming distance $d(x, y) = 1$.

It is a **distance-regular graph**.

Levenshtein, 1997: For any $n \geq 2$, $q \geq 2$ and $r \geq 1$, we have

$$N(L_n(q), r) = q \sum_{i=0}^{r-1} \binom{n-1}{i} (q-1)^i$$

The Hamming graph $L_2(q)$ is the *lattice graph*:

$$\begin{aligned} N(L_2(q), 1) &= q \\ N(L_2(q), 2) &= q^2 \end{aligned}$$

Intersections in error graphs: Johnson graph

Johnson graph J_e^n : is defined on the subset $V = J_e^n \subseteq F_2^n$ consisting of all vectors of Hamming weight e . On J_e^n the Johnson distance is defined as half the (even) Hamming distance, and two vertices x, y are joined by an edge if and only if they are at Johnson distance 1 from each other. It is a **distance-regular graph**.

Levenshtein, 1997: For any $n \geq 2$, $e \geq 1$ and $r \geq 1$,

$$N(J_e^n, r) = n \sum_{i=0}^{r-1} \binom{e-1}{i} \binom{n-e-1}{i} \frac{1}{i+1}$$

The Johnson graph J_2^n is the *triangular graph* $T(n)$, $e = 2$ and $n \geq 4$:

$$\begin{aligned} N(T(n), 1) &= n \\ N(T(n), 2) &= \frac{n(n-1)}{2} \end{aligned}$$

Other metrics and other error graphs? Bioinformatics!

A single chromosome is represented by a permutation π on the integers $1, \dots, n$. A genome is presented by a map that provide the location of genes along a chromosome.

Genome rearrangements: local **mutations** (deletions, insertions, substitutions of nucleotides) as well as global rearrangements (inversions and transpositions of fragments).

The study of genome rearrangements are useful for studying *evolution*.

Problems:

- reconstructing evolutionary trees;
- reconstructing ancestral genomes;

⇒ A reconstruction problem in molecular biology

Evolutionary distances: the reversal distance

One of the ways to compare genomes is to compare the order of appearance of identical genes in the two species.

In: J. D. Palmer, L. A. Herbon. Tricircular mitochondrial genomes of brassica and Raphanus: reversal of repeat configurations by inversion, Sequence alignment in molecular biology, *Nucleid Acids Research*, 14 (1986) 9755-9764.

it was shown that the difference in order may be explained by a small number of *reversals*. These reversals correspond to evolutionary changes during the history of the two genomes, so the number of reversals reflects the evolutionary distance between the species.

The *evolutionary distance* between two genomes is measured by the *reversal distance* of two permutations that is the minimum number d of reversals needed to transform one permutation to another.

Example. $\pi = [4135\underline{2}] \rightarrow [\underline{4}1325] \rightarrow [1\underline{4}3\underline{2}5] \rightarrow [12345] = I$
 $d(\pi, I) = 3$

ZiF-2003: Reconstruction of permutations from their erroneous patterns

$B_r(\pi)$ is the set of permutations at reversal distance at most r from π .

The reconstruction of an unknown permutation π from its erroneous reversal patterns is to find, for a fixed integer r , the minimal number $N(n, r)$ such that for the given $N(n, r)$ permutations one can determine whether there exists a permutation π for which all these permutations belong to $B_r(\pi)$ and determine this permutation π in the affirmative case.

The problem was investigated for $r = 1$ ($r = 2$)

The answer is $N(n, 1) = 4$ (permutations) (a big surprise!)

The answer is $N(n, 1) = 3$ (signed permutations) (a big surprise!)

The Reversal and Transposition graphs

Reversal graph over the symmetric group

is the Cayley graph over the symmetric group Sym_n generated by all reversals on intervals $[i, j]$, $1 \leq i < j \leq n$.

Reversal graph over the hyperoctahedral group

is the Cayley graph over the hyperoctahedral group $B_n = \mathbb{Z}_2 \wr \text{Sym}_n$ generated by sign-change reversals on intervals $[i, j]$, $1 \leq i < j \leq n$.

Transposition graphs

is the Cayley graph over Sym_n ($B_n = \mathbb{Z}_2 \wr \text{Sym}_n$) generated by all transpositions (sign-change transpositions).

All graphs are not distance-regular.

Results in vertex reconstruction in Cayley graphs

Let M be the number of vertices at distance at most r from an unknown vertex x which are sufficient to reconstruct it in a Cayley graph.

The reversal Cayley graphs:

if $r = 1$ then

- $M = 4$ (for permutations)
- $M = 3$ (for signed permutations)

if $r = 2$ then

- $M \geq \frac{3}{2}(n-2)(n+1) + 1$ (for permutations)
- $M \geq n(n+1) + 1$ (for signed permutations)

E. V. Konstantinova. Reconstruction of permutations, Bayreuther Mathematische Schriften, 73 (2005) 213–227.

E. V. Konstantinova. Reconstruction of permutations distorted by reversal errors, Discrete Applied Mathematics, 155(18) (2007) 2426–2434.

E. V. Konstantinova. On reconstruction of signed permutations distorted by reversal errors, Discrete Mathematics, 308 (2008) 974–984.

Results in vertex reconstruction in Cayley graphs

The transposition Cayley graphs:

if $r = 1$ then

- $M = 3$ (for permutations)
- $M = 2$ (for signed permutations)

if $r = 2$ then

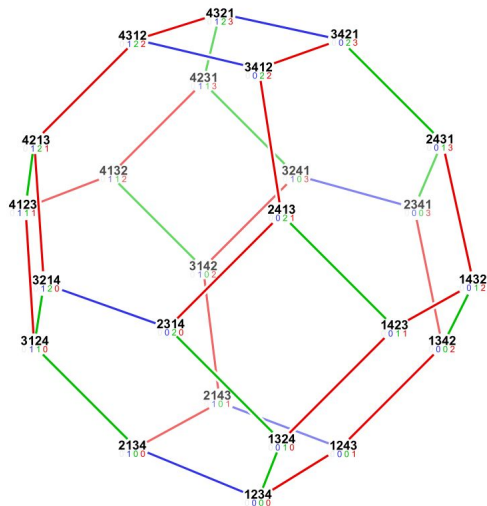
- $M = \frac{3}{2}(n-2)(n+1) + 1$ (for permutations)
- $M = n(n+1) + 1$ (for signed permutations)

E. V. Konstantinova, V. I. Levenshtein, J. Siemons, Reconstruction of permutations distorted by single transposition errors, 2007, arXiv:0702191

E. V. Konstantinova, Vertex reconstruction in Cayley graphs, Discrete Mathematics, 309 (2009) 548–559.

V.I. Levenshtein, J. Siemons, Error graphs and the reconstruction of elements in groups, Journal of Combinatorial Theory, Series A, 116(4) (2009) 795–815.

From Bubble-sorting via Permutohedron...



Cayley graph over Sym_n generated by transpositions of two neighbours

... to results on reconstructions

The distance in this graph is called as

- the bubble-sort distance (in computer science, [D. E. Knuth, Sorting and Searching, 1994])
- the Kendall's τ -metric (in statistics, [M. Kendall, J. D. Gibbons, Rank Correlation Methods, 1990])

E. Yaakobi, M. Schwartz, M. Langberg, J. Bruck, Sequence reconstruction for Grassmann graphs and permutations, Proc. Int. Symp. Inform. Theory, 2013, 874–878.

In: [J. Lu, X. Wang, New Results on the Reconstruction of Permutations Distorted By Single Kendall \$\tau\$ -Errors, IEEE Access, 10 \(2022\) 98761-98768](#)
it is shown that for any $n > 8$:

$$\frac{n^3 - n - 3}{3}$$

erroneous patterns are needed to reconstruct an unknown permutation.

Improved Permutation Arrays for Kendall Tau Metric, by S. Bereg, W. Bumpass, M. Haghpanah, B. Malouf, I. H. Sudborough,
Submitted on 26 Jan 2023, <https://arxiv.org/abs/2301.11423>

New Bounds on the Size of Permutation Codes With Minimum Kendall τ -distance of Three, by A. Abdollahi, J. Bagherian, F. Jafari, M. Khatami, F. Parvaresh, R. Sobhani,
Submitted on 21 Jun 2022, <https://arxiv.org/abs/2206.10193>

The sequence reconstruction problem for permutations with the Hamming distance, by X. Wang, E. V. Konstantinova,
Submitted on 21 Oct 2022, <https://arxiv.org/abs/2210.11864>

The sequence reconstruction problem for permutations with the Hamming distance

Let S be a set of all sequences of length n , ρ be a metric in S , and $B_r(x) = \{y \in S \mid \rho(x, y) \leq r\}$ be a metric ball of radius r centered at $x \in S$. For any integer $d \geq 1$, the minimum number of transmission channels has to be greater than the maximum intersection of two metric balls centered at elements of S denoted as follows:

$$N(n, d, r) = \max_{x_1, x_2 \in S, \rho(x_1, x_2) \geq d} |B_r(x_1) \cap B_r(x_2)|. \quad (1)$$

The problem of determining $N(n, d, r)$ is the *sequence reconstruction problem*.

Cayley graph on permutations with the Hamming distance

Let $\text{Sym}_n(H)$, $n \geq 2$, be the Cayley graph over the symmetric group Sym_n generated by cycles of length at least two from the following set:

$$H = \{\gamma \in \text{Sym}_n \mid \text{disc}(\gamma) = [1^{n-i}i^1], i \in [n] \setminus \{1\}\}, \quad (2)$$

where $\text{disc}(\pi) = [1^{h_1}2^{h_2} \dots n^{h_n}]$ is the cycle type of π and h_i is the number of cycles of length i .

For example, the cycle type of $\pi = (12)(345)(678)$ is written as $\text{disc}(\pi) = [2^13^2]$.

It follows from the known fact:

$$|H| = \sum_{i=2}^n \frac{n!}{(n-i)! \cdot i}.$$

Cayley graph on permutations with the Hamming distance

Property 1

The distance between two permutations in $\text{Sym}_n(H)$, $n \geq 2$, is the Hamming distance.

For any two permutations π, τ of the graph $\text{Sym}_n(H)$, the distance between π and τ is the least sum of lengths of disjoint cycles transforming π into τ , where each cycle contributes a value with the length of the cycle to the Hamming distance between π and τ .

Property 2

The Cayley graph $\text{Sym}_n(H)$, $n \geq 4$, is not distance regular.

The sequence reconstruction problem for permutations with the Hamming distance

Main results

For any $n \geq 2r$, we have:

$$N(n, 2r, r) = \begin{cases} \binom{2m}{m}, & \text{if } r = 2m; \\ 2\binom{2m-2}{m-1}, & \text{if } r = 2m + 1. \end{cases}$$

For any $n \geq 2r - 1$, we have:

$$N(n, 2r - 1, r) = \begin{cases} \binom{2m-2}{m}, & \text{if } r = 2m; \\ \binom{2m-1}{m-1}, & \text{if } r = 2m + 1. \end{cases}$$

The sequence reconstruction problem for permutations with the Hamming distance

Main results

For any $n \geq 4$, we have:

$$N(n, d, 4) = \begin{cases} 7n^2 - 31n + 36, & \text{if } d = 2; \\ (9n^2 - 27n + 12)/2, & \text{if } d = 3; \\ n^2 + 15n - 52, & \text{if } d = 4; \\ 6n + 14, & \text{if } d = 5; \\ 38, & \text{if } d = 6; \\ 10, & \text{if } d = 7; \\ 6, & \text{if } d = 8; \\ 0, & \text{if } d \geq 9, \end{cases}$$



Happy birthday!