Search problems on Cayley graphs

Elena Konstantinova

Sobolev Institute of Mathematics

Novosibirsk, Russia

e_konsta@math.nsc.ru

Search Methodologies, ZIF Cooperation Group 2010, October 25–29

Search Methodologies

Short description of the project:

In the three decades which passed since 1979 there has been an explosion of developments in search for instance in Computer Science, Image Reconstruction, Machine Learning, Information Theory, and Operation Research.

A search structure is defined by a space of objects searched for and a space of tests (questions). In specifying a search problem performance criteria have to be chosen. Furthermore we distinguish combinatorial and probabilistic models.

$\Longrightarrow$ **Search combinatorial problems on Cayley graphs**

Space of objects - Cayley graphs

Let G be a group, and let $S \subset G$ be a set of group elements as a set of generators for a group such that $e \notin S$ and $S = S^{-1}$.

Definition. In the *Cayley graph* $\Gamma = Cay(G, S) = (V, E)$ vertices correspond to the elements of the group, i.e. $V = G$, and edges correspond to the action of the generators, i.e. $E = \{(g, gs) : g \in G, s \in S\}$.

Properties:

(i) Γ is a connected regular graph of degree $|S|$;

(ii) Γ is a vertex–transitive graph.

Examples: hypercube graph, pancake network (de Bruijn graph)

Space of questions - combinatorial problems

Open combinatorial problems on Cayley graphs:

- Diameter problem (Pancake problems);
- Sorting by reversals;
- Hamiltonian problem;
- Vertex reconstruction problem;

Applications:

- computer science (networks);
- molecular biology;
- coding theory;

Applications in Computer science

SIAM International Conference on Parallel Processing, 1986: it was suggested to use Cayley graphs as a "tool to construct vertex–symmetric interconnection networks."

Interconnection networks are modeled by graphs: the vertices correspond to processing elements, memory modules, or just switches; the edges correspond to communication lines.

<u>Advantages in using Cayley graphs as network models</u>:

- *vertex–transitivity* (the same routing algorithm is used for each v);
- *hierarchical structure* (allows recursive constructions);
- *high fault tolerance* (the maximum number of vertices that need to be removed and still have the graph remain connected);
- *small degree and diameter*.

Cayley graphs in computer science

The transposition Cayley graphs:

• *transposition networks* $Sym_n(T)$ on the symmetric group Sym_n generated by the transpositions from the set $T = \{(i,j),\ 1 \le i < j \le n\}$;

• *star Cayley graphs* $Sym_n(ST)$ generated by $(1,i),\ 1 < i \le n$;

• *bubble sort Cayley graphs* $Sym_n(t)$ generated by $(i,i+1),\ 1 \le i < n$,

The pancake Cayley graphs:

• *pancake graph* $Sym_n(PR)$ on Sym_n generated by prefix–reversals on intervals $[1,i],\ 1 < i \le n$;

• *burnt pancake graph* $B_n(PR^\sigma)$ on the hyperoctahedral group $B_n = \mathbb{Z}_2 \wr Sym_n$ generated by sign–change prefix–reversals on intervals $[1,i]$.

$\Rightarrow$ well–known open combinatorial pancake problem

Pancake problem

The original pancake problem was posed in 1975 in the *American Mathematical Monthly* by *Jacob E. Goodman* writing under the name "Harry Dweighter" (or "Harried Waiter") and it stated as follows:

> "The chef in our place is sloppy, and when he prepares a stack of pancakes they come out all different sizes. Therefore, when I deliver them to a customer, on the way to the table I rearrange them (so that the smallest winds up on top, and so on, down to the largest on the bottom) by grabbing several pancakes from the top and flips them over, repeating this (varying the number I flip) as many times as necessary. If there are n pancakes, what is the maximum number of flips (as a function of n) that I will ever have to use to rearrange them?"

Pancake problem

A stack of n pancakes is represented by a permutation on n elements and the problem is *to find the least number of flips (prefix–reversals) needed to transform a permutation into the identity permutation*.

This number of flips corresponds to the diameter D of the pancake Cayley graph on the symmetric group generated by prefix–reversals.

Currently, exact values of D are known for $n \leq 17$:

n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
D	1	3	4	5	7	8	9	10	11	13	14	15	16	17	18	19

Asai S., Kounoike Y., Shinano Y. and Kaneko K., Computing the diameter of 17–pancake graph using a PC cluster, LNCS **4128** (2006) 1114–1124

Pancake problem: bounds

Bill Gates, *Papadimitriou*, 1979, for $n \leq 14$:

$$17n/16 \leq D \leq (5n + 5)/3$$

Heydari, *Sudborough*, 1997:

$$15n/14 \leq D$$

Sudborough, etc., 2007:

$$D \leq 18n/11$$

Open problem: What is the diameter of the pancake graph?

Burnt pancake problem

$Gates, Papadimitriou,$ 1979: here one side of each pancakes is burnt, and the pancakes must be sorted with the burnt side down.

Two–sided pancakes can be represented by a signed permutation on n elements with some elements negated: $-I = [-1, -2, \ldots, -n]$.

Open problem: What is the diameter of the burnt pancake graph?

Currently, exact values of the diameter D are known for $n \leq 18$:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	4	6	8	10	12	14	15	17	18	19	21	22	23	24	26	28	29

$Cohen, Blum,$ 1995: $3n/2 \leq D \leq 2n - 2$ (the upper bound for $n \geq 10$)

The diameter problem

$Even$, $Goldreich$, 1981: Computing the diameter of an arbitrary Cayley graph over a set of generators is NP–hard.

General upper and lower bounds are very difficult ro obtain.

There is a fundamental difference between Cayley graphs of abelian and non–abelian groups.

$Babai$, $Kantor$, $Lubotzky$, 1989: *Every non–abelian finite simple group G has a set of ≤ 7 generators such that the resulting Cayley graph has diameter* $O(\log |G|)$.

This property does not hold for Cayley graphs of abelian groups.

The diameter problem

Conjecture. ($Babai$, $Kantor$, $Lubotzky$, 1988) There exist a constant C such that for every non–abelian finite simple group G, the diameter of every Cayley graph of G is $\leq (\log |G|)^C$.

If the conjecture is true, one would expect to find Cayley graphs of these groups with small degree and diameter $\Rightarrow$

$\Rightarrow$ Applications:

- computer science;
- theory of intercommunication networks;

Other applications of Cayley graphs

Molecular biology: A single chromosome is presented by a permutation π on the integers $1, \ldots, n$ as well as a signed permutation when a direction of a gene is important. A genome is presented by a map that provide the location of genes along a chromosome. To compare two genomes, we often find that these two genomes contain the same set of genes. But the order of the genes is different in different genomes.

For example, it was found that both human X chromosome and mouse X chromosome contain eight $1, 2, \ldots, 8$ genes which are identical.
In human, the genes are ordered as $[4, 6, 1, 7, 2, 3, 5, 8]$
and in mouse, they are ordered as $[1, 2, 3, 4, 5, 6, 7, 8]$

It was also found that a set of genes are in cabbage as $[1, -5, 4, -3, 2]$ and in turnip, they are ordered as $[1, 2, 3, 4, 5]$.

Permutations in molecular biology

The comparison of two genomes is significant because it provides us some insight as to how far away genetically these species are.

One of the ways to compare genomes is to compare the order of appearance of identical genes in the two species. $Palmer$(1986) has shown that the difference in order may be explained by a small number of *reversals*: Genome X: $(3, \underline{1, 5, 2}, 4) \longrightarrow$ Genome Y: $(3, 2, 5, 1, 4)$

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

Example. $\pi = [413\underline{52}] \to [\underline{41}325] \to [1\underline{432}5] \to [12345] = I$

$d(\pi, I) = 3$

Cayley graphs in molecular biology

The reversal Cayley graphs:

• *reversal graph* $Sym_n(R)$ on Sym_n generated by reversals on intervals $[i, j]$, $1 \leq i \leq j \leq n$.

• *reversal graph* $B_n(R)$ on the hyperoctahedral group $B_n = \mathbb{Z}_2 \wr Sym_n$ generated by sign–change reversals on intervals $[i, j]$, $1 \leq i \leq j \leq n$.

The prefix–reversal Cayley graphs:

• *prefix–reversal graph* $Sym_n(PR)$ on Sym_n generated by reversals on intervals $[1, i]$, $1 < i \leq n$; (Pancake graph)

• *prefix–reversal graph* $B_n(PR)$ on the hyperoctahedral group $B_n = \mathbb{Z}_2 \wr Sym_n$ generated by sign–change reversals on intervals $[1, i]$, $1 < i \leq n$.

$\Longrightarrow$ sorting by reversals

Sorting permutations by reversals

The problem of sorting permutations by reversals is to find, for a given permutation π, a minimal sequence d of reversals that transforms π to the identity permutation I.

Mathematical analysis of the problem was initiated by $Sankoff, 1990$.

- find the reversal distance between two permutations (a linear–time algorithm, $D.Bader, 2001$);
- find a sequence of reversals which realizes the distance;
  - solutions are far from unique ($A.Bergeron, 2002$);
  - NP–hard for the unsigned permutations (1.5–approximation algorithm, $D.A.Christie, 1998$);
  - polynomial for the signed permutations ($O(n^2)$, $H.Kaplan, 1999$)

Hamiltonicity

Let $\Gamma = (V, E)$ be a connected graph where $V = \{v_1, v_2, \ldots, v_n\}$.

A *hamiltonian cycle* in Γ is a spanning cycle $(v_1, v_2, \ldots, v_n, v_1)$.

A *hamiltonian path* in Γ is a path $(v_1, v_2, \ldots, v_n)$.

A graph is *hamiltonian* if it contains a hamiltonian cycle.

Hamiltonian paths and cycles naturally arise

- in computer science;
- in the study of word–hyperbolic groups and automatic groups;
- in combinatorial designs.

Example: hypercube H_n is hamiltonian (the binary Gray code)

Testing whether a graph is hamiltonian is an NP-complete problem.

Hamiltonicity: conjectures

Conjecture ($Lovász$,1970) Every connected vertex–transitive graph has a hamiltonian path.

Conjecture ($Babai$,1996) For some $\varepsilon > o$, there exist infinitely many connected vertex–transitive graphs (even Cayley graphs) Γ without cycles of length $\geq (1-\varepsilon)|V(\Gamma)|$.

There are only 4 vertex–transitive (not Cayley) graphs which do not have a hamiltonian cycle, and have a hamiltonian path.

Conjecture Every connected Cayley graph on a finite group has a hamiltonian cycle.

Hamiltonicity of Cayley graphs on Sym_n

Conjecture Every connected Cayley graph on a finite group has a hamiltonian cycle.

It is true for:

- abelian groups (and hence for circulants) ($Maru\check{s}i\check{c}$, 1983);
- nonabelian groups of special type;
- dihedral groups, ($Alspach$, $Zhang$, 1989)
- symmetric group generated by transpositions, ($Liskovets$, 1975)

Thus, all transposition Cayley graphs are hamiltonian $\Rightarrow$

- star networks, bubble–sort networks are hamiltonian;
- pancake networks are hamiltonian.

Other applications of Cayley graphs

Coding theory: sequences (or any other information) are considered as a vertex set V 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.

There are situations when encoding is absent and any vertex (sequence) of the graph can be used for information transmission. In this case the problem of reconstructing an unknown vertex (sequence) $x \in V$ from the minimal number of vertices (erroneous patterns) in $B_r(x)$ arises (Levenshtein, 1997).

$\Longrightarrow$ Vertex reconstruction problem

Vertex reconstruction problem

For given $r \geq 1$ denote by $N(\Gamma, r)$ the largest number N such that there exist a subset $A \subseteq V$ of size N and two vertices $x \neq y$ with $A \subseteq B_r(x)$ and $A \subseteq B_r(y)$. Thus any $N+1$ distinct vertices are contained in $B_r(x)$ for at most one vertex x, while this statement is wrong for any $N < N(\Gamma, r)$.

This means that an unknown vertex of Γ can be reconstructed uniquely by any subset of $N(\Gamma, r) + 1$ or more distinct vertices at distance at most r from the vertex, if such a subset exists.

Thus, for a given graph Γ and integers $r = 1, ..., d(\Gamma)$ it is required to find

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

Intersections in a given graph

Hamming graph $L_n(q)$: is presented by the vertex set given by the vector space 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 distance–regular, and the Cayley graph on the additive group F_q^n when we take the generator set $S = \{xe_i : x \in (F_q)^{\times}, 1 \leq i \leq n\}$ where the $e_i = (0, ..., 0, 1, 0, ...0)$ are the standard basis vectors of F_q^n.

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*. This graph is strongly regular with parameters $v = q^2$, $k = 2(q-1)$, $\lambda = q - 2$, $\mu = 2$,

$$N(L_2(q), 1) = q \text{ and } N(L_2(q), 2) = q^2.$$

Intersections in a given 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 distance–regular but not Cayley 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 $n \geq 4$. This graph is strongly regular with parameters $v = \frac{n(n-1)}{2}$, $k = 2(n-2)$, $\lambda = n-2$, $\mu = 4$; $N(T(n), 1) = n$ and $N(T(n), 2) = \frac{n(n-1)}{2}$.

Intersections in Cayley graphs

Reversal graph $Sym_n(R)$: is defined on Sym_n and generated by the reversals from the set $R = \{r_{i,j} \in Sym_n, 1 \leq i < j \leq n\}$, $|R| = \binom{n}{2}$, where reversals $r_{i,j}$ inverse segments $[i,j]$ of a permutation, i.e., $[\ldots, \pi_i, \pi_{i+1}, \ldots, \pi_{j-1}, \pi_j, \ldots] r_{i,j} = [\ldots, \pi_j, \pi_{j-1}, \ldots, \pi_{i+1}, \pi_i, \ldots]$.
It is not distance–regular graph.

Konstantinova, 2007: $N(Sym_n(R), 1) = 3,\ n \geq 3$

$$N(Sym_n(R), 2) \geq \frac{3}{2}(n-2)(n+1),\ n \geq 3$$

A permutation:

- is uniquely reconstructible from 4 its distinct 1-neighbors;
- is reconstructible from 3 its 1-neighbors with $p_3 \to 1$ as $n \to \infty$;
- is reconstructible from 2 its 1-neighbors with $p_2 \sim \frac{1}{3}$ as $n \to \infty$;

Intersections in Cayley graphs

The transposition Cayley graph $Sym_n(T)$: is defined on Sym_n and generated by all transpositions. It is a connected bipartite $\binom{n}{2}$-regular graph of order $n!$ and diameter $(n-1)$. It is not distance–regular and hence not distance–transitive graph;

Levenshtein, Siemons, 2009:

$$N(Sym_n(T), 1) = 3,\ n \geq 3$$

$$N(Sym_n(R), 2) = \frac{3}{2}(n-2)(n+1),\ n \geq 3$$

A permutation:

- is uniquely reconstructible from 4 its distinct 1-neighbors;

Intersections in Cayley graphs

The reversal graph $B_n(R^\sigma)$: is defined on B_n and generated by the sign–change reversals $r^\sigma_{i,j}$ flipping the signs of elements on the segments $[i,j]$, i.e. $[\ldots,\pi_i,\overline{\pi}_{i+1},\ldots,\pi_{j-1},\pi_j,\ldots]r^\sigma_{i,j}=[\ldots,\overline{\pi}_j,\overline{\pi}_{j-1},\ldots,\pi_{i+1},\overline{\pi}_i,\ldots]$. It is not distance–regular graph.

Konstantinova, 2008: $N(B_n(R^\sigma),1)=2,\ n\geq 2$

$$N(B_n(R^\sigma),1)\geq n(n+1),\ n\geq 2$$

A signed permutation:

- is uniquely reconstructible from 3 its distinct 1-neighbors;
- is reconstructible from 2 its 1-neighbors with $p_2\sim\frac{1}{3}$ as $n\to\infty$;

Open problems

- reconstruction of unsigned and signed permutations distorted by block-changed errors when two adjacent or nonadjacent parts of permutations switch places, for example, $[1\underline{23}\overline{45}6] \to [1\overline{45}\underline{23}6]$;

- reconstruction of unsigned and signed permutations distorted by reversal block-changed errors when two adjacent or nonadjacent parts of permutations switch places and one is reversed, for example, $[1\underline{23}\overline{45}6] \to [1\overline{45}\underline{32}6]$;

- reconstruction of elements of another groups:
  - what are the groups?
  - what are the generating sets?

Phylogenetic tree of human