A New Tight Upper Bound on the Transposition Distance

Anthony Labarre[a]

Université Libre de Bruxelles

alabarre@ulb.ac.be

October 4, 2005

Fifth Workshop on Algorithms in Bioinformatics

[a]Funded by the "Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture" (F.R.I.A.).

Context and motivation

- Genome rearrangement: find how (much) several genomes are related;
- Models vary depending on assumptions:
  - gene order known or not;
  - orientation of genes known or not;
  - mutations taken into account;
- In our model:
  - gene order is known, but their orientation is ignored;
  - all genomes share the same set/number of genes;
  - we consider only transpositions (see next slide);

Presentation of the problem

- A *transposition* exchanges two adjacent blocks in a permutation:

$$(\ 2\ 1\ \boxed{7\ 3\ 5\ 6}\ \boxed{4\ 10\ 9}\ 8\ )$$
$$\downarrow$$
$$(\ 2\ 1\ \boxed{4\ 10\ 9}\ \boxed{7\ 3\ 5\ 6}\ 8\ )$$

- The *transposition distance* $d(\pi, \sigma)$ is the minimal number of transpositions transforming permutation π into permutation σ;

- *Sorting by transpositions* is the problem of transforming a permutation into the *identity* $\iota = (1\ 2\ \cdots\ n)$ using transpositions;

Sorting by transpositions

Example 1 *The following permutation can be sorted using two transpositions <u>and no less</u>:*

$$\pi = \quad (\; 3 \; \boxed{\mathit{1}} \; \boxed{\mathit{4}} \; 2 \;)$$

$$\downarrow$$

$$(\; \boxed{\mathit{3 \; 4}} \; \boxed{\mathit{1 \; 2}} \;)$$

$$\downarrow$$

$$\iota = \quad (\; 1 \; 2 \; 3 \; 4 \;)$$

Therefore $d(\pi, \iota) = d(\pi) = 2.$

Status of the problem

- Introduced in [Bafna and Pevzner, 1998];
- Both the complexities of:
  - sorting by transpositions;
  - computing the transposition distance;

  ... are apparently unknown;
- So is the maximal value $d(\pi)$ can reach (the *diameter*);
- Best polynomial-time approximation has a ratio of $\frac{11}{8}$ or 1.375 [Elias and Hartman, 2005];

Our results

1. Use of another approach than the traditional one (see next slide);

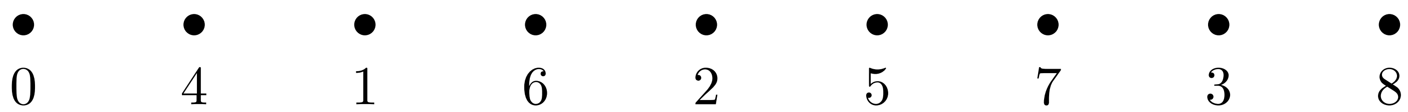
2. A nice correspondence between the cycles of our graph and that introduced in [Bafna and Pevzner, 1998] for a certain class of permutations called $\gamma-$*permutations*;

3. $O(n)$ time and space computation of the transposition distance of $\gamma-$permutations, without the need of any graph structure;

4. A new upper bound on the transposition distance, tight for $\gamma-$permutations.

The cycle graph [Bafna and Pevzner, 1998]

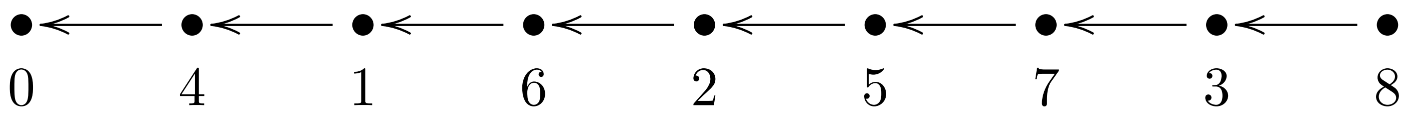
- Given a permutation π, construct the *cycle graph* $G(\pi)$ as follows:



1. $V(G) = \{\pi_0 = 0, \pi_1, \pi_2, ..., \pi_n, \pi_{n+1} = n + 1\}$;

The cycle graph [Bafna and Pevzner, 1998]

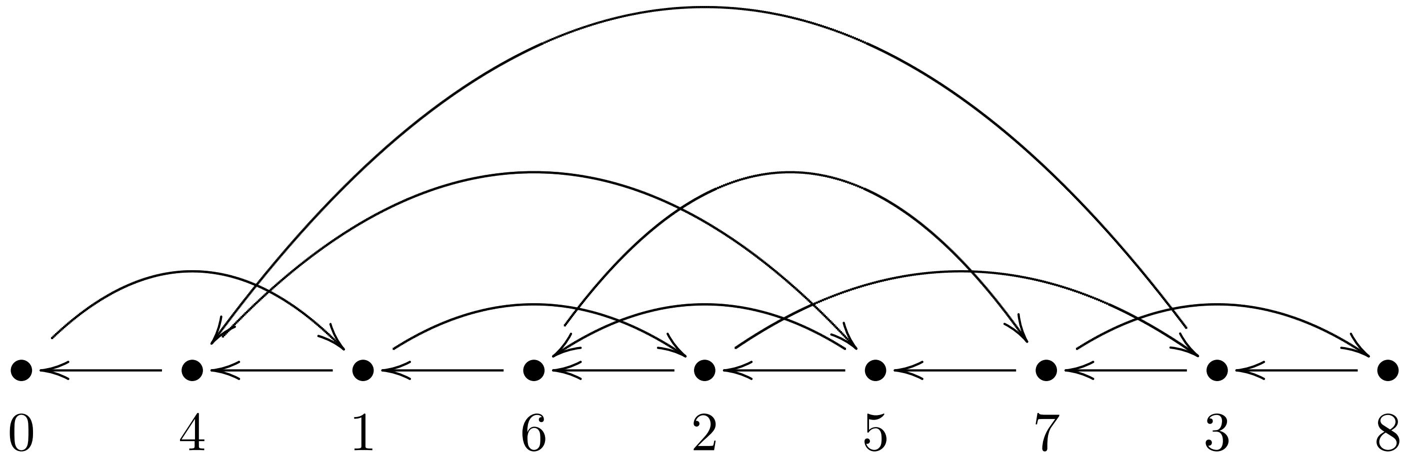
- Given a permutation π, construct the *cycle graph* $G(\pi)$ as follows:



1. $V(G) = \{\pi_0 = 0, \pi_1, \pi_2, ..., \pi_n, \pi_{n+1} = n+1\}$;
2. $E(G) = \underbrace{\{(\pi_i, \pi_{i-1}) \mid 1 \leq i \leq n+1\}}_{\textit{black}\ \text{edges}}$

The cycle graph [Bafna and Pevzner, 1998]

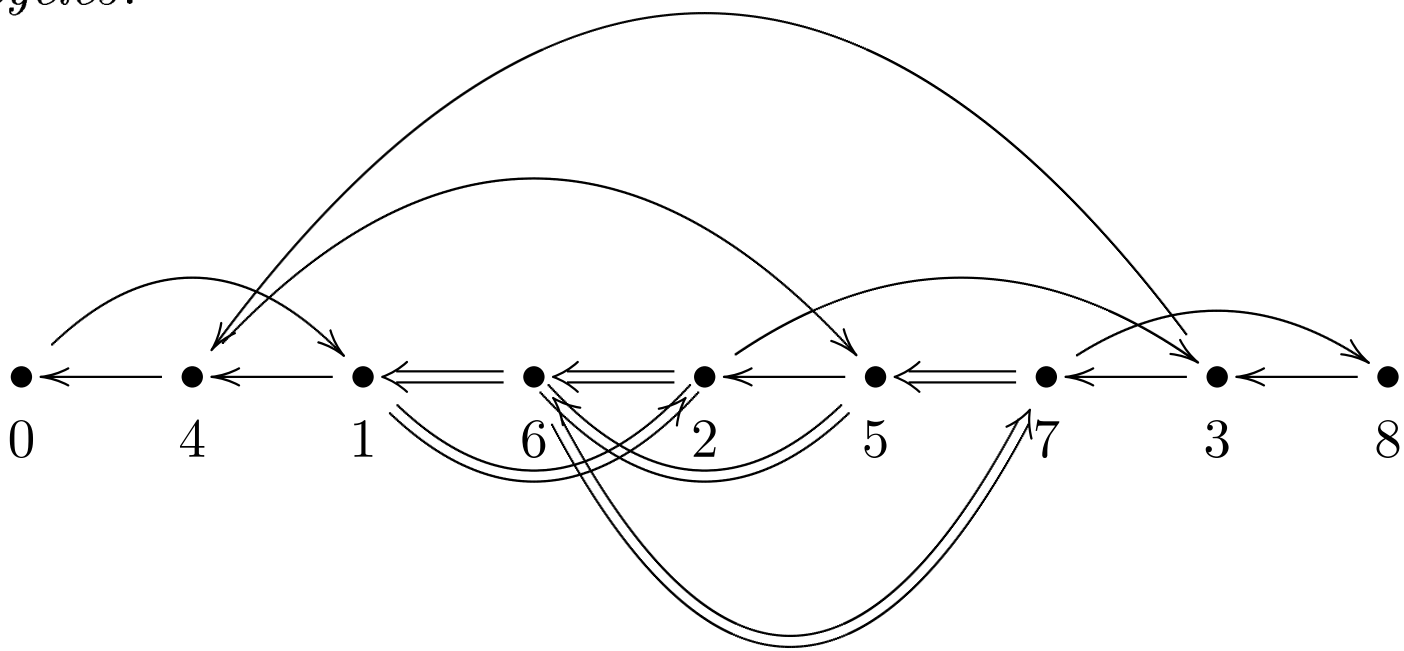
- Given a permutation π, construct the *cycle graph* $G(\pi)$ as follows:



1. $V(G) = \{\pi_0 = 0, \pi_1, \pi_2, ..., \pi_n, \pi_{n+1} = n+1\}$;
2. $E(G) = \underbrace{\{(\pi_i, \pi_{i-1}) \mid 1 \leq i \leq n+1\}}_{\textit{black}\text{ edges}} \cup \underbrace{\{(\pi_i, \pi_i + 1) \mid 0 \leq i \leq n\}}_{\textit{gray}\text{ edges}}$;

Alternate cycles

- The cycle graph decomposes in an unique way into *alternate cycles*:

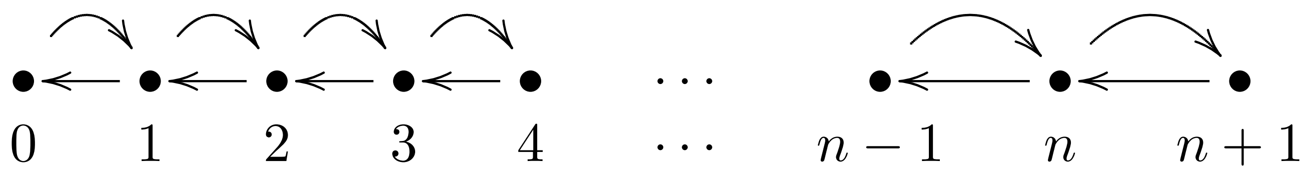


- *Parity* of a cycle = that of the number of black edges it contains; here $c(G(\pi)) = 2 = c_{odd}(G(\pi))$.

Alternate cycles

- The identity $\iota = (1\ 2\ \cdots\ n)$ is the only permutation with $c(G(\iota)) = n + 1 = c_{odd}(G(\iota))$;

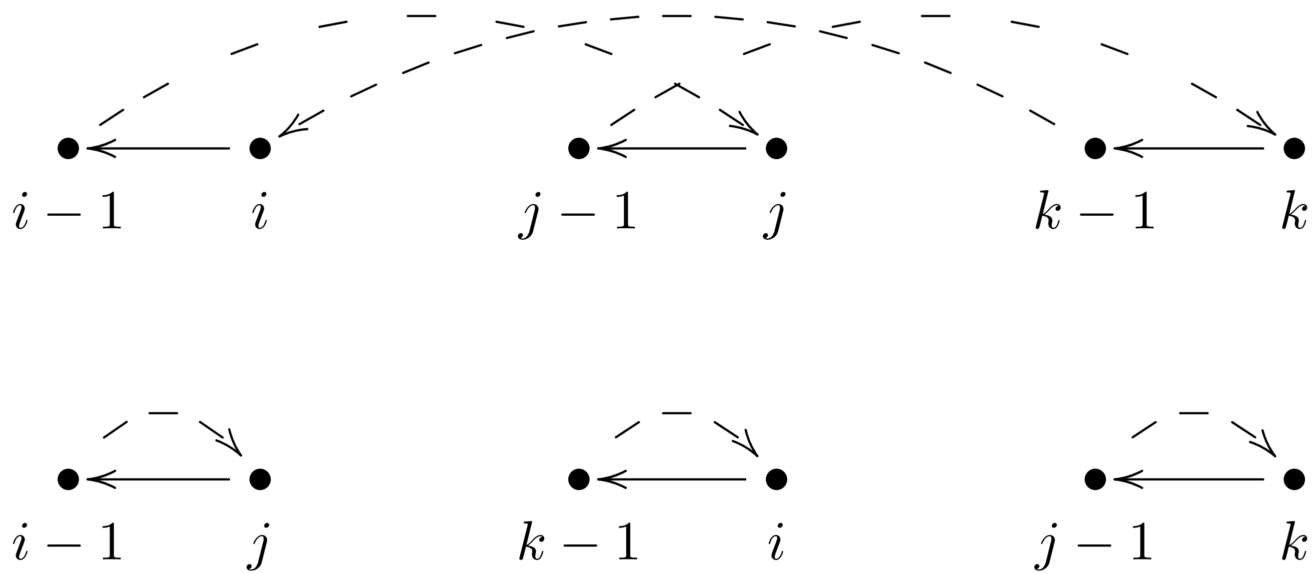
$G(\iota)$:



- Therefore sorting by transpositions comes down to creating odd alternate cycles "as fast as possible";
- How fast can it be done?

A lower bound on the transposition distance

- Best case: two new cycles in one move:



Theorem 1 *[Bafna and Pevzner, 1998]* $\forall\ \pi \in S_n$:

$$d(\pi) \geq \frac{n+1-c_{odd}(G(\pi))}{2} .$$

Reduced permutations

- *Breakpoint* in a permutation π: ordered pair (π_i, π_{i+1}) such that $\pi_{i+1} \neq \pi_i + 1$;
- $b(\pi)$ denotes the number of breakpoints of π;
- π is *reduced* if $b(\pi) = n - 1$, $\pi_1 \neq 1$, and $\pi_n \neq n$.

Example 2

$$(\ 4\ 2\ 1\ 3\ ) \quad \textit{is reduced}$$

$$(\ \underline{1}\ 4\ 3\ 2\ ) \quad \textit{is not reduced}$$

$$(\ 3\ 2\ 1\ \underline{4}\ ) \quad \textit{is not reduced}$$

$$(\ 4\ \underline{2\ 3}\ 1\ ) \quad \textit{is not reduced}$$

Reduced permutations

- Every permutation π is *reducible* to a permutation $gl(\pi)$.

Example 3

$$\pi = \quad (\; 1 \; 2 \mid 9 \mid 4 \; 5 \; 6 \mid 3 \mid 7 \; 8 \mid 10 \;)$$

$$\downarrow$$

$$(\; \boxed{\mathit{1\;2}} \; \boxed{\mathit{9}} \; \boxed{\mathit{4\;5\;6}} \; \boxed{\mathit{3}} \; \boxed{\mathit{7\;8}} \; \boxed{\mathit{10}} \;)$$

$$\downarrow$$

$$(\; \boxed{\mathit{9}} \; \boxed{\mathit{4}} \; \boxed{\mathit{3}} \; \boxed{\mathit{7}} \;)$$

$$\downarrow$$

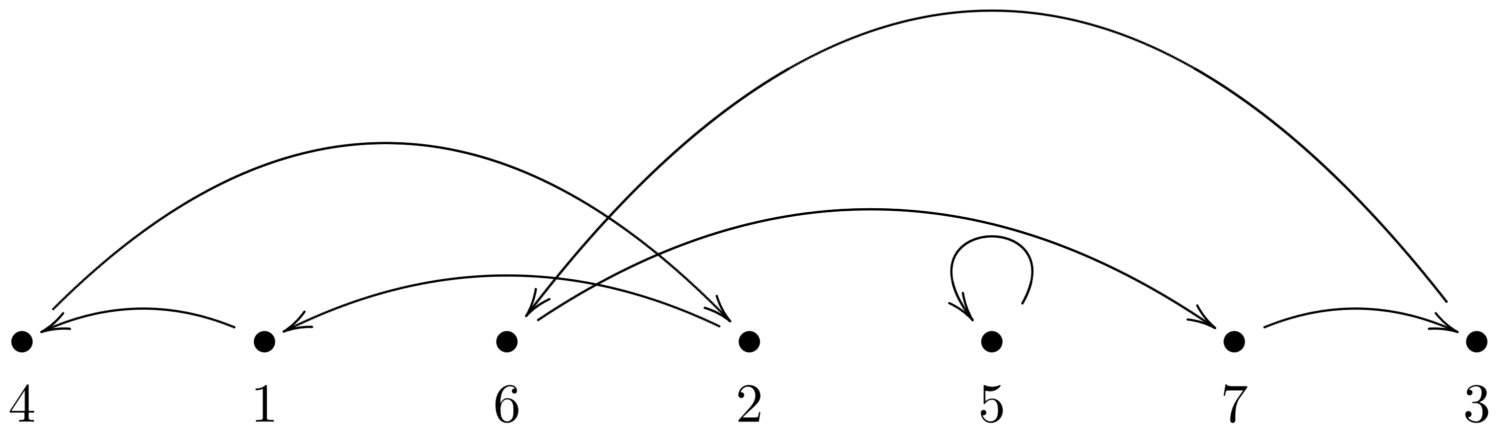
$$gl(\pi) = \quad (\; 4 \; 2 \; 1 \; 3 \;)$$

Theorem 2 *[Christie, 1998]* $\forall \; \pi \in S_n : d(\pi) = d(gl(\pi))$.

Γ−graph

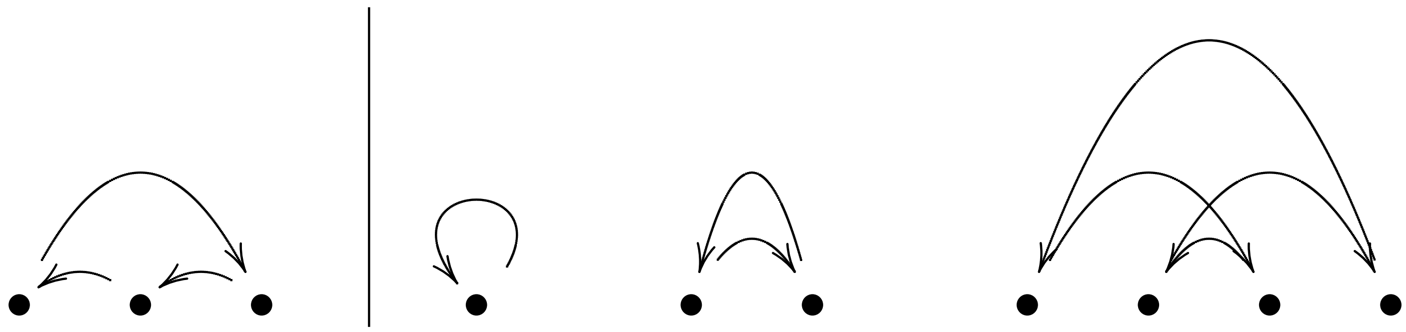
- Every permutation is the product of disjoint cycles;
- The Γ−*graph* is the *graph of the permutation* with a total order on the vertices.

Example 4 *Let* $\pi = (4\ 1\ 6\ 2\ 5\ 7\ 3) = (1,\ 4,\ 2)\ (3,\ 6,\ 7)\ (5)$*; then* $\Gamma(\pi)$ *is:*



Some definitions on the cycles of the Γ−graph

- A k−*cycle* of $\Gamma(\pi)$ is a cycle on k vertices;
- Such a cycle is *odd* (resp. *even*) if k is odd (resp. even);
- *Orientation* of a cycle:

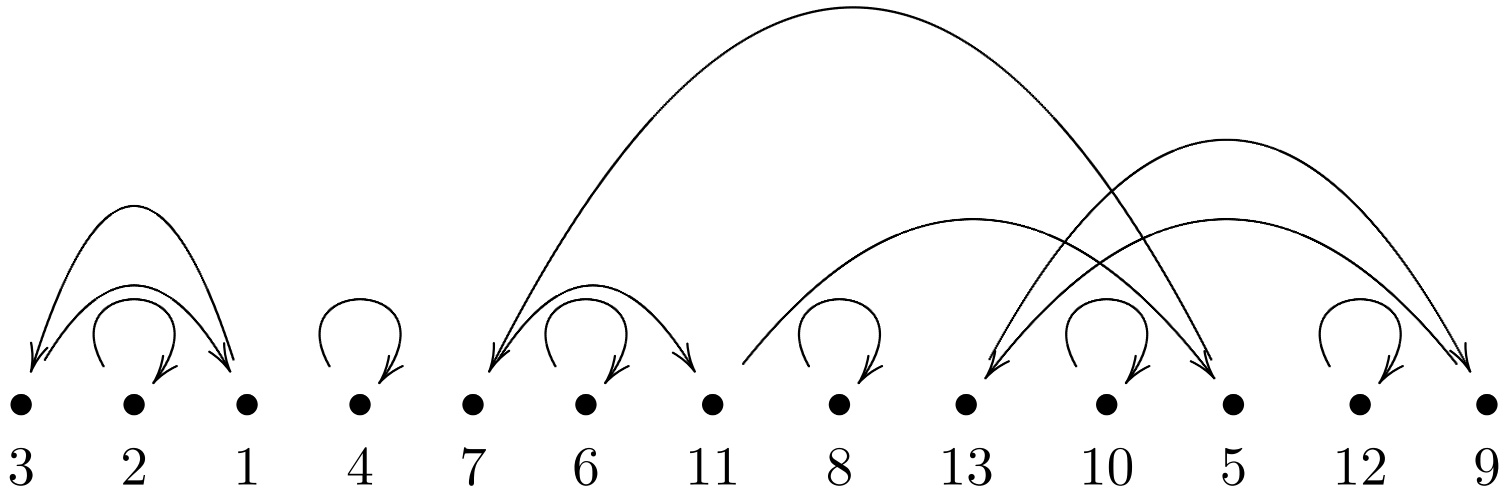


oriented | unoriented

An explicit formula for some permutations

Definition 1 *A* $\gamma-$permutation *is a reduced permutation that fixes all even elements.*

Example 5

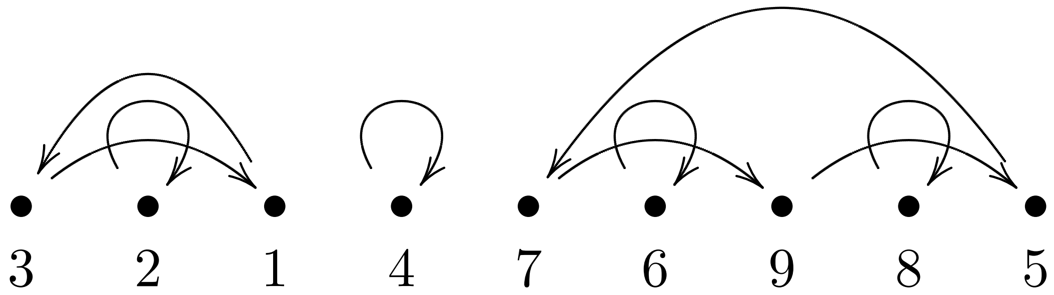


Correspondence between G and Γ for γ−permutations

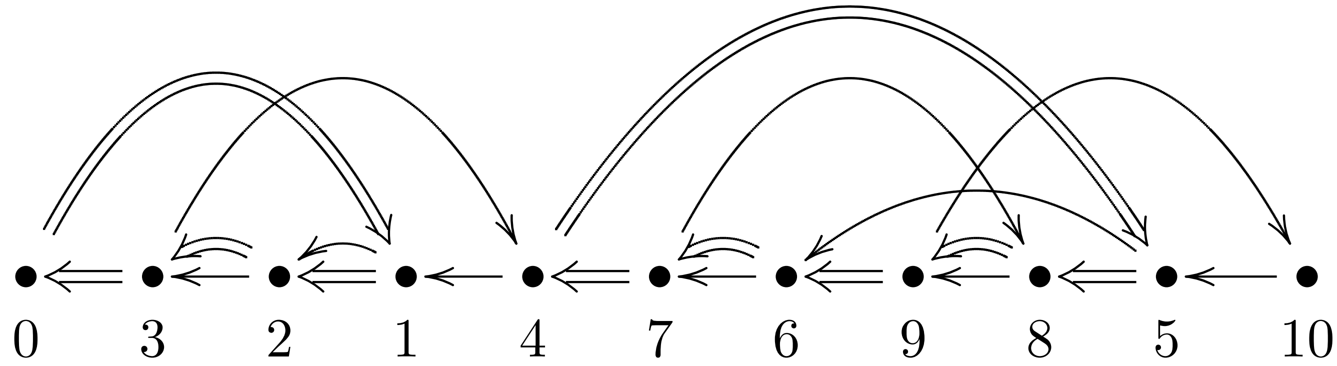
Proposition 1 *For every γ−permutation π in S_n:*

$$\begin{cases} c_{even}(G(\pi)) & = & 2\ c_{even}(\Gamma(\pi))\ ; \\ c_{odd}(G(\pi)) & = & 2\left(c_{odd}(\Gamma(\pi)) - \frac{n-1}{2}\right)\ . \end{cases}$$

$\Gamma(\pi)$:



$G(\pi)$:



Correspondence between G and Γ for $\gamma-$permutations

- Recall that $d(\pi) \geq \frac{n+1-c_{odd}(G(\pi))}{2}$ (Theorem 1);
- This and Proposition 1 yield the following result:

Lemma 1 *For every $\gamma-$permutation π in S_n:*

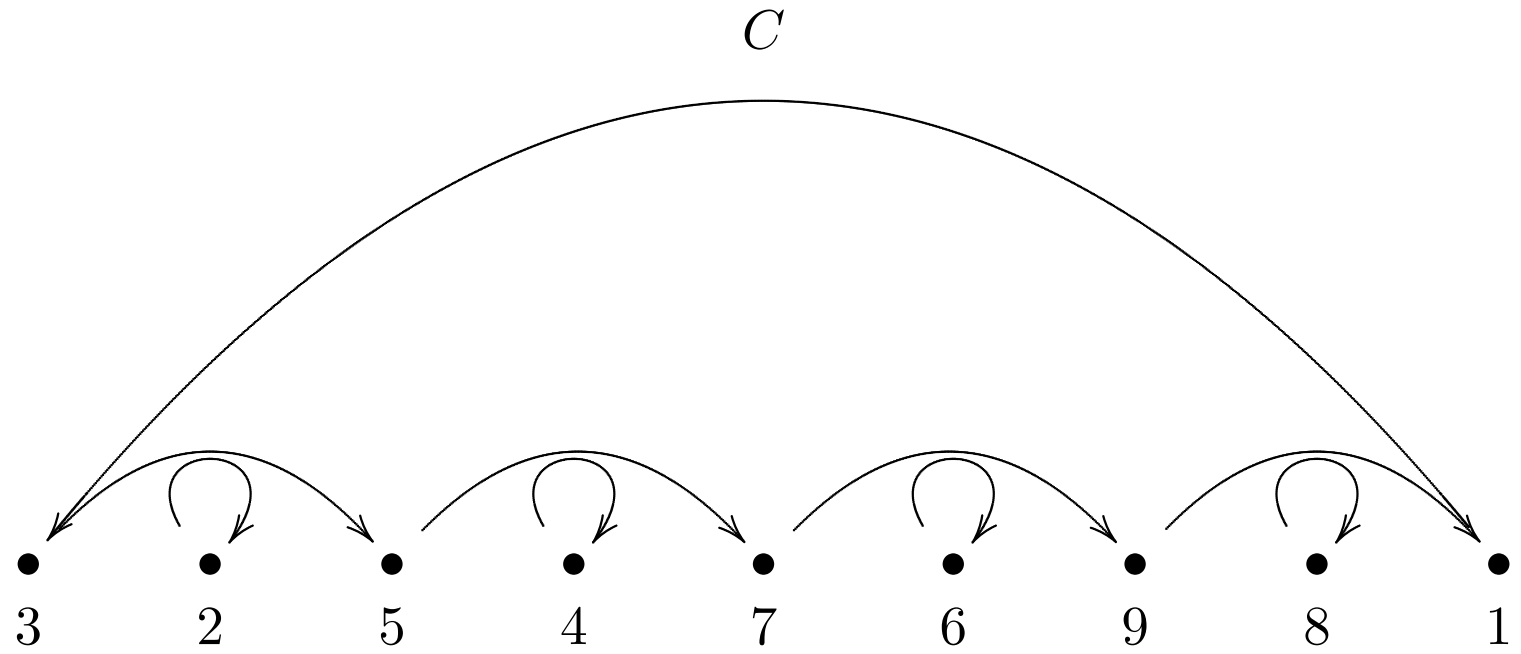
$$d(\pi) \geq n - c_{odd}(\Gamma(\pi)) \ .$$

- This lower bound is actually reached, as will be shown through the separate analysis of:
  1. oriented cycles of Γ, and
  2. unoriented cycles of Γ.

α−permutations

Definition 2 *An* α−permutation *is a reduced permutation that fixes all even elements and whose odd elements form one oriented cycle in* Γ*.*

Example 6



Distance of $\alpha-$permutations

Proposition 2 *For every $\alpha-$permutation π in S_n:*

$$d(\pi) = n - c_{odd}(\Gamma(\pi)) = |C| - (|C| \bmod 2) \quad .$$

Example 7

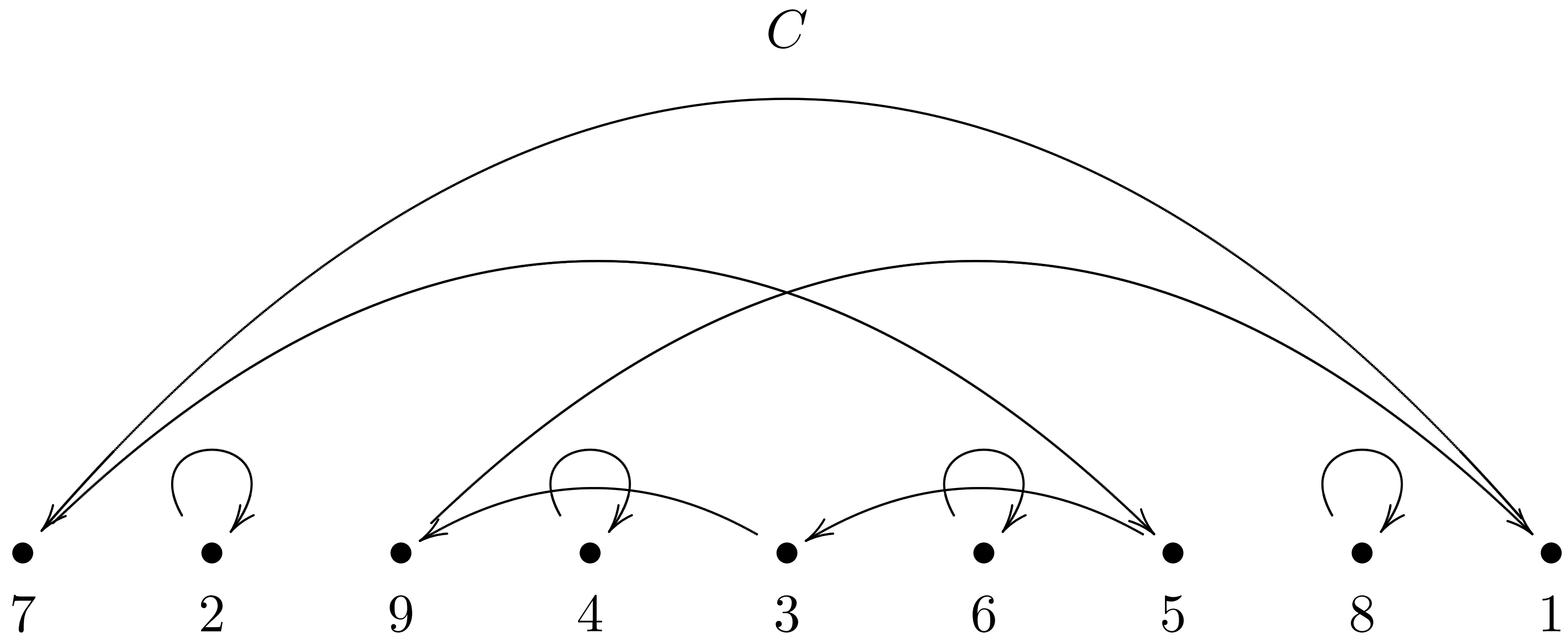
$$\begin{array}{ll}
(\; 3 \; \boxed{\mathit{2}\;\mathit{5}} \; \boxed{\mathit{4}\;\mathit{7}\;\mathit{6}\;\mathit{9}\;\mathit{8}\;\mathit{1}} \;) & 1 \\
(\; \boxed{\mathit{3}\;\mathit{4}} \; \boxed{\mathit{7}\;\mathit{6}\;\mathit{9}\;\mathit{8}\;\mathit{1}\;\mathit{2}} \; 5 \;) & 2 \\
(\; 7 \; \boxed{\mathit{6}\;\mathit{9}} \; \boxed{\mathit{8}\;\mathit{1}\;\mathit{2}\;\mathit{3}\;\mathit{4}\;\mathit{5}} \;) & 3 \\
(\; \boxed{\mathit{7}\;\mathit{8}} \; \boxed{\mathit{1}\;\mathit{2}\;\mathit{3}\;\mathit{4}\;\mathit{5}\;\mathit{6}} \; 9 \;) & 4 \\
(\; 1\;2\;3\;4\;5\;6\;7\;8\;9 \;) &
\end{array}$$

$\Rightarrow d(\pi) = 4 = 5 - (5 \bmod 2)$.

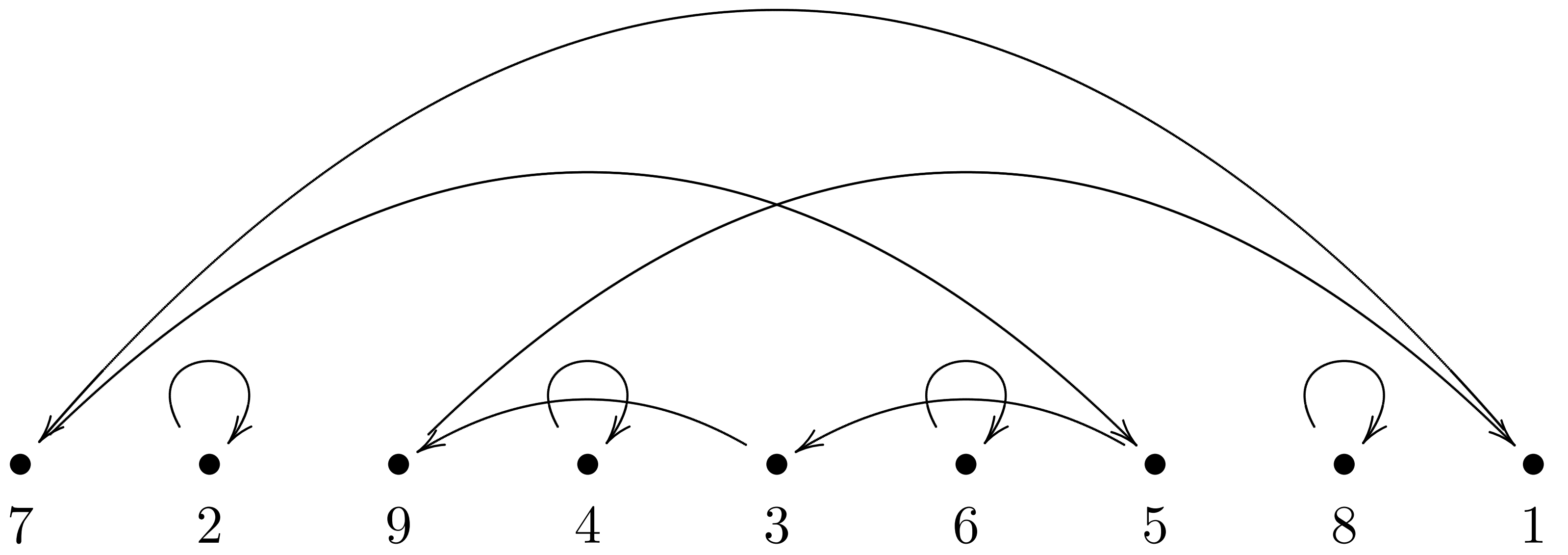
β−permutations

Definition 3 *A* β−permutation *is a reduced permutation that fixes all even elements and whose odd elements form one unoriented cycle in* Γ*.*

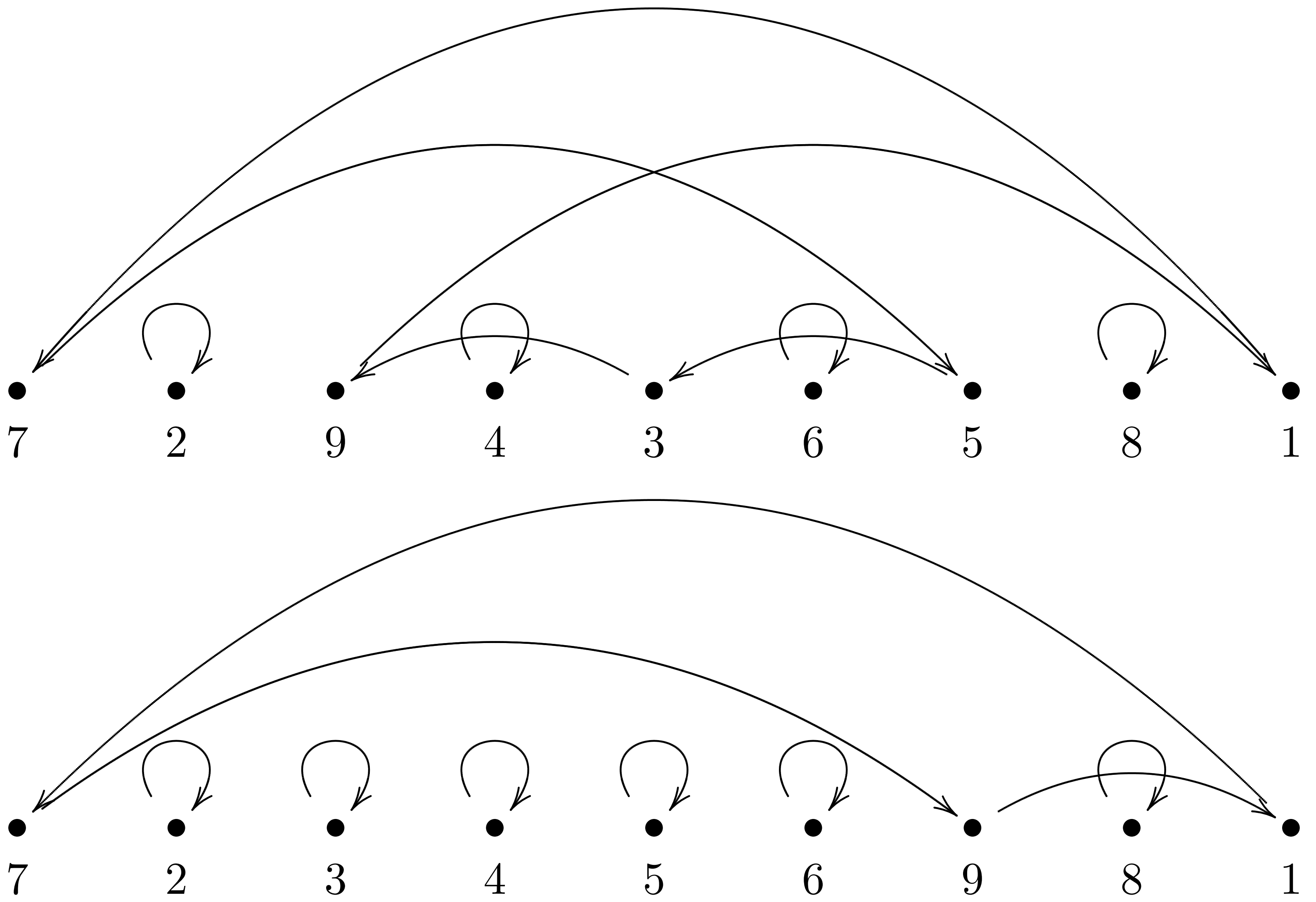
Example 8



Path contraction



Path contraction



Distance of β−permutations

General strategy:

1. π (β−permutation) $\longrightarrow \pi'$ (no crossing);
2. π' reduces to an α−permutation σ whose distance is known (Proposition 2);
3. $d(\pi) \leq d(\pi, \pi') + d(\pi')$;
4. $d(\pi) = d(\pi, \pi') + d(\pi')$ because the lower bound of Lemma 1 is reached.

Proposition 3 *For every β−permutation π in S_n:*

$$d(\pi) = n - c_{odd}(\Gamma(\pi)) = |C| - (|C| \bmod 2) \quad .$$

Distance of γ−permutations

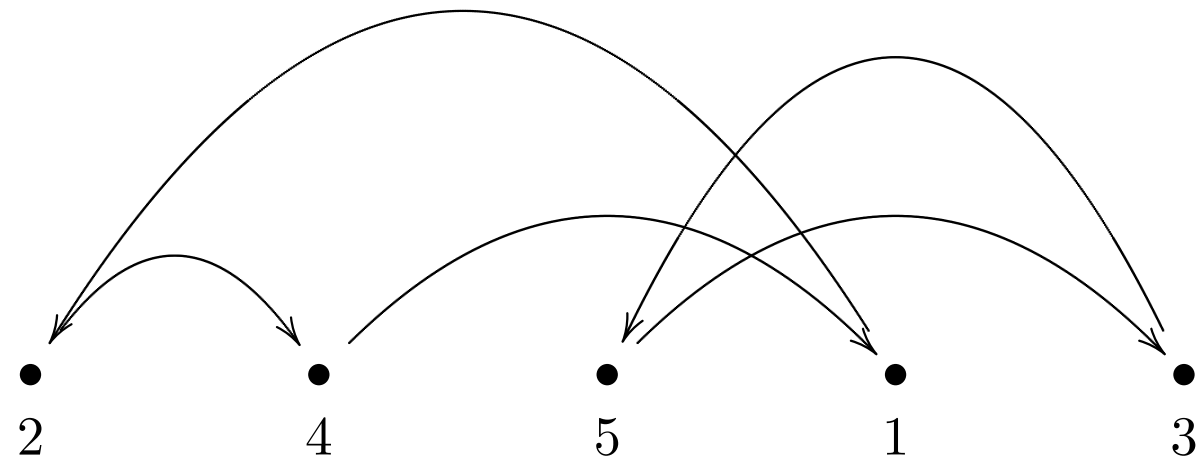
Every cycle of $\Gamma(\pi)$ is either oriented or unoriented and can be sorted independently – one by one. We have:

$$\begin{aligned} d(\pi) &\leq \sum_{i=1}^{c(\Gamma(\pi))} |C_i| - (|C_i| \bmod 2) \\ &= \sum_{C_{i_1} \in \, odd(\Gamma(\pi))} (|C_{i_1}| - 1) + \sum_{C_{i_2} \in \, even(\Gamma(\pi))} |C_{i_2}| \\ &= \sum_{i=1}^{c(\Gamma(\pi))} |C_i| - c_{odd}(\Gamma(\pi)) \\ &= n - c_{odd}(\Gamma(\pi)) \end{aligned}$$

which equals the lower bound of Lemma 1.

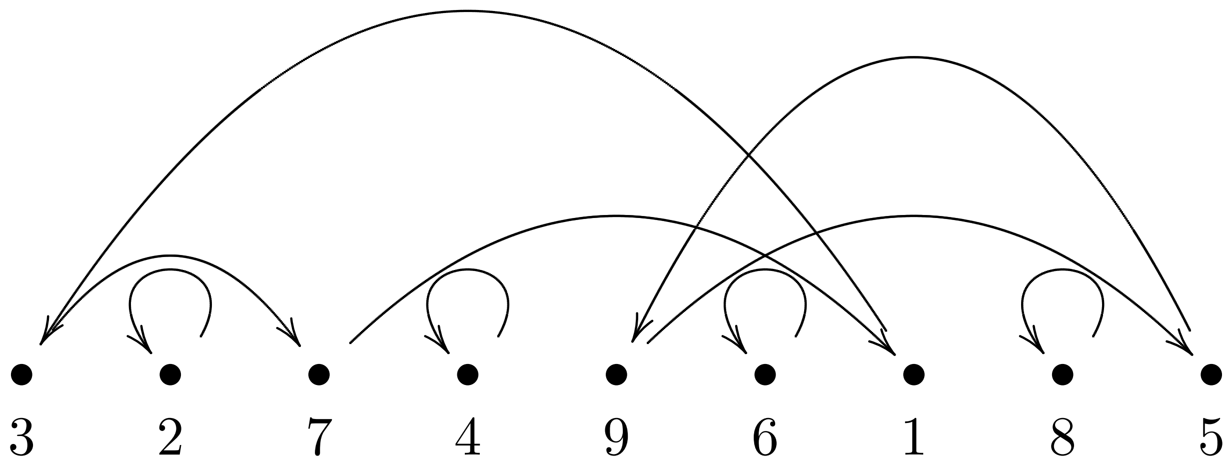
A new upper bound

- Every permutation $\pi \neq \iota$ can be obtained by removing k 1−cycles from the Γ−graph of a γ−permutation σ.



A new upper bound

- Every permutation $\pi \neq \iota$ can be obtained by removing k 1−cycles from the Γ−graph of a γ−permutation σ.



A new upper bound

- Every permutation $\pi \neq \iota$ can be obtained by removing k $1-$cycles from the $\Gamma-$graph of a $\gamma-$permutation σ.
- The independent cycle elimination that worked for σ still works for π (but is not necessarily optimal anymore).
- Therefore:

$$\begin{aligned} d(\pi) \leq d(\sigma) &= n + k - c_{odd}(\Gamma(\sigma)) \\ &= n + k - c_{odd}(\Gamma(\pi)) - k \\ &= n - c_{odd}(\Gamma(\pi)) \ . \end{aligned}$$

Theorem 3 $\forall\ \pi \in S_n$:

$$d(\pi) \leq n - c_{odd}(\Gamma(\pi)) \ . \qquad (1)$$

Comparison with other bounds

- Other upper bounds have been found by:
  - [Bafna and Pevzner, 1998]:

$$d(\pi) \leq \frac{3(n + 1 - c_{odd}(G(\pi)))}{4} \; . \tag{2}$$

  - [Dias et al., 2000]:

$$d(\pi) \leq \frac{3}{4}\, b(\pi) \; . \tag{3}$$

  - [Eriksson et al., 2001]:

$$d(\pi) \leq \begin{cases} \left\lceil \frac{2n}{3} \right\rceil & \text{if } n < 9 \; ; \\ \left\lfloor \frac{2n-2}{3} \right\rfloor & \text{if } n \geq 9 \; . \end{cases} \tag{4}$$

- How does our upper bound compare with earlier results?

Comparison with other bounds

n	Number of permutations	(1)≤(2)	(1)≤(3)	(1)≤(4)
3	6	2	1	[6]
4	24	8	[8]	15
5	120	45	24	31
6	720	304	⁞49⁞	495
7	5040	2055	722	1611
8	40320	[17879]	3094	4355
9	362880	⁞104392⁞	60871	⁞10243⁞
		(28–44%)	(6–33%)	(2–100%)

([x] = best case, ⁞x⁞ = worst case; (1) = the new upper bound, (2) = Bafna and Pevzner's, (3) = Dias et al.'s, (4) = Eriksson et al.'s)

Future plans

- Further improvement of the upper bound of Equation (1);
  - a good improvement can be achieved partly through torism [Hultman, 1999], but it is heuristic;
- Complexity?
  - Increase the number of polynomial-time solvable instances;
  - Characterize hardest cases;
- Help find diameter;
- Extension of those results to other rearrangement problems?
  - the graph of a permutation has proved useful when additionally dealing with fusions and fissions [Dias and Meidanis, 2001].

References

[Bafna and Pevzner, 1998] Bafna, V. and Pevzner, P. A. (1998). Sorting by transpositions. *SIAM J. Discrete Math.*, 11(2):224–240 (electronic).

[Christie, 1998] Christie, D. A. (1998). *Genome Rearrangement Problems*. PhD thesis, University of Glasgow, Scotland.

[Dias and Meidanis, 2001] Dias, Z. and Meidanis, J. (2001). Genome Rearrangements Distance by Fusion, Fission, and Transposition is Easy. In *Proceedings of SPIRE'2001 - String Processing and Information Retrieval*, pages 250–253.

[Dias et al., 2000] Dias, Z., Meidanis, J., and Walter, M. E. M. T. (2000). A New Approach for Approximating The Transposition Distance. In *Proceedings of SPIRE'2000 - String Processing and Information Retrieval*, La Coruna, Espagne.

[Elias and Hartman, 2005] Elias, I. and Hartman, T. (2005). A 1.375−Approximation Algorithm for Sorting by Transpositions. In Casadio, R. and Myers, G., editors, *Proceedings of the Fifth Workshop on Algorithms in Bioinformatics*, volume 3692 of *Lecture Notes in Computer Science*, pages 204–215, Mallorca, Spain. Springer-Verlag.

[Eriksson et al., 2001] Eriksson, H., Eriksson, K., Karlander, J., Svensson, L., and Wästlund, J. (2001). Sorting a bridge hand. *Discrete Mathematics*, 241(1-3):289–300. Selected papers in honor of Helge Tverberg.

[Hultman, 1999] Hultman, A. (1999). Toric Permutations. Master's thesis, Dept. of Mathematics, KTH, Stockholm, Sweden.