On the average number of reversals needed to sort signed permutations

[1]**Thaynara Arielly de Lima** & [1,2]**Mauricio Ayala-Rincón**

Departamentos de Matemática[1] & Ciência da Computação[2]

Universidade de Brasília

Authors funded by CAPES and CNPq

"XI Seminário Informal, mas Formal!"

November 2013

Genome Rearrangement

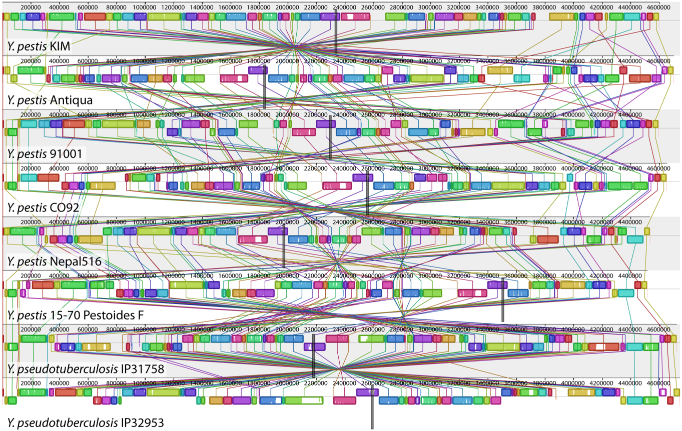


Figure: A genome alignment of eight Yersinia (Figure in [DMR08]).

Reversals

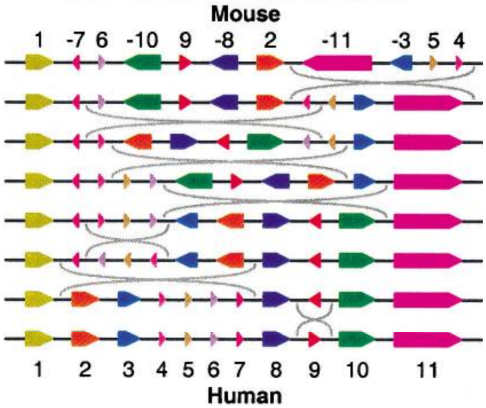


Figure: A most parsimonious rearrangement scenario for human and mouse X-chromosomes (Figure in [PT03]).

Genome Rearragement Problem

Restricted to reversals...

Finding the MINIMUM number of reversals needed to transform a permutation into identity permutation.

Complexities

Operation	Example	Complexity
Reversals on signed permutations	$+1+2\underline{-5-4-3}+6$ $+1+2+3+4+5+6$	Polynomial
Reversals on unsigned permutations	$12\underline{543}6$ 123456	$\mathcal{NP}$-hard

This work is based in sorting signed permutations by reversals.

Average number of block interchange

Block Interchange	1 $\boxed{45}$ 3 $\boxed{2}$ 6 123456	**Polynomial**

Consider an unsigned permutation $\pi = \pi_1\ \pi_2 \cdots \pi_n$

Miklós Bóna & Ryan Flynn [BF09] showed that:

$$a_n = \frac{n - \frac{1}{\lfloor (n+2)/2 \rfloor} - \sum_{i=2}^{n} \frac{1}{i}}{2}$$

where a_n = average number of Block Interchange needed to sort permutations of length n.

Our goal

Search for the average number of reversals needed to sort signed permutations.

Breakpoint Graph

Consider $\pi = \pi_1\ \pi_2 \cdots \pi_n$. Extend π by adding $\pi_0 = +0$ e $\pi_{n+1} = -0$. Associate to each π_i the pair $-\pi_i \quad +\pi_i$.

+0 -2 +2 -3 +3 +1 -1 -4 +4 +5 -5 -0

Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

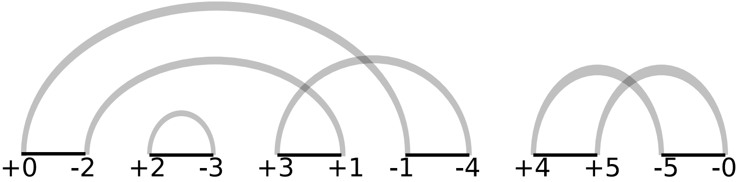


Figure: Breakpoint Graph of permutation$\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

i) **there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;**

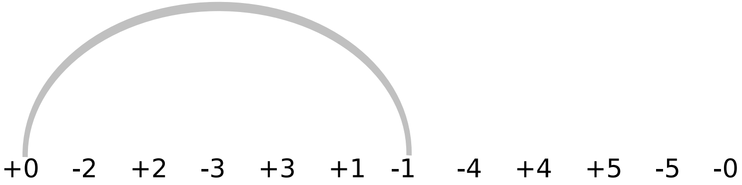


Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

i) **there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;**

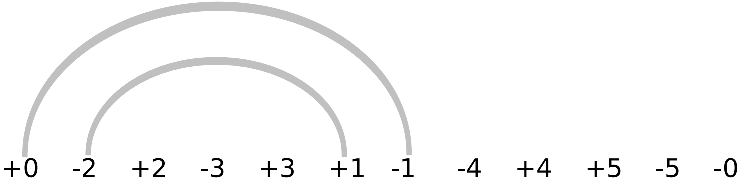


Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

i) **there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;**

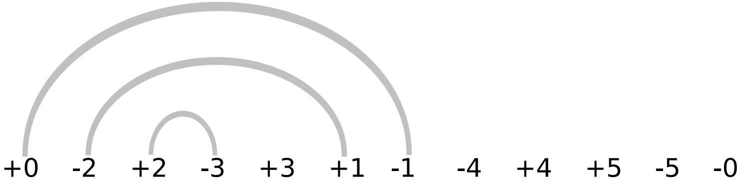


Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

i) **there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;**

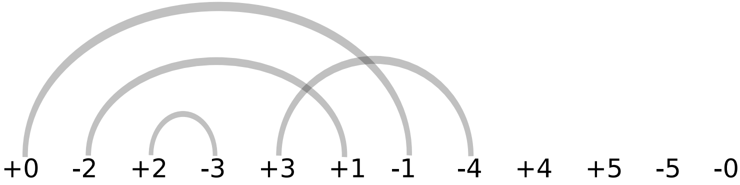


Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

i) **there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;**

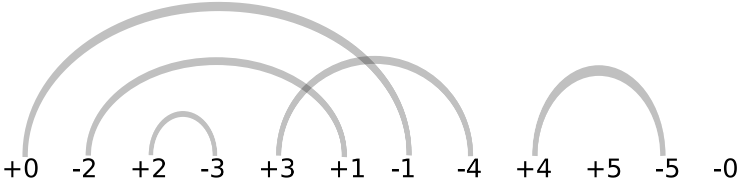


Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

i) **there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;**

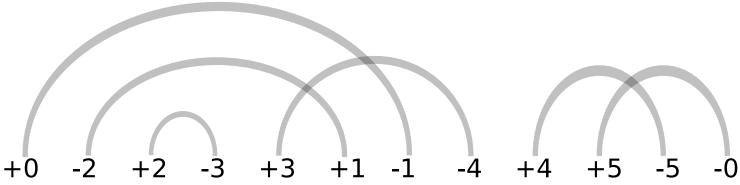


Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n+2$ vertices such that:

i) *there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;*

ii) **there is a black edge between vertices with labels π_i and $-\pi_{i+1}, 0 \leq i < n$ and π_n and π_{n+1}.**

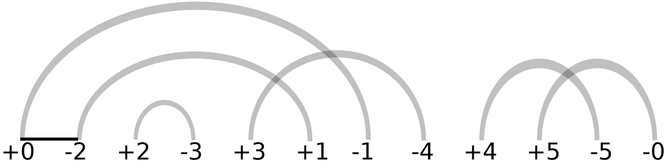


Figure: Breakpoint Graph of permutation $\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Breakpoint Graph

Definition (Breakpoint Graph)

The Breakpoint Graph $G(\pi)$ of a permutation π is a bi-colored graph with $2n + 2$ vertices such that:

i) *there is a gray edge between vertices with labels $+i$ and $-(i+1)$, $0 \leq i < n$ and $+n$ and -0;*

ii) **there is a black edge between vertices with labels π_i and $-\pi_{i+1}, 0 \leq i < n$ and π_n and π_{n+1}.**

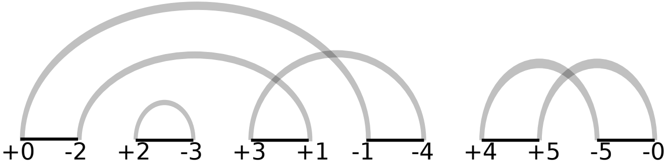


Figure: Breakpoint Graph of permutation$\pi = +2 \quad +3 \quad -1 \quad +4 \quad -5$

Reversal distance and Breakpoint Graphs

$$b(\pi) - c(\pi) \leq d(\pi) \leq b(\pi) - c(\pi) + 1$$

where $b(\pi)$ = number of black edges in $G(\pi) = n + 1$
$c(\pi)$ = number of alternating cycles in $G(\pi)$ and
$d(\pi)$ = reversal distance.

Finding the average number of reversals needed to sort permutations is equivalent to find the average number of alternating cycles in Breakpoint Graphs of all permutations of length n.

Plan:

Associate to each permutation π a specific permutation θ, such that the number of cycles in θ is related with the number of alternating cycles in Breakpoint Graph of π.

Cycles in Breakpoint Graphs and cycles in permutations

Given $\pi = \pi_1\ \pi_2 \ldots \pi_n$

Associate $\pi^\circ = (+0\ \pi_1 \ldots \pi_n)(-\pi_n\ \ldots -\pi_1\ -0)$

Fix $\gamma_n = (+0\ -0)(+1\ -1)\ldots(+i\ -i)\ldots(+n\ -n)$

$\sigma_n = (+0\ +1\ \ldots + n)(-n\ \ldots -1\ -0)$

Note that $\boldsymbol{\gamma_n \pi^\circ = (+0\ -\pi_1)(\pi_1\ -\pi_2)\ldots(\pi_j\ -\pi_{j+1})\ldots(\pi_n\ -0)}$

$\gamma_n\sigma_n = (+0\ -1)(+1\ -2)\ldots(+i\ -(i+1))\ldots(+n\ -0)$

Cycles in Breakpoint Graphs and cycles in permutations

For $\pi = +2 \ +3 \ -1 \ +4 \ -5$

$$\pi^\circ = (+0 \ +2 \ +3 \ -1 \ +4 \ -5)(+5 \ -4 \ +1 \ -3 \ -2 \ -0)$$

$$\boldsymbol{\gamma_5 \pi^\circ = (+0 \ -2)(+2 \ -3)(+3 \ +1)(-1 \ -4)(+4 \ +5)(-5 \ -0)}$$

$$\gamma_5 \sigma_5 = (+0 \ -1)(+1 \ -2)(+2 \ -3)(+3 \ -4)(+4 \ -5)(+5 \ -0)$$

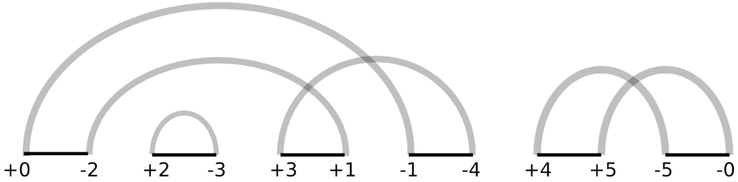


Figure: Breakpoint Graph of permutation $\pi = +2 \ +3 \ -1 \ +4 \ -5$

Cycles in Breakpoint Graphs and cycles in permutations

If the number of alternating cycles in $G(\pi)$ is k then the number of cycles in permutation $(\gamma_n\pi^\circ)(\gamma_n\sigma_n)$ is $2k$.

For $\pi = +2\ +3\ -1\ +4\ -5$

$$\boldsymbol{(\gamma_5\pi^\circ)}(\gamma_5\sigma_5) = (+0\ -4\ +1)(-1\ -2\ +3)(+2)(-3)(-0\ +4)(+5\ -5)$$

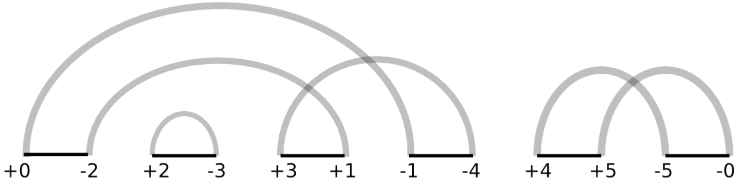
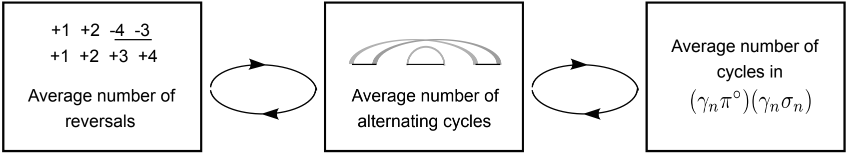


Figure: Breakpoint Graph of permutation $\pi = +2\quad +3\quad -1\quad +4\quad -5$

Finding the average number of alternating cycles in Breakpoint Graphs is equivalent to find the average number of cycles in permutations $(\gamma_n\pi^\circ)(\gamma_n\sigma_n)$ over all permutations π of length n.



Building a permutation of $n + 1$ elements

Consider a permutation $\pi = \pi_1 \dots \pi_n$.

One can build a permutation π' by inserting the element $\pi'_{i+1} = \pm(n+1)$ between two specific entries $\pi_i = a$ and $\pi_{i+1} = b$ of π.

$$\pi'_{i+1}$$

$$\pi = \pi_1 \dots a \quad b \dots \pi_n$$

$$\Downarrow$$

$$\pi' = \pi_1 \dots a \ \pi'_{i+1} \ b \dots \pi_n$$

Figure: The permutation $\pi = \pi_1 \dots \pi_i \ a \ b \dots \pi_n$ and the permutation $\pi' = \pi_1 \dots \pi_i \ a \ \pi'_{i+1} \ b \dots \pi_n$.

Notation

Denote as

$$c(\Gamma(\pi)) = \text{ number of cycles of a permutation } \pi$$

$$(\gamma_n \pi^\circ)(\gamma_n \sigma_n) = \theta$$

$$(\gamma_{n+1} \pi'^\circ)(\gamma_{n+1} \sigma_{n+1}) = \theta'$$

Behavior of $c(\Gamma((\gamma_n\pi^\circ)(\gamma_n\sigma_n)))$

Proposition (T.A. de Lima & M.A. Rincón)

Let a, b, π *and* π' *be as before,* $\theta = (\gamma_n\pi^\circ)(\gamma_n\sigma_n)$ *and* $\theta' = (\gamma_{n+1}\pi'^\circ)(\gamma_{n+1}\sigma_{n+1})$. *Thus* $c(\Gamma(\theta')) =$

1. $c(\Gamma(\theta)) - 2$,
2. $c(\Gamma(\theta))$,
3. $c(\Gamma(\theta)) + 2$,

Behavior of $c(\Gamma((\gamma_n\pi^\circ)(\gamma_n\sigma_n)))$

Proposition (T.A. de Lima & M.A. Rincón)

Let a, b, π *and* π' *be as before,* $\theta = (\gamma_n\pi^\circ)(\gamma_n\sigma_n)$ *and* $\theta' = (\gamma_{n+1}\pi'^\circ)(\gamma_{n+1}\sigma_{n+1})$. *Thus* $c(\Gamma(\theta')) =$

1. $c(\Gamma(\theta)) - 2$, *if* $\begin{cases} a \text{ and } +n \text{ are not in the same cycle, } -b \text{ and } +n \text{ are not in the} \\ \text{same cycle in } \theta; \end{cases}$
2. $c(\Gamma(\theta))$,
3. $c(\Gamma(\theta)) + 2$,

Behavior of $c(\Gamma((\gamma_n\pi^\circ)(\gamma_n\sigma_n)))$

Proposition (T.A. de Lima & M.A. Rincón)

Let a, b, π *and* π' *be as before,* $\theta = (\gamma_n\pi^\circ)(\gamma_n\sigma_n)$ *and* $\theta' = (\gamma_{n+1}\pi'^\circ)(\gamma_{n+1}\sigma_{n+1})$*. Thus* $c(\Gamma(\theta')) =$

1. $c(\Gamma(\theta)) - 2$, *if*
   $\begin{cases} a \text{ and } +n \text{ are not in the same cycle, } -b \text{ and } +n \text{ are not in the same cycle in } \theta; \end{cases}$
2. $c(\Gamma(\theta))$, *if*
   $\begin{cases} -b \text{ and } +n \text{ are not in the same cycle, } a \text{ and } +n \text{ are in the same cycle in } \theta \text{ and, } \pi'_{i+1} = -(n+1); \text{ or} \\ -b \text{ and } +n \text{ are in the same cycle, } a \text{ and } +n \text{ are not in the same cycle in } \theta \text{ and, } \pi'_{i+1} = +(n+1); \end{cases}$
3. $c(\Gamma(\theta)) + 2$,

Behavior of $c(\Gamma((\gamma_n\pi^\circ)(\gamma_n\sigma_n)))$

Proposition (T.A. de Lima & M.A. Rincón)

Let a, b, π and π' be as before, $\theta = (\gamma_n\pi^\circ)(\gamma_n\sigma_n)$ and $\theta' = (\gamma_{n+1}\pi'^\circ)(\gamma_{n+1}\sigma_{n+1})$. Thus $c(\Gamma(\theta')) =$

1. $c(\Gamma(\theta)) - 2$, *if*
   $\begin{cases} a \text{ and } +n \text{ are not in the same cycle, } -b \text{ and } +n \text{ are not in the same cycle in } \theta; \end{cases}$
2. $c(\Gamma(\theta))$, *if*
   $\begin{cases} -b \text{ and } +n \text{ are not in the same cycle, } a \text{ and } +n \text{ are in the same cycle in } \theta \text{ and, } \pi'_{i+1} = -(n+1)\text{; or} \\ -b \text{ and } +n \text{ are in the same cycle, } a \text{ and } +n \text{ are not in the same cycle in } \theta \text{ and, } \pi'_{i+1} = +(n+1); \end{cases}$
3. $c(\Gamma(\theta)) + 2$, *if*
   $\begin{cases} -b \text{ and } +n \text{ are in the same cycle in } \theta \text{ and } \pi'_{i+1} = -(n+1)\text{; or} \\ a \text{ and } +n \text{ are in the same cycle in } \theta \text{ and } \pi'_{i+1} = +(n+1). \end{cases}$

Recurrence formula for the average

Lemma (T.A. de Lima & M.A. Rincón)

Denote as:

- $P_i, 1 \leq i \leq 3$ *the probability that the item i in previous proposition occurs;*
- a_n *the average number of cycles in permutations* $(\gamma_n \pi^\circ)(\gamma_n \sigma_n)$, $|\pi| = n$.

So,

$$\mathbf{a}_{n+1} = P_1(\mathbf{a}_n - 2) + P_2\mathbf{a}_n + P_3(\mathbf{a}_n + 2)$$

Recurrence formula for the average

By mathematical computations

$$\begin{aligned} \mathbf{a}_{n+1} &= P_1(\mathbf{a}_n - 2) + P_2\mathbf{a}_n + P_3(\mathbf{a}_n + 2) \\ &= \mathbf{a}_n + \frac{3}{n+1}\sum_{i=-n}^{+n} P(\mathcal{A}^j_{+n}) - 2 \end{aligned}$$

where $P(\mathcal{A}^j_{+n})$ is the probability that the event "given $a = j$, a and $+n$ are in the same cycle" occurs.

Goal

To obtain or estimate the average number of cycles in θ, and consequently the average number of alternating cycles in $G(\pi)$, the expression $\sum_{i=-n}^{+n} P(\mathcal{A}^{j}_{+n})$ should be either solved or bounded.

Computational experiments

a \ n	1	2	3	4	5	6	7	8
+0	0	2	12	104	1072	13224	188624	3064000
-1	1	3	18	140	1384	16428	228248	3628960
+1	2	2	16	128	1280	15368	215072	3441248
-2		3	16	128	1280	15356	215024	3440336
+2		8	12	128	1272	15336	214736	3437856
-3			18	128	1284	15372	215192	3442032
+3			48	104	1280	15336	214976	3440256
-4				140	1280	15372	215072	3440832
+4				384	1072	15368	214736	3440256
-5					1384	15356	215192	3440832
+5					3840	13224	215072	3437856
-6						16428	215024	3442032
+6						46080	188624	3441248
-7							228248	3440336
+7							645120	3064000
-8								3628960
+8								10321920

Table: Frequency of occurrence of a and $+n$ in the same cycle in θ

An interesting property

$|A^{j}_{+n}|$ = how many times the elements i and $+n$ are in the same cycle in θ

$$\sum_{j=-n}^{+n} |\mathcal{A}^{j}_{+n}| = |\mathcal{A}^{-(n+1)}_{+(n+1)}|$$

$$\sum_{j=-1}^{+1} |\mathcal{A}^{j}_{+1}| = 0 + 1 + 2 = |\mathcal{A}^{-2}_{+2}| = 3$$

$$\sum_{j=-2}^{+2} |\mathcal{A}^{j}_{+2}| = 2 + 3 + 2 + 3 + 8 = |\mathcal{A}^{-3}_{+3}| = 18$$

An interesting property

$$
\begin{aligned}
P(\mathcal{A}_{+(n+1)}^{-(n+1)}) &= \frac{\sum_{j=-n}^{+n} |\mathcal{A}_{+n}^{j}|}{2^{n+1}(n+1)!} \\
&= \frac{1}{2(n+1)} \sum_{j=-n}^{+n} \frac{|\mathcal{A}_{+n}^{j}|}{2^n n!} \\
&= \frac{1}{2(n+1)} \sum_{j=-n}^{+n} P(\mathcal{A}_{+n}^{j})
\end{aligned}
$$

$$
\sum_{j=-n}^{j=+n} P(\mathcal{A}_{+n}^{j}) = 2(n+1) P(\mathcal{A}_{+(n+1)}^{-(n+1)})
$$

Average number of reversal distance

Proposition (T.A. de Lima & M.A. Rincón)

For all positive integer n

$$a_{n+1} = a_n + 6P(\mathcal{A}_{+(n+1)}^{-(n+1)}) - 2$$

Theorem (T.A. de Lima & M.A. Rincón)

For all positive integer n

$$d_n = n + 2 - \frac{a_{n-1}}{2} - 3P(\mathcal{A}_{+n}^{-n})$$

where d_n denote the average number of reversals needed to sort signed permutations.

Average number of reversal distance

Proposition (T.A. de Lima & M.A. Rincón)

For all positive integer n

$$a_{n+1} = a_n + 6P(\mathcal{A}_{+(n+1)}^{-(n+1)}) - 2$$

Theorem (T.A. de Lima & M.A. Rincón)

For all positive integer n

$$d_n = n + 2 - \frac{a_{n-1}}{2} - 3P(\mathcal{A}_{+n}^{-n})$$

where d_n denote the average number of reversals needed to sort signed permutations.

Results for small permutations

Table: Average number of cycles in $(\gamma_n\pi^\circ)(\gamma_n\sigma_n)$, for $1 \leq n \leq 9$.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
3	3,25	3,5	3,6875	3,85	3,9891	4,1119	4,2214	4,3205

Table: Average number of reversal distance, for $1 \leq n \leq 9$.

d_1	0,5 or 1,5
d_2	1,375 or 2,375
d_3	2,25 or 3,25
d_4	3,15625 or 4,15625
d_5	4,075 or 5,075
d_6	5,00545 or 6,00545
d_7	5,94405 or 6,94405
d_8	6,8893 or 7,8893
d_9	7,83975 or 8,83975

Conclusion

- Genome rearrangement is an important tool to study mutations in live organisms and, consequently, reconstruction of evolutionary chains;
- The average number of *operations* needed to sort permutations is an important problem, because this average shows the quality of approximate solutions;

Conclusion

For unsigned permutations:

Perm. length	[JGJM13]	Average (%)
10	5,810	58,10
20	12,940	64,70
30	20,589	68,63
40	28,254	70,64
50	36,291	72,58
60	44,633	74,39
70	52,949	75,64
80	60,887	76,11
90	69,555	77,28
100	78,096	78,10
110	86,702	78,82
120	95,258	79,38
130	104,582	80,45
140	113,539	81,10
150	122,671	81,78

Table: Average number of reversals using a set of 100 permutations (Table of [JGJM13])

Conclusion

- We transform a graph problem (finding the average number of alternating cycles in a graph) into an algebraic problem (finding the average number of cycles in $(\gamma_n\pi^\circ)(\gamma_n\sigma_n)$);
- We obtain the recurrence formula for the average number of cycles in $(\gamma_n\pi^\circ)(\gamma_n\sigma_n)$:

$$a_{n+1} = \mathbf{a}_n + 6P(\mathcal{A}_{+(n+1)}^{-(n+1)}) - 2$$

- Consequently, we obtain an expression for average number of reversals needed to sort signed permutations.

Future work

- Obtain or estimate the expression $P(\mathcal{A}_{+(n+1)}^{-(n+1)})$;

$\Rightarrow$ This way, obtain or estimate the average number of reversals needed to sort signed permutations;

- Study the average number of reversals needed to sort unsigned permutations.

References

A.E.Darling, I. Miklós & M.A.Ragan.
Dynamics of Genome Rearrangement in Bacterial Populations.
PLoS Genetics, 4(7): e1000128, 2008.

J.L. Soncco-Álvarez, G.M. Almeida, J. Becker & M. Ayala-Rincón.
Parallelization and Virtualization of Genetic Algorithms for Sorting Permutation by Reversals.
Preprint submitted to NaBIC 2013 .

J.P.Doignon & A.Labarre.
On Hultman numbers.
Journal of Integer Sequences 10. Article 07.6.2, 2007.

M. Bóna & R. Flynn.
The average number of block interchanges needed to sort a permutation and a recent result of Stanley.
Information Processing Letters. pages 927-31, 2009.

References

P. Pevzner & G. Tesler
Genome Rearrangements in Mammalian Evolution: Lessons from Human and Mouse Genomes.
Genome Research Journal, 13:37-45, 2003.

Z.Dias & J.Meidanis.
Rearranjo de Genomas: Uma coletânea de Artigos.
Tese de Doutorado,Instituto de Computação, UNICAMP, 2002.

Thank you!

thaynaradelima@gmail.com