



The binary matroid of genome arrangements

Eric Tannier, INRIA, LBBE, Lyon

Dedicated to András Sebő, Grenoble, April 25, 2014

Université Grenoble I – Joseph Fourier

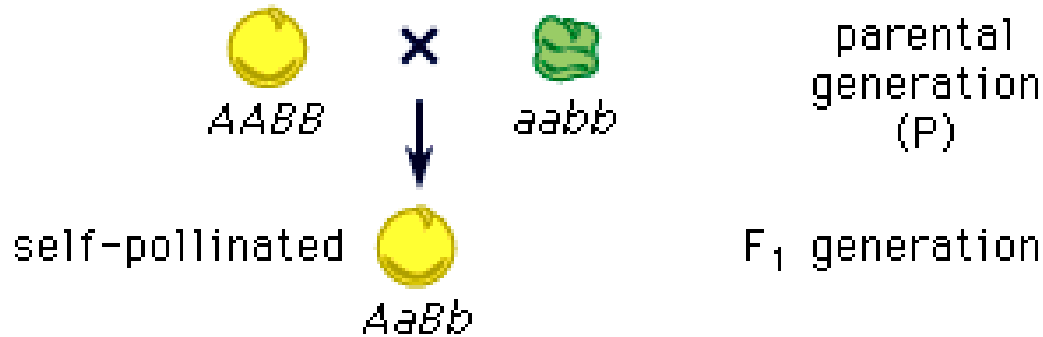
Sur quelques problèmes de recouvrement et empilement dans les graphes et les matroïdes

Éric TANNIER

Thèse présentée pour l'obtention du titre de Docteur de l'Université Joseph Fourier
École Doctorale Mathématiques, Informatique, Sciences et Technologies de l'Information
Formation Recherche Opérationnelle, Combinatoire et Optimisation

1/ Mendel law

2 values for 2 characters



| ♀ \ ♂ | | pollen | | | |
|--------|----|--------|------|------|------|
| | | AB | Ab | aB | ab |
| ovules | AB | AABB | AABb | AaBB | AaBb |
| | Ab | AABb | AAbb | AaBb | Aabb |
| | aB | AaBB | AaBb | aaBB | aaBb |
| | ab | AaBb | Aabb | aaBb | aabb |

9:3:3:1 ratio for the 4 different combinations

F₂ generation

2/ Genetic linkage



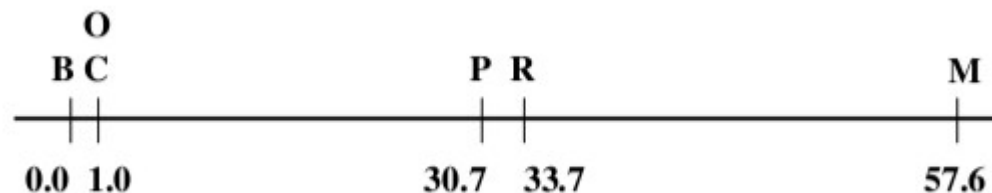
| Traits | NUMBER OF PROGENY | |
|------------------------------------|-------------------|-----------------------------|
| | Observed | Expected from 9:3:3:1 ratio |
| purple, long ($P/- \cdot L/-$) | 4831 | 3911 |
| purple, round ($I- P \cdot III$) | 390 | 1303 |
| red, long ($p/p \cdot L/-$) | 393 | 1303 |
| red, round ($p/p \cdot III$) | <u>1338</u> | <u>435</u> |
| | 6952 | 6952 |

3/ Linear embedding of genetic linkage

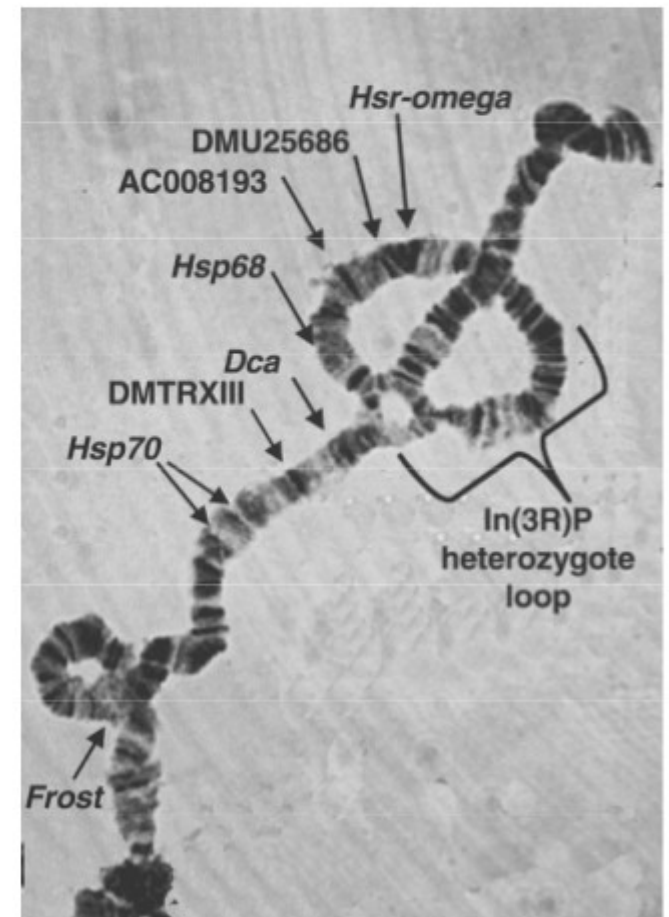
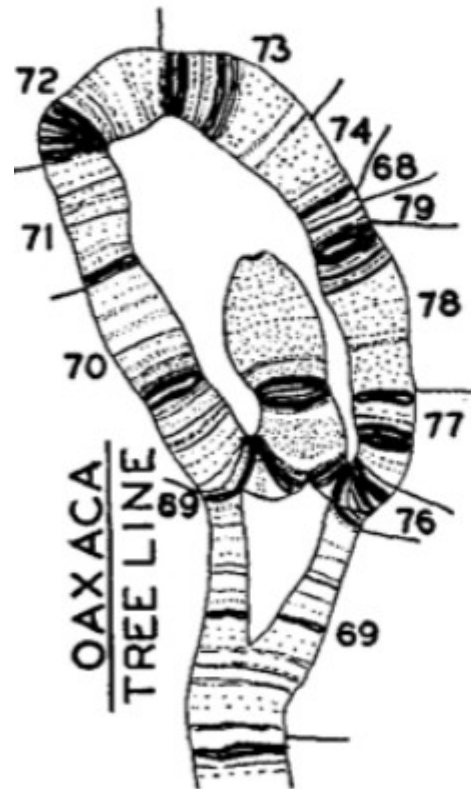
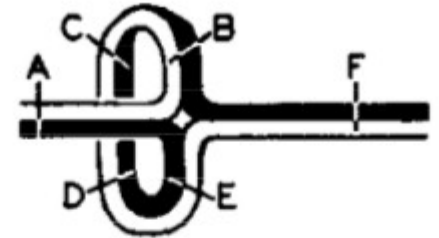
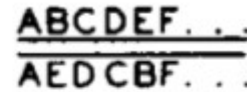
The linkage measure is "close" to a linear distance

$$d(a,c) = d(a,b) + d(b,c)$$

Factors are organized linearly (prediction of the structure of chromosomes)



Inversions were the first genomic mutations predicted, then discovered



Mathematical modeling of inversions, Sturtevant, Tan, 1937

If the *pseudoobscura* sequence in each arm is arbitrarily taken as an alphabetical one (A B C . . .), then the *melanogaster* sequences become:

| | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|-----|-----|---|---|---|---|---|-----|
| X | L | H | F | E | B | A | D | C | K | I | J | G | M | (7) |
| II L | D | E | F | A | C | B | (2) | | | | | | | |
| II R | A | C | E | B | F | D | (4) | | | | | | | |
| III L | C | F | E | B | A | D | (3) | | | | | | | |
| III R | A | E | B | C | F | D | G | (3) | | | | | | |

The numbers in parentheses represent the numbers of successive inversions necessary to turn these sequences into alphabetical ones (in the case of X we are not yet certain that six inversions may not be sufficient). The mathematical properties of series of letters subjected to the operation of successive inversions do not appear to have been worked out, so that we are so far unable to present a detailed analysis. It does appear, however, that the five arms (taken together) are definitely more alike in the two species than could result from chance alone.

Sorting by Reversals

0 7 5 3 -1 -6 -2 4 8

0 1 2 3 4 5 6 7 8

Sorting by Reversals

0 7 5 3 -1 -6 -2 4 8



0 1 -3 -5 -7 -6 -2 4 8

0 1 2 3 4 5 6 7 8

Sorting by Reversals

0 7 5 3 -1 -6 -2 4 8



0 1 -3 -5 -7 -6 -2 4 8



0 1 -3 -5 -4 2 6 7 8

0 1 2 3 4 5 6 7 8

Sorting by Reversals

0 7 5 3 -1 -6 -2 4 8



0 1 -3 -5 -7 -6 -2 4 8



0 1 -3 -5 -4 2 6 7 8



0 1 -3 -2 4 5 6 7 8

0 1 2 3 4 5 6 7 8

Sorting by Reversals

0 7 5 3 -1 -6 -2 4 8



0 1 -3 -5 -7 -6 -2 4 8



0 1 -3 -5 -4 2 6 7 8

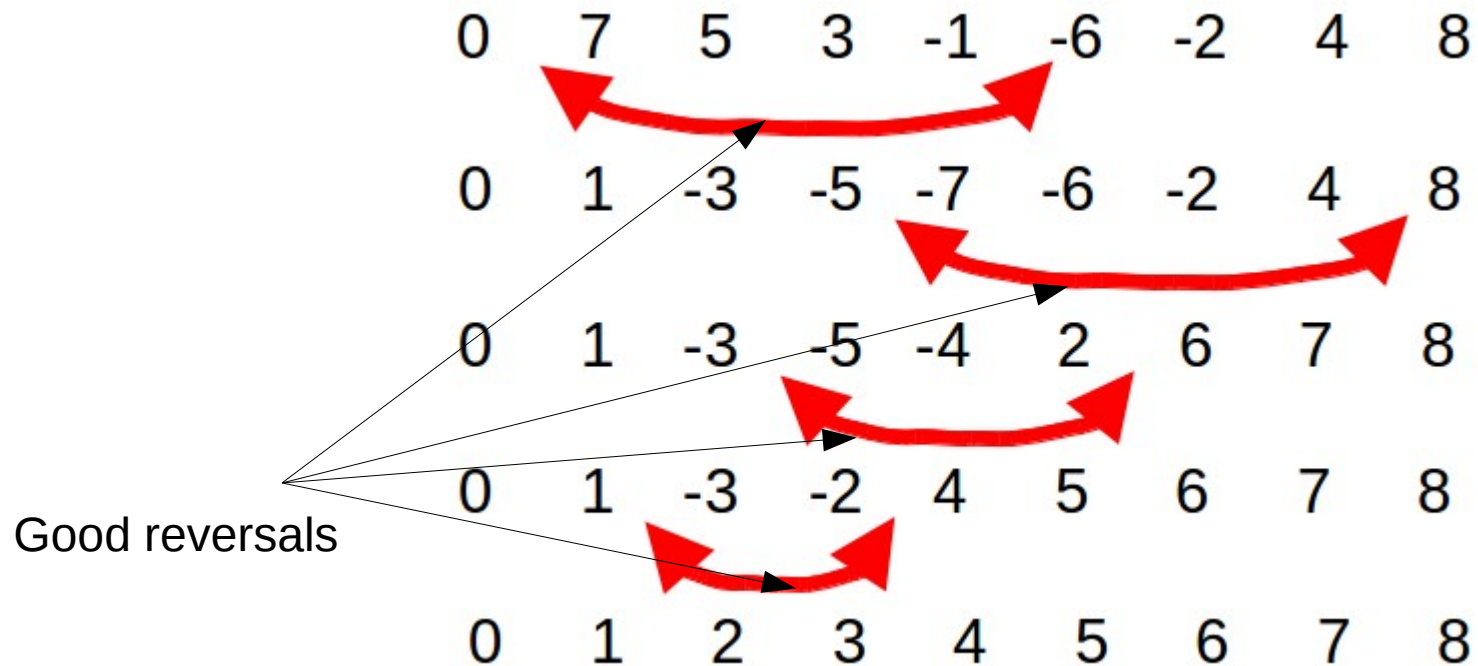


0 1 -3 -2 4 5 6 7 8



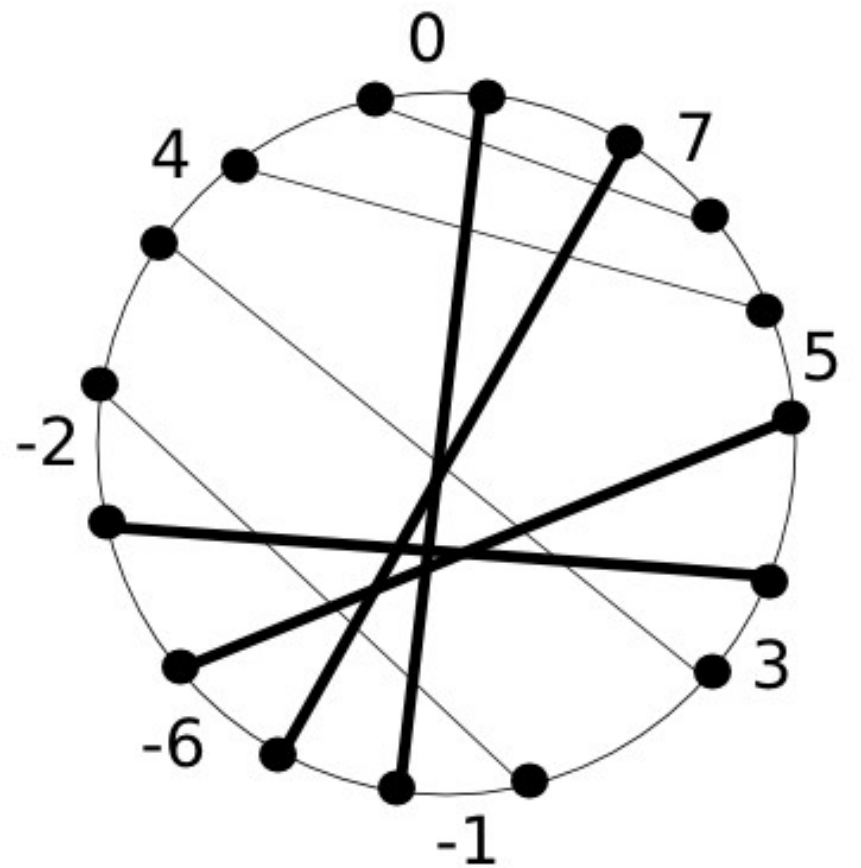
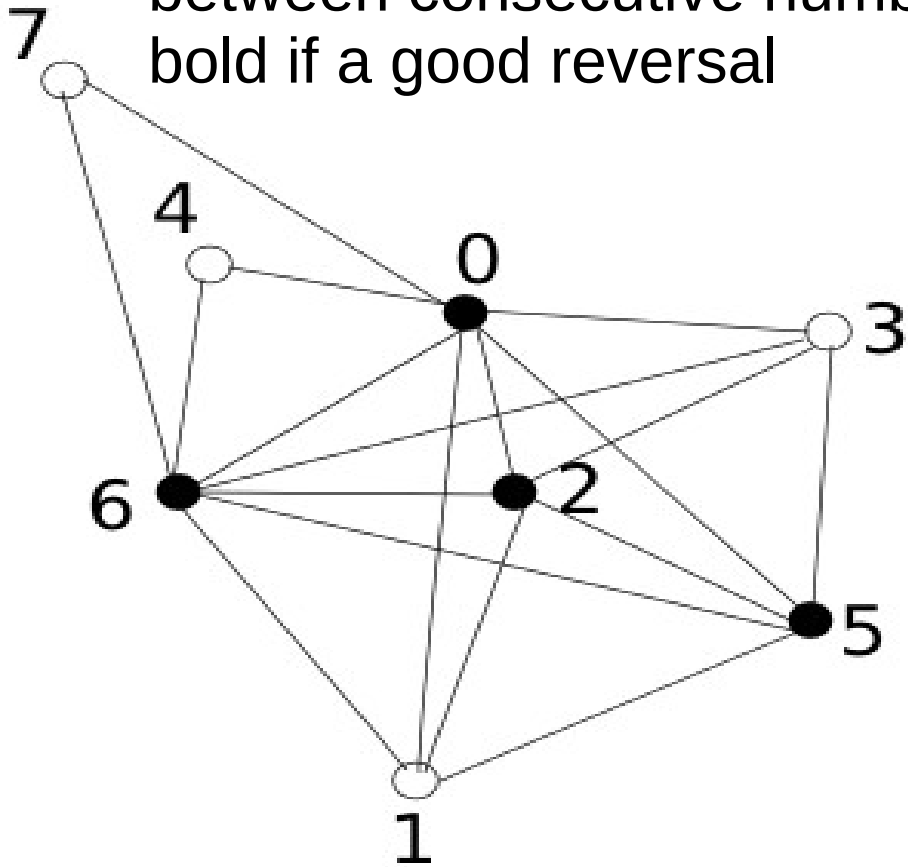
0 1 2 3 4 5 6 7 8

Good reversal : results in the contiguity of two consecutive numbers



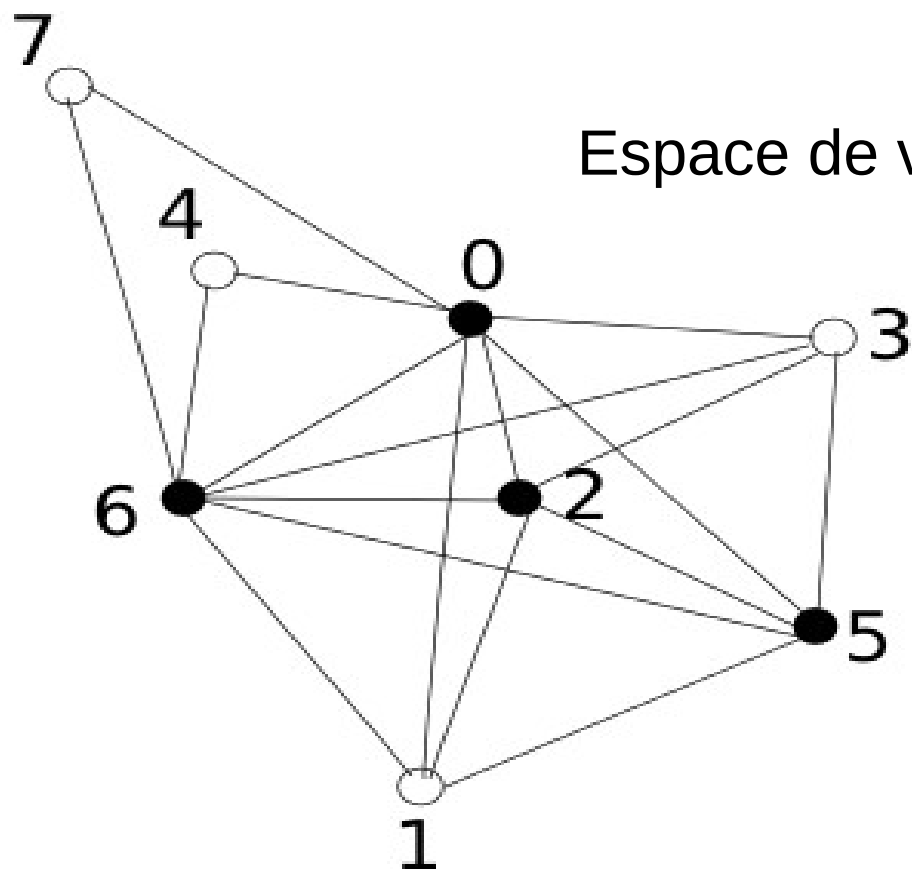
0 7 5 3 -1 -6 -2 4 8

Circle graph of a permutation:
 intersection graph of the chords
 between consecutive numbers,
 bold if a good reversal



Unoriented components are all white
 connected components

Espace de voisinage (Jaeger, 1983)



| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 2 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 3 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 4 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 5 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

Graphes de Cordes et Espaces Graphiques

F. JAEGER

Soit $G = (V, E)$ un graphe simple; $\forall v \in V$, soit N_v l'ensemble des sommets adjacents à v . $\mathcal{P}(V)$ étant supposé muni de sa structure canonique d'espace vectoriel sur $\text{GF}(2)$, nous appellerons *espace de voisinages de G* tout sous-espace de $\mathcal{P}(V)$ engendré par une famille de parties de V de la forme: $(\varepsilon(v)\{v\} + N_v, v \in V)$, où ε est une application quelconque de V dans $\text{GF}(2)$.

Un graphe est dit *de cordes* s'il est isomorphe au graphe de croisement d'un ensemble de cordes d'un cercle.

The adjacency matroid of a signed permutation

Elements = elements of the permutation

Cycles = linear dependency of the neighborhood vectors

Theorem (F. Jaeger, 1983)

The adjacency matroid of a permutation (more generally, of a circle graph) is cographic.

The adjacency matroid of a permutation
 $M(P)$

Lemma

A good **reversal** on a permutation P corresponds to **contraction** of an element in $M(P)$

Theorem

The **reversal distance** of a permutation P without unoriented component is the **rank** of $M(P)$
(the binary rank of the adjacency matrix of the circle graph)


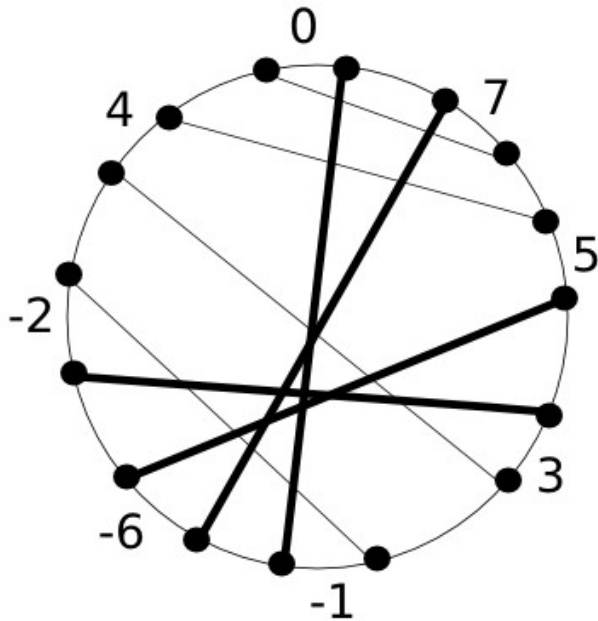
$$d(P) = \text{rank}(P)$$

Lemma

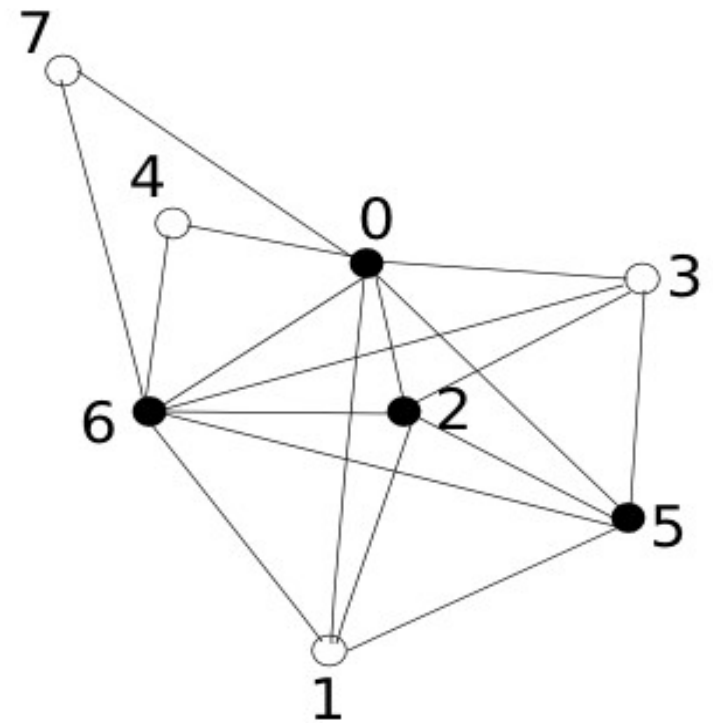
A good **reversal** on a permutation P corresponds to **contraction** of an element in $M(P)$

A sorting reversal complements locally the circle graph

0 7 5 3 -1 -6 -2 4 8
 0 1 -3 -5 -7 -6 -2 4 8

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 2 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 3 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 4 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 5 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |



Conclusion

There are simpler ways to compute the reversal distance, as $M(P)$ is also graphic and eulerian.

But the link with matroids may lead to more general results:

For example,

G is a looped graph (not necessarily a circle graph), it is possible to find a **sequence of complementation of loop vertices** that transforms G into an empty graph. and the size of the sequence is the rank of the adjacency matroid.

Any equivalent for (binary) matroids ?

