

Exercises – Algorithms for Genome Rearrangement

Universität Bielefeld, SS 2014, Dr. Pedro Feijao

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014summer/gr>

Exercise List X – review — 07.07.2014

Hand in exercises by: 14.07.2014 (only discussion, not necessary to hand in)

Reversals

Exercise 1 Given the permutation $\pi = (+2 +1 +3 +5 +4)$

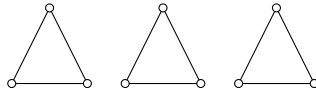
- Draw the BP graph and find the reversal distance $d(\pi) = n - c + h + f$.
- Apply one Hurdle Merging operation, redraw the BP graph and find the new reversal distance.
- Apply one Hurdle Cutting operation, redraw the BP graph and find the new reversal distance.
- After applying the operation in (b), draw the overlap graph, find the oriented reversal with maximum score and apply it.

Exercise 2 Given permutation $\pi = (+2 +1 +3 +5 +4)$

- Draw the components in the permutation, and the component tree. Find an optimal tree cover and the reversal distance according to $d(\pi) = n - c + T$, where T is the optimal tree cover cost.

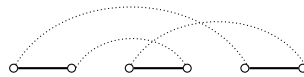
Exercise 3 Given a signed permutation π , an unsigned permutation π' can be obtained by simply ignoring the signs of π . If the signed reversal distance of π is $d_R(\pi) = k$, what can you say about the unsigned distance $d_U(\pi')$, in relation to k ? (is it smaller, greater, equal...)

Exercise 4 Consider the following overlap graph $O(\pi)$:

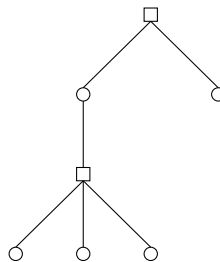


- How many hurdles does the breakpoint graph $BP(\pi)$ may have?
- What are the possible values for the reversal distance $d(\pi)$?

Tip: a white triangle (unoriented component with three white vertices) in the BP graph can only be something similar to this:



Exercise 5 Consider the following tree T_P :



Find a permutation π that has a component tree like T_P above.

Translocations

Exercise 6 Consider the genomes $A = (4\ 3)(1\ 2\ -7\ 5)(6\ -8\ 9)$ and $Id = (1\ 2\ 3)(4\ 5)(6\ 7\ 8\ 9)$.

- Draw the elementary interval graph, the intrachromosomal components, the forest F_A and find the translocation distance between A and Id .
- Apply bad translocations that destroy all the intrachromosomal components, redrawing the elementary interval graph, the components and the forest F_A , also recalculating the translocation distance.
- After all the intrachromosomal components are destroyed, apply one proper translocation, increasing the number of cycles by one.

Exercise 7 The cost of an interchromosomal forest F_A is given by three different possible cases, depending on the parity of the number of leaves L , and the number of trees T .

- Draw a forest F_A where L is even and $T = 1$.
- Draw a forest F_A where L is even and $T \neq 1$.
- Draw a forest F_A where L is odd.
- What is the cost t of each of these trees? Explain this extra cost, in terms of number of bad translocations and number of decreased cycles in the elementary interval graph.

DCJ

Exercise 8 After transforming one genome into the other, the adjacency graph is composed of only cycles of length 2 and paths of length 1.

A known result is that each component in the adjacency graph contributes to the DCJ distance independently. That is, each component has a fixed “cost” to sort, the minimum number of DCJ operations that are needed to transform this component in cycles of length 2 and paths of length 1 (length is the number of edges).

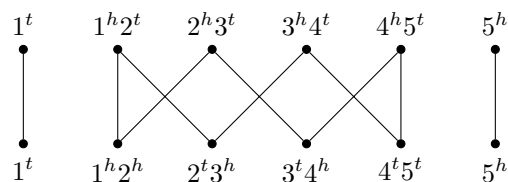
- What is the cost of sorting a cycle of length 4? And a cycle of length 8?
- What is the cost of sorting an odd path of length 3? And length 5?
- Find a general cost equation for cycles, depending on the length.
- Find a general cost equation for odd paths, depending on the length.

Exercise 9 Consider an adjacency graph with a AA-paths, b BB-paths, and c AB-paths. How many linear chromosomes each genome has?

Exercise 10 In a rearrangement problem, the *diameter* is the maximum possible distance between two genomes of size n .

- What is the DCJ diameter?
- Find two genomes with maximum distance for $n = 4$.

Exercise 11 Consider the following adjacency graph:



Find a DCJ operation that breaks the cycle of length 8 in two cycles of length 4. What is the type of this operation? Is it optimal?

Breakpoint Models

Exercise 12 The SCJ and BP distances are given by

$$d_{SCJ} = 2n - 2c_2 - p \text{ and } d_{BP} = n - c_2 - p_2/2$$

where n is the number of genes, and c_2, p_2 and p are the number of cycles of length 2, paths of length 2, and paths of the adjacency graph.

The expected relationship between the SCJ and BP distance is $d_{SCJ} = 2d_{BP}$. Using the above equations, describe in which type of adjacency graph this relationship is true.

Algebraic Theory

Exercise 13 Consider permutations $\pi = (1\ 6\ 4\ 3\ 2\ 5)$ and $\sigma = (1\ 2\ 3\ 4\ 5\ 6)$.

- (a) What is the Fission, Fusion and Transposition distance between π and σ ?
- (b) Find all possible good transpositions that can be applied as a first step to sort π into σ .

General

Exercise 14 The number of unsigned permutations of the set $\{1, \dots, n\}$ is $n!$, and signed are $2^n n!$.

- (a) How many unsigned permutations exist where all the even numbers appear before the odd numbers? And signed?
- (b) How many unsigned permutations exist where the parity of the elements is alternating between odd and even? For instance, $\pi = (1\ 4\ 3\ 6\ 5\ 2)$.
- (c) How many signed permutations exist, where all the odd numbers are positive?
- (d) How many signed permutations exist, where all the odd numbers are positive and the even are negative?

Exercise 15 Suppose that you have a model where a rearrangement operation can reduce at most 1 breakpoint at each step. You also have an algorithm that guarantees that when it can not find an operation that removes 1 breakpoint, it is always possible to apply an operation that does not change the number of breakpoints, followed by 2 operations where each one decreases the number of breakpoint by 1. This can be called a (0-1-1)-operation.

What is the approximation ratio of this algorithm?