Exercises – Phylogenetics

Universität Bielefeld, WS 2015/2016,
Dipl.-Inform. Damianos Melidis · B. Sc. Kevin Lamkiewicz
http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015winter/Phylogenetik

Exercise List 10 — 12.01.2016

Due to: 19.01.2016

Exercise 1 Programming exercise: Splitstree. (4 Points)

Implement a function that calculates the "isolation index" for the *split decomposition*: For two given sets J and K calculate $\alpha_{J,K}(d)$ w.r.t a matrix d:

$$\alpha_{J,K}(d) = \frac{1}{2} \min_{\substack{i,j \in J \\ k,l \in K}} \left(\max\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\} - d_{ij} - d_{kl} \right)$$

Send a version of your program to your TA[1] and describe how to use the program. Make it as easy as possible to calculate all the α in task 3. You'll find another matrix and some solutions on the back of this sheet if you want to test your implementation.

Exercise 2 Calculation of *d*-Splits. (4 Points)

Look at the algorithm that calculates d-Splits (lecture notes, page 62).

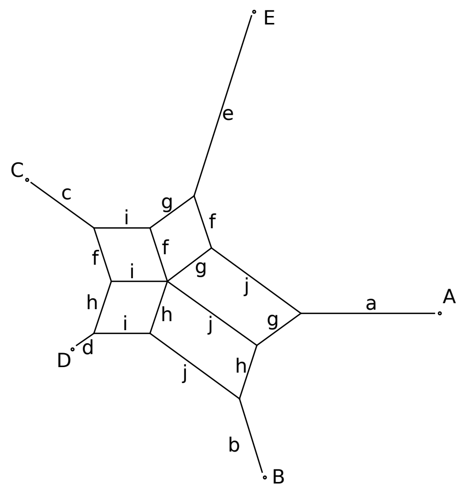
(a) Explain: If a split J, K is extended by one taxon (J', K, where $J' := J \cup \{i\}$), $\alpha_{J',K}$ does not have to be recalculated completely to decide whether J', K is a valid split ($\alpha_{J',K} \neq 0$). A complete recalculation would take $\mathcal{O}(n^4)$. Without the recalculation it is possible to do in $\mathcal{O}(n^3)$. How?

(b) Explain the running time of $\mathcal{O}(n^6)$.
Hint: There are $\binom{n}{2}$ splits for n taxa maximally (you don't need to show this.)

Exercise 3 Splitstree. (3 Points)

Consider the following distance matrix and the corresponding splitstree:

	A	B	C	D	E
A :	0	9	13	12	13
B :		0	12	7	15
C :			0	6	10
D :				0	12
E :					0



Calculate the length of the edges a bis j in the given network.
(Hint: $a = \alpha_{\{A\},\{B,C,D,E\}}$, $g = \alpha_{\{A,E\},\{B,C,D\}}$, etc.)
Solve this exercise with your implementation. You can also use the software *Splitstree* (`www.splitstree.org`) — describe the format of your input and your proceeding.

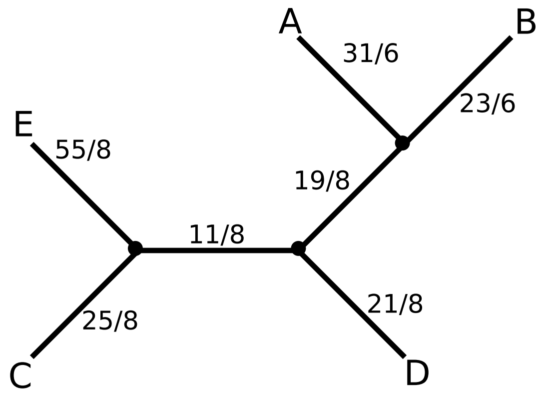
Turn around!

[1] jeschulz@techfak.uni…
dmelidis@cebitec.uni…

Exercise 4 Comparison of methods. **(2 Points)**

If one uses the *Neighbor Joining* algorithm on the distance matrix from task 3, the result is the following tree. Compare the tree with the result from task 3.

Neighbor Joining-Tree:



Matrix and some solution to test your implementation in task 1 (optional!):

	A	B	C	D	E
A :	0	6	8	5	10
B :		0	5	8	10
C :			0	4	8
D :				0	7
E :					0

J	K	$\alpha_{J,K}$
$\{A,B,C,E\}$	$\{D\}$	0.5
$\{A,B,D,E\}$	$\{C\}$	0.5
$\{A,B,E\}$	$\{C,D\}$	0.5
$\{A,C,D,E\}$	$\{B\}$	1.5
$\{A,C,E\}$	$\{B,D\}$	0.0
$\{A,D,E\}$	$\{B,C\}$	1.5
$\{A,E\}$	$\{B,C,D\}$	0.0
$\{B,C,D,E\}$	$\{A\}$	1.5
$\{B,C,E\}$	$\{A,D\}$	1.0
$\{B,D,E\}$	$\{A,C\}$	0.0
$\{B,E\}$	$\{A,C,D\}$	0.0
$\{C,D,E\}$	$\{A,B\}$	2.0
$\{C,E\}$	$\{A,B,D\}$	0.0
$\{D,E\}$	$\{A,B,C\}$	0.0
$\{E\}$	$\{A,B,C,D\}$	4.5