Exercise Sheet 8 — 20.12.2016

Due: 10.1.2017

Task 1 Agglomerative Clustering. (4 points)

	A	B	C	D	E	F
$A:$	0	4	8	18	18	6
$B:$		0	6	12	8	8
$C:$			0	18	18	12
$D:$				0	10	8
$E:$					0	12
$F:$						0

Use the following variants of agglomerative clustering to reconstruct phylogenetic trees from the above matrix.

(a) *Single linkage clustering.*

(b) *WPGMA.*

State the corresponding matrix in every intermediate step and write down the final tree. If there are more possibilities, show all of them.

(c) Compare the results and decide whether the matrix is ultrametric. Explain!

Task 2 Reconstruction of Additive Trees. (4 points)
The distance matrix on the right is *additive*.

Use the algorithm of Waterman (lecture notes, section 7.3.1) to reconstruct the corresponding tree.

Proceed lexicographically, i.e., start with the edge $\{A, B\}$ and add the taxa C, D and E in that order. Try pair $\{A, B\}$ to add a new taxon first. Sometimes you might have to choose another pair.

Write down all steps.

	A	B	C	D	E
$A:$	0	9	10	11	7
$B:$		0	3	8	6
$C:$			0	9	7
$D:$				0	8
$E:$					0

Xmas Exercises
Points of the Xmas exercises are bonus points.

Task 3 Pseudocode Perfect Phylogeny. (2 points)
Write down a pseudo code that solves the *Perfect Phylogeny Problem* by using Gusfield's theorem. The set operations should be written in pseudo code as well.

Task 4 Ultrametrics. (4 points)
Given a path metric d^T which states the distances of n taxa the following property holds:
If d^T forms an ultrametric, there are at most $n - 1$ different entries $d^T_{i,j}$ for all $i \neq j$.

(a) Justify this statement. (Hint: For each internal node of an ultrametric tree, think about what distances two taxa could have whose LCA is such a node.)

(b) Find a counterexample of four taxa which shows that the reverse is not true. ("Each path metric with at most $n - 1$ different entries $d^T_{i,j}$ for all $i \neq j$ forms an ultrametric.")

Please turn over! Bitte wenden!

Task 5 Agglomerative Clustering – Sequel.   **(4 points)**
Use the following variants of agglomerative clustering to reconstruct phylogenetic trees from the matrix
given in Exercise 1.

(a) *Complete linkage clustering.*

(b) *UPGMA.*

State the corresponding matrix in every intermediate step and write down the final tree. If there are
more possibilities, show all of them.