Exercises – Phylogenetics

Universität Bielefeld, SS 2019
Dr. Roland Wittler, M. Sc. Tizian Schulz
https://gi.cebitec.uni-bielefeld.de/Teaching/2019summer/Phylogenetik

Exercise Sheet 6 — 16.05.2019

Due: 23.05.2019

Task 1 Additive Metric, Ultrametric. **(4 points)**
Given the following matrices:

i)	A	B	C	D	E
A	0	10	10	10	4
B		0	2	6	10
C			0	6	10
D				0	10
E					0

ii)	A	B	C	D	E
A	0	3	7	9	10
B		0	8	10	11
C			0	4	5
D				0	3
E					0

Show that the matrix i) describes an ultrametric and that the matrix ii) describes an additive metric, but not an ultrametric. Explain your result.

Task 2 Properties of Distances. **(4 points)**
For every distance function d applies the following relation:

"d is *ultrametric*" $\Rightarrow$ "d is *additive*" $\Rightarrow$ "d satifies the triangle inequality"

Prove this relation by showing the following items:

(a) "d satifies the *three point condition*" $\Rightarrow$ "d satifies the *four point condition*"
Proof sketch: Pick four arbitrary elements and proceed: Denote the two elements with the smallest distance as a and b. Denote the other two elements as c and d such that $d_{ac} \leq d_{ad}$. Now we can order a, b, c and d in two different, binary, ultrametric trees. **Which?** In both cases several three point conditions are fulfilled (**Which?**) that can be combined/transformed to the wanted four point condition. **How?** (The case where the distances don't correspond to a binary topology is not considered here. We cover this case by allowing edges of length 0. For example: ((a:2,b:2,c:2):1,d:3); would be the binary tree (((a:2,b:2):0,c:2):1,d:3);)

(b) "d satifies the four point condition" $\Rightarrow$ "d satifies the triangle inequality"
Hint: The four point condition can be used on three points as well.

Task 3 Agglomerative Clustering. **(4 points)**
Given the following distance matrix:

	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 methods to reconstruct phylogenetic trees from the matrix.

(a) *Single linkage clustering.*

(b) *WPGMA.*

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

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