TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Exercise Sheet 4
Computational Biology (Part 2), WS 12/13

Hand In: *Until Monday, 10.12.2012, 10:00 am, email to* `wild@cs...` *or in lecture.*

Exercise 9
3 Points

We consider the WPGMA algorithm which is essentially the UPGMA algorithm as introduced in the script on page 162. However in step 3, new distances are computed as follows:

$$dist(W, X) = dist(X, W) = \frac{|R_1| \cdot dist(R_1, X) + |R_2| \cdot dist(R_2, X)}{|R_1| + |R_2|} \qquad \text{für alle } X \in \Gamma \setminus W.$$

Let $dist(X, Y)$ be the distance between node X and Y computed by the WPGMA algorithm. Show that

$$dist(X, Y) = \frac{1}{|X| \cdot |Y|} \sum_{x \in X, y \in Y} \delta(x, y) \, ,$$
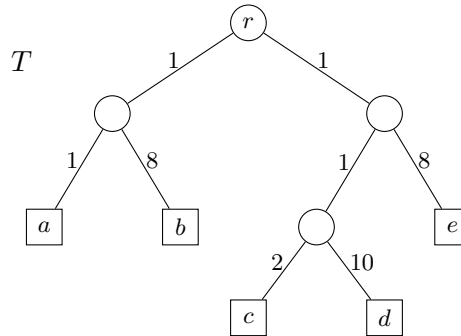
where δ is the input metric.

Exercise 10

<div align="right">2 + 4 Points</div>

We consider additive, binary and rooted phylogenetic trees.

a) Let T be the following additive phylogenetic tree for taxa $A = \{a, b, c, d, e\}$:

Write down the distance matrix δ encoded by T explicitly. Then, apply the UPGMA algorithm on A and δ — even though δ is not an ultrametric.

b) Let T be any additive, binary and rooted phylogenetic tree for taxa A where the taxa correspond exactly to the *leaves* of T. Let r be the root of T and call $dist_T(x, y)$ the length of the unique path from x to y in T.

Now consider the following distances for all $x, y \in A$, $x \neq y$

$$\delta'(x, y) \quad := \quad \hat{\delta}_r \; + \; \frac{dist_T(x, y) - dist_T(x, r) - dist_T(y, r)}{2} \, ,$$

where $\hat{\delta}_r := \max_{a \in A} dist_T(a, r)$ is the height of T. Of course, we set $\delta'(x, x) := 0$.

(i) Compute δ' for the tree T from a) and apply the UPGMA algorithm on δ'. Compare the resulting ultrametric tree to T.

(ii) Show that δ' is always an ultrametric.

Note: This statement is equivalent to the fact that the UPGMA algorithm on δ' always reconstructs the topology of the underlying additive phylogenetic tree T.