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 5 — 17.11.2015

Due to: 24.11.2015

Exercise 1 Maximum Parsimony Branch-and-Bound **(3 Points)**

Consider the given matrix. Find a binary tree with labels for the internal nodes where the costs are minimal. Assume unit costs for this task. For the four taxa we know this three topologies:

1. ((A,C),(B,D)); 2. ((A,B),(C,D)); 3. ((A,D),(B,C));

	1	2	3	4
A	a	c	e	y
B	b	d	x	y
C	a	c	e	z
D	b	d	x	z

Use the *Column-wise Branch-and-Bound* algorithm in order to calculate the topology with minimal cost. Write down all intermediate steps. If you can stop at some point explain why. (**Hint:** The order of the given topologies and states makes it a bit easier.)

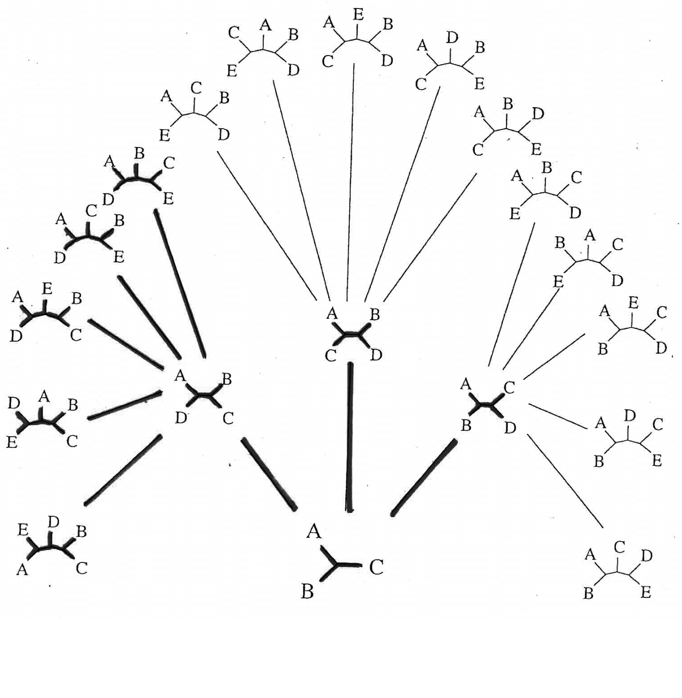
Exercise 2 Greedy Sequencial Addition. **(3 Points)**

Consider the *Greedy Sequencial Addition* heuristic (GSA) to approach a solution of the *maximum parsimony problem.* For given taxa $t_1, t_2, t_3, \ldots, t_n$ the algorithm works as follows:

GSA($[t_1, t_2, t_3, \ldots, t_n]$)

1. construct tree $T_3 := (t_1, t_2, t_3)$;
2. for $i = 4...n$:
   a. for each edge e in T_{i-1}:
      i. connect taxon t_i to edge e
      ii. compute the weight of this tree
   b. select T_i with minimal weight among all above considered trees of size i.
3. return T_n

The figure below shows the search space for an example with $n = 5$ taxa: A, B, C, D and E. Three trees and their costs are calculated by adding the fourth taxon in the first run of the loop (second line). A tree with the lowest cost is chosen. In the next step five new trees are analyzed based on the chosen tree. Those trees are marked in the figure.



(a) How many trees are considered before choosing one, if you add the i-th taxon?

(b) How many trees are considered in all steps $i = 4...n$? Show that this number increases quadratically with n. (Hint: Use the formula of the "small Gauß")

Exercise 3 Nearest Neighbor Interchange. **(3 Points)**

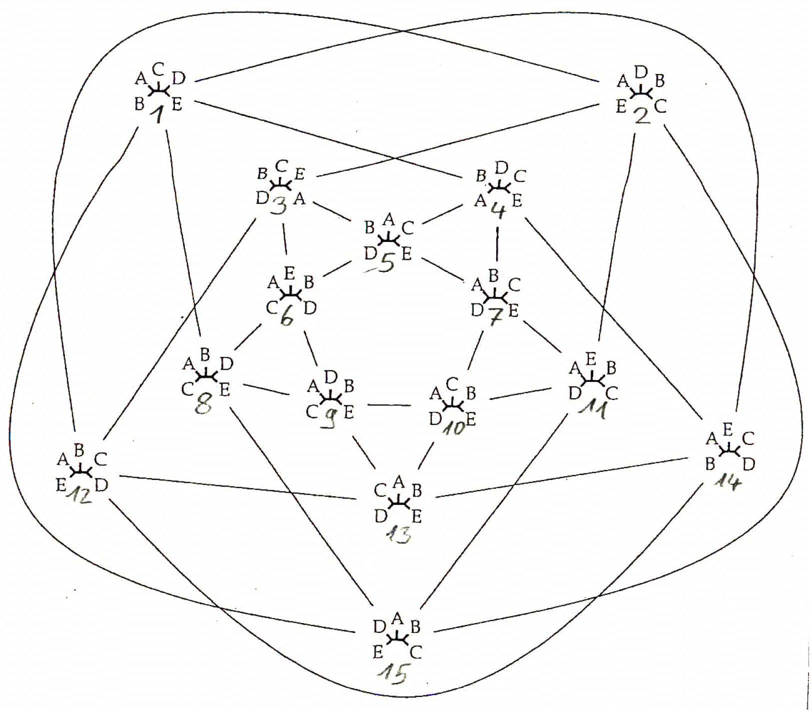
Use the *Nearest Neighbor Interchange* heuristic (NNI) to approach a solution of the *maximum parsimony problem*. For a given tree T the algorithm works as follows:

NNI(T)

1. compute $W(T)$, the parsimony score of tree T
2. compute neighbors $T_1, \ldots, T_n$ of T
3. compute score $W(T_i)$ for all neighbors $T_i, 1 \leq i \leq n$
4. if$\big(W(T_i) \geq W(T)$ for all $1 \leq i \leq n\big)$
5. then return T
6. else
   a. select any $T_{next} \in \{T_1, \ldots, T_n\}$ with $W(T_{next}) = \min_i(W(T_i))$
   b. return NNI(T_{next})

All 15 unrooted topologies for the species A to E are given in the figure below. Every tree is labeled with a number and connected with his neighbors. Use this relations and the *parsimony scores* in the table to find an optimal tree with the NNI-algorithm. Start with tree 8. Write down all intermediate steps. If there are more than one possibilty in step 6a, follow all of them.
Is the optimal tree found in every case (not starting at 8 aswell)?

Tree No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Cost	5	4	5	5	5	5	5	5	3	2	4	3	1	3	3



The figures in exercises 2 and 3 are from: J. Felsenstein, Inferring Phylogenies, 2004