

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 6 — 24.11.2015

Due to: 01.12.2015

Exercise 1 Properties of spanning trees.

(3 Punkte)

Let T be a minimum spanning tree in a graph $G = (V, E)$. Let e be an edge in T that divides T into two subtrees T_1 and T_2 . Then e is of least weight among all the edges in E that connect a node of T_1 and a node of T_2 . (page 36)

Prove this property.

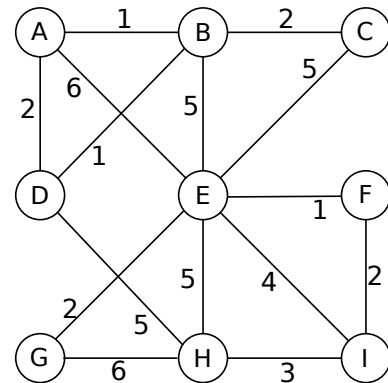
Exercise 2 Spanning Trees.

(3 Punkte)

In the figure on the right is a graph with weighted edges. Find all minimum spanning trees in that graph. Write down the order of the edges you've chosen. If there are several edges that can be chosen in one step, state every edge but choose one (arbitrary) edge to continue.

(a) Use Kruskal's algorithm.

(b) Use Prim's algorithm. Start with node F.



Exercise 3 Spanning tree heuristic.

(5 Punkte)

Find the *most parsimonious tree* of the taxa A to E with regard to the following sequences.

A : G A C A A
B : T A T A T
C : G C A A A
D : T A A A T
E : G A A C A

Let G be an *DNA grid graph* that contains all sequences of length 5 and therefore particularly the nodes that correspond to the taxa A to E.

(a) Use the spanning tree heuristic to approximate a *Steiner tree* for the nodes. Proceed like this:

Step 1: Shortest paths. Calculate all Hamming distances for all taxa and create the graph G' with weighted edges.

Step 2: Spanning tree. Create a minimum spanning tree T_1 in G' that contains the edge $A - B$. Create another minimum spanning tree T_2 that does **not** contain the edge $A - B$.

Step 3: Map back to G . Draw the part of the grid graph G that contains T_1 and T_2 respectively. (for each tree one graph!) Add all sequences as nodes into the trees such that the Hamming distance between all nodes is exactly 1. Try to add as few nodes as possible and reuse some nodes for different edges. (Hint: The nodes GAAAA and TAAAA are quite useful (especially for T_2))

(b) Compare the costs of your solutions. Discuss the differences.

(If you don't have different costs in task (a), discuss the accuracy of the spanning tree heuristic.)