

Exercises – Phylogenetics

Universität Bielefeld, WS 2017/2018,
 Dr. Roland Wittler, Prof. Dr. Jens Stoye, M. Sc. Tizian Schulz
<https://gi.cebitec.uni-bielefeld.de/Teaching/2017winter/Phylogenetik>

Exercise Sheet 6 — 16.11.2017

Due: 23.11.2017

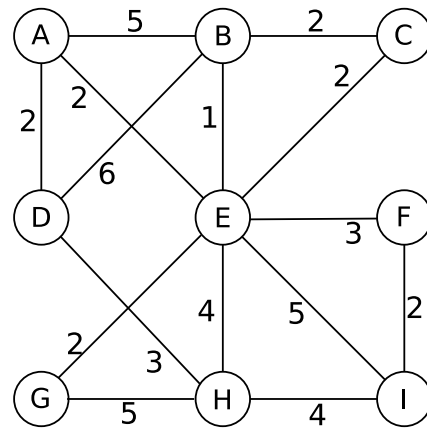
Task 1 Properties of Spanning Trees. **(3 points)**

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.

Task 2 Spanning Trees. **(3 points)**

Find all minimum spanning trees in the given graph. Write down the order of the edges you choose. If there are several edges that can be chosen in one step, list them all but choose one (arbitrary) edge to continue.



- (a) Use Kruskal's algorithm.
- (b) Use Prim's algorithm. Start with node A.

Task 3 DNA Grid Graph. **(2 points)**

What is the number of edges $|E|$ for a DNA grid graph $G = (V, E)$ for sequence length m ? Derive a formula for $|E|$ that only depends on m and explain it.

Task 4 Spanning Tree Heuristic. **(3 points)**

Search for a *most parsimonious tree* of the taxa A to E with regard to the following sequences.

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

Let G be a *DNA grid graph* that contains all sequences of length 5 and therefore particularly the nodes that correspond to the taxa A to E. Use the spanning tree heuristic to approximate a *Steiner tree* for the nodes. Proceed like this:

Step 1: Shortest paths. Calculate all pairwise Hamming distances for the given taxa A to E and create the edge-weighted graph G' .

Step 2: Spanning tree. Create a minimum spanning tree T' in G' .

Step 3: Map back to G. Draw the part of the grid graph G that contains T' . 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.