

Exercises – Phylogenetics

Universität Bielefeld, SS 2018
Dr. Roland Wittler

<https://gi.cebitec.uni-bielefeld.de/Teaching/2018summer/Phylogenetik>

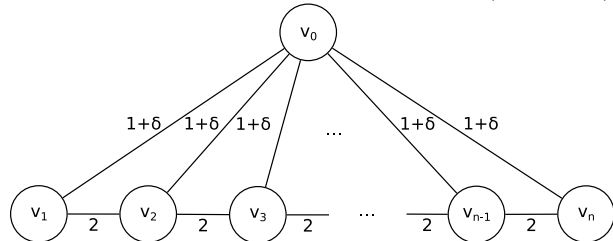
Exercise Sheet 6 — 24.05.2018

Due: Monday, 04.06.

Task 1 Spanning Tree Heuristic: 2-Approximation.

(3 points)

We have shown that the approximation factor of the spanning tree heuristic is at most 2. Use the given graph (with terminal nodes v_1, \dots, v_n) to show that this value is *tight*, i.e., the approximation factor is not smaller than 2.



Task 2 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.

Task 3 Additive Metric, Ultrametric.

(4 points)

Given the following matrices:

i)	A	B	C	D	E
A	0	5	5	5	2
B		0	1	3	5
C			0	3	5
D				0	5
E					0

ii)	A	B	C	D	E
A	0	9	7	9	7
B		0	5	5	9
C			0	5	7
D				0	9
E					0

Decide for i) and ii) whether the matrix describes an additive metric or even an ultrametric. Explain your result.

Task 4 Properties of Distances.

(2 points)

For every distance function d , the following relation applies:

d is *additive* $\Rightarrow d$ satisfies the triangle inequality

Prove this relation by showing: d satisfies the four point condition $\Rightarrow d$ satisfies the triangle inequality. Hint: The four point condition can be used on three points as well.