Exercises – Phylogenetics

Universität Bielefeld, WS 2016/2017,
M. Sc. Tizian Schulz

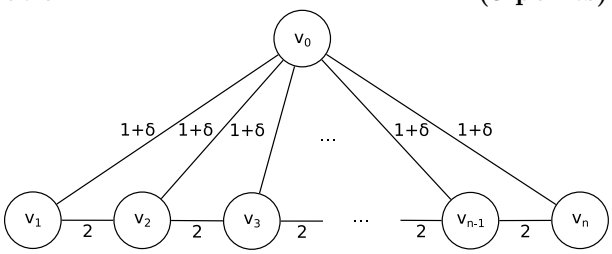
https://gi.cebitec.uni-bielefeld.de/Teaching/2016winter/Phylogenetik

Exercise Sheet 7 — 13.12.2016

Due: 20.12.2016

Task 1 Spanning Tree Heuristic: 2-Approximation. **(3 points)**

Use the given graph to show that the approximation factor of the spanning tree heuristic is *tight*, i.e., it has no smaller factor than 2. Terminal nodes should be $v_1, ..., v_n$.



Task 2 Spanning Tree Heuristic: Application. **(3 points)**

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

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

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.

Step 4: Extract T. Finally, draw the resulting phylogenetic tree T and specify its cost.

Task 3 Additive Metric, Ultrametric. **(4 points)**

Given the following matrices:

i)	A	B	C	D	E
A	0	1	9	9	5
B		0	9	9	5
C			0	9	9
D				0	9
E					0

ii)	A	B	C	D	E
A	0	4	2	9	10
B		0	3	6	7
C			0	8	9
D				0	2
E					0

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

Task 4 Properties of Distances. **(2 points)**

For every distance function d, the following relation applies:

"d is *additive*" ⇒ "d satisfies the triangle inequality"

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