

# Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2013/2014, Dr. Roland Wittler, M.Sc. Kai Stadermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013winter/Phylogenetik>

## Blatt 6 vom 18.11.2013

Abgabe in einer Woche zu Beginn der Vorlesung oder vorab bei der Tutorin oder beim Veranstalter.

### Aufgabe 1 Eigenschaften von Spannbäumen.

(3 Punkte)

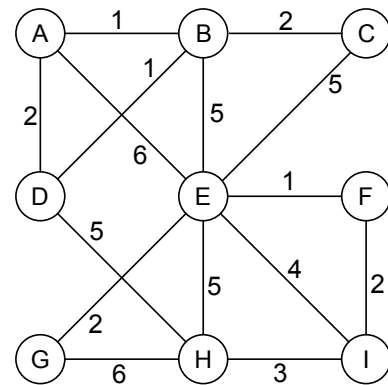
Ein minimaler Spannbaum  $T$  in einem Graph  $G = (V, E)$  hat die Eigenschaft: Sei  $e$  eine Kante in  $T$ , die  $T$  in zwei Teilbäume  $T_1$  und  $T_2$  aufteilt. Dann hat  $e$  das geringste Gewicht aller Kanten aus  $E$ , die in  $G$  einen Knoten aus  $T_1$  und einen Knoten aus  $T_2$  miteinander verbinden. (Skript Seite 36)

Beweise diese Eigenschaft.

### Aufgabe 2 Spannbäume.

(3 Punkte)

Im nebenstehenden Bild ist ein Graph mit Kantengewichten dargestellt. Finde in diesem Graphen minimale Spannbäume. Gib dabei an, in welcher Reihenfolge die Kanten ausgewählt werden. Sollte es in einem Schritt mehrere Kanten zur Auswahl geben, gib alle an, aber wähle nur eine (beliebige) aus um fortzufahren.



- Verwende den Algorithmus von Kruskal.
- Verwende den Algorithmus von Prim und beginne dabei mit Knoten H.

### Aufgabe 3 Spannbaumheuristik.

(5 Punkte)

Im Folgenden soll der 'most parsimonious tree' der Taxa  $A$  bis  $E$  auf Basis der folgenden Sequenzen approximiert werden.

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

Gegeben sei der DNA-Gridgraph  $G$ , der alle Sequenzen der Länge 5 enthält und damit auch Knoten, die den Taxa  $A$  bis  $E$  entsprechen.

- Verwende die Spannbaumheuristik um in  $G$  einen Steinerbaum für diese Knoten zu approximieren. Gehe dazu wie folgt vor:

**Schritt 1: Kürzeste Pfade.** Ermittle zunächst die Hammingdistanzen zwischen allen Taxa und konstruiere damit den vollständigen Graphen  $G'$  mit gewichteten Kanten.

**Schritt 2: Spannbaum.** Konstruiere einen minimalen Spannbaum  $T_1$  in  $G'$ , der die Kante  $C - D$  enthält und einen minimalen Spannbaum  $T_2$ , der die Kante  $C - D$  *nicht* enthält.

**Schritt 3: Zurückführen auf  $G$ .** Zeichne jeweils den Teil des Gridgraphen  $G$ , der  $T_1$  bzw.  $T_2$  enthält. Füge dazu Sequenzen als weitere Knoten in die vorhandenen Kanten des Baums ein, so dass die Hammingdistanz zwischen allen Knoten genau 1 ist. Versuche insbesondere für  $T_2$  möglichst wenige Knoten einzufügen und gegebenenfalls für verschiedene Kanten wiederzuerwenden (Tipp: Die Knoten GAAAA und TAAAA können dir hier helfen).

- Vergleiche die Kosten der beiden Lösungen für  $T_1$  und  $T_2$  und diskutiere die Unterschiede. (Solltest du in Aufgabenteil (a) keine unterschiedlichen Kosten erhalten haben, diskutiere kurz allgemein die Exaktheit der Spannbaumheuristik.)