

# Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2012/2013, Dr. Roland Wittler  
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2012winter/Phylogenetik>

## Blatt 6 vom 22.11.2012

Abgabe in einer Woche zu Beginn der Vorlesung oder vorab im Briefkasten bei U10-151.

### 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 35)

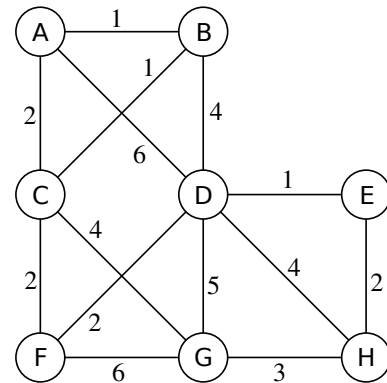
Beweise diese Eigenschaft.

### 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.

- (a) Verwende den Algorithmus von Kruskal.
- (b) Verwende den Algorithmus von Prim und beginne dabei mit Knoten D.



### 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.

- (a) 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.** Finde einen minimalen Spannbaum  $T$  in  $G'$ .

**Schritt 3: Zurückführen auf  $G$ .** Zeichne den Teil des Gridgraphen  $G$ , der  $T$  enthält. Füge dazu Sequenzen als weitere Knoten in die vorhandenen Kanten von  $T$  ein, so dass die Hammingdistanz zwischen allen Knoten genau 1 ist. Erstelle nur unbedingt benötigte Knoten.

- (b) Ist der so entstandene Baum optimal (most parsimonious)? Begründe!  
Tipp: Füge die Knoten GAAAA und TAAAA in deine Zeichnung des Gridgraphen ein.