

Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2016/17

Martin Vingron · Annalisa Marsico · Alena van Bömmel · Edgar Steiger · Thimo Wellner

Blatt 2 · Ausgabe am 24.10.2014

Abgabe am 31.10.2014 vor Beginn der Vorlesung

Name:

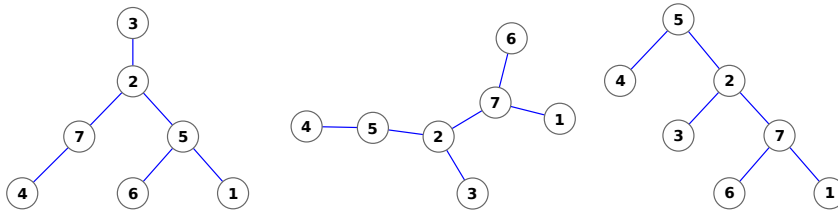
Matrikelnummer:

Übungsgruppe:

Aufgabe 1 (30 Punkte; Rechnen/Theorie; Baumtopologie). In folgender Aufgabe sollen Sie sich mit der graphischen Darstellung, der Implementierung und der Abzählbarkeit von Baumtopologien beschäftigen.

A) Entscheiden und begründen Sie, welche der Bäume in den folgenden zwei Gruppen äquivalent sind.

- Gruppe 1:



- Gruppe 2:



B) Beweisen Sie die in der Vorlesung vorgestellte Formel zur Berechnung der Anzahl an unterschiedlichen Topologien von nicht gewurzelten binären Bäumen mit n Blättern:

$$U_n = \prod_{i=3}^n (2i - 5).$$

C) Machen Sie Vorschläge für Datenstrukturen um phylogenetische Bäume zu implementieren. Welche Operationen und Suchanfragen wird man beim Arbeiten mit phylogenetischen Bäumen häufig brauchen?

Aufgabe 2 (20 Punkte; Theorie; Baumtopologie). Gegeben seien fünf Taxa a, b, c, d, e und das folgende multiple Alignment von Nukleotidsequenzen mit drei Spalten (Merkmalen):

	1	2	3
a	G	G	A
b	T	A	A
c	A	A	A
d	G	G	A
e	C	T	T

sowie eine Baumtopologie (im Newick Format): $((a, b), (e, (c, d)))$.

- Bestimmen Sie mit Hilfe des in der Vorlesung vorgestellten Fitch-Algorithmus für obige Baumtopologie eine Belegung der inneren Knoten mit Nukleotiden, so dass der Baum minimale Länge hat. Wie viele Mutationen braucht man?
- Welche Spalten des Alignments sind in Bezug auf Maximum Parsimony phylogenetisch informativ? Informativ bedeutet hier, dass die Spalte eine Entscheidung zwischen verschiedenen Baumtopologien ermöglicht.

Aufgabe 3 (10+10 Punkte; Programmieren). In der 3. Aufgabe vom Übungsblatt 1 haben Sie eine zufällige DNA-Sequenz und Mutationen in 10.000 Generationen generiert.

- Berechnen Sie in jeder Generation die Hamming-Distanz zur Originalsequenz und stellen Sie die Entwicklung der Distanz über die Generationen graphisch dar. Was sind Ihre Beobachtungen? Sie können gerne die graphische Darstellung in einem anderen Programm als die Simulation und Berechnung ausführen. Die Grafik soll als separate Datei oder ausgedruckt auch abgegeben werden.
- Bonusfrage: Haben Sie eine Erklärung für Ihre Beobachtungen der Entwicklung der Hamming-Distanz?*

Aufgabe 4 (40 Punkte; Rechnen/Theorie; Markovketten). In der Mensa gibt es jeden Tag entweder ein Fleischgericht (M), ein Fischgericht (F) oder ein Gemüsegericht (V). Aus langjähriger Beobachtung wissen die Studierenden, dass mit einer gewissen Wahrscheinlichkeit zum Beispiel auf einen Fleischtag ein Fischtag ($P(F|M)$) oder ein Gemüsetag ($P(V|M)$) folgt. Es wurden die folgenden Übergangswahrscheinlichkeiten für alle möglichen Kombinationen $P_{i,j}, i, j \in \{M, F, V\}$ an Gerichten beobachtet:

$$P(V|M) = 0.1, P(F|M) = 0.1, P(M|V) = 0.03,$$

$$P(F|V) = 0.02, P(M|F) = 0.2, P(V|F) = 0.75.$$

- Modellieren Sie die Essensausgabe als Markovkette erster Ordnung und geben Sie die Transitionsmatrix P an. Skizzieren Sie die Markovkette.
- Berechnen Sie den Wahrscheinlichkeitsvektor π des stationären Zustandes.
- Sie gehen an einem beliebigen Tag in die Mensa und wissen nicht, was für ein Essen es am vorherigen Tag gab. Mit welcher Wahrscheinlichkeit wird Gemüse angeboten?
- Sie gehen am Montag in die Mensa und wissen nicht, was es am Freitag zu Essen gab. Mit welcher Wahrscheinlichkeit gibt es diese Woche $VVVMF$?
- Wenn es an einem Montag Fleisch gab, mit welcher Wahrscheinlichkeit gibt es dann welches Essen am folgenden Freitag?