

Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2009/2010

Dipl.-Inform. Roland Wittler · Dipl.-Inform. Peter Husemann

<http://wiki.techfak.uni-bielefeld.de/gi/GILectures/2009winter/Phylogenetik>

Blatt 5 vom 18.11.2009

Abgabe in einer Woche zu Beginn der Vorlesung.

Aufgabe 1 Greedy Sequential Addition.

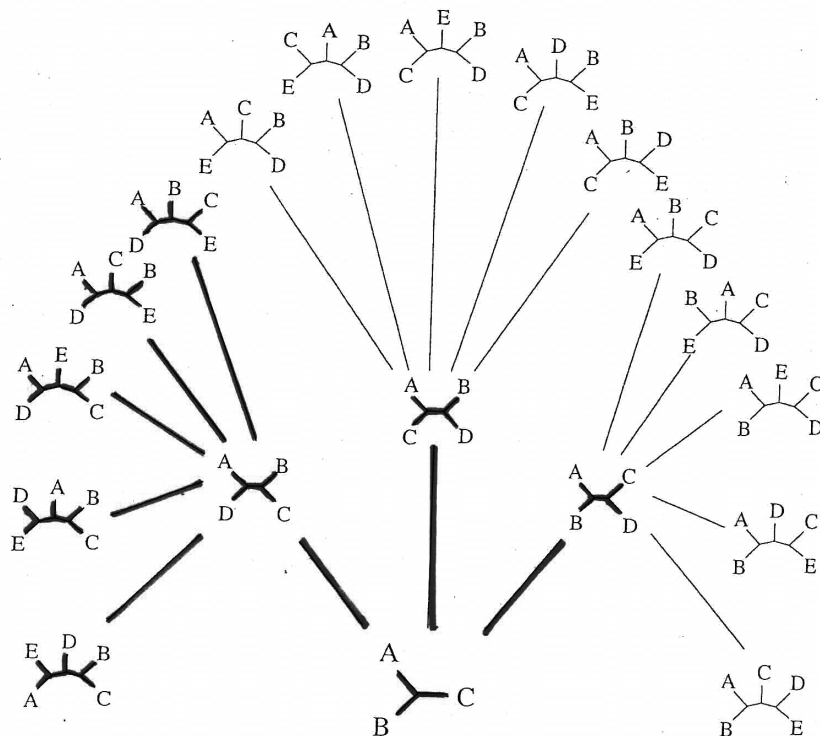
(3 Punkte)

Betrachte die *Greedy Sequential Addition*-Heuristik (GSA) zur Annäherung einer Lösung des *Maximum Parsimony Problems*. Der Algorithmus verfährt für gegebene Taxa t_1, t_2, t_3 bis t_n wie folgt:

GSA($[t_1, t_2, t_3, \dots, t_n]$)

1. construct tree $T_3 := (t_1, t_2, t_3)$;
2. for $i = 4 \dots n$:
 - a. for each edge e in T_{i-1} :
 - i. connect taxon t_i to edge e
 - ii. compute the weight of this tree
 - b. select T_i with minimal weight among all above considered trees of size i .
3. return T_n

Die untenstehende Abbildung zeigt den gesamten Suchraum für ein Beispiel mit $n = 5$ Taxa: A, B, C, D und E . Im ersten Durchlauf der Schleife (Zeile 2 des Pseudocodes) werden beim Hinzufügen des 4. Taxons drei Bäume erzeugt und deren Kosten berechnet. Ein Baum mit minimalen Kosten wird ausgewählt und im nächsten Schleifendurchlauf werden ausgehend von diesem Baum fünf weitere analysiert. Die betrachteten Bäume sind in der Abbildung hervorgehoben.



- (a) Wie viele Bäume werden allgemein beim Hinzufügen des i -ten Taxons zunächst berücksichtigt bevor einer ausgewählt wird?
- (b) Wie viele Bäume werden in allen Schritten $i = 4 \dots n$ insgesamt betrachtet? Zeige, dass diese Anzahl quadratisch mit n zunimmt. (Tipp: Verwende die Formel des kleinen Gauß.)

Aufgabe 2 Nearest Neighbor Interchange.

(3 Punkte)

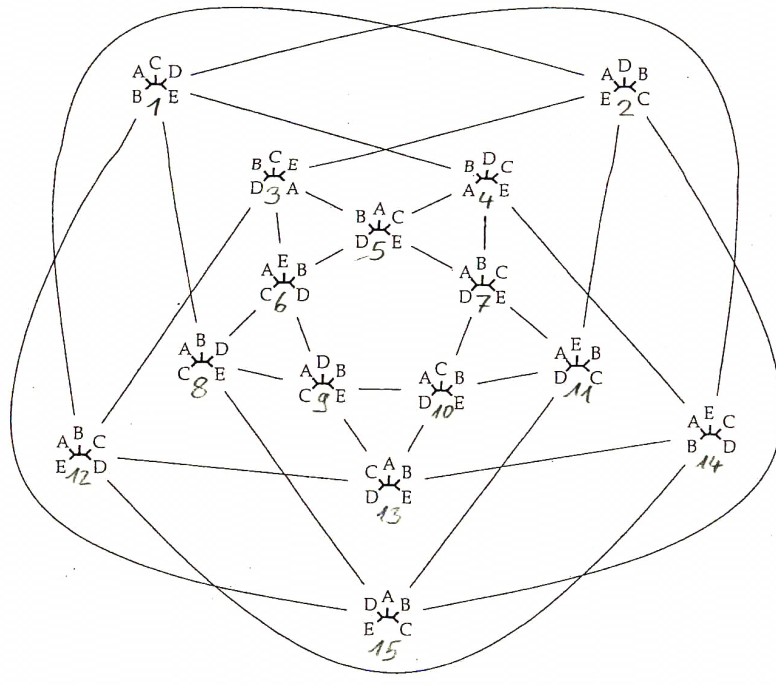
Verwende die *Nearest Neighbor Interchange*-Heuristik (NNI) zur Annäherung einer Lösung des *Maximum Parsimony Problems*. Der Algorithmus verfährt für einen gegebenen Baum T wie folgt:

NNI(T)

1. compute $W(T)$, the parsimony score of tree T
2. compute neighbors T_1, \dots, T_n of T
3. compute score $W(T_i)$ for all neighbors $T_i, 1 \leq i \leq n$
4. if $(W(T_i) \geq W(T))$ for all $1 \leq i \leq n$
5. then return T
6. else
 - a. select $T_{next} \in \{T_1, \dots, T_n\}$ with $W(T_{next}) = \min_i(W(T_i))$
 - b. return NNI(T_{next})

In der untenstehenden Abbildung sind alle 15 ungewurzelten Topologien für die Spezies A, B, C, D und E und abgebildet. Darin ist jeder Baum durch eine Nummer gekennzeichnet und durch Linien mit seinen Nachbarn verbunden. Verwende diese Nachbarschaften und die in der untenstehenden Tabelle gegebenen *Parsimony Scores* um mit Hilfe des NNI-Algorithmus einen optimalen Baum zu finden. Gehe von Baum 8 aus und gib die Zwischenschritte und die Lösung an. Gibt es in Schritt 6a des Algorithmus mehrere Alternativen, verfolge alle Möglichkeiten. Wird der optimale Baum in jedem Fall gefunden?

Baum Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kosten	11	9	11	11	11	11	11	11	9	9	9	10	8	10	9

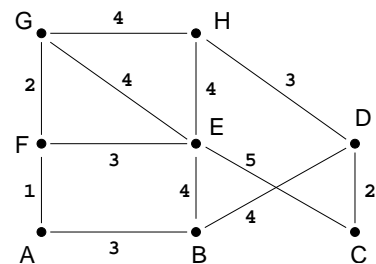


Aufgabe 3 Steinerbäume.

(3 Punkte)

Im nebenstehenden Bild ist ein Graph mit Kantengewichten dargestellt. Finde in diesem Graphen minimale Steinerbäume

- (a) auf den Knoten A, C und H, sowie
- (b) auf den Knoten B, C und G.



⁰Die Bilder zu den Aufgaben 1 und 2 stammen aus: J. Felsenstein, *Inferring Phylogenies*, 2004