

Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2011/2012, Dr. Roland Wittler
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/Phylogenetik>

Blatt 9 vom 8.12.2011

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

Aufgabe 1 Programmieraufgabe: Splittree.

(14 Punkte)

Programmiere in einer Programmiersprache deiner Wahl die folgenden Funktionen aus dem Kontext *Split Decomposition*:

- (a) Eine Funktion, die den 'isolation index' $\alpha_{J,K}(d)$ für zwei gegebene Mengen J und K bezüglich einer Matrix d berechnet: $\alpha_{J,K}(d) = \frac{1}{2} \min_{\substack{i,j \in J \\ k,l \in K}} (\max\{d_{ij} + d_{kl}, d_{ik} + d_{jl}, d_{il} + d_{jk}\} - d_{ij} - d_{kl})$
- (b) Eine Funktion für die 'split metric':

$$\delta_{J,K}(i, j) = \begin{cases} 0 & \text{wenn } i \text{ und } j \text{ beide in } J \text{ oder beide in } K \text{ enthalten sind} \\ 1 & \text{sonst} \end{cases}$$

- (c) Eine Funktion, die die Matrix d^1 berechnet: $d^1(i, j) = \sum_{d\text{-splits } J,K} \alpha_{J,K} \delta_{J,K}(i, j)$

Du musst nicht effizient berechnen, welche d-Splits es wirklich gibt. Summiere einfach über alle möglichen Splits. Dies ist zwar nicht effizient, aber das Ergebnis ist korrekt, da $\alpha_{J,K} = 0$ für Splits, die keine d-Splits sind.

Tipp: Zur Representation/Aufzählung von Teilmengen eignen sich Zahlen im Binärformat. Zum Beispiel kann die Menge $\{B, D, E\}$ durch 01011 dargestellt werden.

- (d) Eine Funktion, welche die 'splittable percentage' ρ berechnet: $\rho := \left(\frac{\sum_{\text{taxa } i,j} d_{ij}^1}{\sum_{\text{taxa } i,j} d_{ij}} \right) \cdot 100\%$

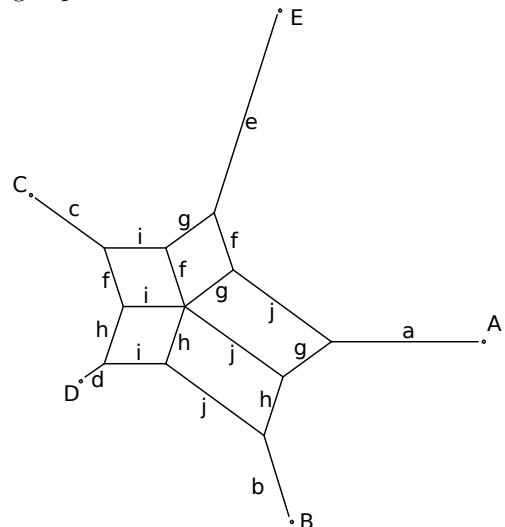
Schicke eine lauffähige Version deines Programmes an deine(n) Tutor(in) und beschreibe in der E-Mail¹, wie das Programm aufgerufen werden muss, um α und ρ in der folgenden Aufgabe zu berechnen. Auf der Rückseite befindet sich eine Matrix und einige Lösungen zum Testen deines Programms.

Aufgabe 2 Anwendung deines Programmes.

(3 Punkte)

Gegeben seien die folgende Distanzmatrix sowie der zugehörige Splittree:

	A	B	C	D	E
A:	0	6	9	8	9
B:		0	8	5	10
C:			0	4	7
D:				0	8
E:					0



- (a) Verwende deine Implementierung, um ρ für die gegebene Matrix zu berechnen.
- (b) Berechne in dem gegebenen Splittree-Netz die Kantenlängen a bis j.
Hinweis: Zum Beispiel $a = \alpha_{\{A\},\{B,C,D,E\}}$
bzw. $g = \alpha_{\{A,E\},\{B,C,D\}}$

¹Erfrage die Emailadresse im Tutorium oder schick die Lösung an mich: roland(ät)cebitec.uni-...

Hier eine Matrix und einige Teillösungen zum (freiwilligen!) Testen deiner Implementierung:

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>A</i> :	0	6	8	5	10
<i>B</i> :		0	5	8	10
<i>C</i> :			0	4	8
<i>D</i> :				0	7
<i>E</i> :					0

<i>J</i>	<i>K</i>	$\alpha_{J,K}$
{ <i>A, B, C, E</i> }	{ <i>D</i> }	0.5
{ <i>A, B, D, E</i> }	{ <i>C</i> }	0.5
{ <i>A, B, E</i> }	{ <i>C, D</i> }	0.5
{ <i>A, C, D, E</i> }	{ <i>B</i> }	1.5
{ <i>A, C, E</i> }	{ <i>B, D</i> }	0.0
{ <i>A, D, E</i> }	{ <i>B, C</i> }	1.5
{ <i>A, E</i> }	{ <i>B, C, D</i> }	0.0
{ <i>B, C, D, E</i> }	{ <i>A</i> }	1.5
{ <i>B, C, E</i> }	{ <i>A, D</i> }	1.0
{ <i>B, D, E</i> }	{ <i>A, C</i> }	0.0
{ <i>B, E</i> }	{ <i>A, C, D</i> }	0.0
{ <i>C, D, E</i> }	{ <i>A, B</i> }	2.0
{ <i>C, E</i> }	{ <i>A, B, D</i> }	0.0
{ <i>D, E</i> }	{ <i>A, B, C</i> }	0.0
{ <i>E</i> }	{ <i>A, B, C, D</i> }	4.5

$$\sum_{\text{taxa } i,j} d_{ij}^1 = 128$$

$$\sum_{\text{taxa } i,j} d_{ij} = 142$$

$$\rho \approx 90,141\%$$