Universität Greifswald
Institute für Mathematik and Informatik
Lecturer: Marc Hellmuth Tutor: Nikolai Nøjgaard

11. Exercise "Bioinformatics", SS 17

Aufgabe 4: (5 Credits)
Let T be a binary rooted tree with leaf set L. Let $R(T)$ be the set of all rooted triples that are displayed by T. Proof that $R(T) \cup \{r\}$ is not consistent for all $r \notin R$.
Explain whether the latter property is also true for non-binary trees in general.

Aufgabe 1: (5 Credits)
Show that the two definitions for cographs are indeed equivalent:

Def 1:

- K_1 is a cograph.
- The disjoint union of two cographs is a cograph.
- The complement of a cograph is a cograph.

Def 2:

- K_1 is a cograph.
- The disjoint union of two cographs is a cograph.
- The join of two cographs is a cograph.

Aufgabe 2: (5+5=10 Credits)
Let A, B, C, D be four different species from which we extracted some genetic material. In particular, we observed two independent gene families (sets of homologous genes) G^1 and G^2 consisting of the respective subsets G^1_X and G^2_X for each species $X \in \{A, B, C, D\}$.

Assume for the first family that $G^1_A = \{a_1, a_2\}$, $G^1_B = \{b_1, b_2, b_3\}$, $G^1_C = \{c_1\}$, $G^1_D = \{d_1\}$, and the second family that $G^2_A = \{a'_1, a'_2\}$, $G^1_B = \emptyset$, $G^1_C = \{c'_1, c'_2\}$, $G^1_D = \{d'_1, d'_2\}$.

Using multiple sequence alignments we obtained the following (symmetric) genetic distance scores for the genes in G^1 and G^2:

G^1	b_1	b_2	b_3	c_1	d_1
a_1	1	1	2	2	3
a_2	2	2	1	2	3
b_1				2	2
b_2				2	2
b_3				2	2
c_1					1

G^2	c'_1	c'_2	d'_1	d'_2
a'_1	2	3	4	4
a'_2	3	2	4	4
c'_1			2	2
c'_2			2	2

For simplicity, no two genes from different gene families will be estimated as orthologs or paralogs.

(a) Based on the similarity scores determine the estimated orthology relation $\widehat{R}^1_\bullet$ and $\widehat{R}^2_\bullet$ (for both gene families G^1 and G^2 independently).
Show that both estimated orthology relation are "valid" orthology relations.

(b) Determine the two gene trees together with its event-label that you can derive from $\widehat{R}^1_\bullet$ and $\widehat{R}^2_\bullet$.

Deadline: Tuesday - June 27, 2017