Universität Greifswald
Institute für Mathematik and Informatik
Marc Hellmuth

6. Übung "Bioinformatik", SS 16

Aufgabe 1: (5+1+2+2 (+10)=10(+10) Credits)
Given the strings $u =$ `GTTTAAG` $v =$ `GAAGA` and the scoring matrix S with entries

$$S[i,j] = \max \begin{cases} S[i-1,j] & + \quad \delta(u_i,-) \\ S[i-1,j-1] & + \quad \delta(u_i,v_j) \\ S[i,j-1] & + \quad \delta(-,v_j) \end{cases}$$

where

$$\delta(a,b) = \begin{cases} 1 & \text{if } a = b \\ -1 & \text{if } a \neq b \text{ and } a,b \neq - \\ -3 & \text{else} \end{cases}$$

(a) Compute S for the strings u and v.

(b) What is the optimal alignment score?

(c) Give one possible optimal alignment for u and v.

(d) How many optimal alignments are there for u and v - Explain shortly your results.

(e) *Optional Exercise:* Implement the Alignment Algorithm with Trace Back in `C++` based on the given scoring matrix S and δ. Sent the source files via email to Nikolai Nøjgaard (nnoej10[at]student.sdu.dk).

Aufgabe 2: (5+5=10 Credits)
Get familiar with BLAST (`http://blast.ncbi.nlm.nih.gov/Blast.cgi`),
see e.g. `http://digitalworldbiology.com/dwb/BLAST` for a short tutorial.

Assume you have sequenced the string ">Sequence_Xy" in the file seq.txt (see extra material on teaching homepage) where Xy = the first two letters of your first name. Use the first best hit of BLAST program *nucleotide blast* with database option *Others / nucleotide collection* and optimization option *discontiguous megablast* to determine the organism the sequence is taken from.

(a) Which organism was your sequence from?

(b) Prepare a short presentation ($\leq$ 5min) about the organism you have sequenced - to be held next tutorial.

Aufgabe 3: (2+2+2+2+2=10 Credits)
Let $G = (V, E)$ be an undirected graph.

(a) Show that G has an even number of vertices with odd degree.

(b) Let $|V| \geq 2$. Show that G has two vertices of the same degree.

(c) Two edges $e = (a, b) \in E$ and $f = (x, y) \in E$ are in Relation Θ, in symbols $e\Theta f$, if and only if

$$d(a, x) + d(b, y) \neq d(a, y) + d(b, x).$$

Show that no two distinct edges on a shortest path in G are in relation Θ.

(d) Show that if G is a forest, then G has $|V| - |E|$ connected components.

(e) Assume now that $V = \{0, 1, 2, 3\}$ and $E = \{(0, 1), (1, 2), (1, 3)\}$. Show that $H \simeq G$ for $H = (\{a, b, c, d\}, \{(d, b), (a, d), (c, d)\})$. How many isomorphism are there?

Aufgabe 4: (2+2+2+2+2=10 Credits)
Let χ denote the chromatic number and Δ the maximum degree. The complement of a graph $G = (V, E)$ is the graph $\overline{G} = (V, \overline{E})$ with $\overline{E} = \{(u, v) \mid (u, v) \notin E, u, v \in V, u \neq v\}$. Prove or disprove

(a) If G is bipartite then $\overline{G}$ is bipartite.

(b) There are graphs with $\chi(G) = \chi(\overline{G})$.

(c) If $\chi(G) = |V|$ then $\chi(G) = \Delta(G) + 1$.

(d) If $\chi(G) = \Delta(G) + 1$ then $\chi(G) = |V|$.

Deadline: Monday - June 6, 2015 - 2pm