Universität Greifswald
Institute für Mathematik and Informatik
Marc Hellmuth

3. Übung "Bioinformatik", SS 16

Aufgabe 1: (4+6+3=13 Credits)
Given the set of strings $P = \{\texttt{CTTA}, \texttt{TGAT}, \texttt{TACT}, \texttt{GATG}\}$.

(a) Draw the overlap graph (omit edges with weight 0)

(b) Apply the algorithm `MGreedy` with input P. Give for each execution-step the sets P and T as well as the final superstring.

(c) Determine $S^*(P)$.

Aufgabe 2: (15 Credits)
Let $E = \{(S_1, S_2), (S_1, S_3), (S_1, S_4), (S_2, S_5), (S_3, S_5), (S_4, S_5)\}$ be the edge set of the overlap graph $G = (\{S_1, \ldots, S_5\}, E, \mathrm{ov}(,))$, where edges with weight 0 are omitted.
Find sequences $S_1, \ldots, S_5$ that give rise to this graph - the particular weights you come up with are not important.

Aufgabe 3: (12 Credits)
For every $m > 0$ find a substring-free set P such that the algorithm `MGreedy` returns a superstring at least $m - 2$ characters longer than the shortest possible $S^*(P)$. Explain your results.

Deadline: Monday - April 25, 2016 - 2pm