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

3. Exercise "Bioinformatics", SS 17

Aufgabe 1: (2+3+2=7 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: (4 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: (4+5=9 Credits)
Let S be a string that has periods of length p and q where $q \leq p$ and $|S| > p + q$.

(a) Show that S has a period of length $p - q$.

(b) Show that S has a period of length $\gcd(p, q)$.
Hint: Euclidean Algorithm

Deadline: Tuesday - April 25, 2017