



Genomanalyse

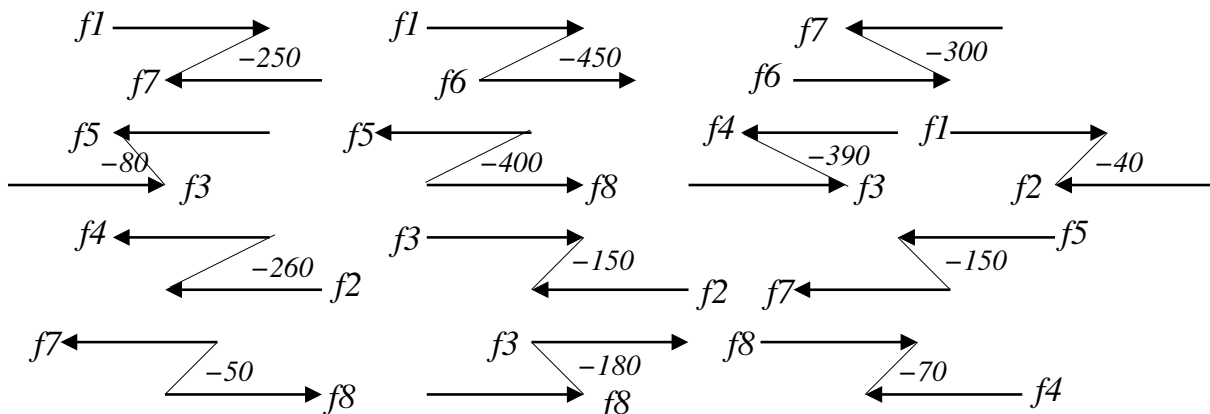
4. Übungsblatt WS 2009

Abgabe: Dienstag 10.11.2009 vor der Übung, Besprechung: in der Übung

Bitte geben Sie auf Ihren abgegebenen Lösungen Name und Matrikelnummer an und arbeiten Sie in Gruppen von 2 oder 3 Personen zusammen.

Aufgabe 7: [3+2+3+2=10 Punkte] Overlap-Graph

Gegeben seien die Reads $\mathcal{F} = \{f_1, f_2, \dots, f_8\}$ der Länge 500. Die Reads überlappen wie folgt:



- Zeichnen Sie den Overlap-Graphen mit den entsprechenden Kantenbeschriftungen.
- Bestimmen Sie in diesem Overlap-Graphen einen minimalen aufspannenden Baum der alle Read-Kanten enthält.
- Zeichnen Sie dann das durch den minimalen aufspannenden Baum beschriebene Layout der Reads. Was sind also die globalen Koordinaten der einzelnen Reads?
- Sind alle gegebenen Overlaps mit dem Layout konsistent? Falls nicht, wo und warum gibt es Inkonsistenzen?

Aufgabe 8: [3+3=6 Punkte] Die *arrival statistic* für *unitigs*

Gegeben sei eine *source sequence* der Länge $G = 500\,000$ und eine Sammlung von $R = 7000$ *fragment reads* der Länge 600.

Wir haben ein *unitig* der Länge $\rho = 5000$, welches aus 90 *fragment reads* besteht.

- Was sagt uns die *arrival statistic* in diesem Fall? Ist das *unitig* mit höherer Wahrscheinlichkeit *unique* oder *repeat affected*? Und sind wir auch „genügend“ sicher?¹
- Bei welchem Wert von k „kreuzt“ die *arrival statistic* den Schwellenwert von 10?

¹Anm.: Die Aufgabenstellung, wurde hier nachträglich noch klarer formuliert. Ebenso das Skript.